RNA STRUCTURE

Loops and Pairs

from our Molecular Biology Correspondent

WHEN a few years ago the sequences of low-molecular weight RNAs began to appear, they were seized on by eager individuals, who, by pure contemplation, by model building or by computer methods, set off in pursuit of convincing base-pairing schemes. Before long a horrible murrain of hypothetical structures had spread through the literature. So far as any lessons were to be drawn from them, it was that in general—and the cloverleaf model by tRNA is probably an exception—there are as many feasible base-pairing schemes as there are authors. If then one's interest in predicting the base-pairing schemes from sequences is anything more than frivolous, there is little choice but to lay the ground rules by the use of specially tailored models. This of course is hard work, but as two new articles from Doty's laboratory at Harvard demonstrate, the effort is amply justified.

The difficulty that has beset the study of base-pairing in oligomers of A and of U, for example, has been the preferential formation of three-stranded structures. To overcome this problem Martin, Uhlenbeck and Doty (*J. Mol. Biol.*, **55**, 201; 1971) have synthesized block oligomers, in which a short run of U residues is grafted on to a similar run of A. The formation of three-stranded helices, (A+2U), from these species is scarcely feasible if the lengths of the blocks are similar, and indeed the spectroscopic evidence is strong in favour of double strands. Moreover, the melting curves in these systems show a strong concentration dependence, which indicates that association occurs between two molecules, and not by the formation of hairpins, which would preclude the pairing of as many residues as are required to make the turn. From the concentration-dependence of the melting temperature, a theoretical treatment worked out by Applequist can be used to derive the thermodynamic parameters associated with base-pairing. These include the enthalpy per base-pair, which turns out to depend on the helix length—a result which is not altogether unexpected, and may arise from a lesser contribution from the bases at the ends of the short duplex, or from the increased magnitude of the stacking interactions in the melted state, when the melting occurs at lower temperature.

When a solitary C is introduced between the A and U blocks, the double helix is interrupted in the middle by an unpaired residue in each strand, and there is accordingly a sharp decrease in stability. This type of effect is further pursued in a second paper (Uhlenbeck, Martin and Doty, *ibid.*, 217). In the first place they show that the effect of an unpaired G is much less severe than that of a C. This at last lends concrete substance to the widespread belief in the existence of G–U base-pairs: two antiparallel strands of the oligomer A_nGU_n can form two G–U pairs and six A–U pairs with an overlap of one unpaired A at either end—a scheme which is supported by the similar stability of chains in which the G is off-centre. The argument is that if the G residues were looped out of the helix, the effect of two single-base loops would assuredly be different from that of one bulge in the centre with two G residues opposed. When, on the other hand, a G–C pair is inserted, as in complexes of A_nGU_n with U_nCA_n, there is a large increase in stability, and this increases further in self-paired A_nGCU_n. A G–C pair is evidently worth 3–4 kcal/mole more than an A–U, and it is composition more than length which dominates the helix stability. A graphic illustration of the latter principle is also given by Coutts (*Biochim. Biophys. Acta*, **232**, 94; 1971), who finds a very high melting temperature for a fragment of tRNA containing four G–C pairs only.

To appreciate the value of data of this kind, one need only now examine the paper by Tinoco, Uhlenbeck and Levine in *Nature* last month (**230**, 362; 1971). Putting numbers to the stabilization brought about by each structural feature—+1 unit for each A–U pair, +2 for G–C and 0 for G–U, and various values from −2 to −7 for hairpin loops, looped out bases in one strand of duplex, and "bulges" of unpaired bases in both strands (these numbers being derived from melting data on models, processed according to straightforward statistical mechanical arguments)—they are able to estimate the relative stabilities of different pairing schemes for a given sequence. Some of the data are still tentative, and will no doubt be improved as more apposite model compounds are synthesized and examined, but it is

Detecting Proviruses

IF the RNA tumour viruses replicate by making with reverse transcriptase a DNA provirus—a DNA copy of the RNA viral genome—and if this provirus is integrated into the chromosomes of infected cells, it should, in theory at least, be possible to detect a provirus by DNA/RNA hybridization experiments. Radioactively labelled viral RNA should hybridize specifically with the proviral DNA of an infected cell. There are, of course, enormous technical difficulties, most of which stem from the fact that the proviral DNA must constitute only a minute fraction of the total DNA in an infected cell, and the past years have witnessed a series of claims and counter-claims; the optimists have claimed to detect such specific hybridization and the pessimists have argued the hybridization detected is not specific. But as the report by Baluda and Markham in next Wednesday's *Nature New Biology* suggests, the pendulum is decidedly swinging in favour of the optimists.

Baluda and his colleagues have repeatedly detected some four to ten-fold greater hybridization between the DNA of transformed chick cells and the RNA of the transforming virus—either avian myeloblastosis virus or avian sarcoma virus—than between mouse fibroblast DNA and the RNA of these avian tumour viruses. It seems therefore that transformed chick cells have DNA sequences complementary to the transforming virus's RNA. But the provirus hypothesis demands more than that. If the DNA provirus is to act as a template from which progeny RNA molecules, destined to become the genomes of progeny virus particles, can be synthesized, then the provirus must contain the entire viral genome.

How can one test whether the viral RNA which hybridizes with a fraction, the putative provirus, of the DNA of a transformed cell comprises the whole virus genome rather than a few regions of that genome? The obvious answer is to recover from the DNA/RNA hybrid the viral RNA and characterize it. But this presents a fresh set of technical difficulties, for it is crucially important not only to recover all the RNA which has specifically and completely hybridized to the cellular DNA, but also to exclude the RNA which has only partially and non-specifically hybridized to the DNA. Faced with these problems Baluda and Markham first fragmented the viral RNA to small pieces, then they hybridized these to DNA of cells transformed by the particular virus, washed the hybrids with saline detergent solutions to eliminate partial hybrids and finally recovered the RNA fragments which had specifically hybridized to the DNA.

To show that this collection of RNA fragments includes the entire viral genome, Baluda and Markham compared the average base composition of intact viral RNA genomes with that of the fragments of RNA recovered from the hybrids. The two compositions are almost identical. The result is therefore at least consistent with the notion, although it does not prove it, that the provirus is a DNA copy of the complete tumour virus RNA genome.