Macromolecular Structure and Properties of Ribonucleic Acids

By R. A. Cox
NATIONAL INSTITUTE FOR MEDICAL RESEARCH, MILL HILL, LONDON, N.W.7

1 Introduction

The variety of roles that ribonucleic acids play in the living cell became apparent during the past decade. Living organisms develop and reproduce themselves with great fidelity and geneticists have long been able to deduce laws governing the inherited characteristics of the organism. However, growth, development, and reproduction depend on directed chemical reactions which are brought about by specific catalysts called enzymes. Each enzyme is a unique protein molecule which is synthesised by the cell, presumably with the aid of other enzymes. The individual polypeptide chains for a protein are polymers of 20 common amino-acids so that the number of possible amino-acid sequences is unlimited, yet the living cell is able precisely to reproduce a particular sequence.[1]

Biologists have known for many years that ribonucleic acid (RNA) is implicated in the biosynthesis of proteins.[2] During the past decade methods of isolating it were so improved that studies of the biological, chemical, and macromolecular properties of RNA have been possible. At least 4 classes of RNA are now recognised; these are 'messenger', ribosomal, transfer, and viral RNA. Each class of RNA has unique properties which reflects its function in the cell. Studies of the biological activity of RNA have led to the elucidation of the mechanism of protein biosynthesis.

Abbreviations.—The abbreviation RNA for 'ribonucleic acids' is used generically for polymers of nucleotides linked by diesterified phosphate bonds joining the C(3′) hydroxyl of one ribose residue with the C(5′) hydroxyl of the next ribose residue. Particular species of ribonucleic acid are referred to thus: t-RNA, transfer RNA; phenylalanyl t-RNA, etc., is the ester formed between phenylalanine, etc., and the t-RNA specific for phenylalanine, etc.; m-RNA, messenger

[1] M. F. Perutz, 'Proteins and Nucleic Acids: Structure and Function', Elsevier, Amsterdam, 1962; J. N. Davidson, 'The Biochemistry of Nucleic Acids', Methuen and Co., London, 5th ed., 1965; J. D. Watson, 'Molecular Biology of the Gene', Benjamin, New York, 1965; 'The Living Cell', Readings from *Scientific American*, W. H. Freeman and Co., San Francisco, 1965.

[2] J. Brachet, 'The Biological Role of Ribonucleic Acids', Elsevier, Amsterdam, 1960; H. Chantrenne, 'The Biosynthesis of Proteins', Pergamon Press, London, 1961; H. R. V. Arnstein, *Brit. Med. Bull.*, 19, **21**, 217; P. N. Campbell and J. R. Sargent, 'Techniques in Protein Biosynthesis', Academic Press, London, 1967; R. Schweet and R. Heintz, *Ann. Rev. Biochemistry*, 1966, **35**, 723.

RNA; r-RNA, ribosomal RNA; A, C, G, and U refer respectively to adenine, cystosine, guanine, and uracil residues; AMP, CMP, GMP, UMP, respectively are adenylic acid, cytidylic acid, guanylic acid, and uridylic acid. Poly-A (poly-adenylic acid) etc. refers to a polymer of AMP etc. having the same inter-nucleotide linkages as RNA: poly-AU etc. refers to a copolymer of AMP and UMP etc.: poly(A + U) etc. refers to a complex formed between one strand of poly-A and one strand of poly-U etc.: poly(A + 2U) etc. refers to a complex formed between one strand of poly-A and two strands of poly-U etc.

2 Biological Properties

The genetic information which specifies the sequence in which the amino-acids must occur in order to produce a particular protein is stored within the chromosome of the cell and is used as required. The chromosome consists of very long chains of DNA[3] (deoxyribonucleic acid) in combination with protein. Each chromosome is made up of many genes and each gene has the information essential for the synthesis of a particular protein.[4] The DNA component of the chromosome is the repository of genetic information. DNA is synthesised from four principal deoxyribonucleotides[5] and the sequence in which these occur specifies the amino-acid sequence of the protein which is under the control of a gene. A sequence of three deoxyribonucleotides (a codon) is required to specify an amino-acid. The code words have now been identified and a genetic dictionary exists.[6] Thus a linear sequence of the nucleotides of the gene is translated into the linear amino-acid sequence of the protein.

Whereas genetic information is stored in the chromosome, proteins are synthesised principally in the cytoplasm. The genetic message is carried from one site to another in the form of messenger RNA.[7] The linear deoxyribonucleotide sequence of the gene is copied by means of an enzyme (DNA-dependent RNA polymerase) that synthesises an RNA molecule which has a nucleotide sequence that is complementary to one of the strands of the gene.[8] Messenger RNA is then transported to the ribosomes of the cytoplasm where polypeptide synthesis takes place. Ribosomes are ribonucleoprotein particles[9] having a molecular weight of (3—4) $\times 10^6$ daltons and consisting of 30—50% protein. In general, the 'messenger' interacts with more than one ribosome to form

[3] H. J. F. Cairns, *J. Mol. Biol.*, 1963, **6**, 208; C. I. Davern, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 792.
[4] F. Jacob and J. Monod, *J. Mol. Biol.*, 1961, **3**, 318; N. R. Cohen, *Biological Reviews*, 1966, **41**, 503.
[5] P. A. Edwards and K. V. Shooter, *Quart. Rev.*, 1965, **19**, 369.
[6] 'The Genetic Code', Cold Spring Harbor Symposia, 1966, vol. 31.
[7] S. Brenner, F. Jacob, and M. Meselson, *Nature*, 1961, **190**, 576; F. Gros, H. Hiatt, W. Gilbert, C. G. Kurland, R. W. Riseborough, and J. D. Watson, *ibid.*, 1961, **190**, 581; E. Volkin and L. Astrachan, *Virology*, 1956, **2**, 149.
[8] V. V. Koningsberger and L. Bosch, 'Regulation of Nucleic Acid and Protein Biosynthesis', Elsevier Publishing Co., London, 1967; J. Josse, A. D. Kaiser, and H. L. Kornberg, *J. Biol. Chem.*, 1961, **236**, 864; S. B. Weiss and T. Nakamoto, *Proc. Nat. Acad. Sci. U.S.A.*, 1961, **47**, 1400; M. Chamberlin and P. Berg, *ibid.*, 1962, **48**, 81.
[9] M. L. Petermann, 'The Physical and Chemical Properties of Ribosomes', Elsevier, Amsterdam, 1964; H. R. V. Arnstein, *Ann. Reports*, 1963, **60**, 512.

polyribosomes.[10] An adaptor molecule (transfer RNA) is required to translate the polynucleotide sequence into an amino-acid sequence.[11] The adaptor is able to combine chemically with a specific amino-acid to form a amino-acid ester of transfer (t-)RNA and it is able to recognise the triplet of nucleotides (the codon) which correspond to its specific amino-acid. The principal steps in normal protein biosynthesis (see Figure 1) have been confirmed by Khorana and his colleagues[12] who synthesised oligo-deoxyribonucleotides of known sequence and then showed that the amino-acid sequence of the polypeptide which could be synthesised *in vitro* was determined by the DNA. The genetic code is given in Table 1.

