Journal of Structural Biology 179 (2012) 261–268

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Computational methods for prediction of protein–RNA interactions

Tomasz Puton [a,1], Lukasz Kozlowski [b,1], Irina Tuszynska [b], Kristian Rother [a], Janusz M. Bujnicki [a,b,*]

[a] Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, PL-61-614 Poznan, Poland
[b] Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland

ARTICLE INFO

Article history:
Available online 12 October 2011

Keywords:
RNA
Protein
RNP
Binding site prediction
Macromolecular docking
Structural bioinformatics

ABSTRACT

Understanding the molecular mechanism of protein–RNA recognition and complex formation is a major challenge in structural biology. Unfortunately, the experimental determination of protein–RNA complexes by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) is tedious and difficult. Alternatively, protein–RNA interactions can be predicted by computational methods. Although less accurate than experimental observations, computational predictions can be sufficiently accurate to prompt functional hypotheses and guide experiments, e.g. to identify individual amino acid or nucleotide residues. In this article we review 10 methods for predicting protein–RNA interactions, seven of which predict RNA-binding sites from protein sequences and three from structures. We also developed a meta-predictor that uses the output of top three sequence-based primary predictors to calculate a consensus prediction, which outperforms all the primary predictors. In order to fully cover the software for predicting protein–RNA interactions, we also describe five methods for protein–RNA docking. The article highlights the strengths and shortcomings of existing methods for the prediction of protein–RNA interactions and provides suggestions for their further development.

1. Background

Protein–RNA interactions play an essential role in many cellular processes, such as RNA transcription, reverse transcription, replication, RNA transport, posttranscriptional processing of RNA, mRNA translation, and regulation of RNA levels in the cell (Chen and Varani, 2005; Glisovic et al., 2008). Defects in protein–RNA interactions have been described for a number of diseases, ranging from neurological disorders to cancer (Cooper et al., 2009; Lukong et al., 2008). RNA-binding proteins (RBPs) are a large and heterogeneous group of macromolecules that fulfill their function using a wide range of domain architectures. In particular, there are numerous protein domains involved in RNA binding, with the prevalence of αβ structures, e.g. of the "Rossmannoid" type (Cammer and Carter, 2010). Some common and well-characterized RNA-binding domains include the following: RNA Recognition Motif (RRM), K-homology (KH) domain, RGG box, Sm domain, double stranded RNA-binding domain (dsRBD), cold-shock domain, Pumilio/FBF (PUF) domain, and the Piwi/Argonaute/Zwille (PAZ) domain (reviews: (Chen and Varani, 2005; Lunde et al., 2007)). For some protein domains, exemplified by the RRM and dsRBD, which include "RNA" in their names (Clery et al., 2008), nearly all members show RNA-binding activity. In other families the RNA-binding property is exhibited only by a fraction of members, e.g. in enzyme families such as Rossmann-fold methyltransferase (RFM), related domains can bind RNA, DNA, proteins, or other substrates (Anantharaman et al., 2002; Czerwoniec et al., 2009). The abundance and diversity of RNA-binding proteins is correlated with the complexity of the organism they are found in, with the number of RNA-binding proteins reaching thousands in vertebrates (Anantharaman et al., 2002). It is worth emphasizing that eukaryotic RNA-binding proteins often comprise multiple RNA-binding domains (Glisovic et al., 2008).

The understanding of protein–RNA interactions improves as new structures of RNA–protein (RNP) complexes are solved and the molecular details of interactions analyzed. Unfortunately, the experimental determination of RNP complexes is a slow and

Abbreviation: AUC, area under receiver operating characteristic curve; cryo-EM, cryo-electron microscopy; DARS-RNP, decoys as reference state based potential to assess structure of protein–RNA complexes; FPR, false positive rate (1 – specificity); HMM, Hidden Markov Model; MCC, Matthews Correlation Coefficient; NMR, nuclear magnetic resonance spectroscopy; PDB, the Protein Data Bank (http://www.rcsb.org/); QUASI-RNP, QUASI chemical based potential to assess structure of protein–RNA complexes; RMSD, root mean square deviation; RNP, RNA–protein complex; ROC, receiver operating characteristic; RRM, RNA Recognition Motif; SVM, support vector machine; TPR, true positive rate (sensitivity).

* Corresponding author at: Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Ks. Trojdena 4, PL-02-109 Warsaw, Poland. Fax: +48 22 597 07 15.

E-mail addresses: tputon@genesilico.pl (T. Puton), lukaskoz@genesilico.pl (L. Kozlowski), irena@genesilico.pl (I. Tuszynska), krother@genesilico.pl (K. Rother), iamb@genesilico.pl (J.M. Bujnicki).

[1] The authors wish to be known that in their opinion, T.P. and L.K. should be considered joint first authors.

1047-8477/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2011.10.001

difficult process (Ke and Doudna, 2004; Scott and Hennig, 2008). As of September 2011, 1203 macromolecular complexes involving both protein and RNA components (but excluding RNA/DNA hybrids) were available in the Protein Data Bank (PDB), including 1035 solved by X-ray crystallography, 69 by nuclear magnetic resonance (NMR) spectroscopy, and 99 by other methods. These structures contained 12,753 protein chains interacting with RNAs, but many proteins were highly similar to each other. After removing redundant protein chains with sequence identity >90%/40% only 798/480 proteins remained. Despite the fact that numerous protein–RNA interactions have been experimentally determined, for many RNPs, e.g. the spliceosome, RNA-Induced Silencing Complex (RISC), complexes containing Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and Cleavage/Polyadenylation Specificity Factor (CPSF), complete high-resolution structures are not yet available, and our knowledge is based mostly on atomic structures of isolated components and low-resolution structures of their assemblies obtained with, e.g. cryo-electron microscopy (cryo-EM).

Given the scarcity of experimentally determined structures of RNP complexes, the computational prediction of RNP complex structures would greatly help studying protein–RNA interactions. There exists a wealth of low-resolution experimental data that identify the interacting components and often tie them to particular functional states. These data can be exploited by bioinformatics methods for structure prediction. However, while the methodology for prediction and modeling of proteins and protein–protein complexes is very well established [reviews: (Bujnicki, 2008; Moreira et al., 2010; Wichadakul et al., 2009)], there are much fewer methods for predicting and modeling RNA structure and protein–RNA interactions (Laing and Schlick, 2010; Rother et al., 2011a). In this article we review bioinformatics methods for predicting RNA-binding sites for proteins with either known or unknown three-dimensional structures and docking methods for predicting the structures of RNP complexes.

