Although a detailed study of the accretional time span and profile is not included here, simple calculations show that, because of the moderate core temperature today, the accretional time span could not have been more than 5,000 yr and possibly much less. An accretional model, together with the magnetometer data, suggests that the lunar core is composed of primordial matter, never melted, and that the interior temperature is rising. We consider the mantle to be a primordial mass which differentiated *in situ* and which should display the differentiate spectrum of an isolated volume which has undergone strong thermal working. The lava flows on the surface then correspond to outflow from this reservoir.

We thank C. W. Snyder, M. Neugebauer, J. H. Wolfe and the Vela group at Los Alamos Scientific Laboratory for plasma data for testing the field line confinement problem.

Received February 1, 1971.

[1] Sonett, C. P., Dyal, P., Colburn, D. S., Smith, B. F., Schubert, G., Schwartz, K., Mihalov, J. D., and Parkin, C. W., *Proc. Intern. Astron. Union, Brighton* (Reidel, Dordrecht, in the press).
[2] Sonett, C. P., Dyal, P., Parkins, C. W., Colburn, D. S., Mihalov, J. D., and Smith, B. F., *Science* (in the press).
[3] Sonett, C. P., and Dyal, P., *Commun. Astrophys. and Space Phys.*, **2**, 184 (1970).
[4] Dyal, P., Parkin, C. W., and Sonett, C. P., *Lunar Surface Magnetometer Experiment, NASA SP-235*, 55 (1970).
[5] Mihalov, J. D., Colburn, D. S., Currie, R. G., and Sonett, C. P., *J. Geophys. Res.*, **73**, 943 (1968).
[6] Sonett, C. P., Colburn, D. S., Currie, R. G., and Mihalov, J. D., in *Physics of the Magnetosphere* (edit. by Carovillano, R. L., McClay, J. F., and Radoski, H. R.) (Reidel, Dordrecht, 1968).
[7] Nagata, T., Rikitake, T., and Kono, M., *Proc. COSPAR Space Sci.*, **10** (1970).
[8] England, A. W., Simmons, G., and Strangway, D., *J. Geophys. Res.*, **73**, 3219 (1968).
[9] Wood, J. A., Dickey, J. S., Marvin, U. B., and Powell, B. J., *Proc. Apollo 11 Lunar Sci. Conf.*, **1**, 965 (1970).
[10] Horai, K., Simmons, G., Kanamori, H., and Wones, D., *Proc. Apollo 11 Lunar Sci. Conf.*, **3**, 2243 (1970).
[11] Murase, T., and McBirney, A. R., *Science*, **170**, 165 (1970).
[12] Kaula, W. M., *An Introduction to Planetary Physics* (Wiley, New York, 1968).
[13] Sonett, C. P., and Colburn, D. S., *The Moon*, **1**, No. 4 (1970).
[14] Reynolds, J. H., *Ann. Rev. Nucl. Sci.*, **17**, 253 (1967).
[15] Sonett, C. P., Colburn, D. S., and Schwartz, K., *Nature*, **219**, 924 (1968).
[16] Sonett, C. P., Colburn, D. S., Schwartz, K., and Keil, K., *Astron. Space Sci.*, **7**, 446 (1970).

Estimation of Secondary Structure in Ribonucleic Acids

IGNACIO TINOCO, jun., OLKE C. UHLENBECK & MARK D. LEVINE

Department of Chemistry, University of California, Berkeley, California 94720

A simple method for estimating the secondary structure of an RNA molecule has been proposed on the basis of the knowledge of its sequence.

THE nucleotide sequences of transfer RNA molecules[1], portions of viral messenger RNA molecules[2-8] and ribosomal RNA molecules[9-11] are rapidly becoming known. It is well established that in all RNA molecules, most nucleotide residues interact to form short, intramolecular helical regions and unpaired loops, so defining a secondary structure[12,13]. In many cases, further folding leads to a unique, biologically active tertiary structure. Even for molecules with known sequences, it is difficult to determine the secondary structure. Optical techniques can only estimate the number and type of residues in helical regions[14-16]. Other, more direct, methods often need large amounts of purified material, and frequently necessitate studying the structure in non-physiological conditions[17-20].

We propose here a simple method for estimating the secondary structure of an RNA molecule from knowledge of its sequence. This problem was first approached by Fresco, Alberts and Doty[12]. By allowing looped out bases and putting the sequence in register in order to maximize base pairing, they were able to suggest optimal secondary structures from a given sequence. When this approach was applied to transfer RNA sequences, a wide variety of sequences could all be put into a similar cloverleaf pattern with the anticodon in the same position, thus demonstrating the interchangeability of the tRNAs in protein biosynthesis on the ribosome[21]. Knowledge of the function of the molecule helped in the choice of a likely structure. For the 5S ribosomal RNA molecules, the lack of such functional knowledge and the large number of possibilities for intramolecular pairing led to a number of proposed models with no obvious correct choice[22-26]. Longer sequences are expected to lead to more ambiguous possibilities. More thermodynamic information on the relative stability of the loops and various helical regions is necessary. The estimates of secondary structures presented here are the results of analysis of a limited amount of data on double strand formation in oligonucleotides and complementary polynucleotides.