The normal cycle of protein synthesis may be interrupted when the cell becomes infected by an RNA virus.[13] The viral RNA behaves as messenger and takes over the protein synthetic apparatus of the cell for the production of viral protein (Figure 2). The viral RNA has also to replicate itself (by means of its own RNA-dependent RNA polymerase).[14] The single strand of viral RNA is itself copied and a double-stranded intermediate is produced.[15] Viral RNA may also exist in the virus as a double-helical structure.[16]

A nucleic acid must be isolated before its properties can be studied extensively. The isolation of a pure nucleic acid requires methods that do not damage primary or secondary structure and give preparations that are homogeneous with respect to molecular weight, nucleotide sequence, and biological activity. Before 1956, RNA was generally believed to have a molecular weight of about 15,000 daltons although preparations of about 300,000 daltons were occasionally obtained. The sensitivity of RNA to hydrolysis either by potent degradative enzymes (nucleases) or by alkali was not appreciated. The development of improved methods for the fractionation of subcellular components,[17] and the use of reagents such as phenol[18] for the dissociation of protein from RNA, and

[10] A. Gierer, *J. Mol. Biol.*, 1963, **6**, 148; J. R. Warner, A. Rich, and C. E. Hall, *Science*, 1964 **138**, 1399; J. R. Warner, P. M. Knopf, and A. Rich, *Proc. Nat. Acad. Sci. U.S.A.*, 1963, **49** 122; F. O. Wettstein, T. Staehelin, and H. Noll, *Nature*, 1963, **197**, 430.

[11] (*a*) M. Hoagland, 'The Nucleic Acids', E. Chargaff and J. N. Davidson, Academic Press, New York, 1960, vol. 3, p. 349; (*b*) G. L. Brown and S. Lee, *Brit. med. Bull.*, 1965, **21**, 236; G. L. Brown, 'Progress in Nucleic Acid Research', ed. J. N. Davidson and W. E. Cohn, Academic Press, vol. 11, p. 259; G. L. Brown, S. Lee, and D. H. Metz, 'Genetic Elements: Properties and Function', ed. D. Shugar, Academic Press, New York, 1967, p. 57.

[12] H. Kössel, A. R. Morgan, and H. G. Khorana, *J. Mol. Biol.*, 1967, **26**, 449; H. G. Khorana, 'Genetic Elements; Properties and Function', ed. D. Shugar, Academic Press, 1967, p. 209.

[13] 'Basic Mechanisms in Animal Virus Biology', Cold Spring Harbor Symposium, L. Frisch, Long Island Biol. Assoc., New York, vol. 27; E. M. Martin, *Brit. Med. Bull.*, 1967, **23**, 192.

[14] S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **54**, 919.

[15] E. M. Martin, 'Genetic Elements: Properties and Function', ed. D. Shugar, Academic Press, London, 1967, p. 117; L. Montagnier and F. K. Sanders, *Nature*, 1963, **199**, 664.

[16] P. J. Gomatos and I. Tamm, *Proc. Nat. Acad. Sci. U.S.A.*, 1963, **49**, 707.

[17] 'Methods in Enzymology', Biochem. Soc. Symposium, 'Methods of Separation of Subcellular Structural Compounds', Cambridge Univ. Press, 1963, p. 23.

[18] K. S. Kirby, 'Progress in Nucleic Acid Research,' ed. J. N. Davidson and W. E. Cohn, Academic Press, New York, 1964, vol. 3, p. 17; H. R. V. Arnstein and R. A. Cox, *Brit. Med. Bull.*, 1966, **22**, 158; G. L. Cantoni and D. R. Davies, 'Procedures in Nucleic Acid Research', Harper and Row, New York, 1966; L. Grossman and K. Moldave, 'Methods in Enzymology', Academic Press, New York, 1967, vol. 12.

