Protein-assisted RNA fragment docking (RnaX) for modeling RNA–protein interactions using ModelX

Javier Delgado Blanco[a,1], Leandro G. Radusky[a,1], Damiano Cianferoni[a], and Luis Serrano[a,b,c,2]

[a]Center for Genomic Regulation, Barcelona Institute for Science and Technology, 08003 Barcelona, Spain; [b]Universitat Pompeu Fabra, 08002 Barcelona, Spain; and [c]Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain

RNA–protein interactions are crucial for such key biological processes as regulation of transcription, splicing, translation, and gene silencing, among many others. Knowing where an RNA molecule interacts with a target protein and/or engineering an RNA molecule to specifically bind to a protein could allow for rational interference with these cellular processes and the design of novel therapies. Here we present a robust RNA–protein fragment pair-based method, termed RnaX, to predict RNA-binding sites. This methodology, which is integrated into the ModelX tool suite (http://modelx.crg.es), takes advantage of the structural information present in all released RNA–protein complexes. This information is used to create an exhaustive database for docking and a statistical forcefield for fast discrimination of true backbone-compatible interactions. RnaX, together with the protein design forcefield FoldX, enables us to predict RNA–protein interfaces and, when sufficient crystallographic information is available, to reengineer the interface at the sequence-specificity level by mimicking those conformational changes that occur on protein and RNA mutagenesis. These results, obtained at just a fraction of the computational cost of methods that simulate conformational dynamics, open up perspectives for the engineering of RNA–protein interfaces.

FoldX | ModelX | RNA docking | protein–RNA interface design | protein–RNA interface engineering

RNA-binding proteins (RBPs) play a fundamental role in many cellular processes, including DNA transcription, reverse transcription, replication, pre-mRNA splicing, regulation of intracellular RNA concentration, and mRNA stability and translation. RBPs not only influence each of these cellular processes, but also provide links among them (1–4). The rational engineering of RNA–protein interfaces (RPIs) has many biotechnological and medical applications (5, 6). RBPs engineered to recognize specific RNA sequences (7) could become valuable tools for manipulating posttranscriptional regulatory networks and developing therapeutic agents to treat genetic and infectious diseases.

The functional diversity of RBPs suggests a corresponding diversity of structures responsible for RNA recognition; it is estimated that the universe of RBPs composes between 6% and 8% of the human proteome (8). However, most of the RBPs that have been studied so far are composed of a relatively low number of RNA-binding modules (9). Thus, the large diversity of RNA targets is managed by the presence of multiple copies of these RNA-binding domains in different combinations, thereby expanding the functional repertoire of RBPs. Is important to note that intrinsically disordered regions play key roles in RNA-binding proteins, with an estimated ~30% of proteins marked as hubs (the top 20% of the interacting proteins of an interactome) containing disordered motifs (10). The presented method in combination with FoldX could predict the docking of protein-disordered regions as long as they have been observed in other proteins. The main limitation is the conformational diversity of disordered peptides interacting with RNA. Despite efforts by the structural community, the structural landscape of RNA-binding modules is not fully covered in the Protein Data Bank (PDB)

(11); only 275 X-ray *Homo sapiens* RNA–protein complexes (RPCs) have been deposited to date.

Over the past decade, different in silico approaches have tried to address the challenging tasks of achieving RNA-binding site recognition, RNA docking, and RPI design. Direct readout methods are based mainly on machine learning algorithms and include BindN (12), BindN+ (13), PPRInt (14), and PiRaNhA (15), all of which use support vector machines. On the other hand, NAPS (16), RNABindR (17), PRBR (18), and catRapid (19) use decision trees, the naïve Bayes classifier, and random forest algorithms, respectively, incorporating the physicochemical properties of amino acids along with predictable features, such as secondary structure, sequence conservation, and solvent accessibility. In contrast, indirect readout approaches were initially adapted from existing protein–protein docking algorithms. However, these programs often fail to generate native-like structures due to incomplete sampling of the conformational space and deficiencies of the scoring functions, which are not specifically designed for RPIs (20). Nonetheless, docking algorithms, such as HADDOCK (21), RosettaDock server (22), ClusPro (23), GRAMM-X (24), 3D-Garden (25), HEX server (26), SwarmDock (27), ZDOCK server (28), PatchDock (29), ATTRACT (30), pyDockSAXS (31), InterEvDock (32), NPDock (33), and HDOCK (34), with the exception of NPDock, have been adapted to accept nucleic acids. Rigid solid approaches face