Rules for RNA Structure

The secondary structure which an RNA molecule takes up at equilibrium in any solution is that which minimizes the free energy of the solution. The free energy, ΔG, of a double

stranded structure relative to the single strand is written as the sum of enthalpy, ΔH, and entropy, ΔS, terms.

$$\Delta G = \Delta H - T\Delta S$$

For simplicity, we shall assume that both ΔH and ΔS are independent of the absolute temperature, T. It is convenient to consider that the formation of a double stranded region occurs in two steps: the formation of the first base pair (initiation) and the formation of subsequent base pairs (propagation). Both steps contribute to the total free energy of helix formation.

The initiation of a base paired region in a single stranded RNA molecule involves the formation of a loop. A stranded representation for the free energy of initiation is[27]

$$\Delta G \text{ (initiation)} = -2.3RT \log \gamma_m$$

where $R = 2$ calories mol^{-1} deg^{-1} and γ_m is the probability of the formation of a loop of m unbonded bases. γ_m depends on the identity and number of bases in the loop. As m decreases, the formation of the loop becomes more probable, but finally steric hindrance prevents loop formation.

The expression for the free energy of propagation of a helical region is simplified by the existence of a direct relationship between the enthalpy and entropy of propagation. Because the value of ΔG of initiation for a long double stranded RNA is negligible compared with that of propagation, and the melting point for the polymer, T_m^∞, is defined as the temperature at which equal amounts of double strand and single strand are in equilibrium, it follows that

$$\Delta G \text{ (propagation)} = 0 \text{ at } T_m^\infty$$

$$\Delta S \text{ (propagation)} = \frac{\Delta H \text{ (propagation)}}{T_m^\infty}$$

Therefore, the free energy of propagation for a short RNA double helix at a temperature T is given by

$$\Delta G \text{ (propagation)} = \Delta H\left(1 - \frac{T}{T_m^\infty}\right)$$

The value of T_m^∞ used depends on the guanine–cytosine content of the short helix and the ionic strength of the solvent. The value of ΔH, in principle, depends on the chain length, the base composition and the base sequence of the short helix.

The total free energy of double strand formation in an RNA molecule is the sum of initiation and propagation terms for each different helical region.

$$\Delta G = \Delta H\left(1 - \frac{T}{T_m^\infty}\right) - 2.3\ RT \log \gamma_m$$

The structure with the lowest total free energy will be the most stable and probably the one actually present.

By studying double strand formation in oligonucleotides and polynucleotides of defined base composition and sequence, values can be obtained for the parameters given here in various solvents, and thus we can extrapolate to any proposed structure. The principal difficulty is that there is only a limited amount of thermodynamic information on suitable models. This means that at present a considerable number of approximations must be made, which can be removed when more data become available.

It is efficient to present the conclusions first and then discuss the experiments which led to them. Information about the stability of double stranded regions came from melting temperature and thermodynamic data on double stranded polyribonucleotides[28] and oligoribonucleotides[29,30]. The probability of loop formation has been estimated from the melting curves of DNA[31], from oligo dAT melting behaviour[27] and from work on the formation of loops in $A_6C_mU_6$ (Uhlenbeck and Borer, unpublished results). The stability of bulges was estimated from the melting temperatures of partially mismatched double stranded polynucleotides[32,33]. These experiments suggested assignment of the following stability numbers for predicting the most stable secondary structure in an RNA at 25° C in a neutral buffer of moderate or high ionic strength: (1) A–U pairs, +1; (2) G–C pairs, +2; (3) G~U pairs, 0; (4) hairpin loops, −5 to −7; (5) interior loops, −4 to −7; (6) bulges, −2 to −6.