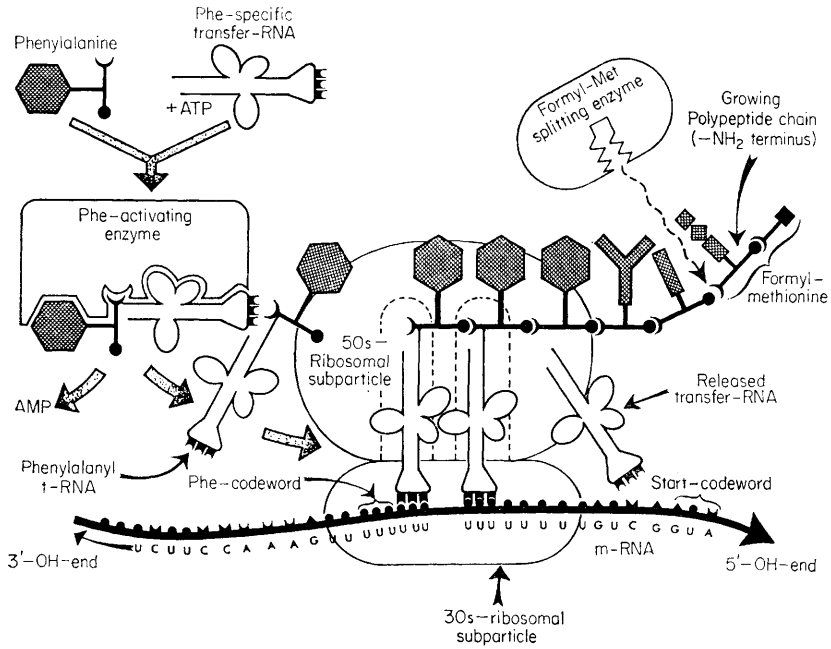

Figure 1 *Mechanisms of protein biosynthesis*
The synthesis of a hypothetical polypeptide with the amino-acid sequence formylmethionine-alanine-valine-phenylalanine-phenylalanine-phenylalanine---- by *E. coli* is illustrated. The next amino-acid to be incorporated is phenylalanine which reacts specifically with phenylalanine t-RNA forming phenylalanyl-t-RNA: the reaction is brought about by activating enzymes. Phenylalanyl-t-RNA is bound to the larger subparticle (W. Gilbert, *J. Mol. Biol.*, 1963, **6**, 360) but interacts with the codon UUU of m-RNA through the anti-codon AAA (H. Kaji and A. Kaji, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 1541; M. Nirenberg and P. Leder, *Science*, 1964, **145**, 1399). The NH_2 group of the phenylalanyl residue of phenylalanyl-t-RNA forms a peptide bond with the CO_2H group of the last amino-acid of the growing polypeptide chain which is now attached to phenylalanyl-t-RNA. The genetic message is read from the 5′ end to the 3′ end (M. Salas, M. A. Smith, W. M. Stanley, A. J. Wahba, and S. Ochoa, *J. Biol. Chem.*, 1965, **240**, 3988; R. Thach, M. A. Cecere, T. A. Sundararajan, and P. Doty, *Proc. Nat. Acad. Sci., U.S.A.*, 1965, **54**, 1167; H. Lamfrom, C. S. McLaughlin, and A. Sarabhai, *J. Mol. Biol.*, 1967, **22**, 355). The codon AUG marks the start of the message (B. F. C. Clark and K. A. Marcker, *J. Mol. Biol.*, 1966, **17**, 394) and the codons UAA and UAG may signal the release of the polypeptide chain (S. Brenner, T. Stretton, and S. Kaplan, *Nature*, 1965, **206**, 994). When the ribosome reaches the 3′ end of m-RNA it is likely that the subparticles fall apart. The smaller subparticle may then form a complex m-RNA (5′ end) and formylmet-t-RNA. The larger subparticle then reacts with the complex to complete the ribosome (G. Mangiarotti and D. Schlessinger, *J. Mol. Biol.*, 1966, **20**, 123; H. P. Ghosh and H. G. Khorana, *Proc. Nat. Acad. Sci., U.S.A.*, 1967, **58**, 2455; R. O. Kaempfer, M. Meselson, and H. J. Raskas, *J. Mol. Biol.*, 1968, **31**, 277). The site for peptide bond synthesis may be in the interior of the ribosome (L. I. Malkin and A. Rich, *J. Mol. Biol.*, 1967, **26**, 329)

Table 1 *The genetic code*

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA OCHRE	UGA ?	A
	UUG Leu	UCG Ser	UAG AMBER	UGG Tryp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA GluN	CGA Arg	A
	CUG Leu	CCG Pro	CAG GluN	CGG Arg	G
A	AUU Ileu	ACU Thr	AAU AspN	AGU Ser	U
	AUC Ileu	ACC Thr	AAC AspN	AGC Ser	C
	AUA Ileu	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GCU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

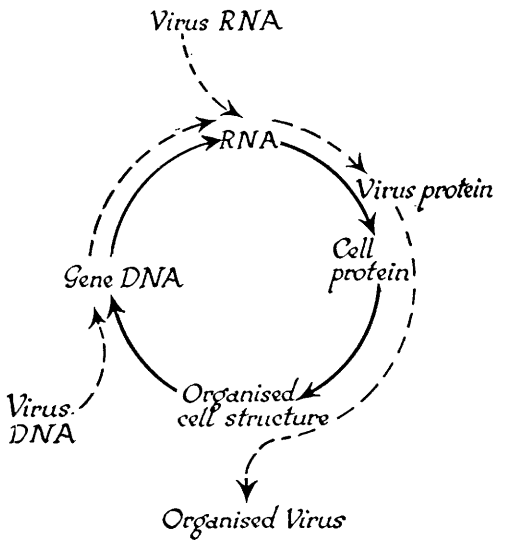

Figure 2 *Diagram showing the way in which infecting virus donates either DNA or RNA to the synthetic machinery of the host cell and thereby diverts the normal course of cell replication* (T. S. Work, *Adv. Sci.*, 1965, **22**, 98)

improvements[19] in the fractionation of RNA have led to the isolation of biologically active RNA.

3 Primary Structure of RNA

The oligonucleotides that were isolated before 1956 proved sufficient[20] for the general features of the chemical structure of RNA to be elucidated. This is illustrated in Figure 3 which shows an oligomer of the 4 common nucleotides, *viz.*, adenylic (AMP), cytidylic (CMP), guanylic (GMP), and uridylic acids (UMP). The sugar was shown to be D-ribose[21] in the furanose form[22] and the nucleosides

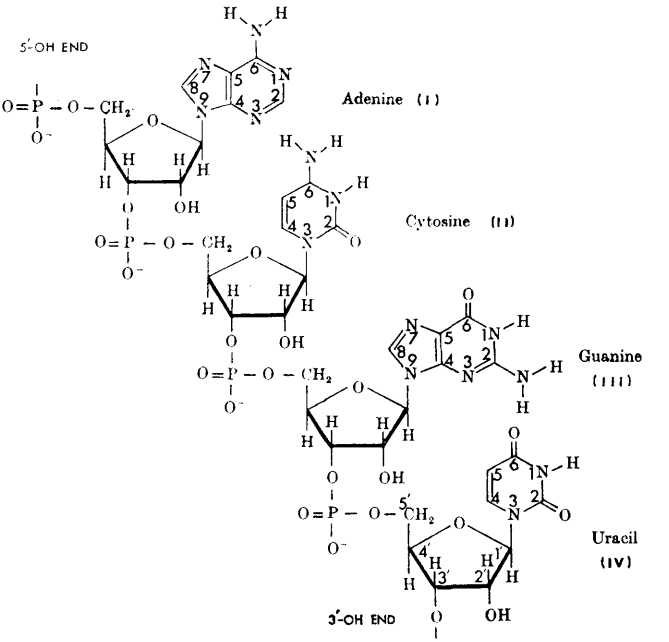

Figure 3 *The chemical structure of ribonucleic acid* (The older numbering system is used for pyrimidine derivatives, in conformity with purines [*cf.* A. M. Michelson, 'The Chemistry of Nucleosides and Nucleotides', Academic Press, New York, 1963])

[19] R. E. Click and D. P. Hackett, *Biochim. Biophys. Acta*, 1966, **129**, 74; J. R. B. Hasting, J. H. Parish, K. S. Kirby, and E. S. Klucis, *Nature*, 1965, **208**, 645; M. D. Dabeva and R. G. Tsanev, *Analyt. Biochem.*, 1966, **17**, 390; D. H. L. Bishop, J. R. Claybrook, and S. Spiegelman, *J. Molec. Biol.*, 1967, **26**, 373; U. E. Loening, *Biochem. J.*, 1967, **102**, 251; R. L. Erikson, *J. Molec. Biol.*, 1966, **18**, 372; J. F. Weiss and A. D. Kelmers, 1967, *Biochemistry*, **1967**, **6**, 2507.

[20] (*a*) P. A. Levene and L. W. Bass, 'Nucleic Acids', American Chemical Society Monograph Series, Chemical Catalog. Co., New York, 1931; E. Chargaff and J. N. Davidson, 'The Nucleic Acids', Academic Press, New York, 1955; D. O. Jordan, 'Chemistry of Nucleic Acids', Butterworths, London, 1960; (*b*) A. M. Michelson, 'The Chemistry of Nucleosides and Nucleotides', Academic Press, New York, 1963.

[21] G. R. Barker, K. R. Farrar, and J. M. Gulland, *J. Chem. Soc.*, 1947, 21; G. R. Barker and J. M. Gulland, *ibid.*, 1943, 625; P. A. Levene and W. A. Jacobs, *Ber.*, 1909, **42**, 2102; 2469; 2472; 2703.

