Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Challenges in structural modeling of RNA-protein interactions

Xudong Liu, Yingtian Duan, Xu Hong, Juan Xie and Shiyong Liu

Abstract

In the past few years, the number of RNA-binding proteins (RBP) and RNA-RBP interactions has increased significantly. Here, we review recent developments in the methodology for protein-RNA and protein−protein complex structure modeling with deep learning and co-evolution, as well as discuss the challenges and opportunities for building a reliable approach for protein-RNA complex structure modelling. Protein Data bank (PDB) and Cross-linking immunoprecipitation (CLIP) data could be combined together and used to infer 2D geometry of protein-RNA interactions by deep learning.

Addresses

School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China

Corresponding author: Liu, Shiyong (liushiyong@gmail.com) (Liu S.)

Current Opinion in Structural Biology 2023, **81**:102623

This review comes from a themed issue on **Protein Nucleic Acid Interactions**

Edited by **James Berger** and **Karl-Peter Hopfner**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2023.102623

RNA-binding proteins (RBP) specifically bind RNA in cells, and their RBP-RNA complexes play important roles in post-transcriptional gene regulation to fine-tune gene expression [1]. Recently, hundreds of RBPs and their binding sites on RNAs have been discovered and investigated using high-throughput CLIP techniques [2−4]. In 2014, the Tuschl group analyzed 1542 RBPs in the human proteome [1]. Since then, the number of identified RBPs has been expanded to over 4000 [5]. This raises the question, how many RBPs are in the human proteome? A sequence-based approach RBPPred encoding global protein sequence descriptors has been proposed, which predicted 6657 possible RBPs in human proteome [6]. Each RBP could bind multiple RNA targets. The number of experimentally solved RBP-RNA complex structures is much smaller than the number of known RBP-RNA interaction pairs. There is a gap in our structural knowledge of protein-RNA complexes, which could be filled by computational RNA-protein complex structure modeling.

What is the current state of RNA-protein complex structure modeling?

The major progress made in recent years involving the development of the RNA-protein complex structure prediction algorithms is reviewed below. Additionally, the possible challenges facing RNA-protein complex structure modeling are analyzed and possible solutions are discussed. The current methods for predicting protein-RNA complex structures include free docking [7−14], template-based docking [15,16], deep learning-based [17] and others [18,19], which are summarized in Table 1.

Free docking for RNA-protein complex structure modeling

Free docking is one of the main strategies for computational modeling of protein-RNA complex structures [7−12,14,23,24]. In early pioneer work, researchers directly built protein-RNA docking decoys with protein-ligand docking [25] and protein−protein docking procedures [7,8]. However, systematic analysis has shown that the characteristics of protein−protein interfaces differ significantly from those of protein-RNA interfaces in terms of atomic packing density, propensity for positively charged residues, π-π stacking interactions, and secondary structure states. In 2013, a protein-RNA docking method 3 dRPC was proposed, which includes the FFT-based sampling algorithm RPDOCK and the coarse-grained scoring function DECK-RP. The prediction success rate of 3 dRPC for the top 10 models on the docking test set is 45.5% [9]. If the output of the top 1000 models is considered, the success rate of docking exceeds 70%. This indicates it is necessary to construct an effective scoring function to select final good models in future work. The neural network energy function [26], DeepRank [27], DeepPotential [28] and DeepRank-GNN [29] could help develop promising scoring functions for ranking of near-native docking structures. Recently, a deep-learning-based scoring function DRPScore was proposed for identifying native-

Table 1

The methods of protein-RNA complexes prediction.

Year	Method	Modeling approach	Webserver/Software	Reference
2010	FTDock	An FFT-based protein−protein docking	–	[7]
2011	ATTRACT	A protein−protein docking with a coarse-grained forcefield for protein−RNA interactions	–	[8]
2013	3 dRPC	An FFT-based protein-RNA docking incorporates the features specific to RNA-protein interfaces	http://biophy.hust.edu.cn/new/resources/3dRPC	[9]
2014	RosettaDock	A genetic algorithm-based strategy with a RosettaDock scoring scheme	http://albios.saclay.inria.fr/rosettadockrna	[10]
2015	NPDock	GRAMM for global macromolecular docking, scoring with a statistical potential	https://genesilico.pl/NPDock/	[11]
2016	PRIME	Template-based modeling	http://rnabinding.com/PRIME.html	[15]
2016	ZDOCK	Integration of physicochemical information about RNA into ZDOCK	–	[12]
2016	HADDOCK	Integrate information derived from biochemical, biophysical or bioinformatics methods to enhance sampling, scoring or both	https://wenmr.science.uu.nl/haddock2.4/	[13]
2017	HDOCK	Hybrid docking algorithm of template-based modeling (Sequence search) and free docking	http://hdock.phys.hust.edu.cn/	[20]
2018	ICM	An FFT-based docking algorithm implemented in the ICM package	–	[14]
2019	RNP-denovo	A Rosetta method with fold-and-dock RNA to protein surface	https://www.rosettacommons.org/	[18]
2019	PRIME2.0	Template-based modeling with RMalign	http://rnabinding.com/RMalign/PRIME2.0.html	[21]
2019	RnaX	A method based on RNA-protein fragment pairs, as well as being integrated into the ModelX tool suite.	https://modelx.crg.eu/	[19]
2020	P3DOCK	A template-based approach PRIME (version 2.1) and an FFT-based docking algorithm 3 dRPC	http://rnabinding.com/P3DOCK/P3DOCK.html	[22]
2022	US-align	Template-based modeling with US-align	https://zhanggroup.org/US-align/	[16]
2022	RoseTTAFoldNA	An end-to-end deep learning approach of modeling protein-nucleic acid complexes.	https://github.com/uw-ipd/RoseTTAFold2NA	[17]