2. Prediction of RNA binding proteins

Computational predictions of RNA binding addresses three connected problems: (i) whether a given protein binds RNA, (ii) if it does – which residues in the protein sequence are directly involved in making contacts with the RNA, and (iii) what is the structure of the protein–RNA complex. The phosphate backbone of RNA is negatively charged, and it preferentially interacts with positively charged proteins, whose surfaces are enriched with residues such as Arg or Lys (Allers and Shamoo, 2001; Jones et al., 1999, 2001; Nadassy et al., 1999). However, not all positively charged residues are involved in RNA-binding. In fact almost all proteins include surface-exposed positively charged residues, and definitely not all of them bind RNA – they may be involved in binding other anionic ligands (in particular DNA that has a very similar backbone), formation of salt bridges, catalysis, and other functions. The relative ratio of positively and negatively charged residues or a theoretical pI that can be calculated from protein sequence is also a poor predictor of RNA binding. There exist negatively charged proteins that bind anionic ligands, including nucleic acids (Ledvina et al., 1996), e.g. in the set of 44 RNA-binding proteins in the PDB analyzed in this work (as of September 2011) six (14%) exhibit a theoretical pI < 7.0. Computational methods have been therefore developed to identify RNA-binding proteins based mostly on charge, but some of them utilize other sequence features such as overall amino acid composition, van der Waals volume, polarity, etc. For instance Cheng et al. (2008) used position-specific scoring matrices (PSSMs), while Kumar et al. (2011), Yu et al. (2005) and Fujishima et al. (2007) used amino acid composition and periodicities as feature vectors to train support vector machines (SVM) for discrimination of DNA- and/or RNA-binding proteins from proteins with other functions. For proteins with known structures, Mandel-Gutfreund and coworkers developed a method to identify patches of positively charged residues on the surface and to discriminate between various types of nucleic acid-binding proteins (Shazman et al., 2007, 2011). There exist other tools for the prediction of the protein–RNA binding function, which rely on machine learning methods, mostly support vector machines (SVMs) (Peng et al., 2011; Shao et al., 2009; Shazman and Mandel-Gutfreund, 2008). Unfortunately, none of these methods are currently available as public web servers or standalone programs for easy local installation. The only such tool that we found available is DRNA (Zhao et al., 2011), which is also used for identification of protein residues interacting with RNA. It has been described in Section 4.

3. Prediction of RNA binding residues from protein sequence

The prediction of RNA-binding residues from protein sequence (usually with an assumption that the protein is known or expected to bind RNA) mainly relies on using machine learning methods such as neural networks, Hidden Markov Models (HMMs), and support vector machines (SVMs). In Table 1 we listed seven publicly available tools for predicting RNA-binding sites from protein sequence alone that do not take any structural information into account. We have also considered the tools RISP (Tong et al., 2008) and PRIP (Maetschke and Yuan, 2009), for predicting RNA-binding sites, but the original URLs were unavailable and we were unable to find alternative websites at the time of writing this manuscript.

All methods listed in Table 1 are based on machine learning algorithms. Four methods use SVMs in order to predict RNA interacting protein residues (BindN, BindNPlus, PPRInt and PiRaNhA). The remaining three methods, RNABindR, NAPS, and PRBR use a Naïve Bayes classifier, decision trees (C4.5 algorithm), and random forest algorithms, respectively. These algorithms typically take into account physicochemical properties of amino acids, in particular charge, hydrophobicity, predicted features such as solvent accessibility and secondary structure, and often also sequence conservation, and the local sequence context.

4. Prediction of RNA binding sites from protein structures

The availability of the tertiary structure of a protein can greatly facilitate the prediction of the RNA-binding site, which is typically formed by surface-exposed residues that are close to each other in space, but not necessarily in sequence. RNA binding sites can often be recognized as positively charged surface patches whose shape is compatible with binding the negatively charged RNA backbone (Shazman and Mandel-Gutfreund, 2008; Shazman et al., 2007). Additionally, visual inspection allows localizing clefts with aromatic or hydrophobic residues that may be involved in stacking interactions with bases of single-stranded RNA. Structure-based predictive methods may exploit the same information as sequence-based methods, but replace the predicted local structural features (e.g. solvent accessibility and secondary structure) by the observed ones. Additionally they may utilize more global features available only on the 3D level, such as surface shape, distribution of the electrostatic potential (which may highlight a region of positive charge in an otherwise negatively charged protein), and spatial proximity of residues with particular features. The prediction of global or local propensity for RNA binding can be also achieved by comparing the query structure to known structures of RNP complexes. Various approaches for predicting RNA-binding proteins based on structural analysis, and for identification of RNA-binding residues in these proteins have been reviewed in the

Table 1
Bioinformatics tools for sequence-based prediction of RNA-binding sites in proteins.

Method	URL	Ref.	Description
BindN	http://bioinfo.ggc.org/bindn/	Wang and Brown (2006)	Uses SVM to predict RNA binding residues based on side chain pKa value, hydrophobicity index and molecular mass of amino acids. It is also capable of predicting DNA binding residues.
BindN+	http://bioinfo.ggc.org/bindn+/	Wang et al. (2010)	An upgraded version of BindN, also using an SVM classifier.
NAPS	http://prediction.bioengr.uic.edu/	Carson et al. (2010)	Uses a combination of machine learning and C4.5 algorithm to predict both RNA and DNA binding residues in protein sequences.
PiRaNhA	http://bioinformatics.sussex.ac.uk/PIRANHA/	Murakami et al. (2010)	Uses an SVM classifier to predict protein residues interacting with either RNA or DNA. The classifier makes use of position specific scoring matrices, residue interface propensity, predicted residue accessibility and residue hydrophobicity.
PPRInt	http://www.imtech.res.in/raghava/pprint/	Kumar et al. (2007)	Uses an SVM classifier trained on a PSSM profile generated by running PSI-BLAST on a non-redundant protein sequence database.
RNABindR	http://bindr2.gdcb.iastate.edu/RNABindR/	Terribilini et al. (2007)	Uses a Naive Bayes classifier trained on interactions observed in structures of protein–RNA complexes in the PDB. Additionally, it can be used as an advanced viewer for known Protein–RNA complexes.
PRBR	http://www.cbi.seu.edu.cn/PRBR/	Ma et al. (2011)	Combines the enriched random forest (ERF) algorithm with a hybrid feature vector, composed of predicted secondary structure, conservation information of the physicochemical properties of amino acids and the information about dependence of amino acids with regard to polarity-charge and hydrophobicity in the protein sequences.

literature (Chen and Lim, 2008a,b; Maetschke and Yuan, 2009), however only a few general-purpose algorithms have been implemented in a form available to the public. In particular, the number of bioinformatics methods for structure-based prediction of RNA-binding residues is much smaller than that of methods for predicting DNA-binding or protein-binding residues. In Table 2 we listed two methods that are available as web servers and one method available as a stand-alone program.

