Hydrolysis of Polynucleotides and the Characterization of their Secondary Structure

A THEORETICAL STUDY

By R. A. COX

National Institute for Medical Research, Mill Hill, London, N.W. 7

(Received 20 July 1967)

1. Single-stranded RNA may be regarded as an assembly of L hairpin loops each stabilized by N base pairs and each containing b unpaired residues; one loop is connected to another by c residues. 2. A theory based on the statistics of the random degradation of linear polymers was developed to relate N , b and c with the probability, p , of hydrolysing a diesterified phosphate bond. 3. The number of residues per hairpin loop, which is $2N + b$, is related to the fraction, f , of the original loops remaining intact by the equation: $2N + b = \log f / \log(1 - p)$. 4. The theory was extended to show that the number of residues per loop may be evaluated by fractionating the RNA after hydrolysis and examining the secondary structure of each fraction. Fragments that are shorter than the hairpin loop cannot reproduce the original secondary structure. The probability that a fragment will form an intact loop increases most rapidly for fragments of between $2N + b$ and $2(2N + b) + c$ residues. 5. The probability of producing a fragment capable of forming one, and only one, hairpin loop was related to N , b and c .

Within the cell different species of RNA each fulfil highly specific functions. The primary structure of the polynucleotide chain of a particular molecule determines the precise three-dimensional conformation (secondary and tertiary structure) that is required for biological activity. A knowledge of the secondary structure of the molecule might increase our understanding of the functions of RNA.

Single-stranded polynucleotides such as transfer RNA, ribosomal RNA, RNA from viruses as well as co-polymers of AMP and UMP (polyAU) or of GMP and CMP (polyGC) in neutral salt solutions may form secondary structures of alternating single-stranded and double-helical regions arising from intramolecular interactions between base residues (see Doty, Boedtke, Fresco, Haselkorn & Litt, 1959). Thus a polynucleotide chain may be regarded as an assembly of L hairpin loops each of $2N + b$ residues (where N is the number of base pairs and b is the number of unpaired residues per loop), linked by a flexible region of c residues. A knowledge of N , b and c is essential for the characterization of a polynucleotide.

The aim of this work is to establish the theoretical basis for the evaluation of N , b and c from degradation studies (Cox & Kanagalingam, 1967). The first section concerns the evaluation of the number of residues per hairpin loop, i.e. $2N + b$; this may be

achieved by relating the probability of chain scission with the fraction of intact hairpin loops that may re-form in the unfractionated hydrolysate. In the second section the analysis is extended to hydrolysis in solvents that do not affect the secondary structure of a polynucleotide. The third section describes another method for the evaluation of $2N + b$ based on the extent to which a particular molecular-weight fraction is able to form an intact hairpin loop. The fourth section deals with the probability of producing a fragment capable of forming a single hairpin loop.

THEORETICAL

Evaluation of the number of residues per hairpin loop from the relation between the probability of main-chain scission and the fraction of intact hairpin loops

The hydrolysis of diesterified phosphate linkages is required to take place essentially at random. Minor differences in the rate of cleavage of different internucleotide linkages will scarcely influence the analysis, provided that the 16 possible internucleotide linkages are distributed in a near-to-random fashion throughout the polynucleotide chain. If this condition is fulfilled hydrolysis by reagents such as ribonucleases, which attack specific internucleotide linkages, will also take place in a near-to-random fashion during the early stages of hydrolysis. The di-, tri- and oligo-nucleotides produced by pancreatic ribonuclease digestion of transfer RNA as well as RNA from ribosomes

and from viruses have a distribution that is close to that predicted for a random co-polymer of the same overall nucleotide composition (Staehelin, 1961), and tracts of more than three or four purine or pyrimidine residues were found to be infrequent. Because the hairpin loop is unlikely to be less than about 15 residues and so will include many different linkages, it is probable that it will be attacked at random to produce fragments within the size range of interest unless the hydrolytic reagent is affected by secondary or tertiary structure.