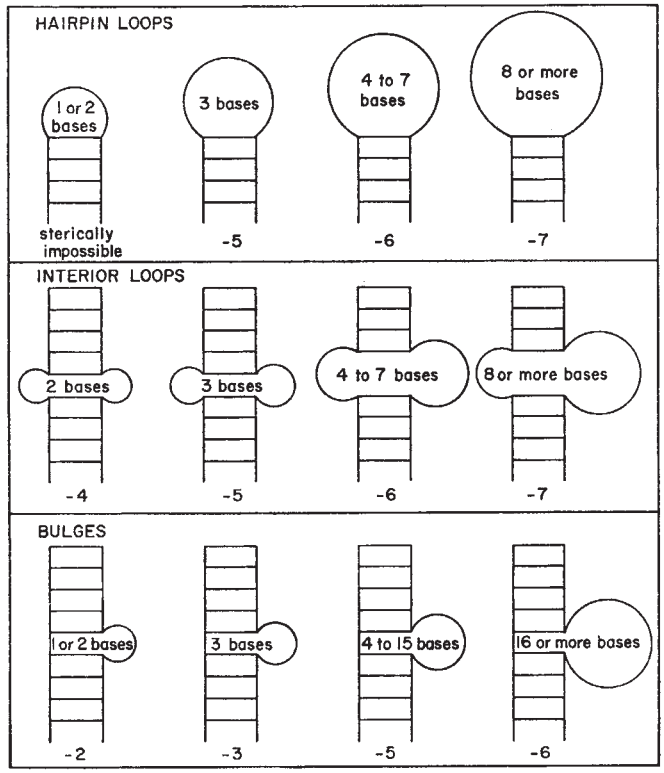

Fig. 1 Giving the contributions of loops and bulges with unbonded bases to the stability of secondary structure in nucleic acids. A positive number represents an increase in stability relative to the single strand; an A–U base pair has a stability number of +1.

The numbers for the different loops and bulges are defined and illustrated in Fig. 1. The stability number for a given RNA secondary structure is the sum of the contributions of the loops, bulges and helices; the structure with the highest number is the most stable. The free energy (and stability number) of the completely single stranded molecule is, by definition, equal to zero. Stable structures have positive numbers measured in units of the free energy of forming an A–U base pair (found to be −1.2 kcalories mol^{-1} at 25° C). A negative stability number characterizes a secondary structure unstable with respect to the single strand. Each base pair formed obviously increases the stability, but each loop decreases the stability. Thus, at least two G–C pairs and one A–U pair are necessary to close a hairpin loop.

As an example, we consider the sequence of fifty-five bases in the coat protein cistron in R17 viral RNA sequenced by Adams *et al.*[3]. All possible pairs between A–U, G–C and G~U in the sequence can be represented by the base pairing matrix in Fig. 2, in which 1, 2 and 0 represent the respective stability numbers. Adjoining numbers on diagonals leading from the lower left to the upper right represent antiparallel base paired regions. Breaks or jogs in the diagonals represent interior loops or bulges. By joining different base paired regions in the pairing matrix, all possible secondary structures can be investigated and their stability numbers computed.

Although a direct search for likely structures is possible for sequences as short as that considered, the base pairing matrix is useful for longer sequences, and the search for structures and their subsequent evaluation is carried out by computer.

Three possible structures for the 55-mer are shown in Fig. 3 and are represented by diagonal lines joining squares in Fig. 2. Each structure was assigned a stability number (and a free energy), formed by summing the contributions listed. Structure I ($\Delta G = -3.6$ kcalories mol^{-1}) is essentially the one given by the authors who determined the sequence[3]. A slightly different structure II ($\Delta G = -9.6$ kcalories mol^{-1}), was predicted to be considerably more stable. In fact, in equilibrium at 25° C, the ratio of the concentration of structure II to structure I was calculated to be $10^{6.0/1.36} = 26{,}000$. Finally, a structure III was calculated to be slightly more stable ($\Delta G = -4.8$ kcalories mol^{-1}) than structure I although it has four fewer base pairs.

Table 1 Stability Numbers of Proposed 5S RNA Models

	5S stability number	No. AU+GC base pairs
Cantor[25]	30	43
Lewis[26]	25	32
Boedtker and Kelling[22]	17	32
Madison[24]	16.5	32
Raake[23]	−22	27

by Lewis and Doty[26] seem to be the most stable, other structures can be found which give comparable or even somewhat higher stability numbers. Many of these structures have an additional difficulty; they are highly compact three dimensional arrangements in which the possible destabilizing effect

