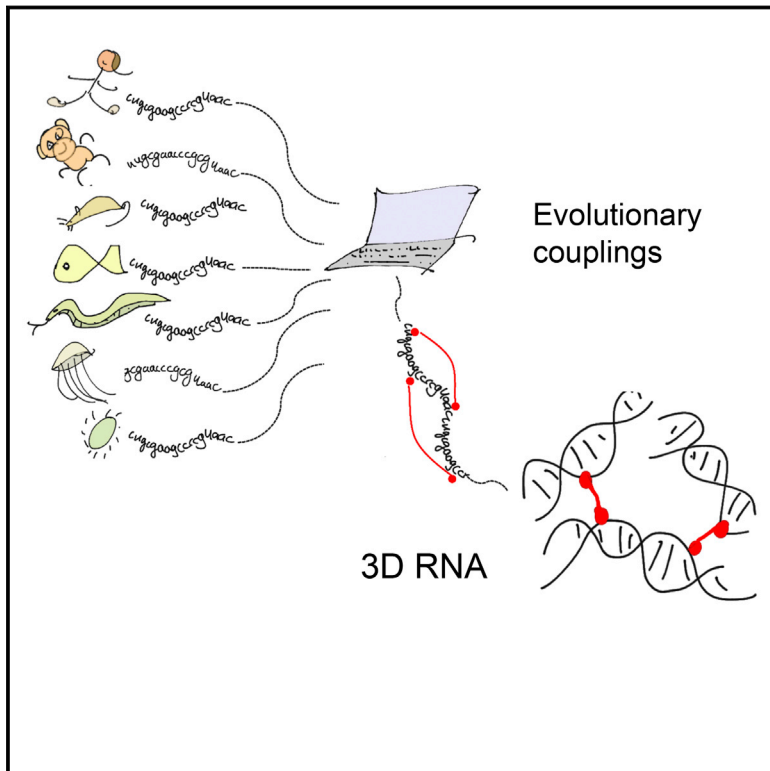


3D RNA and Functional Interactions from Evolutionary Couplings

Graphical Abstract



Authors

Caleb Weinreb, Adam J. Riesselman, John B. Ingraham, Torsten Gross, Chris Sander, Debora S. Marks

Correspondence

debbie@hms.harvard.edu

In Brief

The evolutionary signals for RNA 3D structure and interactions are detected by applying new algorithms to the analysis of thousands of genomes. This approach pinpoints an important interaction for HIV genome transport and predicts 3D contacts for hundreds of RNAs with unknown structure.

Highlights

- Sequence co-evolution reveals tertiary contacts and in vivo 3D RNA structures
- Co-evolved contacts accurately predict the 3D structure of RNA-protein complexes
- Functionally critical contacts show the strongest co-evolutionary signal
- Predictions for 160 RNAs, including viruses and lncRNAs, hint at molecular mechanisms



3D RNA and Functional Interactions from Evolutionary Couplings

Caleb Weinreb,¹ Adam J. Riesselman,^{1,4} John B. Ingraham,¹ Torsten Gross,^{1,2} Chris Sander,³ and Debora S. Marks^{1,*}

¹Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

²Institute of Pathology, Charité – Universitätsmedizin Berlin, 10117 Berlin, Germany

³Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

⁴Program in Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: debbie@hms.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2016.03.030>

SUMMARY

Non-coding RNAs are ubiquitous, but the discovery of new RNA gene sequences far outpaces the research on the structure and functional interactions of these RNA gene sequences. We mine the evolutionary sequence record to derive precise information about the function and structure of RNAs and RNA-protein complexes. As in protein structure prediction, we use maximum entropy global probability models of sequence co-variation to infer evolutionarily constrained nucleotide-nucleotide interactions within RNA molecules and nucleotide-amino acid interactions in RNA-protein complexes. The predicted contacts allow all-atom blinded 3D structure prediction at good accuracy for several known RNA structures and RNA-protein complexes. For unknown structures, we predict contacts in 160 non-coding RNA families. Beyond 3D structure prediction, evolutionary couplings help identify important functional interactions—e.g., at switch points in riboswitches and at a complex nucleation site in HIV. Aided by increasing sequence accumulation, evolutionary coupling analysis can accelerate the discovery of functional interactions and 3D structures involving RNA.

INTRODUCTION

RNAs have diverse known biological roles (Garneau et al., 2007; Huang et al., 2015; Martin and Ephrussi, 2009; McManus and Graveley, 2011; Olsen et al., 1990; Rutherford et al., 2015; Sigova et al., 2015; Warf and Berglund, 2010), and both genetic and biochemical screens suggest reservoirs of functional RNA molecules we know little about. For instance, transcriptional profiling has revealed large numbers of non-coding RNA genes, with unknown functions beyond context-specific expression (Eddy, 2014; Rinn and Chang, 2012). Although much of this transcription may be biological noise, these screens have revealed some novel long non-coding RNAs that may have 3D structures and can act as, for example, protein scaffolds (Quinodoz and

Guttman, 2014). High-throughput biochemical screens, adapted from earlier work on RNA foot-printing (Ehresmann et al., 1987; Latham and Cech, 1989; Moazed and Noller, 1986), have identified transcriptome-wide RNA base pairing in vivo (Ding et al., 2014; Rouskin et al., 2014; Spitale et al., 2015; Wan et al., 2014), including within the coding region of mRNAs, suggesting a function independent of coding potential.

Many of these newly observed RNAs may have specific 3D structures (Mortimer et al., 2014; Novikova et al., 2012). Since high-resolution structure determination remains labor intensive, there is a renewed interest in the computational prediction of RNA 3D structure and identification of functional interactions. One long-standing approach for inferring RNA structure is to search for pairs of positions that show correlated substitutions in alignments of homologous sequences. This approach was used to define the 1969 Levitt model of tRNA (Levitt, 1969), the Fox-Woese model of 5S rRNA (Fox and Woese, 1975) and the Michel-Westhof model of a group 1 ribozyme (Michel and Westhof, 1990). Comparative sequence analyses of RNA continue to contribute to successful RNA sequence alignment (Nawrocki and Eddy, 2013) and secondary structure prediction methods (Hofacker et al., 2002; Nussinov and Jacobson, 1980; Rivas and Eddy, 1999; Zuker, 2003) (review Hofacker and Lorenz, 2014), but existing techniques for identifying correlated positions have been less successful at detecting key tertiary contacts not involved in Watson-Crick base pairing (Dutheil et al., 2010).

The inability of existing methods to detect long-range tertiary contacts from sequence covariation has limited the progress of purely in silico RNA 3D structure prediction, since RNAs with multiple helical segments can assume diverse folds, producing a conformational space that is impossible to search effectively. Thus, despite rapid strides in 3D structure prediction accuracy for small (<40 nt) RNAs (Cao and Chen, 2011; Das and Baker, 2007; Das et al., 2010; Frelsen et al., 2009; Parisien and Major, 2008), predicting the structure of large (>70 nt) RNAs remains challenging (Laing and Schlick, 2010; Miao et al., 2015), unless experimental restraints are available from biochemical probing data (Cheng et al., 2015; Magnus et al., 2014; Ramani et al., 2015).

Why have existing methods that compute pairwise patterns of sequence co-variation in RNA (Mokdad and Frankel, 2008; Pang et al., 2005; Shang et al., 2012) had limited success in detecting tertiary interactions? One possibility is that RNA tertiary contacts often form complex networks (Butcher and Pyle, 2011) in which patterns of sequence constraints can interfere with each other,

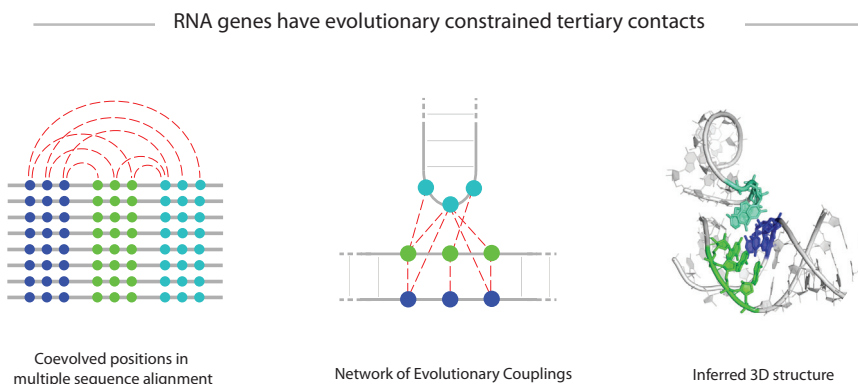


Figure 1. Inferring RNA Structure from Sequence Co-evolution

Non-coding RNAs form 3D structures stabilized by complex networks of secondary and tertiary interactions. In many cases, these interactions leave an evolutionary imprint reflecting epistasis between contacting nucleotides. Computationally detecting these interactions in multiple sequence alignments can reveal RNA 3D structure.

obscuring true interactions and producing spurious transitive correlations when multiple contacts are chained together. A similar problem stymied protein structure prediction until the application of global maximum entropy models that could deconvolve the underlying network of constrained residue-residue interactions (Hopf et al., 2012, 2014; Marks et al., 2011, 2012; Morcos et al., 2011; Ovchinnikov et al., 2014; Weigt et al., 2009).

Here, we adapted the maximum entropy model to RNA sequence alignments and tested the ability of the approach to predict tertiary structure contacts on 180 RNA gene families representing thousands of RNA genes. Comparisons to known structures confirm the accuracy of contact prediction and 3D folding results. We further extend the model to protein-RNA interactions and accurately predict six RNA-protein interactions.