like RNA-protein structures [30]. The DRPScore was trained using a four-dimensional convolutional neural network (4DCNN) on a protein-RNA docking decoys generated by 3 dRPC. The success rate of DRPScore is reported as 43.86% if the top 5 predictions are considered [30]. The performance of DRPScore may be limited by the success rate of the rigid-body FFT-based docking algorithm 3 dRPC.

Template-based modeling for RNA-protein complex structure modeling

Compared to free docking, template-based modeling methods have made great progress. They are based on the hypothesis that similar protein sequences may fold into similar three-dimensional structures. This hypothesis has been applied and confirmed using protein−protein complexes, which demonstrated that similar protein structures may bind in similar ways to form protein−protein complexes [31,32]. In protein−protein complexes, the similarity of binding modes is related to the structural similarity of participating proteins [32]. This hypothesis can also be extended to protein-RNA complexes [15]. As with monomeric proteins and protein−protein complex systems, protein-RNA complex systems do have a transition point, which is defined as the point when a protein-RNA complex transitions from a random to a similar binding mode as the sequence or structural similarity of the protein and RNA increases. Based on this principle, a template-based protein-RNA complex structure prediction method PRIME was developed [15]. The top 1 accuracy of PRIME is about 40%, which is much higher than the previously developed protein-RNA free docking algorithm 3 dRPC [9]. PRIME has successfully predicted some systems where current free docking methods failed due to conformational changes upon binding. For example, the unbound target (protein: 1XPI_A, RNA: 2I82_E) has a large conformational change between its free state and the native state (2I82_AE), so the free docking algorithm fails. However, the template-based method PRIME achieved success in a few cases. The RNA alignment algorithm SARA in PRIME has some flaws. Specifically, its scoring function depends on the size of the RNA and is missing potentially good templates in some cases. To improve the ability of finding remote homologous templates for RNA, a novel RNA 3D structure alignment algorithm RMalign was developed. The similarity scoring function

of RMalign, RMscore, is independent of RNA size [21]. The latest version of PRIME, PRIME2.0 [21], replaces SARA with RMalign, which improves the success rate by about 10%. If no template structure is found, free docking methods are complementary with that of PRIME. This resulted in a combined protein-RNA docking server, P3DOCK, being proposed in 2020 [22]. P3DOCK contains the template-based method PRIME and the free docking algorithm 3 dRPC. It is challenging to integrate the models generated by the two different docking algorithms, however it can be solved based on the protein-RNA complex transition points identified earlier. The success rate of P3DOCK for the top 1 prediction is 58% [22]. Recently US-align [16] was proposed, which can be used to align monomer and complex structures of protein, RNA and DNA, as well as perform template-based modeling of protein−protein and protein-RNA/DNA complex structures. The method has only been compared with earlier methods (free docking algorithm 3 dRPC and template-based approach PRIME) and did not include PRIME2.0. It was reported that the success rates of US-align, 3 dRPC and PRIME are 22.55%, 15.49% and 19.82% respectively for 439 RNA-protein complexes [16]. Furthermore, when the first size-independent RNA alignment method RMalign was developed, it reported the success rate of PRIME2.0 [21]. For the top 1 prediction, the success rates of PRIME and PRIME2.0 (with RMalign) were 39% and 51%, respectively. Previous data reported in Nature Method claimed the success rate of US-align was 45.6% higher than 3 dRPC and 13.8% higher than PRIME [16]. These conflicting results show individual experiments can be misleading and further investigation is needed. Interestingly, US-align is derived from an RNA alignment algorithm RNA-align [33], which is very similar to RMalign. US-align was extended into US-align2.0 with a non-sequential alignment feature [34], which shows that it can find more remote RNA−RNA structure pairs.

Co-evolution signal for RNA-protein complex structure modeling

Following the idea of co-evolution of protein-RNA interactions, Marks' team proposed a maximum entropy probability model to infer protein-RNA residue-base contacts and constructed a protein-RNA complex structural model with the help of the free docking program HADDOCK [35]. Limited by the precision of inferring coevolutionary signals from sequences and the lack of coevolution-based docking methods, they have only succeeded in a few cases. In CASP15, the accuracy of the deep learning-based methods for predicting the 3D structure of RNA was not satisfactory, perhaps because of the small number of available RNA structures [36]. The high accuracy structure prediction of AlphaFold2 is partially based on the co-evolutionary analysis of the Big Fantastic Database [37]. It remains a challenge for RNA. Recent studies have focused on building comprehensive sequence databases and effective homologous sequence search tools [37,38].