Hydrolysis by reagents of low specificity at a pH or temperature where double-helical secondary structure is absent is considered first. Alkali hydrolyses all inter-nucleotide linkages provided that the $C_{(2)}$ -hydroxyl group of ribose is unsubstituted. The available evidence suggests that oligonucleotides are produced by random cleavage of diesterified phosphate bonds (Kaltreider & Scott, 1962). In accord with a random degradation process the reciprocal of the molecular weight of fragments of reticulocyte ribosomal RNA produced by alkaline hydrolysis was found to be proportional to the time of hydrolysis (Cox & Kanagalingam, 1967). Double-helical secondary structure is also absent at elevated temperatures in neutral salt solutions. Thermal hydrolysis takes place at an appreciable rate at about 60° (Eigner, Boedtker & Michaels, 1961) and chain scission appears to take place at random (Boedtker, Möller & Klemperer, 1962). The theoretical analysis described below for random degradation in solvents that eliminate double-helical secondary structure is illustrated by alkaline hydrolysis at pH 13; it may also apply to thermal hydrolysis in neutral salt solutions.

Principle of the method. At pH 7 RNA forms L_0 loops (e.g. ABCDE, Fig. 1a) each of $2N+b$ residues, where N is the number of base-paired residues (ABDE, Fig. 1a) and b

is the number of unpaired residues (BCD, Fig. 1a), and c residues unite one hairpin loop to another (EF, Fig. 1a). The number of loops where $2N+b=i$ is l_i , so that $L_0=\sum l_i$. Double-helical secondary structure is absent at pH 13 (Cox, 1966a,b) (Fig. 1b), but hydrolysis of diesterified phosphate linkages takes place (Figs. 1c and 1d). After time t the solution is neutralized and the original hairpin loop will re-form, provided that the $2N+b$ residues remain intact (Figs. 1e and 1f).

The fraction of intact hairpin loops is related to the probability, p , of main-chain scission. The probability of an internucleotide chain remaining intact is $1-p$ and the probability of the i linkages of the hairpin loop remaining intact is $(1-p)^i$. For the species $2N+b=i$, the number of loops (initially l_i) remaining intact is $l_i(1-p)^i$. Hence the fraction, f_i , of intact loops is:

$$f_i = (1-p)^i \quad (1)$$

Thus i is the slope of the line obtained on plotting $\log f_i$ against $\log(1-p)$. In practice a mean value $\bar{f}$ of f_i will be obtained leading to an average value $\bar{i}$ for the number of residues within the hairpin loop.

Distribution of damaged loops. The fractions, $f_{i-2(N-n)}$, of loops of species i damaged, but able to form n base pairs only (Fig. 2c) because the sequence of $i-2(N-n)$ residues remain intact, although a bond between the n th and $(n+1)$ th base pairs is broken, is the difference in the probability of a sequence of $i-2(N-n)$ and $i-2[N-(n+1)]$ residues remaining intact, i.e.:

$$f_{i-2(N-n)} = (1-p)^{i-2(N-n)} - (1-p)^{i-2[N-(n+1)]} \\ = [1 - (1-p)^2](1-p)^{i-2(N-n)} \quad (2)$$

The fraction of loops damaged but able to form n base pairs

Fig. 1. Principle of the method for measuring the size of the hairpin loop. (a) Conformation of RNA at pH 7; the vertical lines indicate hydrogen bonds formed between base pairs. (b) Double-helical structure is lost in 0.4 N-KOH. (c) and (d) Hydrolysis of diesterified phosphate linkages takes place; hydrolysed linkages are indicated by X. (e) On neutralization the original secondary structure re-forms provided that the $2N+b$ residues of the loop remain intact. (f) Continued hydrolysis leads to small fragments incapable of forming the original secondary structure.

Fig. 2. Fragments produced by hydrolysis that may form hairpin loops different from the original. The short lines indicate the base residues; the thickened lines indicate the base residues that form the n^0 base pairs that are essential for re-formation of part of the original hairpin loop; the dots indicate hydrogen bonds. (a) RNA at pH 13. 1, 2 and 3 mark the sites of hydrolysis. Hydrolysis leads to fragments such as (b), which cannot, and (c), which can, form a 'frayed' loop on neutralization. Fragment (b) may fold to form a hairpin loop fortuitously.