Significance

Protein–RNA interactions, key in biological processes, remained refractory to prediction algorithms. Here we present a new extension of the ModelX tool suite designed for this purpose. RNA–protein complexes in the Protein Data Bank were decomposed into small peptide–oligonucleotide interacting fragment pairs and used as building blocks to assemble big scaffolds representing complex RNA–protein interactions. This method has already been successful for designing DNA–protein and protein–protein interfaces. Areas under the curve up to 0.86 were achieved on binding site prediction showing the accuracy and coverage of our approach over established and in-house benchmarking sets. Together with FoldX protein design tool suite we were able to engineer backbone- and side chain-compatible interfaces using naked protein structures as input.

problems derived from the intrinsic flexibility of the RNA and the electrostatic nature of forces driving the RPIs (35). To account for these properties, some of the methods taken from protein-protein docking have been updated by adding electrostatic terms to the energy function (36). Unfortunately, however, RPIs are still difficult to predict and model with these methods (37).

Previously we showed that by decomposing DNA-protein complexes into pairs of interacting DNA-protein fragments, we could successfully predict DNA-protein interacting interfaces, in addition to the DNA sequence recognized by the protein (38). The assumption behind this method is that the conformational landscape of interacting pairs of DNA-protein small fragments is large but limited, and as such, it should be possible to extract this landscape from the existing structures (39). In this work, we used a similar approach to predict RPIs (Fig. 1 and *SI Appendix*, Fig. *S2B* and *SI Appendix*). Predicting RNA–protein interactions is a bigger challenge than predicting DNA–protein interactions due to the higher intrinsic flexibility of RNA, as well as the smaller number of available RPCs (~3,000 vs. ~5,100 for DNA-protein complexes). Thus, we built a library composed of protein fragments (pepX) and nucleotide fragments (rnaX), plus the spatial relationships between them (intX). These libraries were then integrated into the ModelX software tool (http://modelx.crg.es). Together with the built database of fragments and interactions, we developed a statistical force field (40, 41) derived from the observed distances between RNA and protein atoms. The developed algorithm, called RnaX, was benchmarked against a published nonredundant dataset of RPCs achieving a maximum Receiver Operating Characteristic (ROC) curve accuracy of ~0.86. We also tested the method with a dataset of newly (2018) deposited RPCs to test its discovery capabilities, achieving a coverage closer to 80% of binding sites. We also developed an algorithm to join docked fragments (GlueDocks) into longer strands, which can be used to remove spurious docks. A comparison of our docking algorithm performance with some of the previously

mentioned methods is provided in *SI Appendix*, Table S4. As some of the methods are web servers, no massive runs can be executed on them; thus, we just provide the benchmarking values published in each paper. The methods not included in the table provided only protein-protein performance values.

Finally, we combined the recently published version of the FoldX force field (42), which accounts for RNA molecules in its energy calculations, by performing RNA mutagenesis over the RNA strands docked by our RnaX docking module of ModelX, showing how they can be used together to reengineer a sequence-specific interface. FoldX and ModelX combined enable predictions of RPIs at sequence-specificity level. Over time, the power of this approach will increase as the number of deposited RPCs also increases.