It is worth mentioning that structure-based prediction methods are agnostic with respect to the methodology used for protein structure determination, so in principle they can be used to predict RNA-binding sites for structures obtained with, e.g. X-ray crystallography, NMR, or theoretical modeling. However, all such models may require special treatment. First, predicting RNA-binding residues for crystal structures may require editing of the input file, such as the addition of missing disordered loops by the comparative modeling method, to represent the entire sequence of interest. Second, structure-based methods typically make predictions for single models rather than ensembles of multiple models, which may require the selection of a representative structure or the calculation of a consensus model for NMR ensembles. Third, predicting RNA-binding residues based on theoretical models requires taking the predicted global and local model quality into account, because errors and inaccuracies of theoretical models may propagate in the predicted complexes. Finally, many proteins bind RNA as oligomers, while some methods may accept only single chains, i.e. monomeric structures.

5. Benchmark of RNA binding site prediction methods

In order to measure and compare the performance of methods predicting protein–RNA interactions, we created a comparative benchmark. We tested seven sequence-based methods for prediction of protein–RNA interactions from Table 1 (BindN, BindN+, NAPS, PiRaNhA, PPRInt, PRBR and RNABindR) and three structure-based methods from Table 2 (KYG, OPRA and DRNA). The testing dataset was compiled from 75 records containing RNP complexes released between January 1st and April 28th 2011 from the Protein Data Bank to minimize the likelihood that any of these structures were used for training of the methods tested. All protein–RNA residue pairs with atoms closer than 3.5 Å were considered as interacting**.** As a result, we obtained a redundant dataset comprising 949 protein chains. We removed the redundancy on the protein level with the CD-HIT program (Li and Godzik, 2006), and kept only one representative per set of proteins with more than 40% sequence identity. Moreover, we kept only proteins which were not sequence-similar to proteins used for training individual methods (we again used 40% sequence identity threshold for filtering out close homologs using the CD-HIT program). The final dataset was composed of 44 sequences for which all of the methods returned predictions (see Supplementary Material for the final dataset used in this study and a compilation of proteins used for testing individual methods as provided by their respective authors). There were many cases where some methods failed (e.g. PiRaNhA and NAPS did not return prediction for chain K from PDB record 3PLA). Ideally, the test of structure-based predictors should be carried out for apo variants (i.e. for structures solved in the absence of the RNA), however the small number of proteins with structures solved both without the RNA (for making the prediction) and with the RNA (for testing the prediction's accuracy) prevents us from carrying out such an analysis.

For each of the 10 methods under consideration, predictions were collected for all 44 test sequences. The methods scored each protein residue with respect to its RNA binding propensity, and residues with scores exceeding the default threshold values (set by the methods' authors) were considered as predicted to be

Table 2
Bioinformatics tools for structure-based prediction of RNA-binding sites in proteins.

Method	URL	Ref.	Description
KYG	http://cib.cf.ocha.ac.jp/KYG/	Kim et al. (2006)	Uses a number of scores based on the RNA-binding propensity of individual residues, doublets of spatially close residues, sequence profiles, and combinations thereof.
DRNA	http://sparks.informatics.iupui.edu/yueyang/DFIRE/dRNA-DB-service	Zhao et al. (2011)	Predicts RNA-binding proteins and RNA binding sites based on similarity to known structures: it performs a structural alignment to known protein–RNA complex structures followed by binding assessment with a DFIRE-based statistical energy function.
OPRA	Program available upon request from the authors	Perez-Cano and Fernandez-Recio (2010)	Predicts RNA-binding residues using a predictive score from propensities of residues at known protein–RNA interfaces weighed by their accessible surface.

interacting with the RNA. By comparing predicted interactions with the ones observed in RNP complexes, we established true and false positives and negatives. We used these values to create receiver operating characteristic (ROC) curves by plotting the false positive rate (1 – specificity, FPR; Eq. (1) against the true positive rate (sensitivity, TPR; Eq. (2) for each method. We have also estimated the areas under the ROC curve (AUC) using the composite trapezoidal rule and calculated the Matthews Correlation Coefficient (MCC; Eq. (3)) for each method.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{1}$$

Eq. (1). True positive rate (TPR; sensitivity). TP – number of correctly predicted interacting residues, FN – number of incorrectly predicted non-interacting residues.

$$\mathrm{FPR} = 1 - \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \tag{2}$$

Eq. (2). False positive rate (FPR; 1 – specificity). TN – number of correctly predicted non-interacting residues, FP – number of incorrectly predicted interacting residues.

$$\mathrm{MCC} = \frac{\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{3}$$

Eq. (3). Matthews Correlation Coefficient (MCC). TP – number of correctly predicted interacting residues, TN – number of correctly predicted non-interacting residues, FP – number of incorrectly predicted interacting residues, FN – number of incorrectly predicted non-interacting residues.

6. Results of the RNA binding site prediction benchmark

The results of our benchmark presented in Table 3 and Fig. 1 provide an overview of the performance of 10 third-party tools for the prediction of protein–RNA interactions and one ad hoc created meta-predictor (described in a separate section below). We were not able to perform the ROC analysis in case of the sequence-based method PRBR, and all three structure-based methods (DRNA, KYG and OPRA). The reason was that the output of those methods did not include scores for individual residues describing their RNA-binding propensity, which is a compulsory requirement for such an analysis. For the remaining methods, we performed the ROC analysis within a range of observed scores describing the predicted RNA-binding propensity. In case of RNABindR, the output for each protein sequence contained three

Table 3
Results of a benchmark of 10 methods predicting protein–RNA interactions – seven sequence-based methods listed in Table 1.

Method	MCC	AUC
Meta-predictor**	0.460	0.835
PiRaNhA	0.435	0.822
BindN+	0.397	0.821
KYG*	0.382	N/A
DRNA*	0.382	N/A
PPRInt	0.339	0.779
RNABindR	0.317	0.708
OPRA*	0.296	N/A
BindN	0.297	0.733
PRBR	0.294	N/A
NAPS	0.215	0.679

Methods were sorted in descending order according to MCC. N/A – not available, MCC – Matthews Correlation Coefficient, AUC – area under curve.
* Three structure-based from Table 2 (KYG, OPRA and DRNA).
** An ad hoc meta-predictor developed during this study based on top three sequence-based methods according to our benchmark (PiRaNhA, PPRInt and BindN+).