RESULTS

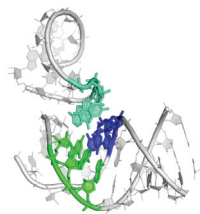
Evolutionary Couplings Accurately Predict 3D Contacts

We adapted the evolutionary couplings (ECs) model, which we previously used to predict contacts within proteins, to calculate ECs for RNA (Figure 1). Briefly, we model each RNA family as the distribution over a sequence space, where the probability of a sequence σ reflects the single-site biases h_i at each position i , and the coupling terms J_{ij} between each pair of positions (i, j) , as follows.

$$P(\sigma) = \frac{1}{Z} \exp \left(\sum_{i=1}^L h_i(\sigma_i) + \sum_{i=1}^{L-1} \sum_{j=i+1}^L J_{ij}(\sigma_i, \sigma_j) \right).$$

We fit this model to natural sequence distributions using an approach based on penalized maximum likelihood estimation. We refer to pairs of positions with the strongest coupling terms as “evolutionary couplings.” See the [Experimental Procedures](#) for details.

Evolutionary couplings (ECs) between pairs of nucleotides were evaluated by their ability to recover RNA 3D contacts in the known structure of a representative sequence from each of 22 RFAM families (Data S1). To assess the accuracy of predicted contacts, we define contacts as true positives if their minimum-atom-distance is $< 8 \text{ \AA}$. All 22 RNAs with known structures had a true positive rate above 70% for the top-ranked $L/2$ ECs (where



L is the number of nucleotides). ECs predicted contacts with greater accuracy than did mutual information (MI), which has been widely used for RNA secondary structure prediction (Freyhult et al., 2005;

Gutell et al., 1992) (Figure 2A). This held even when using enhanced MI that implements two features of the EC statistical model: (1) down-weighting of similar sequences to avoid spurious correlations from phylogeny and (2) an average product correction (APC) (Dunn et al., 2008). We refer to MI without these modifications as raw MI (MI_{raw}) and denote enhanced MI as MI_{E} .

There are two reasons why the global model for ECs may perform better than MI. First, nucleotides that are strongly conserved will not display high mutual information, but may still have a high EC score. Second, false-positive transitive correlations score highly with local methods, such as MI, as each pair is computed without reference to the whole network of interactions, whereas ECs successfully deconvolve transitive correlations.

Evolutionary Couplings Detect Long-Range Contacts and Non-Watson-Crick Base Pairs

Though overall accuracy is important, not all contacts are “created equal.” Often, complex RNA folds are stabilized by a small number of critical long-range contacts that bridge distant parts of the secondary structure. Though ECs of long-range contacts (definition in the [Experimental Procedures](#)) have lower scores than those in secondary structures (Figure S1), we nevertheless robustly detect long-range contacts across 22 RFAM families, with an average of $0.07 * L$ long-range contacts among the top $L/2$ ECs for an RNA of length L . This represents a substantial improvement over previous methods, since ECs contained 2.4 times more long-range contacts than did MI_{E} , (0.8–10.8 times more across 22 individual examples; $p \leq 10^{-5}$ using a paired t test; Figure 2B).

Pairs of RNA bases often form contacts through hydrogen bonding, assuming geometrical configurations that can broadly be divided into Watson-Crick (WC) base pairs and non-WC base pairs. Though covariation has long been used to infer Watson-Crick (WC) base pairs, the strength of coevolution for non-WC base pairs is an area of open investigation (Dutheil et al., 2010). We found that ECs are sensitive to non-WC base pairs, with the top $L/2$ ECs containing 16% of all annotated non-WC base pairs across the 22 structures—1.7 times more on average than MI_{E} (0.5–4.0 times more across 22 individual examples; $p \leq 0.002$ using a paired t test). We expect that ECs will complement existing approaches for detecting non-WC base pairs that rely on the concept of isostericity (Lescoute et al., 2005; Mokdad

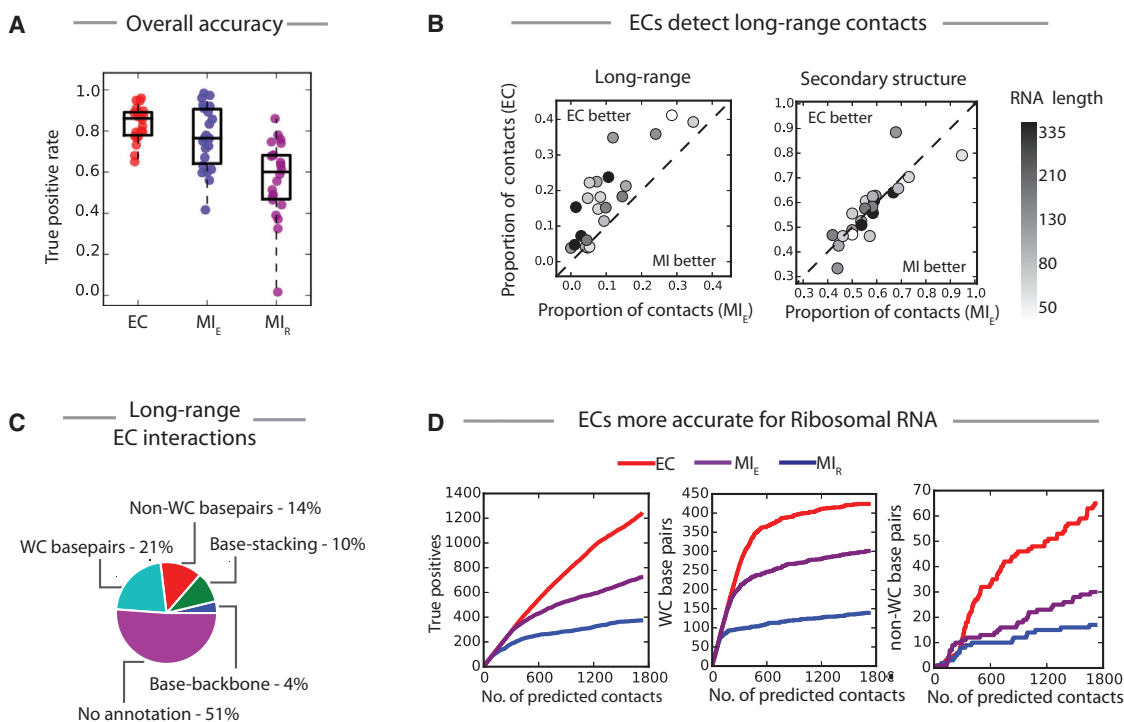


Figure 2. Comparison of EC to MI: Summary of 22 RNA Families

(A) Evolutionary couplings (ECs) predict 3D contacts with a higher overall accuracy than does MI_E or MI_R on a test set of 22 RNA families (Data S1).

(B) ECs and MI_E detect a similar number of secondary structure contacts, but ECs are significantly enriched with long-range contacts.

(C) EC-inferred long-range contacts represent a variety of biochemical interactions annotated from the crystal structure.

(D) ECs for the 40S ribosome (RF01960) are dramatically more accurate than are MI, detecting more true positives overall (left), more Watson-Crick base pairs (middle), and more non-WC base pairs (right). ECs also detect ~50 long-range contacts that bridge distance parts of the secondary structure (Figure S2). See also Figures S1 and S2 and Data S1.

and Frankel, 2008), since they can focus attention on interactions with the strongest co-evolutionary signal.

Evolutionary Couplings Reveal Contacts in the Eukaryotic Ribosome

ECs may be sensitive to interacting nucleotides in large RNAs that form topologically complex folds with abundant long-range contacts. ECs computed on a full alignment of eukaryotic ribosomal sequences (RF01960) are over 90% accurate for the top 900 ($L/2$) contacts and predicted substantially more WC and non-WC base pairs than did MI_E (Figure 2D), as well as more long-range contacts between nucleotides (Figure S2). In addition to increased sensitivity, ECs also have greater specificity than does MI_E, with 2.8-fold fewer false-positive predictions. Capturing complex networks of contacts in large RNAs will be crucial for decoding the structure of long non-coding RNAs (lncRNAs) and structured viral genomes.

Long-Range Contacts Allow Accurate 3D Structure Prediction

We hypothesized that EC-derived long-range contacts could benefit 3D structure prediction for medium-sized RNAs, which has so far not been possible with secondary structure alone. Using coarse-grained molecular dynamics implemented in NAST (Butcher and Pyle, 2011; Jonikas et al., 2009), followed

by simulated annealing with XPLOR (Schwieters et al., 2003), we predicted candidate all-atom structures (Figure 3A) for representatives from five selected RNA families (selection based on length, 70–120 nt and at least one highly long-range contact; see the Experimental Procedures). In each case, the best of four candidate predictions had an all-atom root-mean-square deviation (RMSD) of 7–10 Å, comparable to the state of the art for RNA structure predictions that use tertiary contacts derived from biochemical probing (6.8–11.7 Å) (Miao et al., 2015) and a dramatic improvement over pure in silico structure prediction, where the average RMSD is 20 Å for medium-sized RNAs (50–130 nt) (Laing and Schlick, 2010). Our method shows a high level of precision, with 16/20 (four predicted structures each for five RNA families = 20 total) predicted structures having correct orientation of the helices (Figure S3), and all predictions significantly closer to the experimental structure than controls folded without tertiary constraints (Figure 3B, $p < 0.01$ for all five comparisons; see Data S2 for RMSD and quality structure metrics; see Figure S4 for plots of RMSD versus NAST energy).