[22] A. R. Todd, *J. Chem. Soc.*, 1946, **2**, 647; P. A. Levene and R. S. Tipson, *J. Biol. Chem.*, 1932, **94**, 809; 1932, **97**, 491; 1933, **101**, 529; 1934, **105**, 419; 1934, **106**, 113; 1937, **121**, 131; B. Lythgoe and A. R. Todd, *J. Chem. Soc.*, 1944, **2**, 592; D. M. Brown and B. Lythgoe, *ibid.*, 1950, **3**, 1990.

formed by the reaction of a purine or pyrimidine base with D-ribose were established as ring-N glycosides[23] involving N(3) of pyrimidines or N(9) of the purines. The glycoside linkage has the β configuration.[24] It has been suggested on the basis of studies of the optical rotatory dispersion of the nucleosides that both purine and pyrimidine base-residues adopt the *anti*-conformation[25] as shown in Figure 3.

The nucleotides are linked together through diesterified phosphate residues which are formed by reaction with the C(3′) hydroxyl group of one D-ribofuranose moiety and the C(5′) hydroxyl group of another.[26] The structure has been confirmed by synthesis.[27]

4 Chemical Properties of RNA

The internucleotide bond is readily hydrolysed by alkali because the presence of the C(2′) hydroxyl facilitates the formation of an intermediate 2′,3′-cyclic phosphate during hydrolysis.[28] This intermediate has been isolated.[28b] Whereas hydrolysis of the glycoside linkage between purines and deoxyribose is rapid in mildly acidic solutions (pH 3·5 or less) the corresponding bond between purines and ribose is much more stable. More drastic conditions, *e.g.*, treatment with more concentrated hydrochloric acid (pH 1·6) for 24 hr. at 37°, liberate the pyrimidine as well as the purine bases from DNA, but the resistance of the pyrimidine ribosides is such that it is very difficult to get quantitative yields of the free bases from RNA. The glycoside linkage is relatively stable between pH 3 and 7 but can be hydrolysed in more acidic solutions.[29] The internucleotide bond is hydrolysed at an appreciable rate in neutral aqueous salt solutions at 60° or more.[30] The hydrolysis is catalysed by ions such as magnesium,[31a]

[23] P. A. Levene and R. S. Tipson, *J. Biol. Chem.*, 1934, **104,** 385; S. Furberg, *Acta Chem., Scand.*, 1950, **4,** 751; 1952, **6,** 634; S. Furberg, *Acta Cryst.*, 1950, **3,** 329; J. M. Gulland, *J. Chem. Soc.*, 1938, 1722; G. A. Howard, G. S. Kenner, B. Lythgoe, and A. R. Todd, *ibid.*, 1945, 556; J. Davoll, B. Lythgoe, and A. R. Todd, *ibid.*, 1948, 967; B. Lythgoe, H. Smith, and A. R. Todd, *ibid.*, 1946, 355.

[24] H. Davoll, B. Lythgoe, and A. R. Todd, *J. Chem. Soc.*, 1944, 833; V. M. Clark, A. R. Todd, and J. Zussman, *ibid.*, 1951, 2952.

[25] T. R. Emerson, R. J. Swan, and T. L. V. Ulbricht, *Biochem. Biophys. Res. Comm.*, 1965, **19,** 643; T. L. V. Ulbricht, T. R. Emerson, and R. J. Swan, *Tetrahedron Letters*, 1966, 1561.

[26] P. A. Levene and H. S. Simms, *J. Biol. Chem.*, 1962, **70,** 327.

[27] R. Lohrmann, D. Söll, H. Hayatsu, E. Ohtsuka, and H. G. Khorana, *J. Amer. Chem. Soc.*, 1966, **88,** 819.

[28] A. Fonó, *Arkiv. Kemi. Min. Geol.*, 1947, **24, A,** No. 34, p. 1; D. M. Brown and A. R. Todd, *J. Chem. Soc.*, 1952, 52; D. M. Brown, D. I. Magrath, and A. R. Todd, *ibid.*, p. 2708; R. Markham and J. D. Smith, *Biochem. J.*, 1952, **52,** 552.

[29] E. Visher and E. Chargaff, *J. Biol. Chem.*, 1948, **176,** 715.

[30] J. Eigner, H. Boedtker, and G. Michaels, *Biochim. Biophys. Acta*, 1961, **51,** 165.

[31] (*a*) T. Lindahl, *J. Biol. Chem.*, 1967, **242,** 1970; (*b*) J. J. Butzow and G. L. Eichorn, *Biopolymers*, 1965, **3,** 395; G. L. Eichhorn, P. Clark, and E. D. Becker, *Biochemistry*, 1966, **5,** 245; T. Yamani, and N. Davidson, *J. Amer. Chem. Soc.*, 1961, **83,** 2599; (*c*) G. L. Eichhorn and J. J. Butzow, *Biopolymers*, 1965, **3,** 79; (*d*) A. M. Michelson, 'The Chemistry of Nucleosides and Nucleotides', Academic Press, New York, p. 340; (*e*) P. D. Boyer, H. Lardy, and K. Myrbäck, 'The Enzymes', Academic Press, New York, 1961, vol. 5; T. Uchida and F. Egami, 'Methods in Enzymology', ed. L. Grossmann and K. Moldave, Academic Press, New York, 1967, vol. 12A, pp. 228, 239; S. Linn, *ibid.*, p. 247; W. E. Razzell, *ibid.*, p. 255; K. K. Reddi, *ibid.*, p. 257; M. F. Singer and G. Tolbert, *Biochemistry*, 1965, **4,** 1319; H. O. Robertson, R. E. Webster, and N. Zinder, *J. Biol. Chem.*, 1968, **243,** 82.

copper(II), manganese(II), nickel(II), and zinc;[31b] by lanthanum and other tervalent inner transition metals;[31c] by heavy metal hydroxides;[31d] and by a range of hydrolytic enzymes[31e] (nucleases).

The chemical modification of the base-residues has been studied because of their importance in chemical mutagenesis, *e.g.*, mutants of tobacco mosaic virus have been produced through the action of nitrous acid on the RNA component,[32] and because a modified base-residue provides a useful marker in piecing fragments of RNA together in order to ascertain the nucleotide sequence.[33] The chemical reactivity of the base-residues is modified by conformation so that a knowledge of the rate with which a reagent reacts with a base-residue and of the sites which become modified have proved useful in studies of secondary structure[34] as well as in correlating biological activity with primary and secondary structure (a study of 'active sites').[35] Representative reactions of the nucleotides are given in Table 2.

The sodium and potassium salts of polynucleotides are soluble in formamide and dimethyl sulphoxide, as well as in 80% methanol, and the cetyltrimethylammonium salt is soluble in a wide range of organic solvents.[36] Single-stranded polynucleotides, but not double-helical complexes, tend to precipitate at 0° from M sodium or potassium chloride solutions. The base-residues strongly absorb ultraviolet light below 300 mμ.[37] The ordered oligo nucleotides and polynucleotides are optically active[38] owing to their conformation.