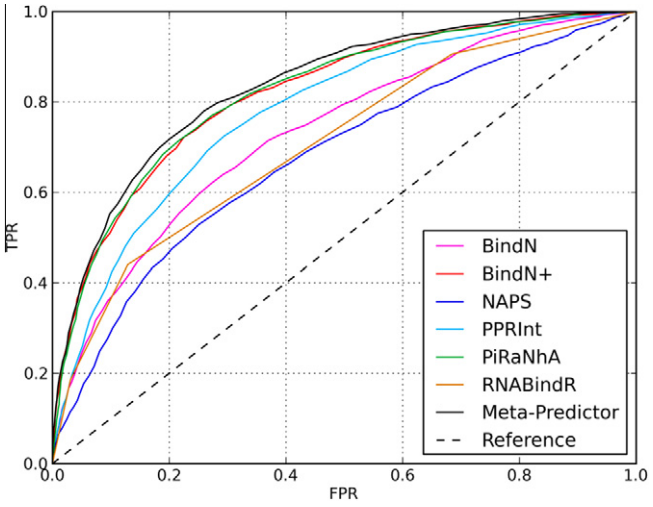

Fig.1. ROC curves of six methods for prediction of RNA-binding residues from sequence (Table 1) and a meta-predictor created during this study using the best three sequence-based methods (PiRaNhA, PPRInt and BindN+).

predictions – optimal, high specificity, and high sensitivity. Because of the fact that no additional scores describing RNA-binding propensities were provided, we calculated scores based on the congruency of predictions. Therefore, as seen in Fig. 1, the RNABindR ROC adopts an unusual shape. In case of all other methods, the ROC analysis was performed based on the scores assigned to protein residues.

According to our benchmark, methods KYG and DRNA were the best among the structure-based methods tested, with MCC values reaching 0.382 in both cases. This value may be overestimated, as the predictions were tested for protein structures taken from protein–RNA complexes (i.e. correspond to the RNA-bound conformation), while in real life the predictions are made for proteins with known structures, but unknown mode of RNA binding. A test of predictions made for unbound variants may be done in the future, when the number of structure pairs with and without RNA grows. Among the sequence-based methods, the ranking was topped by PiRaNhA, for which the MCC reached 0.435 and the AUC was estimated to be 0.822. The next two best-scored methods were BindN+ (MCC: 0.397, AUC: 0.821) and PPRInt (MCC: 0.339, AUC: 0.779).

7. Meta-predictor for protein–RNA interactions

Following the benchmark of primary predictors of RNA-interacting residues, we developed an 'ad hoc' meta-predictor based on three sequence-based primary predictors that ranked highest in our tests (PiRaNhA, PPRInt and BindN+). The meta-predictor works as follows: first, for a query protein sequence, predictions are collected from the three above-mentioned primary predictors. Then, a new meta-score for each residue is calculated as a weighted mean of three scores using the AUC values from the benchmark as weights. As the output, the meta-predictor returns a set of scores for all residues of a given protein sequence query. A threshold to discriminate between RNA-binding and non-binding residues was defined according to the point on the meta-predictor's ROC curve closest to the values of FPR = 0.0 and TPR = 1.0 (upper left corner). Once the threshold value was selected, we were able to calculate the MCC value for the meta-predictions. Our meta-predictor outperformed PiRaNhA only by 1.6% according to AUC (0.835 vs. 0.822) and by 5.7%, according to MCC (0.460 vs. 0.435), which suggests that the predictions of the currently best methods are strongly correlated with each other and combining

them gives a very limited synergy. The meta-predictor is freely available via the GeneSilico protein structure prediction meta-server (http://iimcb.genesilico.pl/meta2/) (Kurowski and Bujnicki, 2003).

8. Recommendations for users interested in predicting RNA-binding residues for their protein sequences and/or structures

The recommendation which program should be used depends mainly on the data available. If a structure is available for the target protein, the user should check whether it exhibits high similarity to other RNA-binding proteins with already known structures in complex with RNA, e.g. by running a BLAST (Altschul et al., 1997) search against the proteins in the PDB database. In such a case, e.g. if the aim is to predict the RNA-binding site in a new RRM protein structure that has many homologs with structures solved in complex with RNA (Clery et al., 2008), the best option is to use structure based methods: KYG, OPRA, and DRNA. For a query protein with known structure that has no close homolog, we recommend to use the top-scoring structure-predictor KYG and to compare its result with the top-performing sequence-based predictors (primary methods: PiRaNhA, PPRInt and BindN+ or our meta-predictor). In the absence of the protein structure, the above-mentioned sequence-based predictors of RNA-binding residues should be used. Additionally, the protein sequence may be used to model the structure computationally, in particular using the comparative approach. For this, we recommend the GeneSilico structure prediction metaserver developed in our laboratory (Kurowski and Bujnicki, 2003). The resulting model may be used as a query for structure-based methods. However, theoretical models usually contain various errors and inaccuracies and their use for predicting binding sites requires caution. In particular, a theoretical model should be first evaluated with Model Quality Assessment Programs (Kryshtafovych and Fidelis, 2009) and entire models and/or individual residues with low scores should be regarded as unreliable for any further predictions. In all cases, where high specificity (i.e. high confidence of positive predictions) is desired, and the presence of false negatives (i.e. missed true RNA-binding residues) is not a problem, we recommend to use not just one, but several methods that scored best in a comparative benchmark, and to select residues predicted to be involved in binding by all or most methods used.

9. Protein–RNA docking

Docking methods are widely used to predict three-dimensional structures of macromolecular complexes, starting from coordinates of their components (Moreira et al., 2010). The larger molecule is usually referred to as the receptor, while the smaller molecule is usually called the ligand. The problem of predicting the structure of a complex can be split into two sub-problems: to search the conformational space of possible orientations and conformations (poses) of the components, and to distinguish near-native structures from all other alternative complex models (decoys) explored through the search algorithm. Many methods combine both tasks, while others specialize only in the assessment of decoys, leaving their generation to the user.