Fig. 3. Fraction of damaged loops calculated from eqn. (2) when $p=0.05$ and $i=30$.

is given in Fig. 3 as a function of n for $i=30$ and $p=0.05$. As expected for random chain scission the fragments are widely distributed.

The total fraction f'_A of damaged or 'frayed' loops is the sum:

$$f'_A = [1 - (1-p)^2] \sum_{i=2(N-n^0)}^i (1-p)^{i-2(N-n)} \quad (3)$$

where $i-2(N-n^0)$ is the length of the hairpin loop that is stabilized by the minimum number n^0 of base pairs required for stability at 25° in the solvent used. When more than one species of hairpin loop is involved i replaces n in eqn. (3).

Abundance of fortuitously formed loops. The possibility exists that hydrolysis may produce fragments having a sequence such as $(n^0-1)+b+N_1+c+N_2+b+n^0+1$ (Fig. 2b), which cannot re-form any part of the original hairpin but may fortuitously form another species. This event will be frequent if the sequences of N_1 and N_2 are partly homologous. The fraction of loops which are so damaged that they are unable to re-form any part of the original hairpin is:

$$1 - (1-p)^i - [1 - (1-p)^2] \sum_{i=2(N-n^0)}^i (1-p)^{i-2(N-n)}$$

where the second and third terms are the fractions of intact (f_i) and 'frayed' (f'_A) loops respectively. The minimum sequence required to form any hairpin loop is $b''+2n''$. Hence the fragments having the potential to form hairpin loops fortuitously constitute a fraction f''_A of the original number of loops.

$$f''_A = [1 - f_i - f'_A] (1-p)^{b''+2n''} \quad (4)$$

Evaluation of p . The probability of main-chain scission at time t is equal to the fraction of bonds hydrolysed. In principle, p may be obtained from end-group measurements. The fraction of bonds hydrolysed is equal to the number of new molecules formed, so that if $\bar{x}_n$ is the number-average chain length of the fragments and $\bar{x}_n^0$ refers to the intact molecule then:

$$p = 1/\bar{x}_n - 1/\bar{x}_n^0 \quad (5)$$

Thus p may be evaluated from the number-average molecular weight by means of eqn. (5).

Evaluation of f_i . The fraction of the original hairpin loops re-formed on neutralization may be obtained from any parameter that reflects the properties of the hairpin loop. The mid-point of the temperature range over which a double-helical structure 'melts' is defined as T_m and is related to N by the equation (Applequist & Damle, 1965; Crothers, Kallenbach & Zimm, 1965; Kallenbach & Crothers, 1966):

$$1/T_m = A/N + 1/T_{m(\infty)} \quad (6)$$

where A is a constant and $T_{m(\infty)}$ is the value of T_m when

$N \rightarrow \infty$. Although factors such as nucleotide sequence (Bautz & Bautz, 1964) or imperfections in the double helix (Fresco, Alberts & Doty, 1960) may affect T_m , eqn. (6) serves to show that the T_m is dependent on N when N is small. In general, single-stranded polynucleotides 'melt' at temperatures that are substantially lower than $T_{m(\infty)}$, indicating that N is small (Cox, 1966a,b). The evaluation of $\bar{f}$ is facilitated because a 'frayed' loop or a fortuitously formed loop would be expected to melt before the corresponding intact loop.

Random degradation of a polynucleotide in solvents where double-helical secondary structure is preserved

The evaluation of the number of residues per hairpin loop by the method described above relies on the re-formation of the hairpin loops on neutralization. The method would be improved if hydrolysis were brought about in solvents that did not affect the original secondary structure, e.g. by the action of enzymes, chemicals or ionizing radiations on polynucleotides in neutral salt solutions at 25° or below. Alkali may also be a suitable reagent since, at 0° , pH 11 may be attained without appreciably affecting double-helical structure (Cox & Littauer, 1963) and the rate of hydrolysis is still appreciable. Provided that degradation takes place at random, eqn. (1) will still apply, so that i may be evaluated.