Materials and Methods

The main command incorporated into the ModelX tool suite, RnaDocking, scans an input protein structure with a sliding window of peptide iterations (with peptide length given by the frag-length parameter), looking for pepX fragments that overlap well with the protein window being scanned. All of the pepX fragments stored in the built database interact with an rnaX fragment, constituting an interacting pair, or intX (Fig. 1). Once a pepX fragment is selected, the corresponding rnaX fragment is placed to interact with the input protein. Compatible pepXs are retrieved by selecting those with geometrically compatible backbones (parameter fit-threshold) and with a sequence similarity given by the pep-mismatches parameter. The length of the retrieved rnaXs is given by the dock-length parameter. A docked rnaX fragment is then evaluated using the developed statistical force field (*SI Appendix*) according to an energy-threshold parameter. The statistical force field thus generated has a biased version that considers the atomic backbone distance distributions of the specific residues and bases in contact with the evaluated fragment, as well as an unbiased version that considers the joint atomic distance distributions without any sequence discrimination (biased parameter). Also, in the atom-type parameter, the atoms to be considered for the energy calculations can be set to BB (only backbone), SC (only side chain), or ALL. The propensity of residues to bind RNA (*SI Appendix*, Fig. S2A) showed a similar profile as previously published DNA propensities, while atomic distances (*SI Appendix*, Fig. S2B) showed similar
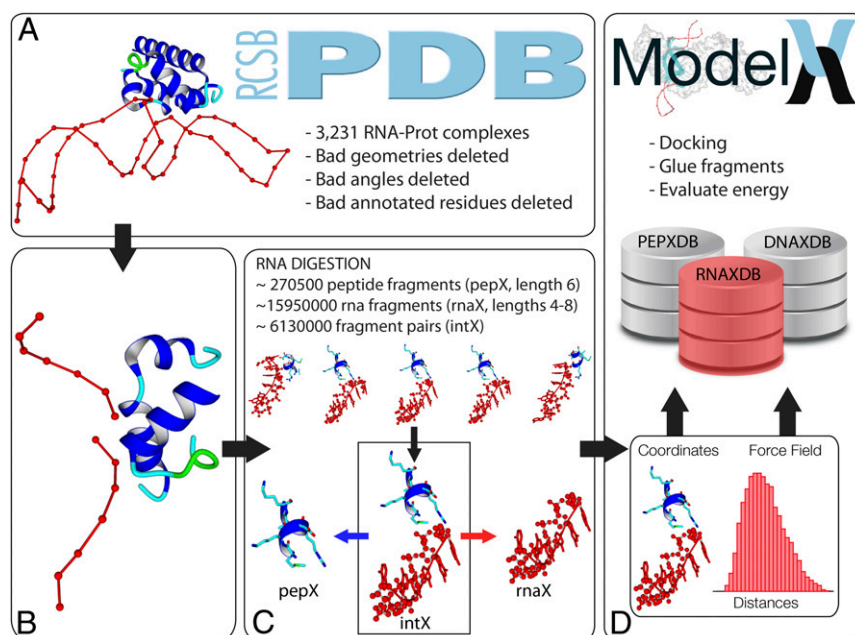

Fig. 1. Database and force field genesis. (*A*) Good-quality RNA–protein complexes are taken from the PDB. (*B*) Detection of RPIs from the complexes. (*C*) Digestion of RPIs into RNA-peptide fragment pairs with peptide fragments of length = 6 and RNA fragments ranging from 4 to 8 (dock length parameter in the RNADocking command). (*D*) Atomic coordinates, peptidic dihedral angles, and distance measurements are stored in RNAXDB, along with the statistical force field generated from these distances. The docking algorithm scans an input protein in a peptide sliding-window fashion, retrieving compatible fragments stored in RNAXDB. Stored RNA fragments interacting with these compatible fragments are placed in the original structure and then evaluated with the statistical force field.

peaks but slightly wider bell distributions. Another command, GlueDocks, can be used to postprocess the docked fragments by joining them into longer RNA strands when they are geometrically compatible. Extensions are accepted according to a min-length parameter, with 2 elongated fragments considered nonredundant extensions when they have overlapping nucleotides up to a defined rmsd-threshold parameter. Details of the database organization, algorithms, and new commands of the ModelX toolsuite are available in *SI Appendix, Materials and Methods*. A step-by-step explanation discussing the applicability of each ModelX and FoldX command used during the modeling process of the cases presented below is provided in the online *SI Running Tutorial*.

Results

RnaDocking Benchmarking. We validated the RNA docking accuracy of ModelX using 2 different benchmarking sets. The first set is a nonredundant dataset of 126 RPCs (43) (*SI Appendix*, Table S1) comprising the archetypal RNA–protein interaction landscape. The second dataset contains 70 X-ray RPCs released in 2018 (*SI Appendix*, Table S2) and is used to show how our method handles novel structures as they are released. Validation experiments were performed using a wide range of combinations of the parameters, including pep-mismatches parameter (sequence mismatches of the superimposed pepX fragments) values of 1 and 2, a frag-length parameter (length of the protein window) values of 4 and 6, and a fixed dock-length (length of the RNA fragment range) of 4 (*Materials and Methods*). Energy evaluation inside the RnaDocking command was carried out using only the protein and RNA backbone atoms and the unbiased force field (*Materials and Methods*), because at this point the aim is to search for RNA docking sites in general, not for a specific RNA sequence. The presented ROC curves (Fig. 2A)

were generated by labeling each amino acid in the crystallized protein as follows: true positives are those contacting both crystallographic RNA and at least 1 docked fragment, true negatives are those not in contact with crystallographic RNA or with any RNA docked fragment, false positives are those contacting at least 1 docked RNA but no crystallographic RNA, and finally false negatives are those in contact with crystallographic RNA but not with any docked fragment. The best predictions were observed using pep-mismatches = 1 and frag-length = 6, with an area under the curve (AUC) of up to 0.86. AUC decreases when pep-mismatches increases due to an increase in the number of false positives; in Fig. 3C the AUC for 3 mismatches on the Bahadur data set for dock-length = 6 falls to 0.52. However, runs allowing higher pep-mismatches could help the user identify the correct binding interface at the expense of more false positives, as shown in Fig. 2 C and D and explained below.