Evolutionary Couplings Identify Intermolecular Contacts in RNA-Protein Complexes

Since ECs can detect contacts in RNA and protein separately, we next investigated whether they can reveal intermolecular

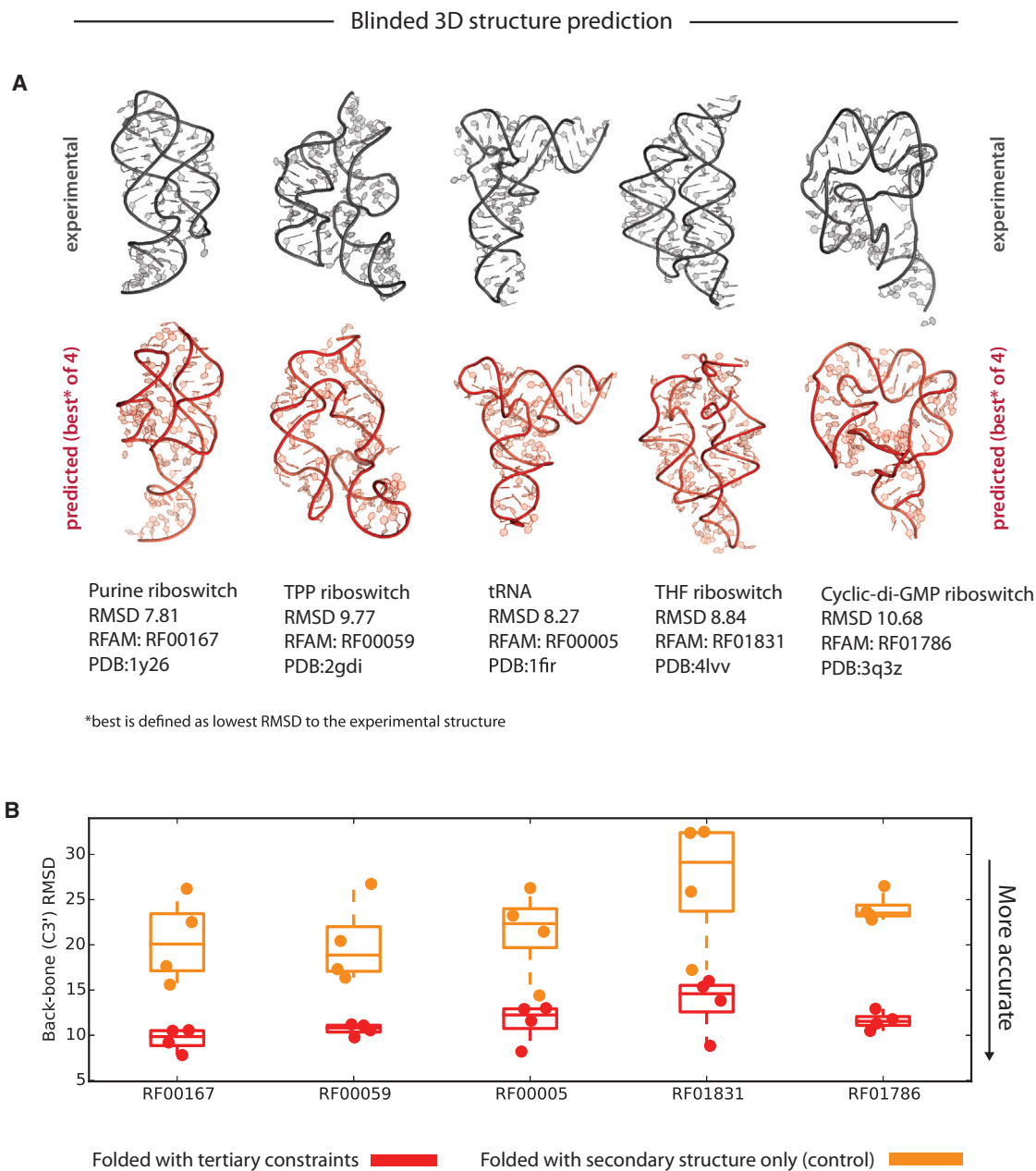


Figure 3. Evolutionary Couplings Significantly Improve 3D Structure Prediction Accuracy

(A and B) We predicted all-atom 3D structures for five RNA families using evolutionary couplings as distance restraints (four candidate structures per family). The candidate (red) with the lowest RMSD to the experimental structure (gray) is shown for each family (A). We performed folding controls with secondary structure only and found that they had significantly higher deviation from the experimental structure than did models folded with EC-derived tertiary contacts (B). For the full results on all predicted models, see [Figures S3 and S4](#) and [Data S2 and S7](#). See also [Figures S3 and S4](#) and [Data S2](#).

contacts in 21 RNA-protein complexes with known structure ([Data S3](#); see the [Experimental Procedures](#) for selection criteria), including 19 ribosomal proteins bound to the bacterial 16S rRNA and the ribonucleoprotein (RNP) complexes RNaseP and tmRNA ([Figure 4](#)). ECs predict RNA-protein contacts with high accuracy, as long as there is sufficient sequence diversity ([Figure S5A](#)). If contact predictions with $\geq 75\%$ true positives for the top four

contacts are defined as “highly accurate,” then only 3/18 complexes with less than one effective sequence per nucleotide ($M_{\text{eff}}/L < 1$) had highly accurate predictions, whereas all (3/3) complexes with ($M_{\text{eff}}/L > 1$) had highly accurate predictions. One striking example is the 5S ribosomal protein RL18—with ~ 4 effective sequences per nucleotide—which had 80% true positives for the top ten intermolecular contacts.

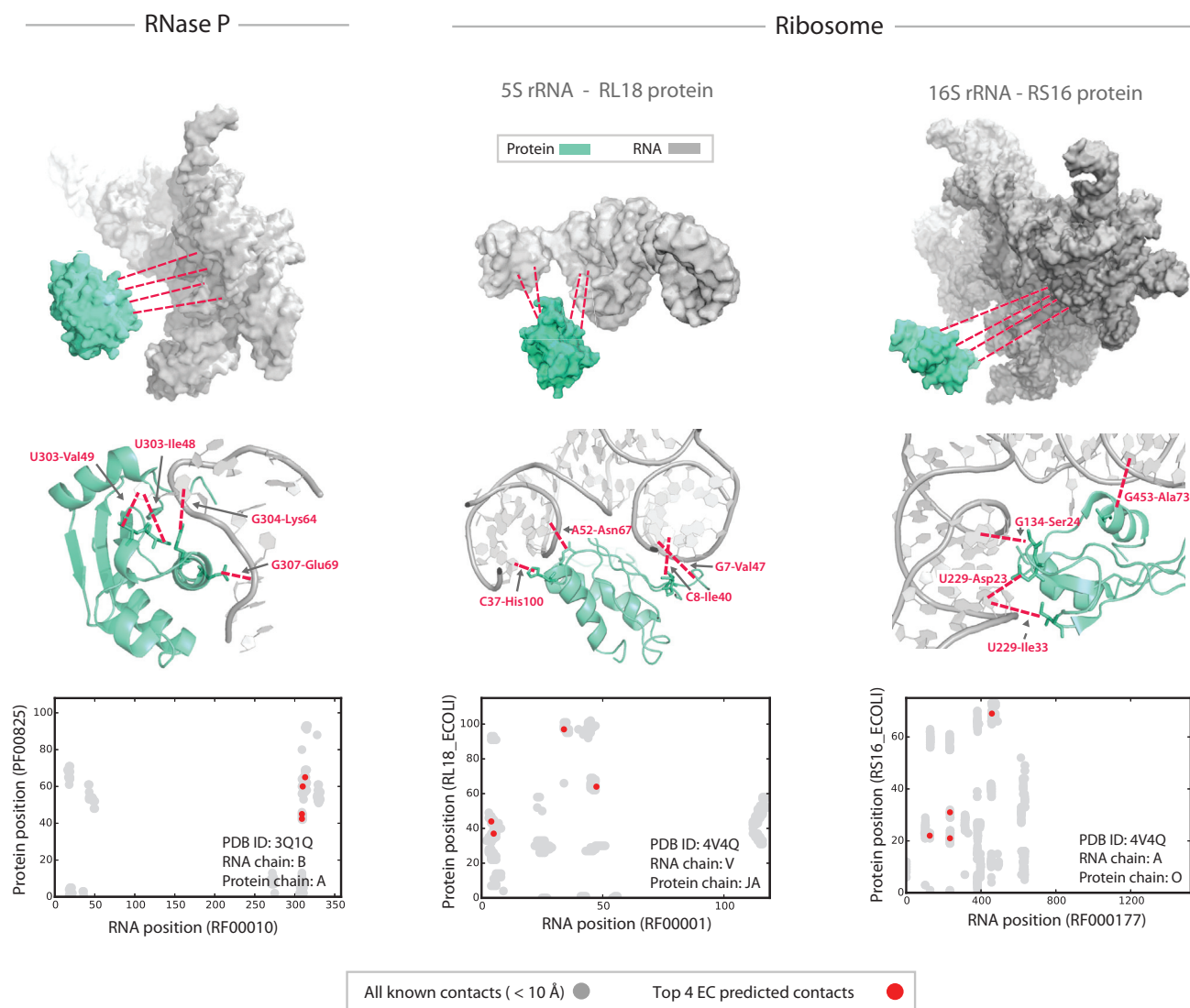


Figure 4. Evolutionary Couplings Detect RNA-Protein Interactions

Functionally important interactions between RNA nucleotides and protein amino acids may be constrained in evolution and detectable in multiple sequence alignments. We calculated ECs from phased alignments of 21 RNA-protein complexes, with three examples shown here (ribonucleoprotein complex RNase P on the left and two proteins from the bacterial ribosome on the right are chosen as illustrative examples). In each case, we plot the top four highest ranking ECs in cartoon style (top), super-imposed on the true structure (middle, in PDB numbering) and as red dots in a contact map (bottom). These interactions anchor their respective RNA-protein interfaces at multiple points of contact and open the door to sequence-based 3D structural studies of RNA protein complexes. For related results, see [Figure S5](#) and [Data S4](#) and [S5](#). For predicted PDB structures and phased alignments, see [Data S7](#). See also [Figure S5](#) and [Data S3](#) and [S4](#).