Hydrolysis under these conditions may provide further information, because chain scissions will be produced that do not immediately affect the molecular weight because the fragments remain held together by double-helical regions (Cox & Kanagalingam, 1967). Chain scissions occurring within the hairpin loop will be 'hidden', provided that the minimum number n^0 of base pairs remain intact so that fragments are united. These 'hidden' breaks will be revealed once double-helical secondary structure is destroyed.

The fraction, β , of chain scissions that are hidden will decrease as p increases, because multiple scissions within a loop will still lead to a single hidden break per loop. A scission will be 'hidden' if it occurs within a sequence of $i-2n^0$ residues. The probability of this sequence remaining intact is $(1-p)^{i-2n^0}$, whereas the probability of this sequence being damaged is $1 - (1-p)^{i-2n^0}$. There will be a 'hidden' break for each loop damaged in this way. The total number of breaks is $L(2N+b+c)p$. Hence:

$$\beta = L[1 - (1-p)^{i-2n^0}]/L(i+c)p \quad (7)$$

$$\beta_{p \rightarrow 0} = (i-2n^0)/(i+c) \quad (8)$$

The molecular-weight distribution of the fragments before hidden breaks are revealed is of interest for two reasons: first, because β may be calculated from the values of M_w obtained before and after 'hidden' breaks are revealed provided that the relation M_n/M_w is the same in both cases; secondly, because in principle $i+c$ may be evaluated. The weight distribution, W_x , is related to the degree of polymerization, x , by the equation:

$$W_x = x(1-p)^x p^2 \quad (9)$$

When RNA is degraded at pH 13 (when double-helical secondary structure is absent) the nucleotide is the repeating unit and the molecular weight of the fragments is obtained from the degree of polymerization by multiplying by the mean residue weight (330).

Fragments differing by about 330 in molecular weight are not readily distinguished by available techniques. If the repeating unit were much heavier, e.g. because of a bulky side chain, it might be possible to distinguish fragments of x and $x+1$ units. When double-helical secondary structure is present, the hairpin loop may be regarded as the repeating unit, since an immediate decrease in molecular weight is only observed when a bond within the region n^0+c+n^0 residues is ruptured. The effective repeating unit is one of $i+c$ residues. Thus x in eqn. (9) will be the number of hairpin loops and the molecular weight of the fragments will be $x(i+c)330$, so that the difference in molecular weight between fragments of x and $x+1$ units will be $330(i+c)$. Fragments differing by 10000 in molecular weight may be separated by polyacrylamide-gel electrophoresis (Richards, Coll & Gratzner, 1965; McPhie, Hounsell & Gratzner, 1966; Gould, 1966), so that a series of discrete bands may be observed (cf. pancreatic ribonuclease digestion of reticulocyte ribosomal RNA; Gould, 1966; Cox & Kanagalingam, 1967).

If the probability of any internucleotide linkage being hydrolysed is p , the probability of the sequence of $2n^0+c$ residues remaining intact is $(1-p)^{2n^0+c}$ and the probability, p_{EF} , of observing an immediate decrease in molecular weight is:

$$p_{EF} = 1 - (1-p)^{2n^0+c} \quad (10)$$

When p is small:

$$p_{EF} = (2n^0+c)p \quad (11)$$

$$W_{x'} = x'(1-p_{EF})^x p_{EF}^2 \quad (12)$$

where x' is the number of hairpin loops.