Deep learning for RNA-protein complex structure modeling

Recently, deep learning and co-evolution features have been widely used to infer residue−residue contacts within single protein structures [39−42], and in protein−protein complex structures [43]. As deep learning has made significant progress in protein structure prediction (AlphaFold2 [40]) and protein−protein complex structure prediction [43], it is plausible that it will also change the way of computational RNA-protein interaction in the near future. AlphaFold is not yet feasible for direct structural modeling of protein-RNA complexes though it has been able to predict protein−protein complex structure with AlphaFold-Multimer [44] or the newer version AlphaFold2 [45]. The success rates of protein−protein complex structure modeling are 63% vs 72% for AlphaFold2 and AlphaFold-Multimer respectively. The mixed results could be better understood following rigorous testing where the redundancy between the test dataset and the training dataset is removed. Recently, an end-to-end deep learning method RoseTTAFoldNA [17] was proposed that can be used to predict nucleic acid and protein-nucleic acid complexes. It was reported that only 27% of 259 monomeric protein/NA complexes were predicted with high confidence [17]. However, the accuracy of protein-RNA complex prediction is lower than the accuracy of protein−protein complex structure prediction by AlphaFold or AlphaFold variations (63%∼72%).

Other methods for RNA-protein complex structure modeling

In addition to the above methods, there are also fold-and-dock [18] and molecular dynamics simulations [46,47] that can be used to predict the structure of protein-RNA complexes. As discussed in a recent review [48], protein docking with conformational changes remains a challenge. FFT-based rigid body docking is limited by the protein or RNA conformational change upon binding. It has been shown that 13.6% of flexible residues belong to the protein-RNA interface [49]. In an attempt to model the large conformational changes of RNA components upon complex formation, an RNP de novo fold-and-dock algorithm was proposed by Das' team [18]. The algorithm consists of two steps: 1) use the FARFAR algorithm to predict the conformation of RNA; 2) use the program RPDock for docking. The average RMSD of the best scoring models was improved from 11.6 Å for 3 dRPC to 6.4 Å for the RNP de novo fold-and-dock approach when testing on 10 systems [18]. This method is able to handle RNA conformational changes, but the overall performance is limited by the accuracy of the RNA structural prediction.

Insights from modeling of a single protein, a single RNA and protein–protein complex structures by deep learning

In single protein and single RNA structure prediction, significant progress has been made in modeling 3D structures using 2D geometry of contact [50–52], distance (AlphaFold [39], AlphaFold2 [40], CopulaNet [41]) and distance and orientation [42,53,54]. The most important findings are introducing co-evolutionary analysis and deep learning algorithms, which outperform those approaches with classical physically based or knowledge-based models [55]. Direct coupling analysis of multiple sequence alignment (MSA) can identify direct residue–residue contacts in protein–protein interactions from a large number of homologous protein sequences [56]. Recently, CDPred [57] was combined with ResNet and an attention-based mechanism using co-evolutionary features generated by MSA transformer [58] and CCMpred [59] as inputs to predict the inter-chain distance map of both homodimers and heterodimers. In addition to using MSA-derived co-evolutionary features as input, the attention-based transformer predicts the 3D structure of the protein directly using MSA as input [40,42]. The outer product in co-evolution aggregator of CopulaNet [41], has been shown to be effective and applied to extract information on residue–residue pairwise interactions from MSA features of protein–protein complexes [42,43]. It has been shown that construction of paired MSAs based on species within different taxonomic ranks can improve the success rate of protein–protein interaction prediction by AlphaFold2 [60]. More recently, DeepFoldRNA [54] and trRosettaRNA [53] in combination with attention-based transformer have been proposed to predict the 3D structure of RNA. Both methods use RNA sequences or MSA of RNA as input to predict distance and orientation distributions as constraints for 3D structure prediction.

2D geometry-assisted modeling of protein-RNA complex structures

Statistical approaches have been used to capture residue co-evolution in protein-RNA complexes [35,61,62]. Residue evolutionary coupling has been proposed to predict residue-base contacts between proteins and RNAs, and then the contacts are used as distance restraints to predict the structure of protein-RNA complexes [35]. Hayashida et al. [62] proposed a mutual information model to predict residue-base interactions between proteins and RNAs. The AUC (Area under ROC Curve) of the model is only 0.59 [61], which is further improved to 0.69 by a novel CRF-based model with pseudolikelihood maximization direct-coupling analysis (plmDCA) [63]. In contrast to the protein–protein complex structure modeling field, few studies have used 2D geometry to model the 3D structure of protein-RNA complexes. Therefore, the development of protein-RNA complex structure modeling algorithms based on deep learning and 2D geometry is still in its infancy. One of the possible reasons for this is that deep learning approaches require lots of protein-RNA complex structures for accurate model training [55]. Due to the limited size of protein-RNA complex structures in PDB, the most promising protein-RNA complex structure modeling approaches may combine classic docking approaches with 2D geometry constrains from deep learning as shown in Figure 1.