Negative Benchmarking. To the best of our knowledge, no negative benchmarking datasets for proteins that do not bind RNA have been published to date. This is because no specific criteria have been established to determine whether a protein can form a complex with RNA. In a review from 2018 (44) that took into account 7 different methods (interactome capture, mRBPome, RBDmap, serial RNA interactome capture, photo-cross-linking and high-resolution mass spectrometry, protein microarrays, and proteomic identification of RNA-binding regions) to experimentally determine RPIs in more than 1,400 *Homo sapiens* genes, not a single gene was tagged as unable to interact with RNA in any of the 7 methods. We assumed that true RNA-binding proteins would score positive in the majority of the methods used, while
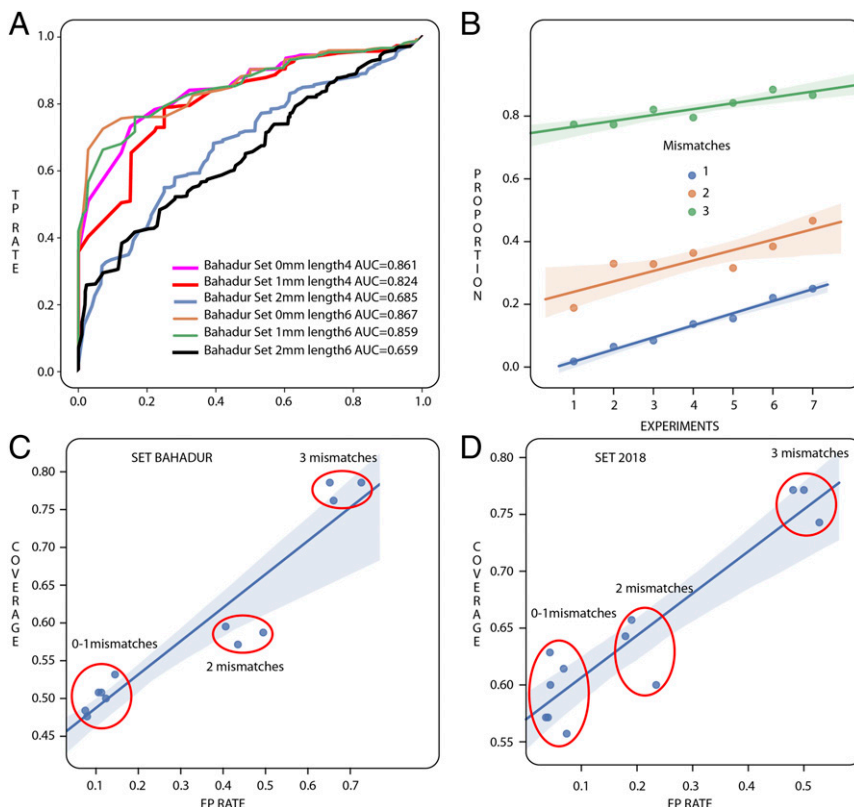

Fig. 2. RNA docking validation. (*A*) ROC curves testing the RMSD accuracy of the docks against crystallographic RNA, allowing 1 and 2 pep-mismatches in the scanned sequence to retrieve compatible fragments using frag-lengths of 4 and 6. (*B*) Proportion of proteins with recovered interfaces by ModelX docking vs. the number of experimental methods reporting RNA binding for a dataset compiled in 2018. More experiments reporting binding correlates with a greater probability of obtaining docks. (*C* and *D*) Coverage of RNA-binding sites vs. the false-positive rate for different pep-mismatches parameter values over the Bahadur published benchmarking dataset (*C*) and the RPCs published in 2018 (*D*).
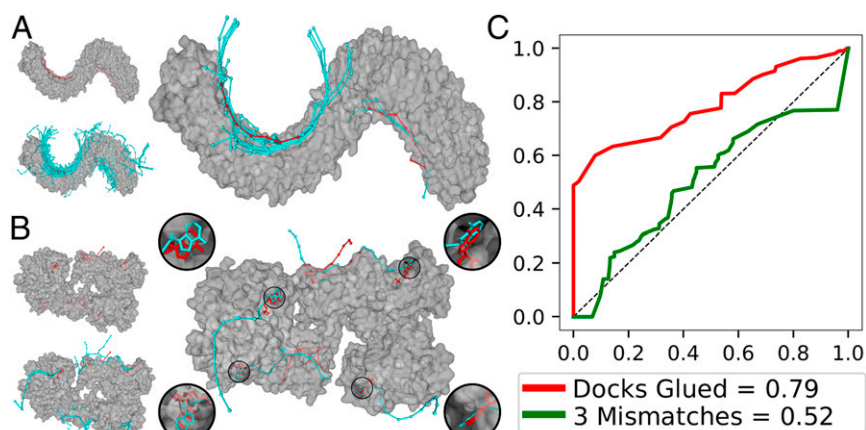
Delgado Blanco et al.