An ideal docking method should be able to assemble the structures of components into a complex, and score the most native-like decoy (a complex structure closest to the native) significantly better than any non-native one. In reality, the structure of the complex is unknown. Structures of binding partners, which are solved individually, usually undergo conformational changes upon association, in a process known as induced fit. "Unbound" docking algorithms must be tolerant to this difficulty. Conformational changes are either modeled explicitly by high-resolution methods, which make such analyses computationally very demanding, or introduce a certain level of 'fuzziness' (reviews: (Moreira et al., 2010; Zacharias, 2010)).

Thus far, a large number of protein–protein docking methods have been developed (reviews: (Janin, 2010; Moreira et al., 2010; Vakser and Kundrotas, 2008)), whereas the number of methods for modeling RNP complexes is still limited. In Table 4 we listed some of the publicly available web resources and standalone docking methods that accept protein and RNA coordinates as an input to generate RNP complex decoys, as well as scoring functions for the selection of presumably native-like models from decoy sets. To our best knowledge, no docking methods have been developed specifically for RNP complexes. Instead, a number of methods for modeling protein–protein complexes have been adapted to deal with nucleic acid molecules as receptors and/or ligands.

One interesting and frequently neglected aspect of RNA structure and protein–RNA interactions is the presence of posttranscriptional modifications, which increase the basic set of four nucleotides (A, U, G, C) to more than 100 variants with altered base and/or ribose moieties (Dunin-Horkawicz et al., 2006). Modified residues in RNA are involved in many processes, including RNA folding and RNA–RNA interactions, but also specific protein–RNA recognition and binding (Grosjean, 2009; Mucha et al., 2001; Soma et al., 2003). It must be noted, that modified residues are often problematic for the available docking methods, because they are not represented in the standard potentials, and must be converted into the unmodified counterparts in RNA structures used for docking, e.g. by using the ModeRNA software (Rother et al., 2011b).

Most protein–RNA docking methods described in Table 4, i.e. GRAMM, PatchDock, and Hex, which are capable of handling modified nucleotides in RNA molecules, do not have appropriate scoring functions to identify near-native structures of RNP complexes, hence they require special extensions for scoring protein–RNA interactions. In the course of the last few years a few statistical potentials for the evaluation of protein–RNA interactions have been proposed. The Varani group developed a distance-dependent all-atom statistical potential (Zheng et al., 2007). It performs well in discriminating models of RNP complexes that are very close to the native structure, i.e. with the root mean square deviation (RMSD) < 5 Å. However, during a real (unbound) docking experiment it may be difficult to obtain many decoys with RMSD < 5 Å, hence this approach is unsuitable for modeling of complexes that may exhibit conformational changes beyond an RMSD of 5 Å between the bound and unbound forms. In most cases of protein–RNA binding, moderate conformational changes of protein and/or RNA molecules occur upon complex formation. There, low resolution methods that apply a coarse-grained energy model to a coarse-grained representation (i.e. without looking at the atomic details that change upon binding) have a chance to be practically useful. Another potential developed by the Fernandez group (Perez-Cano et al., 2010) works on the residue-nucleotide level. It was designed to improve the discriminative power of the FTDock potential and is not available as a standalone program.

We have recently developed two new, medium-resolution, knowledge-based potentials for scoring models of RNP complexes (Tuszynska and Bujnicki, 2011): the quasichemical potential (QUASI-RNP) and the decoys as the reference state potential (DARS-RNP). These potentials are based on a reduced representation of protein and RNA, use the same mathematical base but differ in their reference state. The reduced representation is intermediate between the atom-level Varani potential and the residue-level Fernandez potential. Both statistical potentials comprise a distance and orientation-dependent energy term, a site-dependent energy term, and a penalty for steric clashes. The site-dependent term assesses the probability of interaction of amino acid residues with

Table 4
Examples of publicly available bioinformatics tools for modeling of protein–RNA complexes.

Method	URL	Ref.	Description
Haddock	http://www.nmr.chem.uu.nl/haddock/ http://haddock.science.uu.nl/services/HADDOCK	Dominguez et al. (2003)	Uses biochemical and/or biophysical interaction data as restraints. Enables docking of various molecules including proteins, nucleic acids, and small molecules. Available as a standalone program and a server.
GRAMM	http://vakser.bioinformatics.ku.edu/main/resources_gramm1.03.php	Katchalski-Katzir et al. (1992)	A program for low-resolution docking, performs a 6-dimensional search through the rigid body translations and rotations of the ligand molecule. Does not allow for using restraints during the docking process. Capable of generating decoys for any molecule, but requires specialized external scoring functions for complexes involving molecules other than proteins.
Hex	http://hex.loria.fr/ http://hexserver.loria.fr/	Ritchie and Kemp (2000)	Enables protein–protein and protein–nucleic acid docking. Relies on using spherical polar Fourier (SPF) correlations and graphics processor units (GPUs) to accelerate the calculations. Knowledge of one or both binding sites may be used to focus and shorten the calculation. Decoy scoring includes shape matching and electrostatics. The method does not give a possibility to save all docking decoys and does not have a special function for protein–RNA complexes.
PatchDock	http://bioinfo3d.cs.tau.ac.il/PatchDock/index.html	Schneidman-Duhovny et al. (2005)	A geometry-based molecular docking algorithm available both as a standalone program and a web server, developed for prediction of protein–protein and protein–small molecule complexes. It can generate poses for protein–nucleic acids complexes, but does not have the appropriate scoring function to identify near-native models. It allows to define potential binding sites in both ligand and receptor molecules.
FTDock (3D-Dock)	http://www.sbg.bio.ic.ac.uk/docking/	Gabb et al. (1997)	Performs rigid-body docking. This program was developed for protein–protein docking; it accepts RNA and DNA molecules (without modified nucleotides), but has no specialized scoring function for protein–RNA complexes.
DARS-RNP and QUASI-RNP	http://www.genesilico.pl/RNP/	Tuszynska and Bujnicki (2011)	Statistical and quasi-chemical potentials for scoring of protein–RNA decoys obtained with other methods, e.g. GRAMM, Hex, PatchDock, FTDock, etc.

the Watson–Crick, sugar, and Hoogsteen edges of nucleotide residues, as defined by (Leontis and Westhof, 2001). The DARS-RNP and QUASI-RNP programs also allow for clustering the best scored structures, which helps in identifying ensembles of similar structures with good scores that are more likely to represent near-native conformations.