5 The Principal Conformations of Ribonucleic Acids

The macromolecular properties of polynucleotides are governed by the negatively charged electrostatic field due to the diesterified phosphate residues (pK = *ca.* 1·6) and by the ability of the base-residues to react with one another to form helical structures. The two forces tend to oppose one another so that conformation depends on ionic strength, temperature, and also on pH (owing to the acid–base properties of the purine and pyrimidine residues). The conformation and physicochemical properties of model polynucleotides have been extensively reviewed.[39] Four conformations of RNA can be defined as described below.

[32] H. Schuster, *Biochem. Biophys. Res. Comm.*, 1960, **2**, 320; W. Vielmetter and H. Schuster, *ibid.*, 1960, **2**, 324.
[33] S. W. Brostoff and V. M. Ingram, *Science*, 1967, **58**, 666.
[34] J. T. Penniston and P. Doty, *Biopolymers*, 1963, **1**, 145, 209.
[35] K. Miura, *Progr. Nucleic Acid Res.*, 1967, **6**, 39; P. B. Moore, *J. Mol. Biol.*, 1966, **22**, 145; G. L. Brown, S. Lee, and D. Metz, 'Genetic Elements: Properties and Function', ed. D. Shugar, Academic Press, New York, 1967, p. 57.
[36] P. O. P. Ts'O, G. K. Helmkamp, C. Sander, and F. W. Studier, *Biochim. Biophys. Acta*, 1963, **76**, 54.
[37] G. H. Beavan, E. R. Holiday, and E. A. Johnson, 'The Nucleic Acids', ed. E. Chargaff and J. N. Davidson, Academic Press, New York, 1955.
[38] I. Tinoco, 'Molecular Biophysics', ed. B. Pullman and M. Weissbluth, Academic Press, New York, 1965.
[39] G. Felsenfeld and H. T. Miles, *Ann. Rev. Biochemistry*, 1967, **36**, 407; A. M. Michelson, 'The Chemistry of Nucleosides and Nucleotides', Academic Press, London, 1963, p. 339; M. L. Petermann, 'The Physical and Chemical Properties of Ribosomes', Elsevier, Amsterdam, 1964.

Table 2 *Examples of the reactivity of the common purine and pyrimidine base-residues*

	Site of reaction			
Reagent	*Adenine*	*Cytosine*	*Guanine*	*Uracil*
Formaldehyde[a]	C(6)-NH_2	C(6)-NH_2	C(2)-NH_2	N(1)
Nitrous acid[b]	C(6)-NH_2	C(6)-NH_2	C(2)-NH_2	—
Dimethyl sulphate[c,d]	N(1)	N(3)	N(7)	—
Diazomethane[c,d]	N(1)	N(3)	N(7)	N(3)
Hydrazine[e]	—	C(4)=C(5)	—	C(4)=C(5)
Hydroxylamine[f]	—	C(4)=C(5)	—	C(4)=C(5)
H_2[d,g]	—	C(4)=C(5)	—	C(4)=C(5)
Halogens[d,h]	C(6) or C(8)	C(5)	C(6) or C(8)	C(5)
Osmium tetroxide[i]	—	C(4)=C(5)	—	C(4)=C(5)
Semicarbazide[j]	—	C(6)-NH_2	—	—
Monoperphthalic acid[k]	N(1)	—	—	—
Kethoxal[l,m]	—	—	N(1)-C(2)-NH_2	
CMEC[n,p]	—	—	N(1)	N(1)
Girard-P[q,r]	—	C(6)-NH_2	—	—
H^+[s]	N(1)=C(6)-NH_2	N(1) C(6)-NH_2	C(2)-NH_2	—
OH^-[s]	—	—	-NH·C(6)O	-NH·C(6)O

[a] H. Fraenkel-Conrat, *Biochim. Biophys. Acta*, 1954, **15**, 307; L. Grossman, S. S. Levine, and W. S. Allison, *J. Mol. Biol.*, 1961, **3**, 47; E. J. Eyring and J. Ofengand, *Biochemistry*, 1967, **6**, 2500; H. Boedtker, *Biochemistry*, 1967, **6**, 2718; [b] J. A. Carbon, *Biochim. Biophys. Acta*, 1965, **95**, 550; H. Schüster and G. Schramm, *Z. Naturforsch.*, 1958, **13b**, 697; R. Shapiro and R. S. Klein, *Biochemistry*, 1966, **5**, 2358; R. Shapiro and S. H. Pohl, *Biochemistry*, 1968, **7**, 448; [c] B. E. Griffin, *Methods in Enzymology*, 1967, **12**, *A*, 141; [d] I. Wempen and J. J. Fox, *Methods in Enzymology*, 1967, **12**, *A*, 76; [e] D. M. Brown, *Methods in Enzymology*, 1967, **12**, *A*, 31; [f] J. H. Phillips, *Methods in Enzymology*, 1967, **12**, *A*, 34; [g] A. H. Schein and F. T. Schein, *Methods in Enzymology*, 1967, **12**, *A*, 38; [h] C. T. Yu and P. C. Zamecnik, *Biochim. Biophys. Acta*, 1963, **76**, 209; *Science*, 1964, **144**, 856; J. H. Weil, *Bull. Soc. Chim. biol.*, 1965, **47**, 1303; [i] K. Burton, N. F. Varney, and P. C. Zamecnik, *Biochem. J.*, 1966, **99**, 29*c*—31*c*; K. Burton, *Biochem. J.*, 1967, **104**, 686; [j] H. Hayatsu, K. Takeishi, and T. Ukita, *Biochim. Biophys. Acta*, 1966, **123**, 445; [k] F. Cramer and H. Seidel, *Biochim. Biophys. Acta*, 1964, **91**, 143; [l] β-Ethoxy-α-ketobutyraldehyde (kethoxal); [m] R. Shapiro and J. Hachmann, *Biochemistry*, 1966, **5**, 2799; M. Litt and V. Hancock, *Biochemistry*, 1967, **6**, 1848; [n] *N*-Cyclohexyl-*N*-β-(4-methylmorpholinium)ethylcarbodi-imide (CMEC); [p] P. T. Gilham, *J. Amer. Chem. Soc.*, 1962, **84**, 687; G. Augusti-Tocco and G. L. Brown, *Nature*, 1965, **206**, 683; D. G. Knorre, E. G. Matygin, G. S. Mushinskaya and V. V. Favorov; *Biokhimiya*, 1966, **31**, 334; [q] acetohydrazide pyridinium chloride; [r] K. Kikugawa, H. Hayatsu, and T. Ukita, *Biochim. Biophys. Acta*, 1967, **134**, 221; [s] A. R. Peacocke, *Chem. Soc. Special Publ.*, 1957, No. 8, 139.