The accuracy of evolutionary couplings provides an exciting opportunity to predict the structure of RNA-protein complexes from sequence alone. We used EC-derived contacts as distance restraints for rigid body docking in HADDOCK ([Dominguez et al., 2003](#)), focusing on the 6/21 complexes with $\geq 75\%$ true positives for the top four predicted contacts. In all six cases, structures docked with ECs had significantly lower i-RMSD than did control structures docked with center of mass constraints only ($p < 10^{-4}$ for all six comparisons of mean i-RMSD between cases and controls; see [Figure S5B](#) and [Data S4](#)). In 4/6 cases, one of the top three predicted structures had i-RMSD < 3.5 Å,

whereas the minimum i-RMSD among all controls (docked without EC distance restraints) was 15 Å.

Evolutionary Couplings Highlight Functional Interactions

Among all instances of proximity in structured RNAs, only a small fraction is critical for function. Next, we asked whether evolutionary couplings enrich for these functionally important interactions by investigating the top scoring ECs in riboswitches, which are *cis*-acting regulatory segments of mRNAs that undergo ligand or temperature-dependent conformational changes between at

least two mutually exclusive functional states (Garst et al., 2011; Serganov and Patel, 2012). In four riboswitches from our dataset—S-adenosylmethionine (SAM), active vitamin B₁ (TPP), active folate vitamin (THF), and adenine sensing—we found a cluster of tertiary contacts that stabilize the ligand-bound conformation, but are broken in the ligand-free conformation (Figure 5). For instance, in the TPP riboswitch (Figure 5B), ECs between the L5 loop and J3-2 helix (69A-37G, 69A-23C, and 70A-22C, in numbering from 2gdi.pdb [Serganov et al., 2006]) reveal a set of base-stacking interactions that form when TPP ligand binds, but are broken when TPP is released (Serganov et al., 2006). These functionally important interactions are completely missed by MI_E. Similarly, very high ranking ECs between nucleotides between P2-P4 and P1-P4 in the SAM riboswitch (Figure 5A) reveal contacts that only conditionally form on binding of the ligand S-adenosylmethionine (C25-G89, U26-A88, G27-C87, G28-C86, A9-A84, and A10-A84 in numbering from 4kqy.pdb [Lu et al., 2010]). Although some of these contacts are also detectable by MI_E, a cluster between P1-P4 is not, including base stacking between A9-A84.

Prediction of Contacts for 160 RNAs of Unknown Structure

Encouraged by the accuracy of contact inference for RNAs of known structure, we predicted 3D contacts for 160 RNA genes in the RFAM database that do not have a known 3D structure for any member of the family (detailed results are available online at https://marks.hms.harvard.edu/ev_rna/). These predictions can aid experimental structure prediction and provide 3D constraints for direct folding simulations. In the following sections we show the nucleotide resolution predictions in biological context.

Disambiguating Alternative Structures of the HIV Rev Response Element RNA

To be exported from the nucleus, HIV transcripts assemble into ribonucleoprotein particles that contain multiple copies of the HIV Rev protein bound to the ~350-nt-long Rev response element (RRE) (Bartel et al., 1991; Iwai et al., 1992; Kjems et al., 1992; Peterson and Feigon, 1996; Rausch and Le Grice, 2015). Since the RRE and its interaction with Rev are vital steps in the HIV life cycle, they have become the subject of intense research as potential drug targets (Gallego and Varani, 2001; Luedtke and Tor, 2003; Sreedhara and Cowan, 2001), but important details of Rev-RRE structure and function remain unknown.

The secondary structure of RRE is one area of open investigation. In silico folding and in vivo biochemical foot-printing studies support two stable secondary structures, termed SL4 and SL5, which differ in the region A155-A211 (Figures 6A and 6B). SL5 has a five stem-loop structure (Bai et al., 2014; Kjems et al., 1992; Pollom et al., 2013), and recent work has shown that SL4, a four stem-loop structure, also exists both in vitro and in vivo (Charpentier et al., 1997; Fernandes et al., 2012; Legiewicz et al., 2008; Sherpa et al., 2015; Zimmel et al., 1996).

We found that ECs (Data S5) computed on the RFAM alignment of RRE (RF00036) overwhelmingly support the SL5 structure. Among 34 secondary structure contacts unique to either SL4 or SL5, the ten highest ranking secondary structure contacts support SL5 (Figure 6A, as red lines). In fact almost all of the 18

SL5-exclusive contacts outrank the 16 SL4-exclusive contacts (Figure 6B; $p < 10^{-5}$ based on t test of EC-ranks). Although the evolutionary evidence supports SL5, we cannot rule out the existence of the SL4 secondary structure in an ensemble, but note that the sequence of the clone used by groups reporting SL4 (pNL4-3) (Adachi et al., 1986; Charpentier et al., 1997; Fernandes et al., 2012; Legiewicz et al., 2008; Sherpa et al., 2015; Zimmel et al., 1996) thermodynamically favors SL4 more than 98.6% of the 44,046 sequences in our alignment (Figure 6C; Data S6). We obtained virtually identical results for both the thermodynamic and co-evolutionary analyses (Figure S6) using a separate multiple sequence alignment from the LANL HIV database (<http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>). In summary, our thermodynamic folding energy calculations and comparative sequence analysis indicate that SL5, rather than SL4, is the major RRE secondary structure, without excluding the possibility that SL4 occurs in particular evolutionary contexts, such as the pNL4-3 variant.

Constrained Interactions between HIV Rev Response Element RNA and Rev Protein

Assembly of the Rev-RRE nuclear export complex—a key step in the HIV life cycle—is thought to be driven by early nucleation events dependent on highly specific RNA-protein interactions.

We therefore used evolutionary couplings to infer dominant points of interaction in the interface between the RRE RNA and Rev protein, an HIV-encoded co-factor of nuclear export. Biochemical work shows that Rev protein oligomerizes on the RRE RNA after initial binding to stem IIB (Figure 6A) (Bai et al., 2014; Casu et al., 2013; Kjems et al., 1992). Our co-evolutionary analysis using a concatenated Rev-RRE alignment revealed a few candidate intermolecular contacts (Data S5). Strikingly, the top scoring intermolecular evolutionary coupling (EC) pinpoints a region previously identified as the location of Rev binding to RRE in multiple experiments over the last 25 years (Bai et al., 2014; Bartel et al., 1991; Daugherty et al., 2010; Heaphy et al., 1991; Ippolito and Steitz, 2000; Malim and Cullen, 1991). This evolutionary analysis, which reflects in vivo reality in diverse contexts, thus supports the in vitro experiments that have used only partial and synthetic RNA constructs.

Distant T Box Riboswitch Regions Co-evolve via Cooperative Binding of a tRNA Ligand

Using the inferred evolutionary couplings (ECs), we investigated the mechanism of tRNA sensing by the T box riboswitch (RF00230), a family of RNA elements that ensure homeostasis of charged tRNA by upregulating tRNA synthetases and amino-acid importers when aminoacylation of a specific tRNA is low (Green et al., 2010). T box riboswitches contain two tRNA recognition elements (Figure 7A, right): the “specifier sequence,” which examines tRNA identity by base pairing to its anti-codon sequence (Grundy et al., 2002), and the anti-terminator (T box) domain, which selectively binds the uncharged tRNA acceptor stem. These interactions stabilize the anti-terminator stem, causing a conformational change in the riboswitch that exposes a Shine-Dalgarno (SD) sequence and results in translation of the operon encoded genes (Figure 7A, right). Important details of this process have been elucidated