The theoretical distributions of molecular weight before and after hidden breaks were revealed were calculated from eqns. (9) and (12) on the basis of $p=0.004$, $n^0=4$, $c=5$, $p_{EF}=0.05$ (eqn. 11) and $i=35$. Fig. 4(a) gives the weight fraction as a function of x' calculated by means of eqn. (12). The degree of polymerization x' was multiplied by $330(i+c)$ to give the distribution of molecular weight (Fig. 4b). When the molecular-weight scale is expanded it is clear from Fig. 4(c) that the fragments increase in molecular weight by steps of about 13000. Once hidden breaks are revealed the molecular-weight distribution is obtained by substituting $p=0.004$ in eqn. (9) (Fig. 4d). The degree of polymerization was multiplied by the average weight of a nucleotide (330) to give the molecular weight (Fig. 4e). Although the molecular-weight scale is expanded (Fig. 4f) there is a continuous distribution of molecular weights, in contrast with Fig. 4(c). A suitable procedure for revealing hidden breaks is to expose the RNA to pH 3 at 25° for 30 sec. or less to destroy secondary structure, e.g. RNA (200 μ g.) is added to MN-HCl (4.75 ml.), and m-phosphate buffer, pH 8.0 (50 μ l.), is then added to neutralize the solution. The chain scissions remain hidden if the phosphate buffer is added to neutralize the acid before RNA is introduced. Thus chain scissions remain hidden or are revealed simply by changing the order in which the reagents are added.

Evaluation of the chain length of the hairpin loop from the properties of the fractionated fragments

A fragment of fewer than i residues will not be able to re-form a hairpin loop that was originally i residues long. The probability that a fragment i residues long will form an

Fig. 4. Molecular-weight distribution of RNA fragments before and after hidden breaks are revealed. (a) Weight fraction $W_{x'}$ as a function of x' calculated from eqn. (12) for values of $p_{EF}=0.05$ given by eqn. (11) where $p=0.004$. (b) Values of molecular weight calculated from x' where the hairpin loop $i=35$ and $c=5$ is the repeating unit. (c) Hatched section of (b) expanded. (d) W_x as a function of x calculated from eqn. (9) for $p=0.004$. (e) Values of molecular weight calculated from x where the repeating unit is the nucleotide. (f) Hatched section of (e) expanded.

Fig. 5. Relation between the size of fragments, the number of residues per hairpin loop and the ability to re-form the original secondary structure. Curve I, weight fraction, W_x (ordinate on left-hand side), as a function of x calculated from eqn. (9) for $p=0.05$. Curve II, fraction of the original secondary structure re-formed by a fragment of x residues long (ordinate on right-hand side) calculated from eqn. (14) for a loop of 35 residues, $N-n^0=6$. Curve III, fraction of loops that are 'frayed' formed by fragments x residues long. $\Delta\epsilon$, the increment due to 'frayed' loops, is given as a function of x , and $\Delta\epsilon^0$ has the same significance as for curve II.

intact hairpin loop is $1/(i+c)$, because only one of the $i+c$ ways of starting the fragment will lead to the correct sequence. It is certain that an intact loop will be preserved in a fragment of $i+(i+c)$ residues. A fragment of $i+2(i+c)$ residues will be certain to contain two intact loops. A fragment of $i(l+1)+lc$ residues will be certain to contain l loops. The fraction, h , of residues in double-helical segments of a fragment $i(l+1)+lc$ residues is:

$$h = 2Nl/i(l+1)+lc \quad (13)$$

For the intact molecule $n=h^0=2N/(i+c)$, and thus:

$$f = \frac{h}{h^0} = \frac{l}{l+i/(i+c)} \quad (14)$$

When $c \ll i$ then $f=l/(l+1)$. Hence f will increase rapidly from 0 to 50% when the fragments increase in size from $i-1$ to $2i+c$ residues. If the fragments are fractionated according to size, and if f is evaluated for each fraction, then i may be evaluated. Theoretical plots of W_x against x and of f against x calculated for $p=0.05$ and $i=35$ are given in Fig. 5 as an illustration.