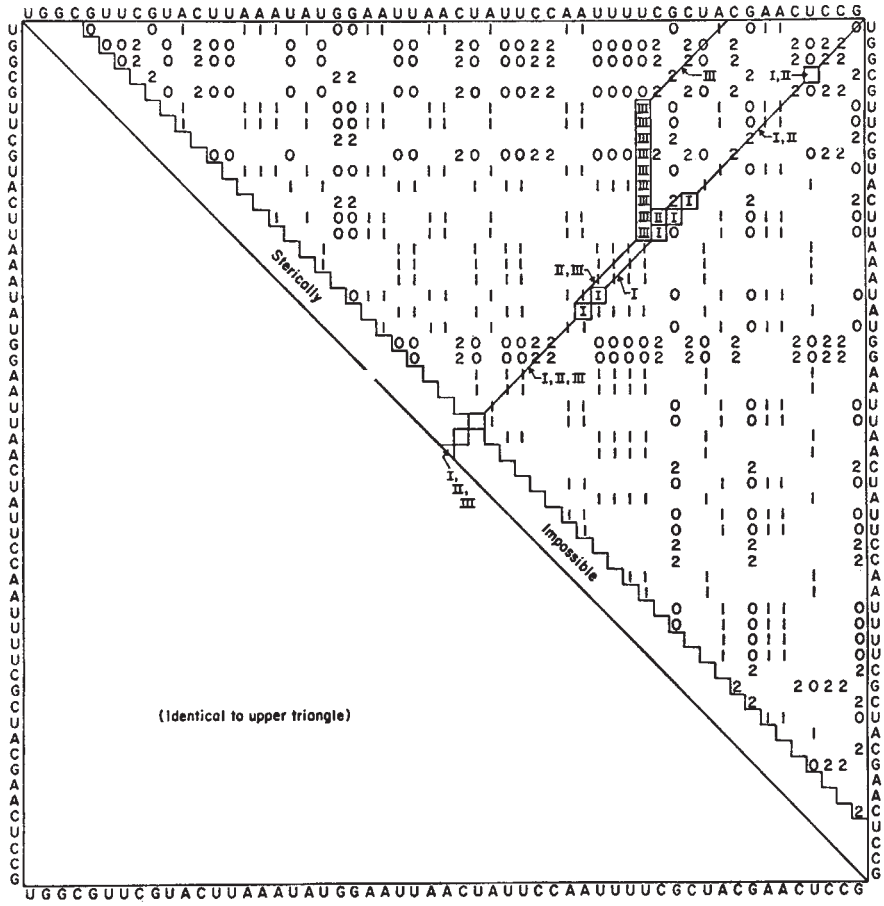

Fig. 2 The base pairing matrix showing all possible base pairs in a sequence of fifty-five bases from R17 viral RNA[3]. An upper right to lower left diagonal line joining base pairs represents an anti-parallel helical region. Squares represent loops between helical regions. A bulge is seen as a vertical or horizontal line. The three most stable structures (I, II, III) are indicated here and in Fig. 3.

Simply breaking the three A–U base pairs in the centre of structure I was calculated to improve its stability by 2.4 kcalories. Furthermore, the two G–C base pairs at the bottom of structures I and II are only marginally stable. No gain in stability was predicted by forming them. Similarly, the bottom helix in structure III is marginally stable. This emphasizes the fact that the practice of simply maximizing base pairs to predict secondary structure must be tempered by also minimizing the number of loops.

These simplified stability calculations have been performed for a number of different RNAs. It is not surprising that the cloverleaf models for tRNAs are essentially the most stable form, although base pairing between the loops of the cloverleaf is sometimes suggested. A comparison of the stability numbers of several proposed structures for *E. coli* 5S RNA is shown in Table 1. Although the structures proposed by Cantor[25] and of double helical regions constrained to be close to each other is not known. Although in this case the method is unlikely to predict an unambiguous stable structure, it is valuable in significantly limiting the number of possibilities. For the longer messenger RNA molecule, in which such complex structures are less likely, the application of the technique is more straightforward. As more data on model compounds are gathered, it should be possible to obtain a more sensitive set of stability numbers.

Thermodynamic Details

The assignment of free energies for the formation of A–U and G–C base pairs is quite direct. The ΔH of formation of double strands by complementary oligonucleotides has been determined for the molecules A_nU_n(n=4–7)[28], A_nGU_n+

$A_mCU_n(n,m=3\text{–}5)$, $A_nCGU_n(n=2\text{–}4)$[28], and $A_nGCU_n(n=2\text{–}4)$ (Uhlenbeck and Dengler, unpublished results) in 1 M NaCl, 0.01 M sodium phosphate buffer, 10^{-4} M sodium EDTA at *p*H 7. ΔH was evaluated from the concentration dependence of T_m with the equation[34]

$$\frac{1}{T_m} = \frac{1}{T_m^{\infty}} + 2.3\,\frac{R}{\Delta H}\log \kappa c$$

where T_m^{∞} is the melting point of a long polynucleotide with the same average base composition as the oligonucleotide, κ is the equilibrium constant for forming the first base pair between the oligonucleotides and c is the oligonucleotide strand concentration. This equation, which assumes that the formation of the double strand is an all or none process, is entirely adequate for the analysis of the oligomers considered. The data for the fifteen different double strands formed could be fitted by the equation

$$\Delta H = (N-1)(-8 \pm 0.7 \text{ kcalories mol}^{-1})$$

where N is the total number of base pairs. The choice of $(N-1)$ assumes that the formation of the first base pair in the helix has a negligible enthalpy, because it is devoid of a stacking interaction. Although the data suggest that ΔH depends on the base composition of the helix, the variation is not conclusive at the present degree of accuracy. This value of $\Delta\bar{H} = -8$ kcalories for adding a mole of base pairs is consistent with direct calorimetric values (-7.4 to -8.2 kcalories mol^{-1}) for the melting of polyadenylic acid : polyuridylic acid double strands in 0.02 M to 0.1 M Na^+ [35]. Also, DNA double strands with G–C contents which varied from 37% to 64% gave a value of -8.0 ± 0.2 kcalories mol^{-1} (ref. 36).