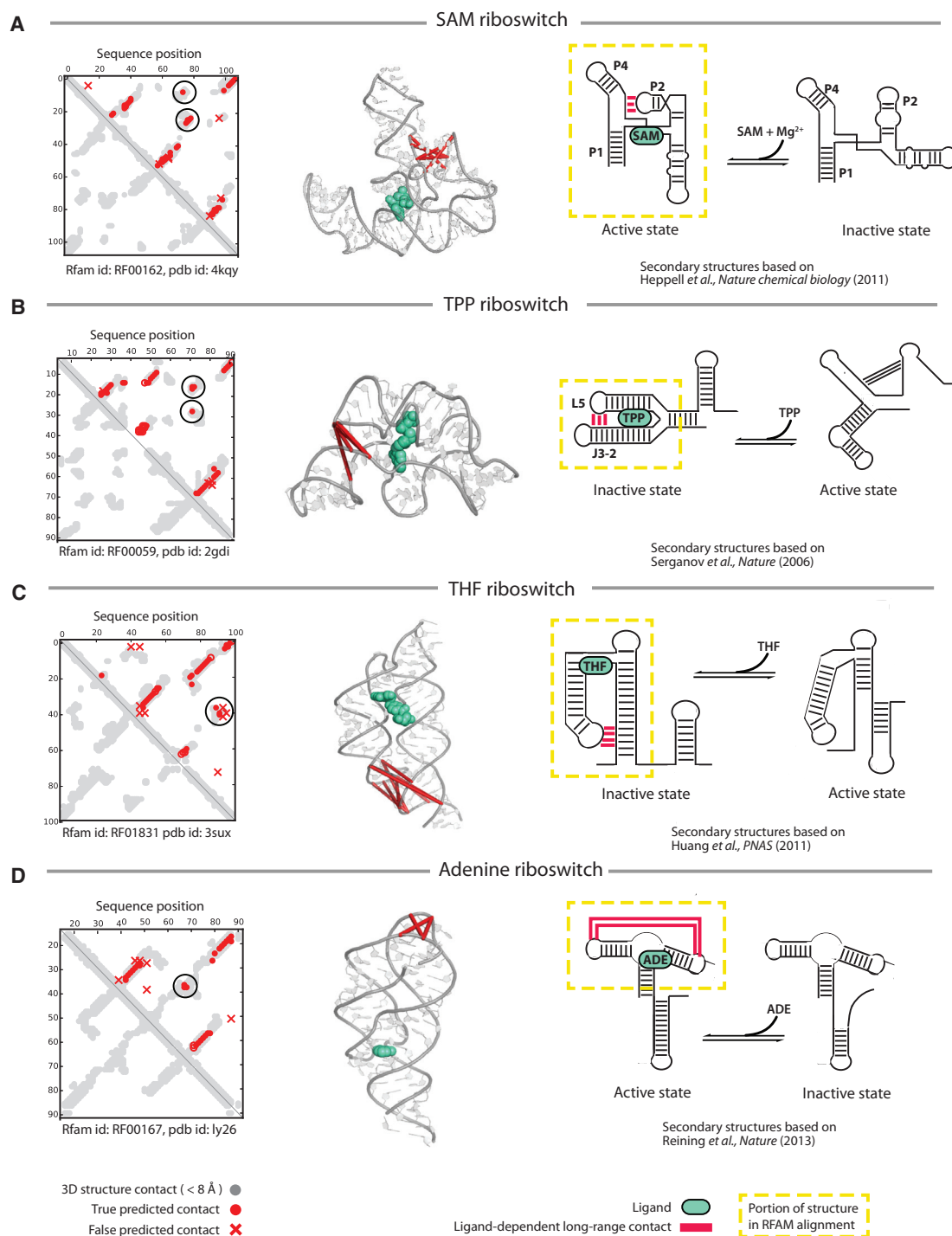


Figure 5. Evolutionary Couplings Identify Functional Interactions in Riboswitches

(A–D) 3D contacts revealed by ECs are conserved across the RNA family and therefore may be functionally important. We found that the top ranking long-range ECs in four riboswitches (A–D) are functionally critical, since they are differentially satisfied in the ligand-bound and ligand-free conformation. In each example, a contact map (left) shows the top L/2 contacts. The circled contacts—which are highlighted red on the 3D structures (middle)—are formed in the ligand-bound state, but are violated in the unbound state. This is illustrated by the schematics (right), which were reproduced from prior studies.

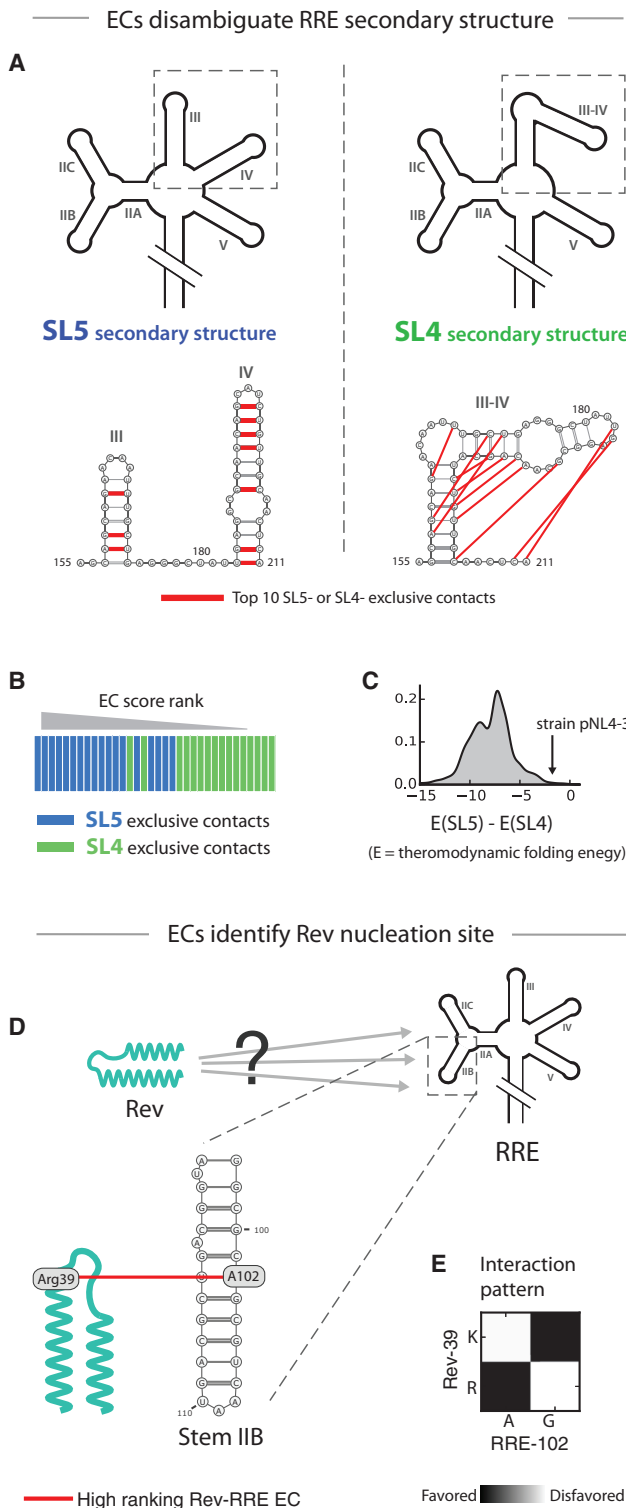


Figure 6. Insight into HIV RRE Structure and Function
 (A–D) The HIV Rev response element (RRE) is an important drug target because its nuclear export function is critical to the HIV life cycle, but essential details of RRE structure and function remain unknown. We first used ECs to disambiguate between two mutually exclusive RRE secondary structures res-

(Caserta et al., 2015; Grigg and Ke, 2013; Zhang and Ferré-D’Amaré, 2013), but only this first stem loop has been observed to date in 3D (Zhang and Ferré-D’Amaré, 2013). It is still unclear how tRNA stabilizes the anti-terminator stem and whether the specifier sequence alone is sufficient to ensure specificity in vivo.

We examined the top-ranking ECs for the T box riboswitch family (see the [Experimental Procedures](#)) to investigate the mechanism of combinatorial tRNA sensing. Remarkably, we found a cluster of six long-range ECs connecting the specifier sequence to the T box region (pink arc in [Figure 7A](#)) that are likely mediated by co-variation with the intervening tRNA. In particular, variation in the nucleotides of the specifier sequence may have led to changes in tRNA specificity, which, in turn, affected the evolution of nucleotides in the T box region, producing the observed co-variation. If true, this hypothesis would imply that the T box region collaborates with the specifier sequence to establish tRNA specificity, which is consistent with mutational studies that showed interdependence between the *Bacillus subtilis* tyrS T box and tRNA(Tyr)-acceptor sequences (Grundy et al., 1994).

Next, we looked at which specific nucleotides from the T box region participate in the six long-range ECs with the specifier sequence. Strikingly, two of the six long-range ECs involve a T box nucleotide (U191 and RFAM reduced numbering) that base pairs to the tRNA discriminator position (Crothers et al., 1972), which has been known for decades to co-vary with the tRNA anti-codon (Klingler and Brutlag, 1993) and is thought to influence the tertiary structure of the acceptor stem (Lee et al., 1993). The other four long-range ECs involve bases G186 and C210, which are paired in the anti-terminator stem, raising the possibility that the tRNA acceptor stabilizes the anti-terminator stem by forming extensive tertiary contacts with bases in and near the T box region. Using an adaptation of our model that previously demonstrated accurate inference of the effects of single-site mutations (Hopf et al., 2015) (see the [Experimental Procedures](#)), we predict that in the *Enterococcus durans* T box, for example, the mutations G186U/C210A and U191A would be most likely to disrupt these interactions.

ported in the literature, termed SL5 (A, left) and SL4 (A, right). We defined RRE contacts as SL4- or SL5-exclusive if they are satisfied in one secondary structure, but not in the other. Strikingly, 10/10 of the top-ranking exclusive contacts (red lines, bottom of A) support SL5, but not SL4. In fact, almost all of the 18 SL5-exclusive contacts outrank the 16 SL4-exclusive contacts (B, $p < 10^{-5}$). Strikingly, all of the studies we found reporting the SL4 structure used the pNL4-3 variant of HIV, which is in the 98.6th percentile for favoring SL4 over SL5 (C) according to thermodynamic folding energy predictions (Data S6). Thus, SL5 is likely the dominant RRE secondary structure in most evolutionary contexts. Next, we scanned for evolutionarily constrained interactions between RRE RNA and Rev protein (D, top). Our top-ranking contact (D, bottom, and E) linking Rev Arg39 to RRE A102 on stem IIB is consistent with biochemical work showing that stem IIB is the major Rev nucleation site. This evolutionary analysis, which reflects in vivo reality in diverse contexts, thus supports the in vitro experiments that have used only partial and synthetic RNA constructs. (See [Figure S6](#) for identical analysis on a different RRE alignment.)