We compared the discriminatory power of the four aforementioned statistical potentials for the identification of native-like RNP complexes among decoys generated by docking and found that the DARS-RNP potentials exhibits the highest discriminatory power for decoy sets that include near-native structures without steric clashes, as well as for decoys generated with the GRAMM method (Tuszynska and Bujnicki, 2011). The advantage of the DARS-RNP potential over the QUASI-RNP potential (as well as over the other two potentials developed by other groups) can be explained by the realistic treatment of "random" protein–RNA interactions. In the DARS-based approach, the statistics of amino acid-nucleotide contacts are inferred from geometrically plausible, but biologically irrelevant decoys. The calculation of a DARS-based potential requires, however, the calculation of a large number of decoys for each complex in the training set, hence it requires a considerably bigger computational effort, which may be prohibitive in case of large training sets.

Recently, a new coarse-grained potential for protein–RNA docking was described (Setny and Zacharias, 2011), which is an extension of the ATTRACT docking method (Zacharias, 2003). The potential was tested on 110 crystallographic structures of protein–RNA complexes; however it was not yet compared directly with other available statistical potentials described above, and it is not yet able to use decoys generated by other methods.

10. Conclusions

In recent years, the number of reported protein–RNA complexes has been rapidly increasing. This growth is visible both in the PDB database as the yearly increase of deposited protein–nucleic acid complexes (258 in 2007 vs. 449 in 2010), and in the PubMed database as the change in the number of publications associated with the term "RNA-binding proteins" (2843 in 2007 vs. 3152 in 2010). Still, the determination of structures for proteins in complex with their partner RNAs is laborious and slow, hence there is a large demand for the development of computational methods for predicting such structures either from structures of the components or directly from sequences. Despite the fact that current predictors are still far from being perfect, they can provide useful hints to guide experimental analyses. The importance of modeling RNP complexes is reflected in their recent inclusion as targets in the Critical Assessment of Prediction of Interactions (CAPRI) experiment (de Vries et al., 2010).

Our benchmark shows that the currently available methods for predicting RNA-binding proteins and RNA-binding sites are far away from the high accuracy desired for practical applications. Small improvements can be achieved by integrating the available methods into meta-predictors. However, the top-performing RNA-binding site predictors based on sequence generate results that are highly correlated with each other, which suggests that further progress can be made by developing methods based on other sequence or structure features than those used so far. On the other hand, the existing structure-based predictors rely to a large extent on detection of global similarity to known structures of RNA-binding proteins, and they do not seem to utilize the full potential of the 3D information. We expect that the most recent developments in the area of protein–RNA docking potentials will prompt the work towards a new generation of predictive methods that utilize both structure and sequence information and will enable accurate predictions for protein structures without obvious similarity to known RNA binders.

Another area for the future development concerns macromolecular docking. None of the methods mentioned in this article are capable of predicting the structures of RNP complexes that involve large conformational changes. In our opinion the main problem of the existing docking methods (typically protein–protein docking methods adapted to handle RNA structures) is their inability to take conformational changes into account on the level of RNA, or both RNA and protein simultaneously. Recently, a number of RNA

3D modeling methods have been described that enable fast RNA folding simulations (Cao and Chen, 2005; Das and Baker, 2007; Ding et al., 2008; Parisien and Major, 2008). We believe that the next useful step would be to accommodate these or similar approaches to re-fold fragments of the RNA molecule predicted to form contacts with the protein partner, and optimize the energy of both internal and mutual interactions. The conceptual similarity of successful algorithms for protein and RNA 3D structure modeling (Rother et al., 2011a) suggests that their combination into unified modeling methods is feasible.

Synergy is also expected from the combination of theoretical predictive methods with low-resolution experimental analyses. Structural probing experiments such as footprinting and cross-linking can provide information about secondary structure, inter- and intramolecular interactions (Weeks, 2010), while SAXS and cryo-EM experiments can be used to obtain information about the shape of macromolecular complexes (Lipfert and Doniach, 2007; Zhou, 2008). For many macromolecular complexes, such as the spliceosome, it has been suggested that the structure may be modeled by using cryo-EM maps, as molecular envelopes into which structures of individual components could be fitted, using restraints from biochemical experiments and other bioinformatics-based predictions (Jurica, 2008). Various methods for modeling RNA structures and RNP complexes based on low-resolution experimental data have been described in the literature (Das et al., 2008; Mertens and Svergun, 2010; Yang et al., 2010) and a number of case studies have been published (e.g. (Tsai et al., 2003)). Nonetheless, dedicated methods for automated prediction of protein–RNA interactions and RNP complex structure modeling based on experimental data remain to be developed.

Note added in proof

We thank Rasna Walia, Vasant Honavar, Drena Dobbs, and Yasser El-Manzalawy (Iowa State University, Ames, USA) for re-analyzing our data and pointing out inconsistencies that allowed us to identify and correct errors in our original manuscript. We also thank Yaoqi Zhou and Yuedong Yang (Indiana University, Indianapolis, USA) for re-analyzing our data and for personal communication. Our discussion with the Yaoqi Zhou group prompted a correction of their scripts, leading to an improvement of DRNA. Following the correction in DRNA, additional 8 proteins in our dataset can be predicted as RNA-binding by their method, increasing the number of predicted RNA-binding proteins to 20, and boosting the MCC of DRNA to 0.49. If all proteins are considered as RNA-binding and a cutoff confidence score of 0.47 is used, the DRNA method achieves MCC of 0.53 for all proteins. For 20 proteins that are predicted as RBPs, the MCC for the binding residues is 0.72 (Yaoqi Zhou, personal communication).

Acknowledgments

We thank Michal Boniecki and Marcin Pawlowski for inspiration and discussions. We also thank Piotr Setny and Juan Fernandez-Recio for explaining details of their method and sending us their software. The research on protein–RNA interactions in our laboratory is funded primarily by the Foundation for Polish Science (FNP, Grant TEAM/2009-4/2), while the development of our software for protein bioinformatics is funded mainly by the Polish Ministry of Science and Higher Education (MNiSW, Grant POIG.02.03.00-00-003/09). T.P. and L.K. were supported by MNiSW (PhD Grants number N N301 035539 to T.P. and N N301 190139 to J.M.B.). I.T. was supported by the 6th Framework Programme of the European Commission (EC FP6, Network of Excellence EURANSET, contract number LSHG-CT-2005-518238). K.R. was supported by the German Academic Exchange Service (Grant D/09/42768). J.M.B. was additionally supported by the European Research Council (ERC, StG Grant RNA+P=123D) and by the "Ideas for Poland" fellowship from the Foundation of Polish Science (FNP).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2011.10.001.