Having assigned a value to the enthalpy, we need only know the values of T_m to estimate the free energy for the formation of A–U and G–C bonds. For natural double stranded RNA, Kallenbach[28] has expressed the reciprocal of T_m as a linear function of the G–C fraction, f_G, as

$$\frac{1}{T_m} = \frac{1}{T_A} + f_G\left(\frac{1}{T_G} - \frac{1}{T_A}\right)$$

where T_A and T_G are the melting temperatures of polymers with 100% A–U and 100% G–C base pairs, respectively. The values of T_A and T_G were extrapolated from a series of natural RNA molecules of different G–C content and also agree with the measured T_m s of the appropriate synthetic polynucleotides[28]. The extrapolated values of T_m obtained from the oligonucleotide data were also consistent with the polymer data. For 1 M NaCl, *p*H 7, T_A is 78° C and T_G is 152° C [29] Thus, at 25° C in 1 M NaCl, *p*H 7, the free energies of base pair formation were estimated as:

$$\Delta G \text{ (addition of A–U base pair)} = \Delta\bar{H}\left(1 - \frac{T}{T_A}\right) = -1.2 \text{ kcalories mol}^{-1}$$

$$\Delta G \text{ (addition of G–C base pair)} = \Delta\bar{H}\left(1 - \frac{T}{T_G}\right) = -2.4 \text{ kcalories mol}^{-1}$$

With stability units of -1.2 kcalories mol^{-1} we obtained the value of $+1$ for an A–U base pair and $+2$ for a G–C base pair (rules 1 and 2). Because the values of T_A and T_G depend on salt concentration, the stability of the G–C pair relative to the A–U pair depends on the solvent. The values given are applicable in high sodium ion concentrations or solvents containing 10 mM magnesium ion or more. At lower ionic strengths, the G–C bond is slightly more stable than the A–U bond. For example, from T_m data in 0.2 M NaCl, we calculated

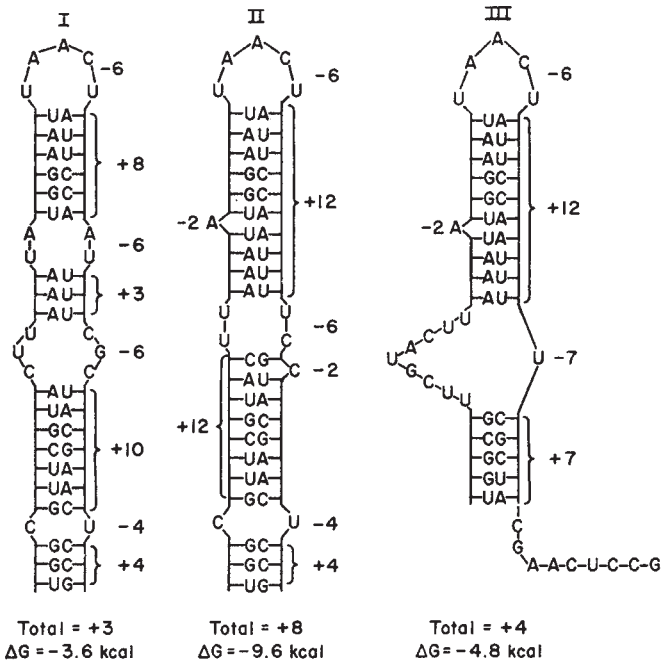

Fig. 3 Gives three possible secondary structures for a sequence of fifty-five bases from R17 viral RNA[3]. Structure I, a proposed structure[3], is calculated to be least stable. Structure II, a slightly different structure, is significantly more stable. Structure III, with fewer base pairs, is slightly more stable than structure I.

ΔG(A–U) $= -1.0$ kcalories mol^{-1} and ΔG(G–C) $= -2.2$ kcalories mol^{-1}. This difference is, however, not large enough to be significant at the present degree of approximation. The temperature at which the secondary structure is considered also affects the free energies. The ratio of G–C to A–U stability increases with increasing temperature, but is constant, to within 20%, in the temperature range 0° to 40° C.

The number zero for G ~ U (rule 3) came from experiments on double strand formation in $A_nGU_m(n,m=4,5)$ and $A_nUGU_n(n=4,5)$[30]. The data are consistent with having G ~ U bonds in the helix with either zero or slightly negative stability, rather than G forming a bulge in the A–U helix.