See also [Figure S6](#) and [Data S5](#) and [S6](#).

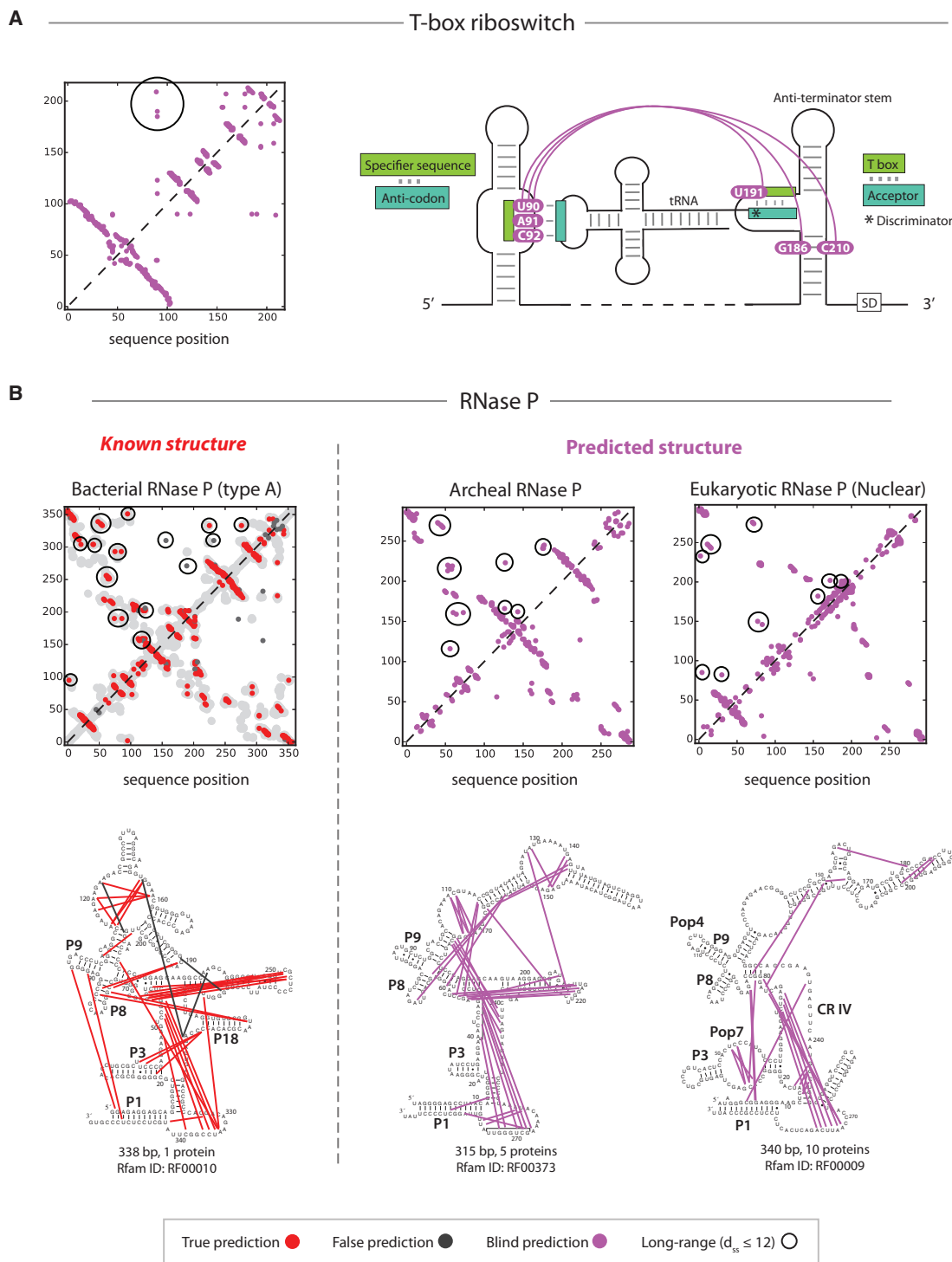


Figure 7. Examples of Evolutionary Couplings Elucidating Structure and Function

(A and B) We provide ECs for 160 RFAM families without a known structure and investigate them in detail for the T box riboswitch (A) and RNAs P (B). In the T box riboswitch, a high-ranking cluster of long-range ECs (A, circled dots in the contact map and pink arcs on the secondary structure) connects the specifier sequence to the T box region. The binding of both of these RNA elements to an uncharged tRNA induces translation of downstream genes by exposing a Shine-Delgarno (SD) sequence. The long-range ECs are probably mediated by co-variation with the intervening tRNA. Supporting this hypothesis, two of the six long-range ECs involve U190, a nucleotide in the T box region that base pairs to the discriminator position in the tRNA (A, marked by the *), which is known to co-vary with the tRNA anti-codon. We also used ECs for the bacterial, archeal, and eukaryotic RNase P to address a hypothesis in the evolution of RNase P, a ribozyme found in all three domains of life (B).

Proteins Displace RNA Tertiary Contacts in the Evolution of RNase P

Next, we used evolutionary couplings (ECs) to address a long-standing hypothesis on the evolution of RNase P, an ancient ribozyme that has become a model system for investigating evolutionary plasticity of molecular structure (Krasilnikov et al., 2004). Whereas the bacterial RNase P (RF00010) can function as pure RNA, the widely diverged archeal (RF00373) and eukaryotic (RF00009) RNase P require their protein subunits (Gopalan, 2007). It is generally believed that during the evolution of RNase P in archaea and eukaryotes, the newly added protein subunits began to replace RNA to stabilize the overall tertiary architecture and establish the active site (Evans et al., 2006). However, a lack of knowledge of RNA-RNA intra-molecular interactions in the archeal and eukaryotic RNase P, which have no known structure, has made it difficult to critically evaluate this hypothesis about their evolution.

Given our high prediction accuracy for RNA-RNA contacts (94% precision for detecting base pairs within 12 Å) in the bacterial RNase P (Figure 7B, left), whose structure is known, we investigated whether our predicted contacts for the archeal and eukaryotic RNase P support the hypothesis that proteins have replaced RNA to establish key tertiary interactions (see the Experimental Procedures for the selection of contacts). Consistent with this hypothesis, we found several co-evolving pairs of nucleotides from the bacterial RNase P that do not correspond to any high-ranking ECs in the archeal or eukaryotic lineage, but emanate from helices that are now the sites of protein binding in eukaryotes (Figure 7B, right). For example, contacts (3A-95G and 95G-341U in the numbering of 3q1q.pdb [Reiter et al., 2010]) connecting P1 to P9 in bacteria are not predicted for the eukaryotic RNase P, where P9 now binds the protein Pop4 (Khanova et al., 2012). Similarly, we predicted contacts (43C-299C, 44C-299C, and 23C-299C) from P3 to P18 in bacteria, but not in eukaryotes, where P3 now binds the protein Pop7 (Khanova et al., 2012).

Alongside predictions that support an increased role for proteins in establishing the tertiary structure of eukaryotic RNase P, we also find high-ranking ECs that do not support this role, but instead suggest that RNA-RNA tertiary contacts have emerged de novo in the eukaryotic lineage to compensate for the loss of key secondary structural elements. These contacts, including G5-A86 (RFAM reduced numbering) from P1 to 3' of P8 and C4-G234 from P1 to conserved region IV (CR IV), significantly co-evolve in eukaryotes, but correspond to no high-ranking ECs in bacteria. Notably, even though the bacterial homologs of these base pairs do not significantly coevolve, they are nevertheless close in 3D space (<2 nm).

DISCUSSION

We show that evolutionary couplings (ECs) derived from sequence co-variation using a global maximum entropy model can predict 3D contacts in RNA, including long-range tertiary contacts and non-Watson-Crick base pairs, with good accuracy. While this paper was in review, De Leonardis et al. (2015) demonstrated a related methodology that retrieves similar contacts. Our predicted 3D contacts allow blinded 3D structure prediction

using only sequence information with accuracy matching state-of-the-art biochemical probing data. ECs not only provide a window into RNA structure but also can reveal functionally important interactions, such as tertiary contacts that distinguish the ligand-bound from the ligand-free conformation of four riboswitches or the potentially complex-nucleating interaction between the RRE RNA and the Rev protein of HIV virus. Since they are computed from broad alignments of homologous sequences, EC-inferred contacts are likely stable in diverse evolutionary contexts and conserved across evolution. Thus, in contrast to unbiased experimental approaches, 3D information from ECs has the unique potential to highlight functionally important interactions in structured RNA.

Building on the success of ECs at detecting contacts in RNA and protein separately, we provide proof of principle for the detection of intermolecular contacts in RNA-protein complexes from evolutionary sequence information alone. High-throughput elucidation of RNA-protein interactions from sequence alone would be a major breakthrough, but requires the development of improved bioinformatics approaches for curating and phasing RNA-protein alignments.