References

Allers, J., Shamoo, Y., 2001. Structure-based analysis of protein–RNA interactions using the program ENTANGLE. J. Mol. Biol. 311, 75–86.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Anantharaman, V., Koonin, E.V., Aravind, L., 2002. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res. 30, 1427–1464.
Bujnicki, J.M., 2008. Prediction of protein structures, functions and interactions. John Wiley & Sons Ltd., Chichester.
Cammer, S., Carter Jr., C.W., 2010. Six Rossmannoid folds, including the Class I aminoacyl-tRNA synthetases, share a partial core with the anti-codon-binding domain of a Class II aminoacyl-tRNA synthetase. Bioinformatics 26, 709–714.
Cao, S., Chen, S.J., 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. RNA 11, 1884–1897.
Carson, M.B., Langlois, R., Lu, H., 2010. NAPS: a residue-level nucleic acid-binding prediction server. Nucleic Acids Res. 38, W431–W435.
Chen, Y.C., Lim, C., 2008a. Predicting RNA-binding sites from the protein structure based on electrostatics, evolution and geometry. Nucleic Acids Res. 36, e29.
Chen, Y.C., Lim, C., 2008b. Common physical basis of macromolecule-binding sites in proteins. Nucleic Acids Res. 36, 7078–7087.
Chen, Y., Varani, G., 2005. Protein families and RNA recognition. FEBS J. 272, 2088–2097.
Cheng, C.W., Su, E.C., Hwang, J.K., Sung, T.Y., Hsu, W.L., 2008. Predicting RNA-binding sites of proteins using support vector machines and evolutionary information. BMC Bioinformatics 9 (Suppl. 12), S6.
Clery, A., Blatter, M., Allain, F.H., 2008. RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. 18, 290–298.
Cooper, T.A., Wan, L., Dreyfuss, G., 2009. RNA and disease. Cell 136, 777–793.
Czerwoniec, A., Kasprzak, J.M., Kaminska, K.H., Rother, K., Bujnicki, J.M., 2009. Folds and functions of domains in RNA modification enzymes. In: Grosjean, H. (Ed.), DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution. Landes Bioscience, Austin.
Das, R., Baker, D., 2007. Automated de novo prediction of native-like RNA tertiary structures. Proc. Natl. Acad. Sci. USA 104, 14664–14669.
Das, R., Kudaravalli, M., Jonikas, M., Laederach, A., Fong, R., Schwans, J.P., Baker, D., Piccirilli, J.A., Altman, R.B., Herschlag, D., 2008. Structural inference of native and partially folded RNA by high-throughput contact mapping. Proc. Natl. Acad. Sci. USA 105, 4144–4149.
de Vries, S.J., Melquiond, A.S., Kastritis, P.L., Karaca, E., Bordogna, A., van Dijk, M., Rodrigues, J.P., Bonvin, A.M., 2010. Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions. Proteins 78, 3242–3249.
Ding, F., Sharma, S., Chalasani, P., Demidov, V.V., Broude, N.E., Dokholyan, N.V., 2008. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. RNA 14, 1164–1173.
Dominguez, C., Boelens, R., Bonvin, A.M., 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. 125, 1731–1737.
Dunin-Horkawicz, S., Czerwoniec, A., Gajda, M.J., Feder, M., Grosjean, H., Bujnicki, J.M., 2006. MODOMICS: a database of RNA modification pathways. Nucleic Acids Res. 34, D145–D149.
Fujishima, K., Komasa, M., Kitamura, S., Suzuki, H., Tomita, M., Kanai, A., 2007. Proteome-wide prediction of novel DNA/RNA-binding proteins using amino acid composition and periodicity in the hyperthermophilic archaeon *Pyrococcus furiosus*. DNA Res. 14, 91–102.
Gabb, H.A., Jackson, R.M., Sternberg, M.J., 1997. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J. Mol. Biol. 272, 106–120.
Glisovic, T., Bachorik, J.L., Yong, J., Dreyfuss, G., 2008. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 582, 1977–1986.
Grosjean, H., 2009. Fine-tuning of RNA functions by modification and editing. Springer Verlag, Berlin Heidelberg, New York.
Janin, J., 2010. Protein–protein docking tested in blind predictions: the CAPRI experiment. Mol. Biosyst. 6, 2351–2362.
Jones, S., van Heyningen, P., Berman, H.M., Thornton, J.M., 1999. Protein–DNA interactions: a structural analysis. J. Mol. Biol. 287, 877–896.
Jones, S., Daley, D.T., Luscombe, N.M., Berman, H.M., Thornton, J.M., 2001. Protein–RNA interactions: a structural analysis. Nucleic Acids Res. 29, 943–954.