Experimental data for the free energies of loops are scarce, and essentially no work has been done for RNA as a function of the number and composition of bases in a loop. Some insight into the stability of loops, however, can be gained from chain statistics. Bases in a loop are less stable than the same bases in a single strand because the ends of the loop are fixed relative to each other, thereby decreasing the entropy of the chain and increasing the free energy. For a chain of m bases, the end to end distances in the single strand of which have a Gaussian distribution, the free energy of a loop relative to the single strand has the form[37]

$$\Delta G \text{ (initiation)} = -2.3RT[B - 1.5\log(m+1)]$$

where B is a parameter which characterizes the formation of the first base pair which closes the loop. The free energy depends on the number of statistical links to the -1.5 power[38]. We equated the number of statistical links to the number of unbonded bases plus one, $(m+1)$, but for a more realistic chain, in which the end to end distances are corrected for excluded volume (that is only self-avoiding conformations are counted), the factor of 1.5 was replaced by 1.75 [39]. The important conclusion from this equation is that, to the extent that entropy determines the loop stability, the free energy changes slowly with loop size; a loop of ten bases is only about 0.5 kcalories mol^{-1} less stable than one of six bases. This conclusion applies equally to interior loops. For small loops (m less than four or five bases), the free energy equation given

here is no longer applicable because significant deviation from Gaussian behaviour is expected and stacking interactions between the residues in the loop may play a considerable role.

It remains to estimate B in order to assign a stability to loops. As a first approximation to RNA loops, we consider the values of B deduced from experimental results for DNA. Crothers and Zimm[31] reported $B=-3.5$ and -3.9 from an analysis of the melting curves of dAT : dAT and dI : dC Crothers[40] obtained a value for B of between -3 and -4 for T2 phage DNA. Scheffler *et al.*[27] found $B=-2.35$ for oligo dAT loops with $m\geq 6$. For complementary double stranded RNA oligonucleotides, we have calculated values of B between -2 and -4 by assuming a correlation between intra and intermolecular initiation processes. A value of B of approximately -3.4 fits our preliminary data for the T_m of $A_6C_mU_6$. We therefore assigned an average value of -3 to B for loops with four or more bases.

We consider now loops with fewer than four bases. Space filling models show that it is impossible to make a hairpin loop with one or two bases. For the minimum hairpin loops of three bases, we estimated the free energy by analogy with the finding of Scheffler *et al.*[27], that the minimum loop is about 1.2 kcalories mol^{-1} more stable than expected from the equation for large loops. This same correction was applied to interior loops of three bases.

Unlike the case of hairpin loops, interior loops of two residues are sterically possible and correspond to two mismatched bases (not A–U, G–C, or G~U) on opposite strands. The oligonucleotides A_nCU_n($n=4$, 5, 7) which form dimers in solution are examples of such loops[18]. If we attribute the differences in T_m for these molecules from the corresponding A_nU_n series solely to ΔG (initiation) of the interior loop, we estimate a value of about $+5$ kcalories mol^{-1} for the formation of interior loops of two bases. This result is consistent with the expected stability of small loops over that of the prediction for Gaussian chains.

Bulges of any number of bases are sterically possible. By analysing the mixing curve for the complex formed between poly U and poly (A, U), Fresco and Alberts[41] showed that bulges are formed in preference to interior loops. This result probably reflects the fact that the formation of an interior loop involves an interruption of the helical region, thus causing the loss of a stacking interaction. This does not occur in the formation of a bulge, because one strand is not interrupted. Other workers have shown that U is not unique in its ability to form small bulges[32,33,42,43]. A rough estimate of the ΔG (initiation) can be obtained from the reported melting temperature of the poly U + poly (A, U)[32] and poly U + poly (A, I)[33] complexes, if the simplifying assumptions are made that the copolymer sequences are random and that each bulge may be treated as though it were of average size, determined from the fraction of mixmatched base in the copolymer. At a mole fraction of 0.1 I in poly (A, I) the average bulge in the complex has slightly more than one base; this yields a value of ΔG (initiation) of $+1.9$ kcalories mol^{-1} at 25° C. At a mole fraction of 0.47 U in poly (A, U), the average bulge in the complex has nearly two bases, but the ΔG (initiation) value is still found to be about $+1.9$ kcalories mol^{-1}.

With the various contributions of helix content, loops and bulges thus estimated, we can now write the expression for the free energy of forming a secondary structure relative to the single strand, as

$$\Delta G = N_{AU}\Delta\bar{H}\left(1-\frac{T}{T_A}\right)+N_{GC}\Delta\bar{H}\left(1-\frac{T}{T_G}\right)-\Delta\bar{H}\left(1-\frac{T}{T_m^{\infty}}\right)-2.3\,RT[B-1.5\log(m+1)]$$