Applications to RBP related biological problems

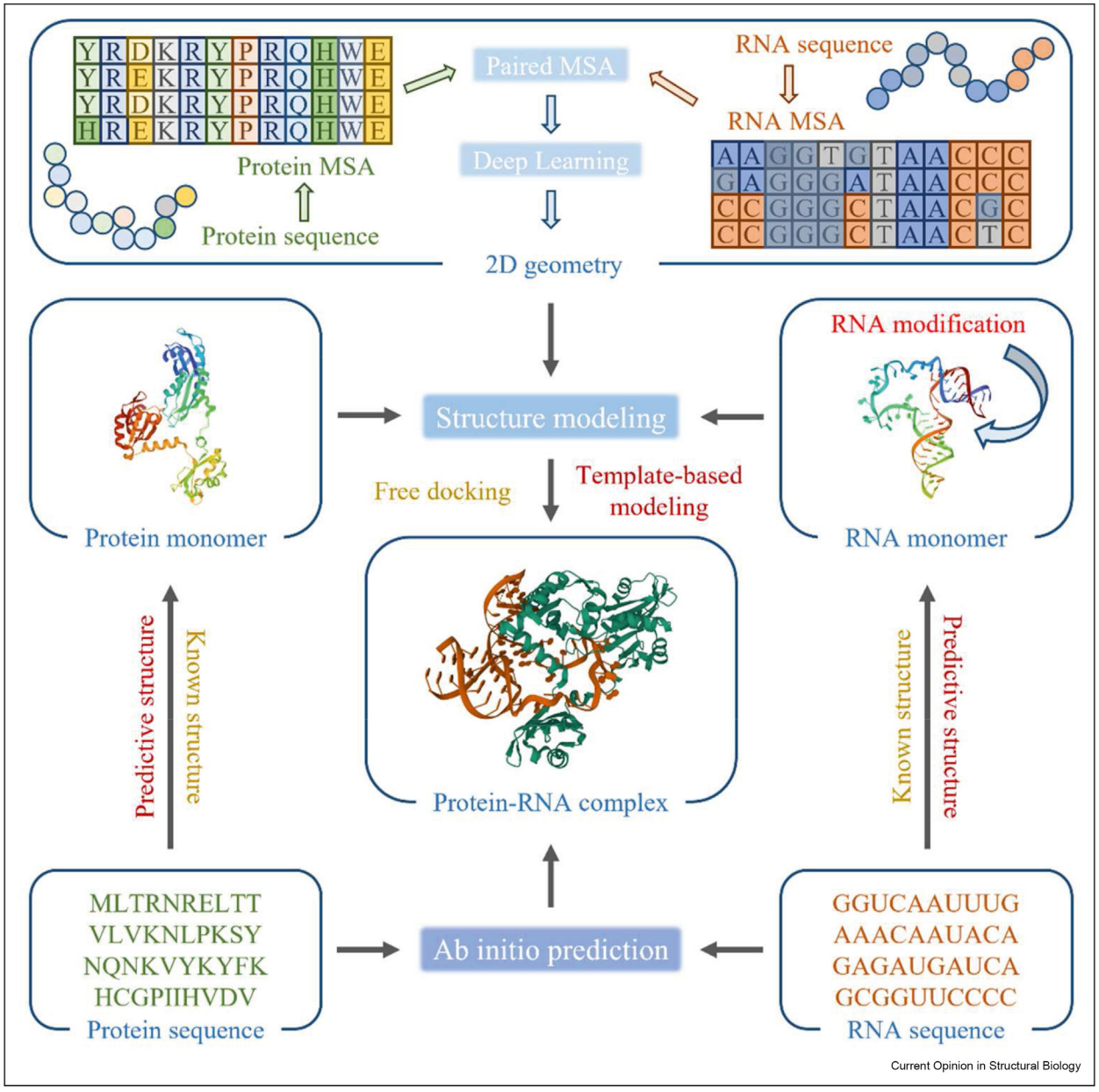
After developing computational tools for predicting RNA-protein interactions, it is important to apply these tools to solve RBP related biological problems. For example, protein-RNA docking was used to build a complex structure model showing that EDAL can specifically bind to EZH2 [64]. CLIP-seq is a high-throughput technique to map protein-RNA interactions transcriptome-wide in vivo, which produces lots of RNA-protein binding sites. Taking full advantage of the protein-RNA binding site data and the RNA secondary structure generated by high-throughput sequencing, a template-based method PRIME-3D2D was proposed to predict binding sites for protein-RNA interactions on a yeast genome [65]. PrismNet [66] uses experimentally derived in vivo RNA structure data and RBP binding data to predict protein-RNA interactions. Several works illustrate that the structure of RNA plays an important role in RBP-RNA recognition [15,65–67]. A large amount of protein-RNA interaction data, in vivo RNA secondary structure data [68] and known protein structures in PDB are available. High precision protein monomer structure prediction methods provide a great deal of predicted protein structures [69]. The data discussed above may compensate for the lack of PDB data and lead to the emergence of new findings.