It is also possible to discern 'frayed' loops from the denaturation spectrum, so that it is possible to evaluate $b+2n^0$. A fragment intermediate between $b+2n^0$ and $i-1$ will have the denaturation properties of a 'frayed' loop but not those of an intact loop. Fragments longer than $2i+c$ will have the denaturation properties of both intact

and damaged loops. The contribution of frayed loops to the denaturation spectrum will decrease as the size of the fragment increases. If the increment in extinction due to frayed loops is $\Delta\epsilon'$ then $\Delta\epsilon'/\Delta\epsilon^0$ will have the form of the broken line in Fig. 5. This method is a useful complement to the analysis given in the first section and provides a possible method for the estimation of $b+2n^0$. For example, when $i=35$ extreme values of N and b are 4 and 27 when $N=n^0$, compared with 16 and 3 when $b=3$. In this instance, the molecular weights of the fragments $b+2n^0$ are respectively 11660 and 3660, the difference being 8000.

Abundance of fragments that form one, and only one, hairpin loop

The probability that a fragment of x residues will include the sequence of one intact loop of i residues is nil when $x \leq i-1$. When $i \leq x \leq 2i+c-1$ the probability of forming a single intact loop is $(x+1-i)/(i+c)$. It is certain that when $2i+c \leq x \leq 3i+2c-1$ at least one loop will be formed, and the probability of forming one and only one intact loop is $(3i+2c-x-1)/(i+c)$. When $x \geq 3i+2c$ it is certain that two intact loops will be formed, so that there is no chance of forming a single loop. Fragments may form 'frayed' as well as intact loops. The probability of a fragment forming one intact hairpin and no other loop whether intact or damaged is nil when $x < i$ and is $(x+1-i)/(i+c)$ when $i \leq x \leq 2i+c-(N-n^0)-1$. When

$$2i+c-(N-n^0) \leq x \leq 2i+c-1$$

the probability of forming a single intact loop and no other is the product of the probability of forming a single intact loop and the probability of not forming a 'frayed' loop and is equal to $(x+1-i)[3i+2c-x-1-2(N-n^0)]/(i+c)^2$. When $2i+c \leq x \leq 3i+2c-1-2(N-n^0)$ the probability that the fragment will form one intact loop and no other is $[3i+2c-x-1-2(N-n^0)]/(i+c)$, because if $x \geq 2i+c$ it is certain that an intact loop will form, but as x increases further the probability of forming one and only one loop decreases and is nil when $x=3i+2c$.

It is possible that an assay for single loops will not be sufficiently sensitive to distinguish intact and 'frayed' species. The probability that a fragment will form a single loop (either intact or frayed) is nil when $x \leq i-2(N-n^0)-1$. Within the range $i-2(N-n^0) \leq x \leq 2i+c-2(N-n^0)-1$ the chance of forming a single loop is

$$[x-i+2(N-n^0)+1]/(i+c)$$

When $x \geq 2i+c-2(N-n^0)$ it is certain that the fragment will have the capacity to form at least one loop, and increasing x still further will increase the chance that it will form a second loop. The probability of forming one and only one loop is nil when $x \geq 3i+2c-2(N-n^0)$. Thus within the range $2i+c-2(N-n^0) \leq x \leq 3i+2c-2(N-n^0)-1$ the probability of forming a single loop is

$$[(3i+2c)-x-1-2(N-n^0)]/(i+c)$$

The distribution of fragments capable of forming a single loop may be related to the probability of chain scission. The weight fraction, W_x , of the fragment x residues long is given by eqn. (9). Thus the weight fraction of fragment that can form one and only one loop is the product of W_x and the probability that a single loop will be formed. The equations

Table 1. *Weight fractions of fragments able to form one and only one hairpin loop*

Species	Wt. fraction (W_x)	Limits	Eqn. no.
Single intact loops	0	$x \leq i-1$	(15)
excluding a second	$W_x = xp^2(1-p)^2(x+1-i)/(i+c)$	$i \leq x \leq 2i+c-(N-n^0)-1$	(16)
loop whether intact	$W_x = xp^2(1-p)^2[3i+2c-x-1-2(N-n^0)](x+1-i)/(i+c)^2$	$2i+c-(N-n^0) \leq x \leq 2i+c-1$	(17)
or damaged	$W_x = xp^2(1-p)^2[3i+2c-x-1-2(N-n^0)]/(i+c)$	$2i+c \leq x \leq 3i+2c-1-2(N-n^0)$	(18)
	0	All other values of x	
Single loops, intact	0	$x \leq i-2(N-n^0)-1$	(19)
or 'frayed', excluding	$W_x = xp^2(1-p)^2(x-i+2(N-n^0)+1)/(i+c)$	$i-2(N-n^0) \leq x \leq 2i+c-2(N-n^0)-1$	(19)
a second loop	$W_x = xp^2(1-p)^2[3i+2c-x-1-2(N-n^0)]/(i+c)$	$2i+c-2(N-n^0) \leq x \leq 3i+2c-2(N-n^0)-1$	(20)
	0	All other values of x