The enthalpy for the formation of one base pair, $\Delta\bar{H}$, is taken to be -8 kcalories mol^{-1}. The first two terms are the free energies of adding N_{AU} A–U and N_{GC} G–C base pairs. The third term takes into account the fact that the free energy of a double strand region depends on $N-1$ rather than N. Each double strand region has a different T_m^{∞} associated with it because of variation in the fraction of A–U and G–C base pairs, but we used an average value of 1.8 kcalories mol^{-1} for this term. The fourth term depends on B, which we chose equal to -3, and on the number of unbonded bases, m, in each loop or bulge. As there must be one loop (not bulge) for each double strand region, we can combine the third term with the factor of -2.3 RTB from the fourth term. Thus, at 25° C we obtain ΔG in kcalories mol^{-1} as

$$\Delta G=-1.2\,N_{AU}-2.4\,N_{GC}+5.9+2.0\log(m+1)\ \text{(for each loop of at least four bases)}$$

or

$$+4.1+2.0\log(m+1)\ \text{(for each bulge of at least four bases)}$$

Dividing all the free energies in this expression by -1.2 kcalories mol^{-1} to transform them into A–U base pairing units, we obtain the numbers (rounded to the nearest integer) given in Fig. 1 for loops and bulges with four or more residues (rates 4–6 in the text). For loops of three bases and bulges of one, two, or three bases, the different values discussed earlier are given in Fig. 1.

We present these estimates of the energetics of forming secondary structure in RNA molecules on the basis of the present, limited data for the formation of loops, bulges, and short helices. More detailed analysis would consider the effect of temperature, ionic strength and base sequence on both ΔH and ΔS. In particular, the sequence of bases in a loop or bulge is likely to be important, because single stranded stacking should have an effect on the stability of such structures[44]. The partitioning of the free energy of the helical region into that of the formation of A–U, G–C and G~U pairs is a simplification, because the sequence dependent stacking between base pairs is the principal source of stability within the helix. Possible non-Watson–Crick base pairs were excluded from consideration and interaction among loops and helical regions within the structure was ignored. In spite of the approximations involved in the assessment of the thermodynamic parameters and the scarcities of data on loops and bulges, these rules should provide a useful framework for the evaluation of RNA secondary structures.

Received January 19, 1971.

1 Zachau, H., *Angew. Chem. Intern. Edit.*, **8**, 711 (1969).
2 Goodman, H. M., Billeter, M. A., Hindley, J., and Weissmann, C., *Proc. US Nat. Acad. Sci.*, **67**, 921 (1970).
3 Adams, J. M., Jeppeson, P. G. N., Sanger, F., and Barrell, B. G., *Nature*, **223**, 1009 (1969).
4 Steitz, J. A., *Nature*, **224**, 957 (1969).
5 Hindley, J. L., and Staples, D. H., *Nature*, **224**, 964 (1969).
6 Billeter, M. A., Dahlberg, J. E., Goodman, H. M., Hindley, J., and Weissmann, C., *Nature*, **224**, 1083 (1969).
7 Nichols, J. L., *Nature*, **225**, 147 (1970).
8 Symp. No. 5, *Nucleic Acid Structure, Eighth Intern. Cong. Biochem.*, 163 (1970).
9 Brownlee, G. G., Sanger, F., and Barrell, B. G., *Nature*, **215**, 735 (1967).
10 Forget, B. G., and Weissman, S. M., *Science*, **158**, 1695 (1967).
11 Fellner, P., Ehresmann, C., and Ebel, J. P., *Nature*, **225**, 26 (1970).
12 Fresco, J. R., Alberts, B. M., and Doty, P., *Nature*, **188**, 98 (1960).
13 Fresco, J. R., Klotz, L. C., and Richards, E. G., *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 83 (1963).
14 Cotter, R. I., McPhie, P., and Gratzer, W. B., *Nature*, **216**, 865 (1967).
15 Cantor, C. R., *Proc. US Nat. Acad. Sci.*, **59**, 478 (1968).
16 Thomas, jun., G. J., *Biopolymers*, **7**, 325 (1969).
17 Uhlenbeck, O. C., Baller, J., and Doty, P., *Nature*, **225**, 508 (1970).
18 Brostaff, S. W., and Ingram, V. M., *Biochemistry*, **9**, 2372 (1970).