Fig. 3. Solving crystallographic RNA gaps with ModelX. (*A*) PDB ID code 1ZBH (a histone mRNA stem-loop) and (*B*) PDB entry 1M8W (a homolog of the Pumilio 1 human domain). Crystallographic RNA is in red; RnaDocking fragments obtained after an exhaustive run (pep-mismatches = 3) are in cyan. Decoy docked fragments could be filtered out by elongating geometrically compatible strands with the GlueDocks new command. (*C*) ROC curve for the Bahadur benchmarking dataset using pep-mismatches = 3 before (green) and after (red) running the GlueDocks command, with the min-length parameter to obtain elongated strands set to 8.

spurious binders would appear in only a few of the methods. If this is so, then proteins to which we docked RNA fragments would be identified as true in the majority of the methods.

To see if this is indeed the case, we selected crystal structures with the highest resolution that cover most of the 1,400 target proteins. We then determined the number of structures with docked fragments versus the number of experimental methods included in the review reporting binding (Fig. 2*B*). We found that the number of experiments reporting RNA binding correlates with what was expected in terms of docked fragments. Of those genes with 6 methods indicating no RNA binding and with a crystal structure deposited in the PDB (212 entries), we predicted RNA docking for only 4 of them (PDB ID codes 1JID, 4KRE, 4QOZ, and 5CCB), all of which have been cocrystallized with RNA. This highlights the importance of constructing a negative data set to aid benchmarking. The protein families included in the mentioned review (850 in total) are summarized in *SI Appendix*, Table S3. Domains known to bind RNA molecules are enriched within this dataset. We show that we could find docking RNA sites for domains sparsely populated in our database, and even for those for which there is evidence that they bind RNA but no protein-RNA complex is available. This is because we are looking at small peptide and RNA fragments, and it is quite likely that similar structural peptide-RNA arrangements are found in different domains with different architecture, just like there are finite ways in which 2 α-helices will interact in a protein.

Fragment Extension and Interface Completion. Another difficulty in terms of benchmarking RNA docking methods is the fact that many crystals have only a portion of the RPI determined. Fig. 3 shows our docking results for PDB ID codes 1M8W (Fig. 3*A*) and 1ZBH (Fig. 3*B*), 2 of the proteins belonging to the validation set. Using these examples, we show how crystallization gaps or incomplete interfaces can be resolved using ModelX. The distant regions of the crystallized RNA fragments were covered and connected with docks that were merged with the GlueDocks command (*Materials and Methods*). This step allowed us to discard spurious docks. In the case of 1ZBH, gaps were filled with rnaXs coming from 4QOZ and 4L8R (45), crystals of the same protein but with their full RPI determined. In the case of 1M8W, the docked RNA fragments come from other proteins, and the interface appears to be correctly covered, extending the actual knowledge provided by the proteins alone.

It is important to note that for validation purposes, some of the fragments that were originally docked over 1ZBH were tagged as false positives, even though in reality they are not. This

is demonstrated by examining the other 2 structures of the same protein, which indicates that our accuracy measurements presented above are in fact a lower bound of the method's actual performance. Fig. 3*C* shows that when docked RNA fragments are annealed with the GlueDocks command, the rate of docked false-positive RNA strands decreases, increasing the AUC of exhaustive runs (pep-mismatches = 3) from 0.52 to 0.79. On the other hand, increasing the number of mismatches allowed us to cover up to 85% of the crystallographic RNA with docks.