Open challenges

RNA modifications are important and missing in protein-RNA structure modeling algorithms

A large number of RBPs were found to interact with modified RNAs and a few of the structures of the RBP-modified RNA complexes have been solved [70]. These structures reveal a new post-transcriptional regulatory mechanism. m^6A modification widely exists in mRNA and non-coding RNA, and many modification sites are evolutionarily conserved between humans and mice [71]. m^6A modification contributes to RBP-RNA binding, acting as structural switches [72] and regulating the expression of nearly 2000 genes [73]. So far, more than 170 RNA modifications have been found, and Chen et al. proposed that each modification may have a reader, writer and eraser system similar to the m^6A modification [74]. Other RNA modifications such as, m^6Am modifications, as well

Figure 1



Protein-RNA complex structure modelling approach. Predicting the structure of protein-RNA complexes directly from the sequences, i.e., ab initio prediction. Otherwise, protein-RNA complex structures are constructed by free docking and template-based modelling using protein and RNA monomer structures known in the PDB or predicted by computational methods, and RNA modifications as well as 2D geometry based on co-evolutionary information are added in order to assist in complex modelling.

as the readers, writers and erasers for other modifications are still unknown [74]. RNA modifications in protein-RNA complexes were ignored in a previous protein-RNA docking study by the Kameda group [12]. Their study integrated physicochemical information about RNA into ZDOCK. Only 20 standard amino acids and 4 nucleotide types were used in their work due to the deletion of modified residues in three widely used protein-RNA docking benchmark datasets [75–77]. When one of the protein-RNA docking benchmark datasets was constructed, they found some modified residues in the unbound and bound RNA structures through the RNA modification database MODOMICS [77,78]. They discussed that these modified residues may be important for protein-RNA docking, but unfortunately these modified residues were removed when the dataset was

constructed [77]. This means that we need to reconstruct the protein-RNA docking benchmark dataset and include the RNA modified residues for method development. In several previous studies, RNA modifications were proven to affect RBP-RNA binding [73,79–81]. This shows that the mechanism of RNA modification in RNA-protein recognition is complex. Which RNA modifications will affect the RBP-RNA interaction? Which RNA modifications play an important role in RBP-RNA free docking? These questions require further investigation. If a protein-modified RNA modeling system can be made, it may help to find these unknown readers, writers and eraser proteins. Although many RBP-modified RNA interactions have been identified, their atomic-scale interaction details are still missing. These details are key to understanding the molecular mechanisms underlying RBP-modified RNA recognition. Therefore, computational methods for structural modeling of RBP-modified RNA complexes are urgently needed to elucidate the atomic details of RBP-modified RNA interactions.

Conclusions

Despite the advances in the field of RBP-RNA complex structure prediction, co-evolutionary signals and modified RNAs have been overlooked in the research process of RBP-RNA complex structure prediction methods. As such, predicting the structure of protein-RNA complexes is still challenging. The question regarding whether or not it is possible to integrate co-evolutionary signals and types of RNA modifications into current protein-RNA docking algorithms remains. Future work includes exploring RNA modifications and co-evolutionary signals for the development of a protein-RNA docking algorithm that could enhance sampling efficiency of protein-RNA docking with PDB and CLIP-based big data.

Author contributions

Conceptualization, SYL; Investigation XDL and SYL; Writing - Original Draft, XDL and SYL; Writing - Review & Editing, XDL, JX, XH, YTD and SYL; Funding Acquisition, SYL; Supervision, SYL.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China [32271267]; This work is also dedicated to the 40th birthday of School of Physics, Huazhong University of Science and Technology and the 125th birthday of Peking University. The authors would like to thank Professor Wei Zhang and Dr. Brandon Bishop at Department of Chemistry, University of Colorado Boulder for improving English text.