[19] Cramer, F., Doepner, H., van de Haar, F., Schlimme, E., and Seidel, H., *Proc. US Nat. Acad. Sci.*, **61**, 1348 (1968).
[20] Ofengand, J., *J. Biol. Chem.*, **242**, 5034 (1967).
[21] Levitt, M., *Nature*, **224**, 759 (1969).
[22] Boedtker, H., and Kelling, D. G., *Biochem. Biophys. Res. Commun.*, **29**, 758 (1967).
[23] Raake, I. D., *Biochem. Biophys. Res. Commun.*, **31**, 528 (1968).
[24] Madison, J. T., *Ann. Rev. Biochem.*, **37**, 131 (1968).
[25] Cantor, C. R., *Nature*, **216**, 513 (1967).
[26] Lewis, J., and Doty, P., *Nature*, **225**, 510 (1970).
[27] Scheffler, I. E., Elson, E. L., and Baldwin, R. L., *J. Mol. Biol.*, **48**, 145 (1970).
[28] Kallenbach, N. R., *J. Mol. Biol.*, **37**, 445 (1968).
[29] Martin, F. H., Uhlenbeck, O. C., and Doty, P., *J. Mol. Biol.* (in the press).
[30] Uhlenbeck, O. C., Martin, F. H., and Doty, P., *J. Mol. Biol.* (in the press).
[31] Crothers, D. M., and Zimm, B. H., *J. Mol. Biol.*, **9**, 1 (1964).
[32] Steiner, R. F., *J. Biol. Chem.*, **235**, 2946 (1960).
[33] Steiner, R. F., *J. Biol. Chem.*, **236**, 842 (1961).
[34] Applequist, J., and Damle, V., *J. Amer. Chem. Soc.*, **87**, 1450 (1965).
[35] Krakauer, H., and Sturtevant, J. M., *Biopolymers*, **6**, 491 (1968).
[36] Bunville, L. G., Geiduschek, E. P., Rawitscher, M. A., and Sturtevant, J. M., *Biopolymers*, **3**, 213 (1965).
[37] Poland, D., and Scheraga, H. A., *Theory of Helix-Coil Transitions in Biopolymers*, 188 (Academic Press, New York, 1970).
[38] Jacobson, H., and Stockmayer, W. H., *J. Chem. Phys.*, **18**, 1600 (1950).
[39] Fisher, M. E., *J. Chem. Phys.*, **45**, 1469 (1966).
[40] Crothers, D., *Biopolymers*, **6**, 391 (1968).
[41] Fresco, J. R., and Alberts, B. M., *Proc. US Nat. Acad. Sci.*, **46**, 311 (1960).
[42] Bautz, E. K. F., and Bautz, F. A., *Proc. US Nat. Acad. Sci.*, **52**, 1476 (1964).
[43] Ichikawa, T., Fujita, H., Matsuo, K., and Tsuboi, M., *Bull. Chem. Soc. Japan*, **40**, 2272 (1967).
[44] Wintermeyer, W., Thiebe, R., Zachan, H. G., Riesner, D., Romer, R., and Maass, G., *FEBS Lett.*, **5**, 23 (1969).

Induction of Directional Chromosome Elimination in Somatic Cell Hybrids

G. PONTECORVO

Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2

The chromosomes of one or other of the "parents" of a hybrid cell are usually gradually lost over successive generations. X-irradiation, γ-irradiation or BUDR labelling of one "parent" before fusion can be used to predetermine which of the two sets of chromosomes will be preferentially lost.

THE fusion of somatic cells of different species produces hybrids which, in most cases, preferentially lose some or all of the chromosomes of one of the parental species on prolonged subculture. So far, however, there has been no general way of predetermining which of the parental sets will be preferentially reduced. I report here a technique to this end in the belief that it will constitute a useful tool in the use of somatic cells for formal genetic analysis.

The idea stems from work carried out more than 30 years ago[1-3]. The problem was how to obtain individuals recombinant with respect to whole chromosomes from two species of *Drosophila* the hybrids of which were sterile so that the recombinants desired could not be produced by means of meiosis and fertilization. The solution was to X-irradiate the haploid sperms of one species and then to use these to fertilize eggs from triploid females of the second species. Triploid females produce a high proportion of eggs with one or more chromosomes in duplicate ("disomic"). Irradiation induces potential breaks in the sperm chromosomes and their subsequent elimination in the cleavage divisions of the zygote. An egg, disomic for a chromosome homologous to one broken in the spermatozoon which fertilized that egg, would give rise to a diploid zygote carrying exclusively the two maternal chromosomes in question. The same procedure was applied successfully at the intraspecific level by using triploid females with multiple recessive genetic markers and diploid males of the same species with the corresponding dominant alleles[4].

The problem of inducing a preferential elimination of chromosomes from interspecific or intraspecific mammalian somatic cell hybrids is similar. In both cases there is redundancy in the chromosome sets and loss of chromosomes is not necessarily lethal. A partial solution for interspecific cell hybrids came in 1967 when Weiss and Green[5] discovered that hybrids between somatic cells of mouse and man would gradually lose the human chromosomes on continued culture. Cell lines would therefore eventually arise which had retained only one or a few human chromosomes. By this means human genes can be assigned to individual human chromosomes and linkages recognized or excluded[5-9]. A similar elimination occurs in Chinese hamster–man hybrids[10]. This directional elimination of chromosomes from interspecific hybrids is very useful for locating genes specifying those human proteins which can be distinguished from their mouse, or Chinese hamster, counterparts. It is of little use, however, for locating human genes known only because of the effects of recessive alleles.

Irradiation of Parent Cells

The technique described here allows the experimenter to choose which of the parental chromosome contributions shall be preferentially eliminated in hybrid somatic cells. It involves X or γ-irradiation of the cells of one "parent" before they are fused with those of the other. An alternative to X-irradiation is to label the chromosomes of one "parent" with bromodeoxyuridine (BUDR) and thus sensitize them to visible light[11,12].

With the first assay system I used, the loss from hybrid clones of the chromosomes of the treated "parent" can be detected unmistakably. All but one of the chromosomes of