Jurica, M.S., 2008. Detailed close-ups and the big picture of spliceosomes. Curr. Opin. Struct. Biol. 18, 315–320.
Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., Vakser, I.A., 1992. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. USA 89, 2195–2199.
Ke, A., Doudna, J.A., 2004. Crystallization of RNA and RNA-protein complexes. Methods 34, 408–414.
Kim, O.T., Yura, K., Go, N., 2006. Amino acid residue doublet propensity in the protein–RNA interface and its application to RNA interface prediction. Nucleic Acids Res. 34, 6450–6460.
Kryshtafovych, A., Fidelis, K., 2009. Protein structure prediction and model quality assessment. Drug Discov. Today 14, 386–393.
Kumar, M., Gromiha, M.M., Raghava, G.P., 2007. Prediction of RNA binding sites in a protein using SVM and PSSM profile. Proteins.
Kumar, M., Gromiha, M.M., Raghava, G.P., 2011. SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. J. Mol. Recognit. 24, 303–313.
Kurowski, M.A., Bujnicki, J.M., 2003. GeneSilico protein structure prediction meta-server. Nucleic Acids Res. 31, 3305–3307.
Laing, C., Schlick, T., 2010. Computational approaches to 3D modeling of RNA. J. Phys. Condens. Matter. 22, 283101.
Ledvina, P.S., Yao, N., Choudhary, A., Quiocho, F.A., 1996. Negative electrostatic surface potential of protein sites specific for anionic ligands. Proc. Natl. Acad. Sci. USA 93, 6786–6791.
Leontis, N.B., Westhof, E., 2001. Geometric nomenclature and classification of RNA base pairs. RNA 7, 499–512.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
Lipfert, J., Doniach, S., 2007. Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu. Rev. Biophys. Biomol. Struct. 36, 307–327.
Lukong, K.E., Chang, K.W., Khandjian, E.W., Richard, S., 2008. RNA-binding proteins in human genetic disease. Trends Genet. 24, 416–425.
Lunde, B.M., Moore, C., Varani, G., 2007. RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell. Biol. 8, 479–490.
Ma, X., Guo, J., Wu, J., Liu, H., Yu, J., Xie, J., Sun, X., 2011. Prediction of RNA-binding residues in proteins from primary sequence using an enriched random forest model with a novel hybrid feature. Proteins 79, 1230–1239.
Maetschke, S.R., Yuan, Z., 2009. Exploiting structural and topological information to improve prediction of RNA–protein binding sites. BMC Bioinformatics 10, 341.
Mertens, H.D., Svergun, D.I., 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J. Struct. Biol. 172, 128–141.
Moreira, I.S., Fernandes, P.A., Ramos, M.J., 2010. Protein–protein docking dealing with the unknown. J. Comp. Chem. 31, 317–342.
Mucha, P., Szyk, A., Rekowski, P., Weiss, P.A., Agris, P.F., 2001. Anticodon domain methylated nucleosides of yeast tRNA(Phe) are significant recognition determinants in the binding of a phage display selected peptide. Biochemistry 40, 14191–14199.
Murakami, Y., Spriggs, R.V., Nakamura, H., Jones, S., 2010. PiRaNhA: a server for the computational prediction of RNA-binding residues in protein sequences. Nucleic Acids Res. 38, W412–W416.
Nadassy, K., Wodak, S.J., Janin, J., 1999. Structural features of protein–nucleic acid recognition sites. Biochemistry 38, 1999–2017.
Parisien, M., Major, F., 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452, 51–55.
Peng, C.R., Liu, L., Niu, B., Lv, Y.L., Li, M.J., Yuan, Y.L., Zhu, Y.B., Lu, W.C., Cai, Y.D., 2011. Prediction of RNA-binding proteins by voting systems. J. Biomed. Biotechnol. 2011, 506205.
Perez-Cano, L., Fernandez-Recio, J., 2010. Optimal protein-RNA area, OPRA: a propensity-based method to identify RNA-binding sites on proteins. Proteins 78, 25–35.
Perez-Cano, L., Solernou, A., Pons, C., Fernandez-Recio, J., 2010. Structural prediction of protein–RNA interaction by computational docking with propensity-based statistical potentials. Pac. Symp. Biocomput. 293, 301.
Ritchie, D.W., Kemp, G.J., 2000. Protein docking using spherical polar Fourier correlations. Proteins 39, 178–194.
Rother, K., Rother, M., Boniecki, M., Puton, T., Bujnicki, J.M., 2011a. RNA and protein 3D structure modeling: similarities and differences. J. Mol. Model. 17, 2325–2336.
Rother, M., Rother, K., Puton, T., Bujnicki, J.M., 2011b. ModeRNA: a tool for comparative modeling of RNA 3D structure. Nucleic Acids Res. 39, 4007–4022.
Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J., 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 33, W363–W367.
Scott, L.G., Hennig, M., 2008. RNA structure determination by NMR. Methods Mol. Biol. 452, 29–61.
Setny, P., Zacharias, M., 2011. A coarse-grained force field for Protein–RNA docking. Nucleic Acids Res.
Shao, X., Tian, Y., Wu, L., Wang, Y., Jing, L., Deng, N., 2009. Predicting DNA- and RNA-binding proteins from sequences with kernel methods. J. Theor. Biol. 258, 289–293.
Shazman, S., Mandel-Gutfreund, Y., 2008. Classifying RNA-binding proteins based on electrostatic properties. PLoS Comput. Biol. 4, e1000146.
Shazman, S., Elber, G., Mandel-Gutfreund, Y., 2011. From face to interface recognition: a differential geometric approach to distinguish DNA from RNA binding surfaces. Nucleic Acids Res. 39, 7390–7399.
Shazman, S., Celniker, G., Haber, O., Glaser, F., Mandel-Gutfreund, Y., 2007. Patch Finder Plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces. Nucleic Acids Res. 35, W526–W530.
Soma, A., Ikeuchi, Y., Kanemasa, S., Kobayashi, K., Ogasawara, N., Ote, T., Kato, J., Watanabe, K., Sekine, Y., Suzuki, T., 2003. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. Mol. Cell 12, 689–698.
Terribilini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V., Dobbs, D., 2007. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res. 35, W578–W584.
Tong, J., Jiang, P., Lu, Z.H., 2008. RISP: a web-based server for prediction of RNA-binding sites in proteins. Comput. Methods Programs Biomed. 90, 148–153.
Tsai, H.Y., Masquida, B., Biswas, R., Westhof, E., Gopalan, V., 2003. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. J. Mol. Biol. 325, 661–675.
Tuszynska, I., Bujnicki, J.M., 2011. DARS-RNP and QUASI-RNP: New statistical potentials for protein–RNA docking. BMC Bioinformatics 12, 348.
Vakser, I.A., Kundrotas, P., 2008. Predicting 3D structures of protein–protein complexes. Curr. Pharm. Biotechnol. 9, 57–66.
Wang, L., Brown, S.J., 2006. BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res. 34, W243–W248.
Wang, L., Huang, C., Yang, M.Q., Yang, J.Y., 2010. BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. BMC Syst. Biol. 4 (Suppl 1), S3.
Weeks, K.M., 2010. Advances in RNA structure analysis by chemical probing. Curr. Opin. Struct. Biol. 20, 295–304.
Wichadakul, D., McDermott, J., Samudrala, R., 2009. Prediction and integration of regulatory and protein–protein interactions. Methods Mol. Biol. 541, 101–143.
Yang, S., Parisien, M., Major, F., Roux, B., 2010. RNA structure determination using SAXS data. J. Phys. Chem. B 114, 10039–10048.
Yu, X., Cao, J., Cai, Y., Shi, T., Li, Y., 2005. Predicting rRNA-, RNA-, and DNA-binding proteins from primary structure with support vector machines. J. Theor. Biol. 12, 6.
Zacharias, M., 2003. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci. 12, 1271–1282.
Zacharias, M., 2010. Accounting for conformational changes during protein–protein docking. Curr. Opin. Struct. Biol. 20, 180–186.
Zhao, H., Yang, Y., Zhou, Y., 2011. Structure-based prediction of RNA-binding domains and RNA-binding sites and application to structural genomics targets. Nucleic Acids Res. 39, 3017–3025.
Zheng, S., Robertson, T.A., Varani, G., 2007. A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. FEBS J. 274, 6378–6391.
Zhou, Z.H., 2008. Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Curr. Opin. Struct. Biol. 18, 218–228.