A. The Amorphous Form.—The properties of polyuridylic acid are consistent with an amorphous structure, *i.e.*, the base-residues have no preferred orientation with respect to one another. When uridylic acid is polymerised to polyuridylic acid a hypochromic effect is observed, *i.e.*, the extinction below 300 mμ is decreased by about 10%, and the optical rotatory dispersion per residue is en-

hanced.[40] These optical properties are scarcely affected[41] by increasing the temperature from 25 to 95°. The base-composition of an oligonucleotide in the amorphous form can be deduced from its optical properties.[42]

At moderate ionic strength the pK of the titratable groups of the base-residues will be slightly higher than the pK of the corresponding nucleotide because of the additional work required to move a proton away from the negative field of the sugar phosphate backbone. The form of the pH-titration curve is given[41,43] by the equation (1) where α is the degree of dissociation of the base residue, pK_m

$$pK = pK_m + \Delta pK = pH + \log(1-\alpha)/\alpha \qquad (1)$$

is the negative logarithm of the dissociation constant of the monomer, and ΔpK is dependent on the electrostatic potential ψ of the polynucleotide. It has been suggested that $\Delta pK = 0{\cdot}434\, e\psi/kT$ where e is the charge of the electron.[44] It appears that the negatively charged diesterified phosphate residues (pK = *ca.* 1·5) make the major contribution to ψ so that there is little variation in ΔpK over the range pH 3—14. Thus, for polyuridylic acid the plot of α against pH has the same form as for UMP although the pK is found to be 9·7 in 0·2M-KCl at 25° compared with 9·4 for UMP. However, ψ (and hence ΔpK) is dependent on the ionic strength of the solution so that pK may increase by as much as 1·5 units when the electrolyte concentration is decreased.

The hydrodynamic properties of poly-U are also influenced by the polyelectrolyte properties of the diesterified sugar phosphate residues. Repulsion between the phosphate residues causes the hydrodynamic volume to expand and, in the absence of added electrolyte, the chain is highly extended as shown by the high viscosity and low sedimentation coefficient. When the concentration of electrolyte (*e.g.*, KCl) is increased above about 0·5mM the viscosity of a flexible polyelectrolyte is found[45] to decrease according to the empirical relation (2),

$$[\eta] = [\eta_0]\,(C_{salt}/C^\circ_{salt})^{-m} \qquad (2)$$

whence $[\eta]$ and $[\eta_0]$ respectively are the limiting viscosity numbers when the electrolyte concentration is C_{salt} and C°_{salt} (a reference state) and m lies within the range 0·4—0·6. The viscosity of poly-U obeys[41] equation (2) where $m = 0{\cdot}38$. In 0·15M-NaCl–0·015M-sodium citrate solutions, the dependence of $S_{20,w}$ and $[\eta]$ on molecular weight are given[43] by the equations (3) and (4).

$$S_{20,w} = 3{\cdot}29 \times 10^{-2} M^{0{\cdot}42} \qquad (3)$$

$$[\eta] = 8{\cdot}98 \times 10^{-5} M^{0{\cdot}75} \qquad (4)$$

[40] H. Simpkins and E. G. Richards, *J. Mol. Biol.*, 1967, **29**, 349.
[41] E. G. Richards, C. P. Flessel, and J. R. Fresco, *Biopolymers*, 1963, **1**, 431.
[42] S. Mandeles and C. R. Cantor, *Biopolymers*, 1966, **4**, 759.
[43] R. C. Warner and E. Breslow, Proc. 4th Int. Congr, Biochem., Vienna, 1958, vol. 9, p. 157.
[44] C. Tanford, 'Physical Chemistry of Macromolecules', J. Wiley, New York, 1961; A. Katchalsky, J. Mazur, and P. Spitnik, *J. Polymer Sci.*, 1957, **23**, 513; A. Katchalsky, Z. Alexandrowicz, and O. Kedem, 'Transactions of the Symposium on Electrolyte Solutions', The Electromechical Society, New York, 1965.
[45] R. A. Cox, *J. Polymer Sci.*, 1960, **47**, 441.

B. Single-stranded Helical Structures.—Many polynucleotides approach the amorphous state at high temperatures. However, the base-residues have an increasing tendency to interact with one another as the temperature is lowered.[46] The term 'stacking' is often used to describe this interaction because the flat purine or pyrimidine rings tend to 'stack' like a pile of pennies with the plane of the base-residues perpendicular to the helix axis.[47] It appears that the distance between adjacent base-residues is about 3·4 Å and that there is a translation of 30—45° between adjacent residues, there being 8—12 residues per turn of the helix.[48] The structure does not seem to be rigid except possibly at low temperatures if the degree of stacking is very high. Interaction takes place principally between nearest neighbours[49] so that the dinucleotide may be regarded as a model for the single-stranded polymer. The dinucleotide adenyl (3′, 5′) adenylate (ApA) has been extensively studied and the enthalpy and entropy for the reaction 'stacked' → amorphous form were calculated[50] to be +8 kcal./mole and +25 to 30 cal. mole^{-1} deg.$^{-1}$. Similar values were calculated for the enthalpy and entropy of polyadenylic acid, confirming that pairs of base-residues contribute independently and non-co-operatively to the equilibrium between 'stacked' and amorphous conformations.[47] Measurements of mass per unit length confirm that poly A is single-stranded.[51]

The optical properties of ApA are typical of poly-A which has a single helical conformation.[47,49,50] Thus the maximum extinction (ϵ_p) per residue (observed at about 260 mμ) is 13,800 at 4° and increases gradually to 14,700 when the temperature is increased to 65° whereas ϵ_p for the monomer is 15,400. The mononucleotide has a small positive rotation owing to the ribose moiety. The dimer has a much larger rotation which arises from the dissymmetry of the conformation. A Cotton effect is observed below 300 mμ within the region of the spectrum where the base-residues absorb and circular dichroism is observed.

'Stacking' tends to suppress the ionisation of the base-residues and may modify the shape of the pH-titration curve[52] which may have a slope (dα/dpH) that is no more than twice that found for mononucleotides. Ionisation of the base-residues tends to reduce 'stacking'. Because polynucleotides are also polyelectrolytes, it might be expected that the stability of the single-helical conformation will depend on the electrolyte concentration. The available data[53] show that the effect is small.

An idea of the hydrodynamic properties of the single-helical conformation is derived from the properties of poly-A. The dependence of $S_{20,w}$ and [η] upon

[46] A. M. Michelson, *J. Chem. Soc.*, 1959, 1371.
[47] J. Brahms, A. Michelson, and K. E. Van Holde, *J. Mol. Biol.*, 1966, **15**, 467.
[48] C. A. Bush and I. Tinoco, jun., *J. Mol. Biol.*, 1967, **23**, 601.
[49] M. Leng and G. Felsenfeld, *J. Mol. Biol.*, 1966, **15**, 455.
[50] K. E. Van Holde, J. Brahms, and A. M. Michelson, *J. Mol. Biol.*, 1965, **12**, 726.
[51] V. Luzzati, A. Mathis, F. Masson, and J. Witz, *J. Mol. Biol.*, 1964, **10**, 28.
[52] R. A. Cox, *Biochem. J.*, 1966, **100**, 146; H. Simpkins and E. G. Richards, *Biochemistry*, 1967, **6**, 25k3.
[53] (*a*) R. A. Cox and K. Kanagalingam, *Biochem. J.*, 1967, *A*, **103**, 431; (*b*) R. A. Cox. and K. Kanagalingam, *ibid.*, 1967, *A*, **103**, 749; (*c*) D. Barszcs and D. Shugar, *Acta Biochim. Polon.*, 1964, **11**, 481.