Interface Engineering. Beyond binding site recognition and RPI landscape completion, ModelX can also be used for RPI engineering. As an example, we chose the small spliceosomal protein family (Pfam accession no. PF00076) (46), for which the number of crystallographic structures allows an exhaustive investigation of the energetics of the mutational landscape. We explicitly excluded 1 protein belonging to this family—the U1A small nuclear ribonucleoprotein (UniProt P09012, PDB 1URN, and all other crystals of this protein)—from the built database and attempted to model it. Modeling was started from a different protein of the same family that presents low backbone variability (Fig. 4*A*), the U2A′ small nuclear ribonucleoprotein (UniProt P09661, PDB 1A9N, and others). Using the FoldX BuildModel command, we introduced the U1A sequence into the U2A′ backbone (U1A model). Then, by running an exhaustive (pep-mismatches = 3, energy-threshold = 2) RnaDocking over the U1A model with parameters to make it RNA protein sequence-specific (biased = true, atom type = ALL), we covered the binding site with compatible crystallographic RNA fragments. We merged these posteriorly with the GlueDocks command to generate longer strands and remove false-positive spurious docks (min-length = 7) (Fig. 4*C*). Then, using the FoldX force field, we determined the RNA sequence specificity of the elongated RNA docked fragments on the U1A model, as well as on the original RNA structure of the U2A′ complex. We found that we could retrieve the original U1A RNA sequence preferences only when using the best docked elongated fragment and not when using the original RNA strand of the crystal U2A′ (Fig. 4*D*). This because the docked fragments capture the conformational differences of the backbone between the RNA molecules bound to the U1A structure and those bound to the U2A′ structure (Fig. 4 *B* and *C*).

Discussion

Here we present a computational approach, termed RnaX, that predicts RPIs and is integrated into the ModelX software suite. This method uses a sliding window over the protein structure in
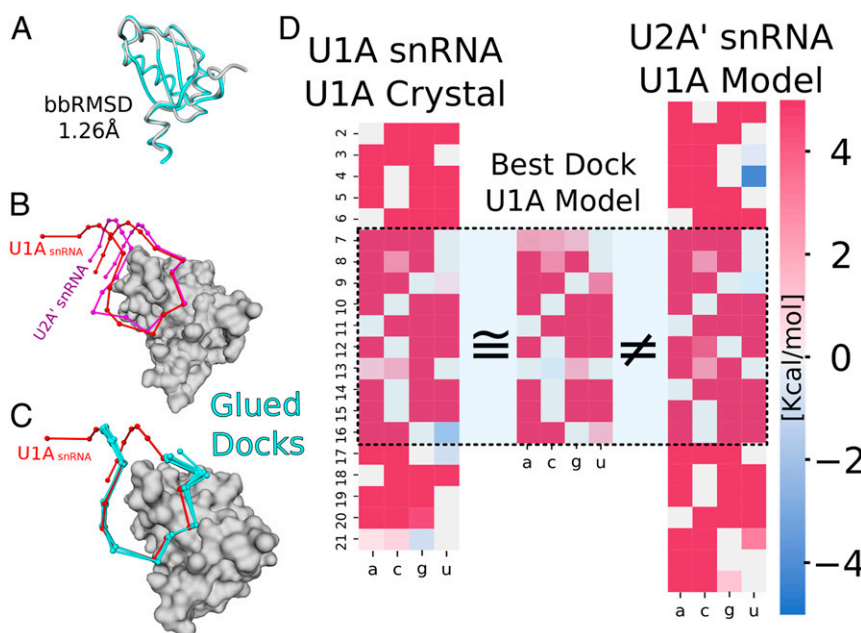
Fig. 4. Sequence-specific interface reconstruction with ModelX and FoldX. (*A*) Superimposition of the backbones of 2 small spliceosomal ribonucleoproteins (PDB ID codes 1URN for the U1A protein and 1A9N for U2A′ protein, in gray and cyan, respectively) enables accurate side chain modeling because of the low RMSD between their protein molecules. (*B*) The U1A sequence was modeled into the U2A′ structure. The U2A′ protein surface is in gray; the 2 snRNA fragments cocrystallized with each protein are in red and pink. The RNA backbones interfacing with the protein have major structural differences starting at position 12. (*C*) The U1A model was subjected to RNA docking, taking into account the sequence-specific–based force field (biased). The docked fragments were elongated by the GlueDocks command (in cyan) and show good superimposition with the crystallographic RNA being modeled (U1A snRNA, in red). (*D*) FoldX ΔΔG RNA Position-Specific Scoring Matrices. On the left are interaction energy values for the U1A crystal as a validation; in the center are the interaction energy values for the elongated docked RNA fragment with the lowest energy, along with the U1A protein model; and on the right are values for the crystallographic U2A′ snRNA strand alongside the U1A protein model. The bases in contact with protein residues are shown within the dashed box. While we recovered the RNA sequence specificity using the docked and elongated fragments, the original U2A′ snRNA strand was not useful for reproducing the U1A specificities due to conformational differences of the RNA backbones, specifically at positions 12 to 16.

