

STRUCTURAL BIOLOGY

RNA structure from sequence

An approach taken from protein structure prediction makes it possible to computationally identify secondary and tertiary contacts in RNA.

Discoveries of new RNA species continue apace, presenting the noncoding RNA field with a growing question: what do these molecules do? The functions and molecular interactions of RNA often depend on its three-dimensional structure. Researchers have long tried to divine structure directly from sequence, but RNA is notoriously flexible, requiring that countless possible spatial configurations be sifted through to find the right structure.

One strategy uses evolutionary patterns to pinpoint residue combinations that have undergone selection for their roles in folding or function. This sequence covariation approach asks whether pairs of nucleotides change in tandem at specific positions of aligned RNA from different organisms, and it has been useful for finding secondary structure—the presence of Watson–Crick base pairs. But scientists have had little faith that it could be used to find tertiary contacts such as long-distance couplings and alternative base interactions. “The dogma was that RNA can only make one friend because it either base pairs or it doesn’t do anything,” says Deborah Marks, a computational biologist at Harvard University.

A key problem is that traditional approaches test for the correlation of each residue pair independently, making them vulnerable to confounding by transitivity. Caleb Weinreb, a graduate student in the Marks lab, gives a minimal example of transitivity. “You have three bases, A, B and C, and A and B are genuinely co-evolved because of biochemical contact, as are B and C. But because they’re both co-evolving in the same sequences simultaneously, you get a spurious correlation between A and C,” he explains. Spurious signal, which can be due to shared evolutionary history or

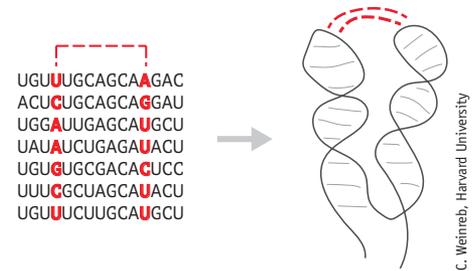
poor sampling, can be higher than signal from true contacts, Marks adds.

To make headway into RNA structure prediction, the team adapted a global evolutionary couplings model that the Marks lab had pioneered to predict amino acid contacts in proteins. “It was a surprise even for proteins, so we thought why not try it,” Marks recalls. The global model considers all possible pairwise contacts in an RNA molecule together, allowing it to deconvolve transitivity.

The researchers used it to generate a wide range of contact predictions, which they fed in some cases into modeling software to create all-atom models. They achieved good precision for 22 RNA families, including the long 40S ribosomal RNA, that had been worked out with painstaking crystallography and nuclear magnetic resonance spectroscopy. They also predicted contacts in riboswitches, tRNA, RNase P and 160 RNAs of unknown structure, among them a long noncoding RNA. In a striking case, they found that only one of two published HIV Rev response element (RRE) structures had strong support. In addition to finding long-range couplings, the algorithm was in fact better at predicting secondary structure than local sequence covariation methods such as mutual information.

Couplings between RNA-binding proteins and RNA can also be used in docking to give structures of the complex and detailed sites of RNA–protein interaction. The most highly co-evolving contacts between Rev protein and RRE were predicted to be precisely where Rev binding initiates an oligomerization event essential for nuclear export and viral function—a demonstration that the approach finds functionally relevant constraints.

Weinreb sees a direction for improving performance. “The biggest limiting factor for us was the bioinformatics challenge of finding enough RNA and protein sequences,” he says. Many long noncoding RNAs are



Bases at RNA structural contact points covary across evolution.

conserved among the vertebrates, for example, with too few sequences currently available for robust prediction. But as there are only four nucleotides for RNA, compared to 20 amino acids for proteins, RNA prediction often requires less than an order of magnitude fewer sequences for prediction. Marks notes that the level of divergence between sequences can also be as important as the number of sequences: “It depends on how much evolution has sampled and how much evolution we’ve sampled,” she says.

High-throughput experimental methods for probing RNA contacts are improving, but they indirectly detect base-pairing and not tertiary contacts, their resolution is typically limited to 5–10 nucleotides, and contacts are not necessarily enriched for functional significance. With evolutionary couplings, “you get information about which pairs were important in evolution for folding or constraints,” says Marks. The researchers are working on unpacking the parameters that affect coupling strength in order to derive quantitative information about the potential impact of mutations.

By making their efficient prediction tool freely available, Marks and Weinreb hope to make RNA structure prediction a routine step in exploring RNA function.

Tal Nawy

RESEARCH PAPERS

Weinreb, C. *et al.* 3D RNA and functional interactions from evolutionary couplings. *Cell* **165**, 1–13 (2016).

C. Weinreb, Harvard University