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Gerstberger S, Hafner M, Tuschl T: **A census of human RNA-binding proteins**. *Nat Rev Genet* 2014, **15**:829–845.
2. König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J: **iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution**. *Nat Struct Mol Biol* 2010, **17**:909–915.
3. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, Blue SM, Nguyen TB, Surka C, Elkins K: **Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)**. *Nat Methods* 2016, **13**:508–514.
4. Ramanathan M, Majzoub K, Rao DS, Neela PH, Zarnegar BJ, Mondal S, Roth JG, Gai H, Kovalski JR, Siprashvili Z: **RNA–protein interaction detection in living cells**. *Nat Methods* 2018, **15**:207–212.
5. Guo JK, Guttman M: **Regulatory non-coding RNAs: everything is possible, but what is important?** *Nat Methods* 2022, **19**:1156–1159.
6. Zhang X, Liu S: **RBPPred: predicting RNA-binding proteins from sequence using SVM**. *Bioinformatics* 2017, **33**:854–862.
7. 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; 2010:293–301.
8. Setny P, Zacharias M: **A coarse-grained force field for Protein–RNA docking**. *Nucleic Acids Res* 2011, **39**:9118–9129.
9. 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**:1–7.
10. Guilhot-Gaudeffroy A, Froidevaux C, Azé J, Bernauer J: **Protein-RNA complexes and efficient automatic docking: expanding RosettaDock possibilities**. *PLoS One* 2014, **9**, e108928.
11. Tuszynska I, Magnus M, Jonak K, Dawson W, Bujnicki JM: **NPDock: a web server for protein–nucleic acid docking**. *Nucleic Acids Res* 2015, **43**:W425–W430.
12. 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 Theor Comput* 2016, **12**:4688–4697.
13. van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, Bonvin A: **The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes**. *J Mol Biol* 2016, **428**:720–725.
14. Arnautova YA, Abagyan R, Totrov M: **Protein-RNA docking using ICM**. *J Chem Theor Comput* 2018, **14**:4971–4984.
15. Zheng J, Kundrotas PJ, Vakser IA, Liu S: **Template-based modeling of protein-RNA interactions**. *PLoS Comput Biol* 2016, **12**, e1005120.
16. Zhang C, Shine M, Pyle AM, Zhang Y: **US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes**. *Nat Methods* 2022, **19**:1109–1115.
17. •• Baek M, McHugh R, Anishchenko I, Baker D, DiMaio F: *Accurate prediction of nucleic acid and protein-nucleic acid complexes using RoseTTAFoldNA*. bioRxiv; 2022.

This article presents the first end-to-end protein-RNA complex structure prediction method RoseTTAFoldNA to obtain 3D structures from only sequences of proteins and RNAs.

18. Kappel K, Das R: **Sampling native-like structures of RNA-protein complexes through Rosetta folding and docking**. *Structure* 2019, **27**:140–151. e145.

19. Delgado Blanco J, Radusky LG, Cianferoni D, Serrano L: **Protein-assisted RNA fragment docking (RnaX) for modeling RNA-protein interactions using ModelX**. In *Proceedings of the national academy of sciences of the United States of America*, **116**; 2019:24568–24573.

20. Li H, Huang E, Zhang Y, Huang SY, Xiao Y: **HDOCK update for modeling protein-RNA/DNA complex structures**. *Protein Sci* 2022, **31**:e4441.

21. Zheng J, Xie J, Hong X, Liu S: **RMalign: an RNA structural alignment tool based on a novel scoring function RMscore**. *BMC Genom* 2019, **20**:1–10.

22. Zheng J, Hong X, Xie J, Tong X, Liu S: **P3DOCK: a protein–RNA docking webserver based on template-based and template-free docking**. *Bioinformatics* 2020, **36**:96–103.

23. 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.

24. Zhang Z, Lu L, Zhang Y, Hua Li C, Wang CX, Zhang XY, Tan JJ: **A combinatorial scoring function for protein-RNA docking**. *Proteins* 2017, **85**:741–752.

25. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM: **Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit**. In *Proceedings of the national academy of sciences of the United States of America*, **100**; 2003: 12648–12653.

26. Huang B, Xu Y, Hu X, Liu Y, Liao S, Zhang J, Huang C, Hong J, Chen Q, Liu H: **A backbone-centred energy function of neural networks for protein design**. *Nature* 2022, **602**:523–528.

27. Renaud N, Geng C, Georgievska S, Ambrosetti F, Ridder L, Marzella DF, Reau MF, Bonvin A, Xue LC: **DeepRank: a deep learning framework for data mining 3D protein-protein interfaces**. *Nat Commun* 2021, **12**:7068.

28. Pearce R, Li Y, Omenn GS, Zhang Y: **Fast and accurate Ab Initio Protein structure prediction using deep learning potentials**. *PLoS Comput Biol* 2022, **18**, e1010539.

29. Reau M, Renaud N, Xue LC, Bonvin A, DeepRank-Gnn: **A graph neural network framework to learn patterns in protein-protein interfaces**. *Bioinformatics* 2023:39.

30. Zeng C, Jian Y, Vosoughi S, Zeng C, Zhao Y: **Evaluating native-like structures of RNA-protein complexes through the deep learning method**. *Nat Commun* 2023, **14**:1060.

31. Vreven T, Hwang H, Pierce BG, Weng Z: **Evaluating template-based and template-free protein-protein complex structure prediction**. *Briefings Bioinf* 2014, **15**:169–176.

32. Kundrotas PJ, Zhu Z, Janin J, Vakser IA: **Templates are available to model nearly all complexes of structurally characterized proteins**. In *Proceedings of the national academy of sciences of the United States of America*, **109**; 2012:9438–9441.

33. Gong S, Zhang C, Zhang Y: **RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA**. *Bioinformatics* 2019, **35**:4459–4461.

34. Zhang C, Pyle AM: **A unified approach to sequential and non-sequential structure alignment of proteins, RNAs, and DNAs**. *iScience* 2022, **25**, 105218.

35. Weinreb C, Riesselman AJ, Ingraham JB, Gross T, Sander C, Marks DS: **3D RNA and functional interactions from evolutionary couplings**. *Cell* 2016, **165**:963–975.