We identify several areas of growth for our approach. One shortcoming is that we always report the top $L/2$ contacts (where L is the length of the RNA), whereas flexibly choosing the number of contacts would better suit the natural diversity of 3D contact abundance in different RNAs. Further improvements can be made in folding pipeline, which is currently not optimized for large RNAs (>120 nt), despite the fact the evolutionary couplings yield highly accurate contacts for RNAs as large as the ribosome. Potential changes include higher sampling density, use of negative restraints, multi-stage annealing with continual addition of contacts, better ranking of decoy structures, and better filtering of false-positive contacts.

Another area for future research is to build on the outstanding work of the RFAM database (Nawrocki et al., 2015) by extending sequence coverage and including more transcripts that potentially have 3D structure. Though detection of accurate ECs for RNA requires far less sequence coverage than for proteins due to their smaller alphabet size and lower average number of contacts per residue, sequence abundance is still the rate-limiting component of our approach. Fortunately, the ongoing explosion of available sequence data means that the outlook for elucidating functional interactions in mRNAs, lncRNAs, and viral genomes, as well as their protein-binding partners, is promising.

EXPERIMENTAL PROCEDURES

Calculating ECs

To calculate ECs, we fit a global probability model to each alignment using pseudo maximum likelihood, as described below. Our full pipeline as single script is available at <https://github.com/debbiemarkslab/plmc>.

Sample Reweighting and Definition of M_{eff}

To prevent phylogenetic correlations between sequences from biasing our analysis, we down-weight each sequence σ by the number $N(\sigma)$ of neighbors it has in a sequence space. We then define the “effective size” of the alignment as the sum of weights:

$$M_{\text{eff}}(A) = \sum_{\sigma \in A} \frac{1}{N(\sigma)}$$

Model Fitting

We fit the following probability model using a pseudo-maximum likelihood approximation (Besag, 1975) and L_2 regularization.

$$P(\sigma) = \frac{1}{Z} \exp \left(\sum_{i=1}^L h_i(\sigma_i) + \sum_{i=1}^{L-1} \sum_{j=i+1}^L J_{ij}(\sigma_i, \sigma_j) \right)$$

h_i represents single-site conservation, and J_{ij} represents co-variation.

Post-processing

Once the parameters \mathbf{h} and \mathbf{J} have been fit to the data, we take the Frobenius norm $FN(i, j)$ of the J_{ij} couplings. In theory, the resultant FN scores represent the strength of co-variation at each position. However, undersampling effects lead to a characteristic distortion of these scores, and we correct for this using an average product correct (APC) (Dunn et al., 2008). APC-corrected FN scores are reported as the final ECs.

3D Structure Prediction

We predicted all-atom structures for five RNA families, representing the subset that (1) have a known structure, (2) have a length between 70–120 nt, and (3) have at least one highly long-range contact, defined as contact with $d_{ss} \geq L/4$, where L is the length of the RNA. We performed structure prediction with Nucleic Acid Simulation Tool (NAST) (Jonikas et al., 2009), a coarse-grained modeler that uses a combination secondary structure and tertiary contacts as inputs (Figure S7). For each RNA family, we created 600–1,000 decoy coarse-grained models and then clustered 20% of the decoys with the lowest energy-per-contact using k-means, where $k = 4$. The final four candidate models were then given all-atom structures and refined in XPLORE (Schwieters et al., 2003). All four models for each family are shown in Figure S3, and the ones with closest match to the experimental structure are shown in Figure 3.

RNA-Protein Structure Prediction

To compute RNA-protein ECs, we used the same approach as for RNA (described above) but now with a full alphabet, including all amino acids. No other changes were made to the model. To determine whether EC-derived RNA-protein contacts improve 3D structure prediction of RNA-protein complexes, we used these contacts as restraints for rigid body docking in HADDOCK (Dominguez et al., 2003). For docking controls, we only applied center of mass restraints.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.03.030>.

AUTHOR CONTRIBUTIONS

C.W. designed and conducted the experiments, performed the analysis, and wrote the paper. A.J.R. and C.W. jointly performed the RNA-protein studies. J.B.I. wrote the code for calculating ECs. T.G. performed the initial exploratory experiments. C.S. discussed the analysis and helped write the paper. D.S.M. designed the experiments, helped with the analysis, and wrote the paper.

ACKNOWLEDGMENTS

We thank the Marks lab for discussions and G. Park for help with figures. A.J.R. was supported by a DOE fellowship (DE-FG02-97ER25308). J.B.I. was supported by an NSF Graduate Research Fellowship. D.S.M., T.G., and C.S. were supported by a NIH grant (R01 GM106303).

Received: October 21, 2015

Revised: January 15, 2016

Accepted: March 18, 2016

Published: April 14, 2016

REFERENCES

- Adachi, A., Gendelman, H.E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M.A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *J. Virol.* *59*, 284–291.
- Bai, Y., Tambe, A., Zhou, K., and Doudna, J.A. (2014). RNA-guided assembly of Rev-RRE nuclear export complexes. *eLife* *3*, e03656.
- Bartel, D.P., Zapp, M.L., Green, M.R., and Szostak, J.W. (1991). HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* *67*, 529–536.
- Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician* *24*, 179–195.
- Butcher, S.E., and Pyle, A.M. (2011). The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* *44*, 1302–1311.
- Cao, S., and Chen, S.-J. (2011). Physics-based de novo prediction of RNA 3D structures. *J. Phys. Chem. B* *115*, 4216–4226.
- Caserta, E., Liu, L.C., Grundy, F.J., and Henkin, T.M. (2015). Codon-anticodon recognition in the *Bacillus subtilis* glyQS T box riboswitch: RNA-dependent codon selection outside the ribosome. *J. Biol. Chem.* *290*, 23336–23347.
- Casu, F., Duggan, B.M., and Hennig, M. (2013). The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophys. J.* *105*, 1004–1017.
- Charpentier, B., Stutz, F., and Rosbash, M. (1997). A dynamic in vivo view of the HIV-1 Rev-RRE interaction. *J. Mol. Biol.* *266*, 950–962.
- Cheng, C.Y., Chou, F.C., Kladwang, W., Tian, S., Cordero, P., and Das, R. (2015). Consistent global structures of complex RNA states through multidimensional chemical mapping. *eLife* *4*, e07600.
- Crothers, D.M., Seno, T., and Söll, G. (1972). Is there a discriminator site in transfer RNA? *Proc. Natl. Acad. Sci. USA* *69*, 3063–3067.
- Das, R., and Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. USA* *104*, 14664–14669.
- Das, R., Karanicolas, J., and Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods* *7*, 291–294.
- Daugherty, M.D., Booth, D.S., Jayaraman, B., Cheng, Y., and Frankel, A.D. (2010). HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proc. Natl. Acad. Sci. USA* *107*, 12481–12486.
- De Leonardis, E., Lutz, B., Ratz, S., Cocco, S., Monasson, R., Schug, A., and Weigt, M. (2015). Direct-coupling analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction. *Nucleic Acids Res.* *43*, 10444–10455.
- Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* *505*, 696–700.
- Dominguez, C., Boelens, R., and Bonvin, A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* *125*, 1731–1737.
- Dunn, S.D., Wahl, L.M., and Gloor, G.B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* *24*, 333–340.
- Dutheil, J.Y., Jossinet, F., and Westhof, E. (2010). Base pairing constraints drive structural epistasis in ribosomal RNA sequences. *Mol. Biol. Evol.* *27*, 1868–1876.
- Eddy, S.R. (2014). Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Annu. Rev. Biophys.* *43*, 433–456.
- Ehresmann, C., Baudin, F., Mougél, M., Romby, P., Ebel, J.P., and Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Res.* *15*, 9109–9128.
- Evans, D., Marquez, S.M., and Pace, N.R. (2006). RNase P: interface of the RNA and protein worlds. *Trends Biochem. Sci.* *31*, 333–341.