molecular weight was found[54] to be as given by equations (5) and (6), indicating

$$S_{20,w} = 2{\cdot}1 \times 10^{-2} M^{0{\cdot}45} \quad (5)$$

$$[\eta] = 6{\cdot}2 \times 10^{-4} M^{0{\cdot}65} \quad (6)$$

that at 20° in 0·15M-NaCl–0·015M-sodium citrate, pH 7·1, owing to base-stacking, poly-A forms a more compact coil than poly-U. The viscosity was found to increase when the electrolyte concentration decreased as expected of a flexible polyelectrolyte. At lower temperatures, when the tendency to 'stack' is strong, the hydrodynamic properties are no longer consistent with a random coil.[55]

The differences in the spectra of 'stacked' and amorphous forms reported[53b] for oligo-A, oligo-C, and oligo-G are distinct so that the hypochromism of oligo- and poly-nucleotides due to 'stacking' depends on base composition. Studies of the hypochromism of di- and tri-nucleotides suggest that base-sequence also influences hypochromism.[56]

The optical properties of the single-stranded form of a polynucleotide can be obtained to a first approximation by preventing the formation of base-pairs, *e.g.*, by reaction with formaldehyde.[57]

C. The Double-helical Conformation.—*X*-Ray diffraction shows[58] that high-molecular-weight RNA in the double-helical form and fragments of RNA, probably of ribosomal origin, have a structure that strongly resembles the A form of DNA: it consists of two antiparallel polynucleotide chains stabilised by Watson–Crick base-pairs, the distance between base-pairs being 3·05 ± 0·05 Å, there being 10 or 11 base-pairs per turn of the helix. The 2′-OH groups of ribose may be free to form intermolecular hydrogen bonds. When native (double-helical) DNA is heated in solution, denaturation (*i.e.*, a transition to a single-stranded form) takes place over a very narrow temperature range; the mid-point of this transition range is termed the melting temperature (T_m). In neutral solutions, the T_m of an RNA double helix is about 10° greater than that of a DNA double helix of similar nucleotide composition.[59] The properties of the RNA double-helix in general resemble those of DNA. Thus the thermal stability of the double-helix form appears to increase as the properties of G:C base-pairs increase; the transition from the double-helical to single-stranded

[54] J. R. Fresco and P. Doty, *J. Amer. Chem. Soc.*, 1957, **79**, 3928.
[55] H. Eisenberg and G. Felsenfeld, *J. Mol. Biol.*, 1967, **30**, 17.
[56] G. B. Zavil'gel'skii, T. V. Venkstern, and A. A. Baer, *Doklady Biochemistry*, 1966, **166**, 38; G. B. Zavil'gel'skii, T. V. Venkstern, and A. A. Baer, *Doklady Akad. Nauk S.S.S.R.*, 1966, **166**, 978; W. M. Stanley, jun., and R. M. Bock, *Analyt. Biochem.*, 1965, **13**, 43.
[57] G. D. Fasman, C. Lindblow, and E. Seaman, *J. Mol. Biol.*, 1965, **12**, 630; H. Boedtker, *Biochemistry*, 1967, **6**, 2718; R. A. Cox and K. Kanagalingam, *Biochem. J.*, 1968, **108**, 599.
[58] P. J. Gomatos and I. Tamm, *Proc. Nat. Acad. Sci. U.S.A.*, 1963, **49**, 707; P. J. Gomatos and R. Langridge, *Science*, 1963, **141**, 694; K. Tomita and A. Rich, *Nature*, 1964, **201**, 1160; S. Arnott, F. Hutchinson, M. Spencer, M. H. F. Wilkins, W. Fuller, and R. Langridge, *Nature*, 1966, **211**, 227.
[59] H. C. Kaerner and H. Hoffman-Berling, *Z. Naturforsch.*, 1964, **19b**, 593; A. R. Bellamy and W. K. Joklik, *J. Mol. Biol.*, 1967, **29**, 19.

forms is abrupt owing to the co-operative nature of the transition, and thermal stability depends on the electrolyte concentration[60a] (the temperature at which denaturation takes place increases by about 12° when the concentration of electrolyte is increased ten-fold within the range 1 mM—M). Multivalent cations[60b] and polyamines[60c] may either stabilise or labilise the double-helical form.

The hydrodynamic properties of double-helical RNA would be expected to be the same as those of native DNA for which the relation between $S^0_{20,w}$ and M was found[61] to be as in equations (7)—(10). The double-helical form can be

$$S^\circ_{20,w} = 0{\cdot}116M^{0{\cdot}325} \text{ (where } M < 4 \times 10^6 \text{ daltons)} \quad (7)$$

$$S^\circ_{20,w} = 0{\cdot}034M^{0{\cdot}405} \text{ (where } M > 4 \times 10^6 \text{ daltons)} \quad (8)$$

$$[\eta] = 1{\cdot}05 \times 10^{-7}M^{1{\cdot}32} \text{ (where } M < 2 \times 10^6 \text{ daltons)} \quad (9)$$

$$[\eta] = 6{\cdot}9 \times 10^{-4}M^{0{\cdot}70} \text{ (where } M > 2 \times 10^6 \text{ daltons)} \quad (10)$$

seen in the electron microscope[62] and its mass per unit length can be deduced by low-angle X-ray scattering.[51]

When solutions are made acid or alkaline T_m is decreased as the pH is made more extreme until denaturation takes place at the ambient temperature.[63] Ionisation is suppressed in the double-helical form so that denaturation (which remains co-operative at all pH values) is also accompanied by an abrupt increase in the degree of ionisation of base-residues. The difference in the spectra of the double-helical and single-stranded forms over the range 220—300 mμ accurately reflects the nucleotide composition of the double-helix.[64]

D. The Hairpin-loop Conformation.—Many species of RNA are single-stranded at high temperatures and low ionic strengths, but as the temperature is decreased or the ionic strength increased the polynucleotide folds upon itself, forming short hairpin loops that are stabilised by interaction principally between complementary base-residues located on different segments of the same molecule.[65] The transitions are illustrated in the Figure 4. The partly double-helical form shown in (*c*) may be over-simplified but serves to illustrate the principal features. Three parameters, the number N of base-pairs per loop, the number b of unpaired residues per loop, and the number c of unpaired residues which link one hairpin loop to another, characterise the structure. Approximate values of N, b, and c can be deduced from degradation studies.[66]

[60] (*a*) A. R. Bellamy and W. K. Joklik, *J. Mol. Biol.*, 1967, **29**, 19; C. Schildkraut and S. Lifson, *Biopolymers*, 1965, **3**, 195; W. F. Dove and N. Davidson, *J. Mol. Biol.*, 1962, **5**, 467; (*b*) G. L. Eichhorn, *Nature*, 1962, **194**, 474; (*c*) E. Gabbay and R. Kleinman, *J. Amer. Chem. Soc.*, 1967, **89**, 7123.