36. Das R, Kretsch RC, Simpkin A, Mulvaney T, Pham P, Rangan R, Bu F, Keegan R, Topf M, Rigden D, *et al.*: **Assessment of three-dimensional RNA structure prediction in CASP15**. *bioRxiv* 2023, https://doi.org/10.1101/2023.04.25.538330: 2023.2004.2025.538330.

37. Chen K, Litfin T, Singh J, Zhan J, Zhou Y: **The master database of all possible RNA sequences and its integration with RNAcmap for RNA homology search**. *bioRxiv* 2023, https://doi.org/10.1101/2023.02.01.526559: 2023.2002.2001.526559.

38. Zhang C, Zhang Y, Pyle AM, rMSA: **A sequence search and alignment algorithm to improve RNA structure modeling**. *J Mol Biol* 2022, 167904.

39. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Zidek A, Nelson AWR, Bridgland A, *et al.*: **Improved protein structure prediction using potentials from deep learning**. *Nature* 2020, **577**:706–710.

40. •• Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.
This paper describes in detail the first high-precision end-to-end protein 3D structure prediction method, AlphaFold2.

41. • Ju F, Zhu J, Shao B, Kong L, Liu T-Y, Zheng W-M, Bu D: **CopulaNet: learning residue co-evolution directly from multiple sequence alignment for protein structure prediction**. *Nat Commun* 2021, **12**:1–9.
This paper presents a new outer product and aggregation idea of predicting the distance between protein residues directly from MSA, which is applied to modeling the 3D structures of proteins and protein–protein complexes.

42. •• Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD: **Accurate prediction of protein structures and interactions using a three-track neural network**. *Science* 2021, **373**:871–876.
This paper presents RoseTTAFold, an end-to-end protein 3D structure prediction method using SE(3) transformer, which is inspired by AlphaFold2.

43. • Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, Zhang J, Ness TJ, Banjade S, Bagde SR: **Computed structures of core eukaryotic protein complexes**. *Science* 2021, **374**, eabm4805.
This paper describes a lighter-weight RoseTTAFold two-track model for protein–protein complex structure prediction.

44. • Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J: **Protein complex prediction with AlphaFold-Multimer**. *bioRxiv* 2022, **2004**, 463034. 2021.2010.
This paper introduces AlphaFold-Multimer, a protein–protein complex structure prediction method extended from AlphaFold2.

45. Bryant P, Pozzati G, Elofsson A: **Improved prediction of protein-protein interactions using AlphaFold2**. *Nat Commun* 2022, **13**:1265.

46. Kim H, Abeysirigunawarden SC, Chen K, Mayerle M, Ragunathan K, Luthey-Schulten Z, Ha T, Woodson SA: **Protein-guided RNA dynamics during early ribosome assembly**. *Nature* 2014, **506**:334–338.

47. Bahadur RP, Kannan S, Zacharias M: **Binding of the bacteriophage P22 N-peptide to the boxB RNA motif studied by molecular dynamics simulations**. *Biophys J* 2009, **97**: 3139–3149.

48. Vakser IA: **Challenges in protein docking**. *Curr Opin Struct Biol* 2020, **64**:160–165.

49. Sankar K, Walia RR, Mann CM, Jernigan RL, Honavar VG, Dobbs D: **An analysis of conformational changes upon RNA-protein binding**. In *Proceedings of the 5th ACM conference on bioinformatics*. Computational Biology, and Health Informatics; 2014:592–593.

50. Adhikari B, Cheng J: **CONFOLD2: improved contact-driven ab initio protein structure modeling**. *BMC Bioinf* 2018, **19**:1–5.

51. Sun S, Wang W, Peng Z, Yang J: **RNA inter-nucleotide 3D closeness prediction by deep residual neural networks**. *Bioinformatics* 2021, **37**:1093–1098.

52. Wang S, Sun S, Li Z, Zhang R, Xu J: **Accurate de novo prediction of protein contact map by ultra-deep learning model**. *PLoS Comput Biol* 2017, **13**, e1005324.

53. Feng C, Wang W, Han R, Wang Z, Ye L, Du Z, Wei H, Zhang F, Peng Z, Yang J: **Accurate de novo prediction of RNA 3D structure with transformer network**. *bioRxiv* 2022.

54. Pearce R, Omenn GS, Zhang Y: **De novo RNA tertiary structure prediction at atomic resolution using geometric potentials from deep learning**. *bioRxiv* 2022.

55. Baek M, Baker D: **Deep learning and protein structure modeling**. *Nat Methods* 2022, **19**:13–14.

56. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T: **Identification of direct residue contacts in protein-protein interaction by message passing**. In *Proceedings of the national academy of sciences of the United States of America*, **106**; 2009:67–72.

57. • Guo Z, Liu J, Skolnick J, Cheng J: **Prediction of inter-chain distance maps of protein complexes with 2D attention-based deep neural networks**. *Nat Commun* 2022, **13**:1–10.

This article presents CDPred, a method for predicting inter-chain distance map of protein–protein complex, which combines residual networks and attention mechanism and takes as input co-evolutionary features generated by direct coupling analysis and MSA transformers.