which we looked for peptide fragments in our peptide-RNA library that superimpose well with the protein peptide in the corresponding window. The peptides in the library are associated to RNA fragments that are positioned automatically at the same time that the peptide is superimposed. Then those RNA fragments that are compatible with the rest of the protein backbone are retained for further analysis. Our validation experiments show that our method has high accuracy over an established benchmarking dataset and over newly deposited RNA–protein structures. By docking fragments originally coupled to highly similar sequences (pep-mismatches = 1), RnaX is able to detect RNA-binding sites with high confidence at the cost of reduced coverage. In contrast, when allowing the method to retrieve docks originally interfacing protein fragments with low sequence identity to the protein part being scanned (pep-mismatches = 3), we obtained close to 85% structural coverage at the cost of many false positives.

The false-positive problem can be partly solved by elongation of the docked RNA fragments. Since this method is based on existing RNA–protein structures, its main limitations are related to coverage of the possible RNA–protein interaction landscape in the PDB database. This implies that RNA-peptide interface configurations not yet observed in deposited crystals cannot be identified by our docking method. Given that the number of protein-RNA structures grows at a linear rate (*SI Appendix*, Fig. S1), we have engineered our software to easily digest new RNA–protein interfaces, increasing the database and thus the prediction

capability. ModelX flexibility modeling generated by backbone-compatible docked fragments can be combined with the fine side chain energetics of FoldX's force field to predict RNA sequence specificity of docked fragments. This is possible because of the conformational flexibility generated by docking RNA fragments, as shown for 2 closely related RNA-binding proteins. Furthermore, ModelX and FoldX can accomplish this at just a fraction of the computational cost of other flexibility simulation methods, such as molecular dynamics.

Data Availability. All structures deposited in the PDB used and analyzed in this paper are mentioned in the *SI Appendix*.

1. A. Kyburz, A. Friedlein, H. Langen, W. Keller, Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3′ end processing and splicing. *Mol. Cell* **23**, 195–205 (2006).
2. S. Millevoi *et al.*, An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.* **25**, 4854–4864 (2006).
3. F. Rigo, H. G. Martinson, Functional coupling of last-intron splicing and 3′-end processing to transcription in vitro: The poly(A) signal couples to splicing before committing to cleavage. *Mol. Cell. Biol.* **28**, 849–862 (2008).
4. P. Hilleren, T. McCarthy, M. Rosbash, R. Parker, T. H. Jensen, Quality control of mRNA 3′-end processing is linked to the nuclear exosome. *Nature* **413**, 538–542 (2001).