Fig. 6. Relation between the ability of a fragment to form one and only one hairpin loop and the original secondary structure. Curve I, weight fraction, W_x , as a function of x calculated by eqn. (9) for $p=0.05$. Curve II, W_x of fragments capable of forming one and only one loop calculated from eqns. (19) and (20) for $i=20$. Curve III, as for curve II but $i=35$.

and the range of values of x for which they are valid are given in Table 1.

The optimum yield of fragments able to form single hairpin loops was evaluated. For a given value of x , W_x is maximum when $p=2/(x+2)$. A fragment of

$$x = 2i + c - 2(N - n^0) - 1$$

is certain to form one and only one hairpin loop, so that the yield of single hairpin loops is maximum when:

$$p = 2/[2i + c - 2(N - n^0) + 1]$$

The calculated values of W_x and of the weight fractions of fragments forming single hairpin loops are given in Fig. 6 for $p=0.05$, for $i=20$, $c=5$ and $N-n^0=4$ and for $i=35$, $c=5$ and $N-n^0=6$.

No technique has yet been established for the measurement of single hairpin loops. However, it is likely (but not proved) that the fragments of ribosomal RNA which crystallize are single hairpin loops. It may be that the size

of the crystals (e.g. Spencer & Poole, 1965) reflects the weight fraction of single hairpin loops.

DISCUSSION

There is a need for methods that increase our knowledge of the secondary structure of RNA. For example, viral RNA functions not only as a genetic messenger but also as a template for its own polymerase, and also it is a component of the virus. It is possible that these different functions are also reflected in the secondary structure of RNA. The melting profiles of RNA from different viruses may have distinctive features (Sprecher-Goldberger, 1967). The relation between hydrodynamic properties and molecular weight is not the same for all RNA species (e.g. Gesteland & Boedtker, 1964), indicating major differences in secondary structure. The secondary structure of transfer RNA is not yet established. Basic data such as the number of hairpin loops per molecule of ribosomal RNA are not available.

For a particular RNA species the evaluation of N , b and c would more closely describe secondary structure. The precision with which $2N+b$ can be evaluated depends on the accuracy of the measurement of p and f . In principle, by measuring p and f for a wide range of fragments, and by calculating f from the changes in extinction at different wavelengths, values of p may be accurate to within $\pm 10\%$ (either by measuring M_n directly or by measuring M_w), and f may also be measured to within $\pm 10\%$: the standard deviation of $2N+b$ will be $\pm 20\%$.

The evaluation of N , b and c is less precise because they represent substantially shorter regions of the chain than $2N+b$.

Evaluation of c . In principle, c may be obtained by means of eqn. (7). A knowledge of $i=2N+b$, β and n^0 is required. The most precise values of β would be obtained by fractionation of RNA (before and after hidden breaks are revealed) by procedures such as preparative polyacrylamide-gel electrophoresis, which would lead to a knowledge of the weight distribution as a function of molecular weight provided that the molecular weights of individual fractions are determined. Though the molecular

weight of an individual fraction may be accurate to within $\pm 10\%$ the accuracy of the molecular-weight distribution should be greater. The accuracy with which c can be determined will depend on the precise value of β (the fraction of breaks hidden by secondary structure) and also on the precision with which i is known. The greatest precision in c will be attained as β approaches unity. Suppose $i = 30 \pm 5$ residues, $n^0 = 5 \pm 1$ and $\beta = 0.6 \pm 0.1$ (i.e. the value of M_n before and after hidden breaks are revealed is accurate to within $\pm 10\%$); then the extreme values of c are 0 and 19, but the average values are $c = 6 \pm 5$ residues.