- Fernandes, J., Jayaraman, B., and Frankel, A. (2012). The HIV-1 Rev response element: an RNA scaffold that directs the cooperative assembly of a homooligomeric ribonucleoprotein complex. *RNA Biol.* 9, 6–11.
- Fox, G.E., and Woese, C.R. (1975). 5S RNA secondary structure. *Nature* 256, 505–507.
- Frellsen, J., Moltke, I., Thiim, M., Mardia, K.V., Ferkinghoff-Borg, J., and Hamelryck, T. (2009). A probabilistic model of RNA conformational space. *PLoS Comput. Biol.* 5, e1000406.
- Freyhult, E., Moulton, V., and Gardner, P. (2005). Predicting RNA structure using mutual information. *Appl. Bioinformatics* 4, 53–59.
- Gallego, J., and Varani, G. (2001). Targeting RNA with small-molecule drugs: therapeutic promise and chemical challenges. *Acc. Chem. Res.* 34, 836–843.
- Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126.
- Garst, A.D., Edwards, A.L., and Batey, R.T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* 3, 3.
- Gopalan, V. (2007). Uniformity amid diversity in RNase P. *Proc. Natl. Acad. Sci. USA* 104, 2031–2032.
- Green, N.J., Grundy, F.J., and Henkin, T.M. (2010). The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett.* 584, 318–324.
- Grigg, J.C., and Ke, A. (2013). Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure* 21, 2025–2032.
- Grundy, F.J., Rollins, S.M., and Henkin, T.M. (1994). Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis* tyrS gene: a new role for the discriminator base. *J. Bacteriol.* 176, 4518–4526.
- Grundy, F.J., Winkler, W.C., and Henkin, T.M. (2002). tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci. USA* 99, 11121–11126.
- Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., and Stormo, G.D. (1992). Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.* 20, 5785–5795.
- Heaphy, S., Finch, J.T., Gait, M.J., Karn, J., and Singh, M. (1991). Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich “bubble” located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. USA* 88, 7366–7370.
- Hofacker, I.L., and Lorenz, R. (2014). Predicting RNA structure: advances and limitations. *Methods Mol. Biol.* 1086, 1–19.
- Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319, 1059–1066.
- Hopf, T.A., Colwell, L.J., Sheridan, R., Rost, B., Sander, C., and Marks, D.S. (2012). Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 149, 1607–1621.
- Hopf, T.A., Schärfe, C.P., Rodrigues, J.P., Green, A.G., Kohlbacher, O., Sander, C., Bonvin, A.M., and Marks, D.S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* 3, 3.
- Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Springer, M., Sander, C., and Marks, D.S. (2015). Quantification of the effect of mutations using a global probability model of natural sequence variation. *arXiv*, arXiv:151004612, <http://arxiv.org/abs/1510.04612>.
- Huang, W., Thomas, B., Flynn, R.A., Gavzy, S.J., Wu, L., Kim, S.V., Hall, J.A., Miraldi, E.R., Ng, C.P., Rigo, F.W., et al. (2015). DDX5 and its associated lncRNA Rmrp modulate TH17 cell effector functions. *Nature* 528, 517–522.
- Ippolito, J.A., and Steitz, T.A. (2000). The structure of the HIV-1 RRE high affinity rev binding site at 1.6 Å resolution. *J. Mol. Biol.* 295, 711–717.
- Iwai, S., Pritchard, C., Mann, D.A., Karn, J., and Gait, M.J. (1992). Recognition of the high affinity binding site in rev-response element RNA by the human immunodeficiency virus type-1 rev protein. *Nucleic Acids Res.* 20, 6465–6472.
- Jonikas, M.A., Radmer, R.J., Laederach, A., Das, R., Pearlman, S., Herschlag, D., and Altman, R.B. (2009). Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 15, 189–199.
- Khanova, E., Esakova, O., Perederina, A., Berezin, I., and Krasilnikov, A.S. (2012). Structural organizations of yeast RNase P and RNase MRP holoenzymes as revealed by UV-crosslinking studies of RNA-protein interactions. *RNA* 18, 720–728.
- Kjems, J., Calnan, B.J., Frankel, A.D., and Sharp, P.A. (1992). Specific binding of a basic peptide from HIV-1 Rev. *EMBO J.* 11, 1119–1129.
- Klingler, T.M., and Brutlag, D.L. (1993). Detection of correlations in tRNA sequences with structural implications. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1, 225–233.
- Krasilnikov, A.S., Xiao, Y., Pan, T., and Mondragón, A. (2004). Basis for structural diversity in homologous RNAs. *Science* 306, 104–107.
- Laing, C., and Schlick, T. (2010). Computational approaches to 3D modeling of RNA. *J. Phys. Condens. Matter* 22, 283101.
- Latham, J.A., and Cech, T.R. (1989). Defining the inside and outside of a catalytic RNA molecule. *Science* 245, 276–282.
- Lee, C.P., Mandal, N., Dyson, M.R., and RajBhandary, U.L. (1993). The discriminator base influences tRNA structure at the end of the acceptor stem and possibly its interaction with proteins. *Proc. Natl. Acad. Sci. USA* 90, 7149–7152.
- Legiewicz, M., Badorrek, C.S., Turner, K.B., Fabris, D., Hamm, T.E., Rekosh, D., Hammarskjöld, M.L., and Le Grice, S.F. (2008). Resistance to RevM10 inhibition reflects a conformational switch in the HIV-1 Rev response element. *Proc. Natl. Acad. Sci. USA* 105, 14365–14370.
- Lescoute, A., Leontis, N.B., Massire, C., and Westhof, E. (2005). Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* 33, 2395–2409.
- Levitt, M. (1969). Detailed molecular model for transfer ribonucleic acid. *Nature* 224, 759–763.
- Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W.M., Henkin, T.M., and Ke, A. (2010). SAM recognition and conformational switching mechanism in the *Bacillus subtilis* yitJ S box/SAM-I riboswitch. *J. Mol. Biol.* 404, 803–818.
- Luedtke, N.W., and Tor, Y. (2003). Fluorescence-based methods for evaluating the RNA affinity and specificity of HIV-1 Rev-RRE inhibitors. *Biopolymers* 70, 103–119.
- Magnus, M., Matelska, D., Lach, G., Chojnowski, G., Boniecki, M.J., Purta, E., Dawson, W., Dunin-Horkawicz, S., and Bujnicki, J.M. (2014). Computational modeling of RNA 3D structures, with the aid of experimental restraints. *RNA Biol.* 11, 522–536.
- Malim, M.H., and Cullen, B.R. (1991). HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* 65, 241–248.
- Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE* 6, e28766.
- Marks, D.S., Hopf, T.A., and Sander, C. (2012). Protein structure prediction from sequence variation. *Nat. Biotechnol.* 30, 1072–1080.
- Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell* 136, 719–730.
- McManus, C.J., and Graveley, B.R. (2011). RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* 21, 373–379.
- Miao, Z., Adamiak, R.W., Blanchet, M.F., Boniecki, M., Bujnicki, J.M., Chen, S.J., Cheng, C., Chojnowski, G., Chou, F.C., Cordero, P., et al. (2015). RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* 21, 1066–1084.
- Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216, 585–610.
- Moazed, D., and Noller, H.F. (1986). Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* 47, 985–994.

- Mokdad, A., and Frankel, A.D. (2008). ISFOLD: structure prediction of base pairs in non-helical RNA motifs from isostericity signatures in their sequence alignments. *J. Biomol. Struct. Dyn.* *25*, 467–472.
- Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., and Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. USA* *108*, E1293–E1301.
- Mortimer, S.A., Kidwell, M.A., and Doudna, J.A. (2014). Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* *15*, 469–479.
- Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* *29*, 2933–2935.
- Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., and Finn, R.D. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* *43*, D130–D137.
- Novikova, I.V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Sizing up long non-coding RNAs: do lncRNAs have secondary and tertiary structure? *Bio-Architecture* *2*, 189–199.
- Nussinov, R., and Jacobson, A.B. (1980). Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. USA* *77*, 6309–6313.
- Olsen, H.S., Nelbock, P., Cochrane, A.W., and Rosen, C.A. (1990). Secondary structure is the major determinant for interaction of HIV rev protein with RNA. *Science* *247*, 845–848.
- Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* *3*, e02030.
- Pang, P.S., Jankowsky, E., Wadley, L.M., and Pyle, A.M. (2005). Prediction of functional tertiary interactions and intermolecular interfaces from primary sequence data. *J. Exp. Zool. B Mol. Dev. Evol.* *304*, 50–63.
- Parisien, M., and Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* *452*, 51–55.
- Peterson, R.D., and Feigon, J. (1996). Structural change in Rev responsive element RNA of HIV-1 on binding Rev peptide. *J. Mol. Biol.* *264*, 863–877.
- Pollom, E., Dang, K.K., Potter, E.L., Gorelick, R.J., Burch, C.L., Weeks, K.M., and Swanstrom, R. (2013). Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.* *9*, e1003294.
- Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* *24*, 651–663.
- Ramani, V., Qiu, R., and Shendure, J. (2015). High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* *33*, 980–984.
- Rausch, J.W., and Le Grice, S.F. (2015). HIV Rev assembly on the Rev response element (RRE): a structural perspective. *Viruses* *7*, 3053–3075.
- Reiter, N.J., Osterman, A., Torres-Larios, A., Swinger, K.K., Pan, T., and Mondragón, A. (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* *468*, 784–789.
- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* *81*, 145–166.
- Rivas, E., and Eddy, S.R. (1999). A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* *285*, 2053–2068.
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* *505*, 701–705.
- Rutherford, S.T., Valastyan, J.S., Taillefermier, T., Wingreen, N.S., and Bassler, B.L. (2015). Comprehensive analysis reveals how single nucleotides contribute to noncoding RNA function in bacterial quorum sensing. *Proc. Natl. Acad. Sci. USA* *112*, E6038–E6047.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* *160*, 65–73.
- Serganov, A., and Patel, D.J. (2012). Molecular recognition and function of riboswitches. *Curr. Opin. Struct. Biol.* *22*, 279–286.
- Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* *441*, 1167–1171.
- Shang, L., Xu, W., Ozer, S., and Gutell, R.R. (2012). Structural constraints identified with covariation analysis in ribosomal RNA. *PLoS ONE* *7*, e39383.
- Sherpa, C., Rausch, J.W., Le Grice, S.F., Hammarskjöld, M.L., and Rekosh, D. (2015). The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Res.* *43*, 4676–4686.
- Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* *350*, 978–981.
- Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., and Chang, H.Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* *519*, 486–490.
- Sreedhara, A., and Cowan, J.A. (2001). Targeted site-specific cleavage of HIV-1 viral Rev responsive element by copper aminoglycosides. *J. Biol. Inorg. Chem.* *6*, 166–172.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., and Chang, H.Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* *505*, 706–709.
- Warf, M.B., and Berglund, J.A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* *35*, 169–178.
- Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* *106*, 67–72.
- Zemmel, R.W., Kelley, A.C., Karn, J., and Butler, P.J. (1996). Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element. *J. Mol. Biol.* *258*, 763–777.
- Zhang, J., and Ferré-D'Amaré, A.R. (2013). Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* *500*, 363–366.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* *31*, 3406–3415.