58. Rao RM, Liu J, Verkuil R, Meier J, Canny J, Abbeel P, Sercu T, Rives A: **MSA transformer**. In *Proceedings of the 38th international conference on machine learning*. Edited by Marina M, Tong Z, *Proceedings of machine learning research: PMLR*, vol. 139; 2021:8844–8856.

59. Seemayer S, Gruber M, Söding J: **CCMpred—fast and precise prediction of protein residue–residue contacts from correlated mutations**. *Bioinformatics* 2014, **30**:3128–3130.

60. Si Y, Yan C: **Protein complex structure prediction powered by multiple sequence alignments of interologs from multiple taxonomic ranks and AlphaFold2**. *Briefings Bioinf* 2022:23.

61. Hayashida M, Kamada M, Song J, Akutsu T: **Prediction of protein-RNA residue-base contacts using two-dimensional conditional random field with the lasso**. *BMC Syst Biol* 2013, **7**: 1–11.

62. Hayashida M, Okada N, Kamada M, Koyano H: **Improving conditional random field model for prediction of protein-RNA residue-base contacts**. *Quantitative Biology* 2018, **6**:155–162.

63. Ekeberg M, Lovkvist C, Lan Y, Weigt M, Aurell E: **Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models**. *Phys Rev E - Stat Nonlinear Soft Matter Phys* 2013, **87**, 012707.

64. Sui B, Chen D, Liu W, Wu Q, Tian B, Li Y, Hou J, Liu S, Xie J, Jiang H, *et al.*: **A novel antiviral lncRNA, EDAL, shields a T309 O-GlcNAcylation site to promote EZH2 lysosomal degradation**. *Genome Biol* 2020, **21**:228.

65. Xie J, Zheng J, Hong X, Tong X, Liu S: **PRIME-3D2D is a 3D2D model to predict binding sites of protein-RNA interaction**. *Commun Biol* 2020, **3**:384.

66. • Sun L, Xu K, Huang W, Yang YT, Li P, Tang L, Xiong T, Zhang QC: **Predicting dynamic cellular protein-RNA interactions by deep learning using in vivo RNA structures**. *Cell Res* 2021, **31**:495–516.

This article describes PrismNet, which uses experimental in vivo RNA structure data and RBP binding data to predict protein-RNA interactions, and applies it to predict how genomic variation affects RBP binding.

67. Wei J, Chen S, Zong L, Gao X, Li Y: **Protein-RNA interaction prediction with deep learning: structure matters**. *Briefings Bioinf* 2022:23.

68. Zhang J, Fei Y, Sun L, Zhang QC: **Advances and opportunities in RNA structure experimental determination and computational modeling**. *Nat Methods* 2022, **19**:1193–1207.

69. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, *et al.*: **AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models**. *Nucleic Acids Res* 2022, **50**: D439–D444.

70. Huang J, Yin P: **Structural insights into N(6)-methyladenosine (m(6)A) modification in the transcriptome**. *Dev Reprod Biol* 2018, **16**:85–98.

71. Huang H, Weng H, Chen J: **m(6)A modification in coding and non-coding RNAs: roles and therapeutic implications in cancer**. *Cancer Cell* 2020, **37**:270–288.

72. Seo KW, Kleiner RE: **Mechanisms of epitranscriptomic gene regulation**. *Biopolymers* 2021, **112**, e23403.

73. Lewis CJ, Pan T, Kalsotra A: **RNA modifications and structures cooperate to guide RNA-protein interactions**. *Nat Rev Mol Cell Biol* 2017, **18**:202–210.

74. Huang H, Weng H, Deng X, Chen J: **RNA modifications in cancer: functions, mechanisms, and therapeutic implications**. *Annu Rev Cell Biol* 2020, **4**:221–240.

75. Huang SY, Zou X: **A nonredundant structure dataset for benchmarking protein-RNA computational docking**. *J Comput Chem* 2013, **34**:311–318.

76. Barik A, Cn, Pm, Bahadur RP: **A protein-RNA docking benchmark (I): nonredundant cases**. *Proteins* 2012, **80**:1866–1871.

77. Perez-Cano L, Jimenez-Garcia B, Fernandez-Recio J: **A protein-RNA docking benchmark (II): extended set from experimental and homology modeling data**. *Proteins* 2012, **80**:1872–1882.

78. Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, de Crecy-Lagard V, Ross R, Limbach PA, Kotter A, *et al.*: **MODOMICS: a database of RNA modification pathways. 2017 update**. *Nucleic Acids Res* 2018, **46**:D303–D307.

79. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T: **N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions**. *Nature* 2015, **518**:560–564.

80. Sun L, Fazal FM, Li P, Broughton JP, Lee B, Tang L, Huang W, Kool ET, Chang HY, Zhang QC: **RNA structure maps across mammalian cellular compartments**. *Nat Struct Mol Biol* 2019, **26**:322–330.

81. Vaidyanathan PP, AlSadhan I, Merriman DK, Al-Hashimi HM, Herschlag D: **Pseudouridine and N(6)-methyladenosine modifications weaken PUF protein/RNA interactions**. *RNA* 2017, **23**:611–618.