Evaluation of N . Eqn. (6) is tentative in view of the absence of data for model hairpin loops. However, a T_m of 55° in 0.1M-sodium phosphate buffer is unlikely to be due to a sequence of more than about 6 base pairs and a T_m of about 100° is unlikely to be due to a sequence of less than 10 base pairs.

Provided that the fraction, h^0 , of residues in the intact double-helical segments is known, an average value of N may be obtained from a knowledge of i and c , since:

$$h^0 = \frac{\sum 2N}{\sum (i + c)}$$

A value of $N - n^0$ may be derived from the denaturation properties of individual fractions of the hydrolysate (e.g. Fig. 5). As yet, the evaluation of N depends on indirect methods. Since $b = i - 2N$, the imprecision in N is reflected in b , and at present the value assigned to b is tentative.

Degradation studies may be most useful in deciding between possible forms of secondary structure based on a knowledge of the primary sequence. For example, three possible structures were proposed (Holley *et al.* 1965) for alanine transfer RNA: only one, the 'clover-leaf' form, has four loops, $i_1 = 73$, $i_2 = 17$, $i_3 = 17$ and $i_4 = 18$. It may be possible to show by degradation studies whether the molecule has the clover-leaf form in solution.

In theory it is possible to evaluate i and c and to obtain tentative values for N and b . Earlier work has shown that the experimental techniques are available to exploit the theory.

The author thanks Miss J. Skegg for her assistance and valuable discussions.

REFERENCES

- Applequist, J. & Damle, V. (1965). *J. Amer. chem. Soc.* **87**, 1450.
 Bautz, E. K. F. & Bautz, F. A. (1964). *Proc. nat. Acad. Sci., Wash.*, **52**, 1476.
 Boedtker, J., Möller, W. & Klemperer, E. (1962). *Nature, Lond.*, **194**, 444.
 Cox, R. A. (1966a). *Biochem. J.* **98**, 841.
 Cox, R. A. (1966b). *Biochem. J.* **100**, 146.
 Cox, R. A. & Kanagalingam, K. (1967). *Biochem. J.* **103**, 431.
 Cox, R. A. & Littauer, U. Z. (1963). *Biochim. biophys. Acta*, **72**, 188.
 Crothers, D. M., Kallenbach, N. R. & Zimm, B. H. (1965). *J. molec. Biol.* **11**, 802.
 Doty, P., Boedtker, H., Fresco, J. R., Haselkorn, R. & Litt, M. (1959). *Proc. nat. Acad. Sci., Wash.*, **45**, 482.
 Eigner, J., Boedtker, H. & Michaels, G. (1961). *Biochim. biophys. Acta*, **51**, 165.
 Fresco, J. R., Alberts, B. & Doty, P. (1960). *Nature, Lond.*, **188**, 98.
 Gesteland, R. F. & Boedtker, H. (1964). *J. molec. Biol.* **8**, 496.
 Gould, H. (1966). *Biochemistry*, **5**, 1103.
 Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965). *Science*, **147**, 1462.
 Kallenbach, N. R. & Crothers, D. M. (1966). *Proc. nat. Acad. Sci., Wash.*, **56**, 1018.
 Kaltreider, H. B. & Scott, J. F. (1962). *Biochim. biophys. Acta*, **55**, 379.
 McPhie, P., Hounsell, J. & Gratzer, W. B. (1966). *Biochemistry*, **5**, 988.
 Richards, E. G., Coll, J. A. & Gratzer, W. B. (1965). *Analyt. Biochem.* **12**, 452.
 Spencer, M. & Poole, F. (1965). *J. molec. Biol.* **11**, 314.
 Sprecher-Goldberger, S. (1967). *Arch. ges. Virusforsch.* **20**, 225.
 Staehelin, M. (1961). *Biochim. biophys. Acta*, **49**, 27.