RNA structure and how to predict it

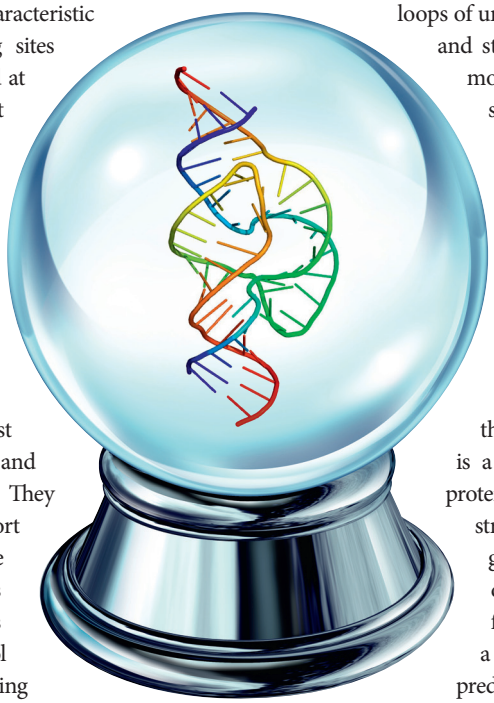
Clare Sansom
(Birkbeck College, London)

The fact that a protein's sequence determines its structure, and sequence and structure together determine function, is almost as widely known in biochemistry as the Central Dogma of Molecular Biology ('DNA makes RNA makes protein' in its shortest and most simplistic version). And thanks to the enormous effort that has gone into solving protein structures over the last few decades, it is possible to imagine a time when most proteins will either have a known or predicted structure, or else be recognized as disordered.

And that statement that sequence determines structure, and sequence and structure together determine function is just as true of nucleic acids as it is of proteins. But the field of nucleic acid structure analysis has, at least until recently, always felt like something of a poor relation. Until the 1990s, nucleic acid structures were not considered particularly interesting. There was only one DNA structure that anyone needed to know much about, the iconic double helix: elegant, essential, but always the same. And there was little point in trying to predict structures for two of the three main classes of RNA: messenger RNAs were simply considered to be linear single strands, and transfer RNAs known to fold into a characteristic 'clover-leaf' shape with binding sites for the codon and its amino acid at opposite ends. The only important RNA thought to have variable structures was the one that makes up the non-protein parts of the ribosome.

We now know that there are many other types of RNA that play subtler, but arguably just as important, roles in gene expression as messenger, transfer and ribosomal RNA. Most of these are found in eukaryotes and archaea or in eukaryotes only. They mainly consist of relatively short nucleotide sequences, as the names of some common types imply. Micro RNA (miRNA) is involved in the negative control of gene expression; small interfering RNA (siRNA) in gene silencing; small nuclear RNA in RNA splicing, and small nucleolar RNA in the chemical modification of other RNAs. Small segments of messenger RNA, which regulates the synthesis of protein from that RNA by binding small molecules , also have definitive secondary (and therefore tertiary) structures.

Unlike DNA, biologically functional RNA molecules almost always exist as single strands. They are, however, able to fold back on themselves and make short sections of helix in which two complementary sets of bases in different parts of the sequence form Watson-Crick base pairs: guanine with cytosine as in DNA, and adenine with uracil. Guanine can also form a non-canonical base pair with uracil, and the 'extra' hydroxyl group in the ribose sugar (as compared to deoxyribose in DNA) allows for a wider range of possible hydrogen bonds with the backbone. Where the sequence separating base-paired complementary sequences is fairly short, the resulting structure is termed a stem-loop. These structures are common in most types of RNA. The stems often contain bulges of unpaired residues on a single strand, or loops of unpaired residues on both strands, and stem-loops can combine to form more complex types of secondary structure including pseudo-knots.



Ribozyme image author: William G. Scott (from Wikipedia)

RNA structure prediction, like protein structure prediction, includes the prediction of secondary as well as tertiary structure. Secondary structure prediction of RNA involves predicting which parts of the sequence will form intra-molecular base pairs (including the stems of stem-loops). This is a complex task, but, unlike with proteins, the pattern of secondary structure elements will generally give a good idea of the three-dimensional tertiary structure or fold of the molecule. Therefore, a RNA secondary structure prediction is likely to be of more use

alone than a prediction of protein secondary structure.

There are a number of different methods of predicting the secondary structure of a RNA molecule, and several groups have set up web servers providing free access to one or more of these. Some, such as RNAfold from the University of Vienna and the 'Predict a Secondary Structure Web Server' from the University of Rochester, will also predict a structure for a single-stranded DNA sequence while recognizing that these are rare in nature. The simplest methods involve searching for the secondary structure – that is, the combination of base-paired and unpaired regions – that has the lowest free energy, known as the 'most stable structure'. Dynamic programming algorithms, which involve breaking the task down into simpler sub-tasks, are popular methods for finding this most stable structure, but they have disadvantages, particularly in that standard methods cannot locate pseudo-knots: a specific method for detecting these is sometimes used alongside this approach. Some programs can also align a set of homologous RNA sequences and generate a consensus sequence for secondary structure determination.

Once the secondary structure of a RNA sequence is known, the next challenge is to predict its tertiary structure. This is largely a matter of predicting the 3D structure of the bulges, loops and other non-base paired regions. Several approaches have been tried: iFoldRNA from the University of North Carolina is based on molecular dynamics, and RNAComposer from the Polish Academy of Sciences in Poznan builds up structures from a database that relates secondary structure elements to fragments of known tertiary structure.

And there is one more way in which the prediction of RNA structure echoes that of protein structure. You may have come across the computer game Fold It in which players fold protein chains into 3D structures. That, too, has its equivalent in the RNA world. The developers of eteRNA (http://eterna.cmu.edu/web/), from Stanford University and Carnegie Mellon Universities, include some of those who worked on Fold It. Players of the RNA game form complex secondary structures from the four ribonucleotide 'building blocks', and researchers synthesize and test the best designs. Even school children have tried it: will any #Bio readers have a go? ■

Best of the web

#RealTimeChem

Mark Burgess (Executive Editor)

You probably know the twitter feed #RealTimeChem, a Twitter-based community project designed to encourage chemists of all kinds to share what they are working on at any given time. It's worth looking at the website of the man who runs it (http://doctorgalacticandthelabcoatcowboy.com), a photochemistry researcher called Jason, now a publishing editor for the RSC.

He is an occasional blogger on all chemistry matters. For instance, in a couple of posts on Dark Knight he speculates on the metal used for Batarangs, how to make Batman's smoke bombs, the chemistry of explosives and of the incapacitating agents which might be used to tip the Bat-Darts. The last discusses general tranquillizers and sedatives before going on to Agent 15, an incapacitating agent weaponized by the US Army in the 1960s.

Jason wrote an especially interesting post on Lance Armstrong and the chemistry and effects of hormones, such as EPO (a glycoprotein hormone that controls red blood cell production), testosterone, corticosteroids and human growth hormone.

#RealTimeChemBanner for January 2015 from @theyakman

He also hands out awards to his 20 favourite #RealTimeChem tweets of the day and at the beginning of this year launched the #RealTimeChem Hub page. From here you can go to the #RealTimeChem: FAQ, RealTimeChemInFocus (a monthly guest blog post from a member of the #RealTimeChem community about their week in chemistry), #RealTimeChemBanners (where you can submit a banner to be used on the site for a month) and the #RealTimeChemPlaylist (chemistry-based tunes and hosted on Spotify). ■