5. S. Hong, RNA binding protein as an emerging therapeutic target for cancer prevention and treatment. *J. Cancer Prev.* **22**, 203–210 (2017).
6. K. E. Lukong, K.-W. Chang, E. W. Khandjian, S. Richard, RNA-binding proteins in human genetic disease. *Trends Genet.* **24**, 416–425 (2008).
7. Y. Chen, G. Varani, Engineering RNA-binding proteins for biology. *FEBS J.* **280**, 3734–3754 (2013).
8. J. Si, J. Cui, J. Cheng, R. Wu, Computational prediction of RNA-binding proteins and binding sites. *Int. J. Mol. Sci.* **16**, 26303–26317 (2015).
9. C. G. Burd, G. Dreyfuss, Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
10. S. Calabretta, S. Richard, Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).
11. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
12. L. Wang, S. J. Brown, BindN: A web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.* **34**, W243–W248 (2006).
13. L. Wang, C. Huang, M. Q. Yang, J. Y. Yang, BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst. Biol.* **4** (suppl. 1), S3 (2010).
14. M. Kumar, M. M. Gromiha, G. P. S. Raghava, Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins* **71**, 189–194 (2008).
15. Y. Murakami, R. V. Spriggs, H. Nakamura, S. Jones, PiRaNhA: A server for the computational prediction of RNA-binding residues in protein sequences. *Nucleic Acids Res.* **38**, W412–W416 (2010).
16. M. B. Carson, R. Langlois, H. Lu, NAPS: A residue-level nucleic acid-binding prediction server. *Nucleic Acids Res.* **38**, W431–W435 (2010).
17. M. Terribilini *et al.*, RNABindR: A server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res.* **35**, W578–W584 (2007).
18. X. Ma *et al.*, 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 (2011).
19. C. M. Livi, P. Klus, R. Delli Ponti, G. G. Tartaglia, catRAPID signature: Identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics* **32**, 773–775 (2016).
20. J. Janin *et al.*; CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins* **52**, 2–9 (2003).
21. S. J. de Vries, M. van Dijk, A. M. Bonvin, The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010).
22. S. Lyskov, J. J. Gray, The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36**, W233–W238 (2008).
23. S. R. Comeau, D. W. Gatchell, S. Vajda, C. J. Camacho, ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, W96–W99 (2004).
24. A. Tovchigrechko, I. A. Vakser, GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* **34**, W310–W314 (2006).
25. V. I. Lesk, M. J. E. Sternberg, 3D-Garden: A system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics* **24**, 1137–1144 (2008).
26. G. Macindoe, L. Mavridis, V. Venkatraman, M.-D. Devignes, D. W. Ritchie, HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **38**, W445–W449 (2010).
27. M. Torchala, I. H. Moal, R. A. G. Chaleil, P. A. Bates, SwarmDock: A server for flexible protein-protein docking. *Bioinformatics* **29**, 807–809 (2013).
28. B. G. Pierce *et al.*, ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).
29. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
30. S. J. de Vries, C. E. M. Schindler, I. Chauvot de Beauchêne, M. Zacharias, A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.* **108**, 462–465 (2015).
31. B. Jiménez-García, C. Pons, D. I. Svergun, P. Bernadó, J. Fernández-Recio, pyDockSAXS: Protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Res.* **43**, W356–W361 (2015).
32. J. Yu *et al.*, InterEvDock: A docking server to predict the structure of protein-protein interactions using evolutionary information. *Nucleic Acids Res.* **44**, W542–W549 (2016).
33. I. Tuszynska, M. Magnus, K. Jonak, W. Dawson, J. M. Bujnicki, NPDock: A web server for protein-nucleic acid docking. *Nucleic Acids Res.* **43**, W425–W430 (2015).
34. Y. Yan, D. Zhang, P. Zhou, B. Li, S.-Y. Huang, HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373 (2017).
35. Y. A. Arnautova, R. Abagyan, M. Totrov, Protein-RNA docking using ICM. *J. Chem. Theory Comput.* **14**, 4971–4984 (2018).
36. L. Pérez-Cano, M. Romero-Durana, J. Fernández-Recio, Structural and energy determinants in protein-RNA docking. *Methods* **118–119**, 163–170 (2017).
37. M. F. Lensink, S. J. Wodak, Docking and scoring protein interactions: CAPRI 2009. *Proteins* **78**, 3073–3084 (2010).
38. J. D. Blanco, L. Radusky, H. Climente-González, L. Serrano, FoldX accurate structural protein-DNA binding prediction using PADA1 (Protein Assisted DNA Assembly 1). *Nucleic Acids Res.* **46**, 3852–3863 (2018).
39. E. Verschueren, P. Vanhee, F. Rousseau, J. Schymkowitz, L. Serrano, Protein-peptide complex prediction through fragment interaction patterns. *Structure* **21**, 789–797 (2013).
40. M. J. Sippl, Calculation of conformational ensembles from potentials of mean force: An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **213**, 859–883 (1990).
41. H. Kono, A. Sarai, Structure-based prediction of DNA target sites by regulatory proteins. *Proteins* **35**, 114–131 (1999).
42. J. Delgado, L. G. Radusky, D. Cianferoni, L. Serrano, FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168–4169 (2019).
43. C. Nithin, S. Mukherjee, R. P. Bahadur, A non-redundant protein-RNA docking benchmark version 2.0. *Proteins* **85**, 256–267 (2017).
44. M. W. Hentze, A. Castello, T. Schwarzl, T. Preiss, A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* **19**, 327–341 (2018).
45. R. Thapar, Contribution of protein phosphorylation to binding-induced folding of the SLBP-histone mRNA complex probed by phosphorus-31 NMR. *FEBS Open Bio* **4**, 853–857 (2014).
46. A. Bateman *et al.*, The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).