[61] J. Eigner and P. Doty, *J. Mol. Biol.*, 1965, **12**, 549.

[62] N. Granboulan and R. M. Franklin, *J. Mol. Biol.*, 1966, **22**, 173.

[63] Ch. Zimmer and H. Venner, *J. Mol. Biol.*, 1963, **7**, 603.

[64] G. Felsenfeld and G. Sandeen, *J. Mol. Biol.*, 1962, **5**, 587; S. Z. Hirschman and G. Felsenfeld, *ibid.*, 1966, **16**, 347; H. R. Mahler, B. Kline, and B. D. Mehrotra, *ibid.*, 1964, **9**, 801.

[65] (*a*) A. S. Spirin, *Progr. Nucleic Acid Res.*, 1963, **1**, 301; P. O. P. Ts'o, *Ann. Rev. Plant Physiol.*, 1962, **13**, 45; (*b*) P. Doty, H. Boedtker, J. R. Fresco, R. Haselkorn, and M. Litt, *Proc. Nat. Acad. Sci. U.S.A.*, 1959, **45**, 482.

[66] R. A. Cox, *Biochem. J.*, 1968, **106**, 725; R. A. Cox, H. J. Gould, and K. Kanagalingam, *ibid.*, p. 733.

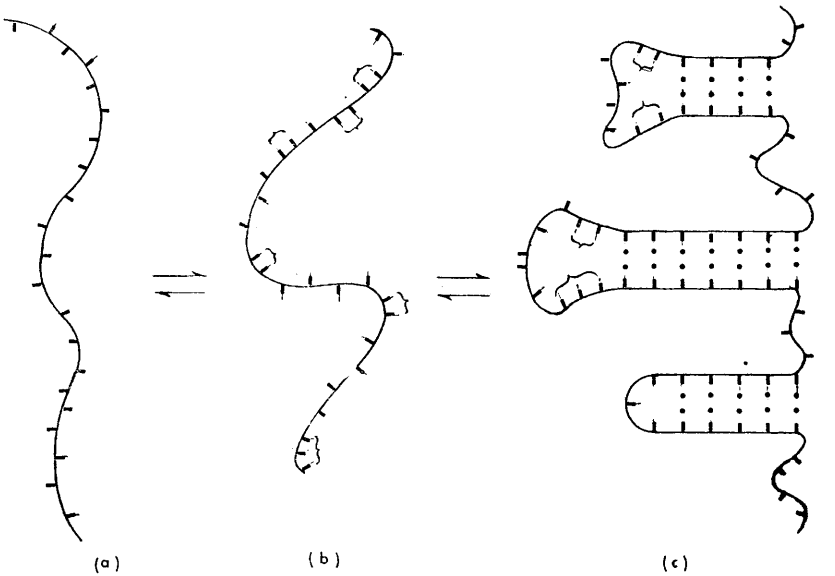

Figure 4 *The principal conformations of a single-stranded polynucleotide* (*cf.* R. A. Cox and K. Kanagalingam, *Biochem. J.*, 1967, **103**, 431). (*a*) the amorphous form in which the base-residues (indicated by the short horizontal lines) have no preferred orientation with respect to one another: (*b*) the single-stranded 'stacked' form in which the base-residues tend to pile one upon another to form a single helix: (*c*) the hairpin-loop conformation where N is the number of base-pairs per loop, b is the number of unpaired residues within the loop, and c is the number of unpaired residues linking one hairpin loop to another

The optical properties[67] and acid-base properties[68] of the partly double-helical form appear to be the sum of the properties of the double-helical and single-stranded regions. However, the temperature at which the double-helical structure reverts to the single-stranded form depends on N, the number of base-pairs per segment, as well as upon pH and electrolyte concentration.[69] The relation between T_m for a particular value of N and $T_{m(\infty)}$ (the value of T_m when $N \rightarrow \infty$) has the form (11) where A is a constant.[70] Although equation (11)

$$1/T_m = 1/T_{m(\infty)} + A/N \qquad (11)$$

may not be precise, it serves to show that T_m depends on N when N is small (*e.g.*, <20).

[67] (*a*) J. R. Fresco, L. C. Klotz, and E. G. Richards, Cold Spring Harbor Symp. Quantitative Biology, 1963, **28**, 83; G. Felsenfeld, and G. L. Cantoni, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **51**, 818; (*b*) W. Guschlbauer, *Biophysik.*, 1966, **3**, 156; C. R. Cantor, S. R. Jaskunas, and I. Tinoco, *J. Mol. Biol.*, 1966, **20**, 39.

[68] R. A. Cox and U. Z. Littauer, *Biochim. Biophys. Acta*, 1963, **72**, 188.

[69] P. Doty, *J. Polymer Sci.*, 1961, **55**, 1; M. N. Lipsett, L. Heppel, and D. F. Bradley, *J. Biol. Chem.*, 1961, **236**, 857; M. N. Lipsett, *ibid.*, 1964, **239**, 1256; E. K. F. Bautz and F. A. Bautz, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 1476.

[70] J. Appliquist and V. Damle, *J. Amer. Chem. Soc.*, 1965, **87**, 1450.

The hairpin-loop form appears to behave as a compact random coil in solution and the mass per unit length approaches that of the double-helical form. Various authors[71] have related the sedimentation coefficients S and molecular weight M by means of equation (12), and values of α within the range 0·40—0·55

$$S = KM^{\alpha} \tag{12}$$

have been reported. Thus far the relations described have limited application.

The formation of the hairpin loops leads to an anomalous contraction of the molecule (Figure 5). The contraction[65,72] is evident from the dependence of $[\eta]$

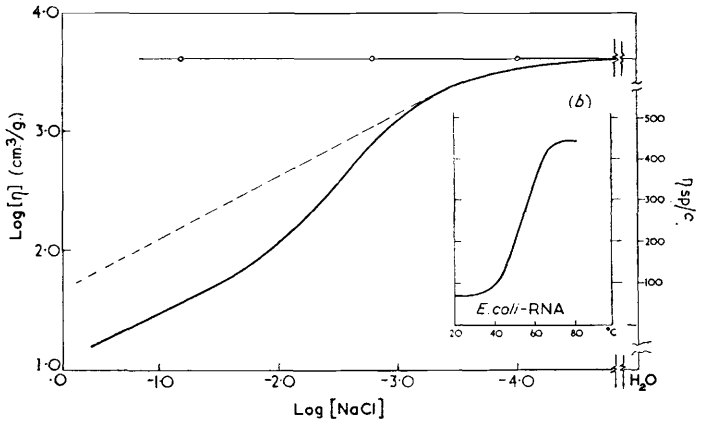

Figure 5a *The dependence of the intrinsic viscosity of RNA and of DNA on the concentration of added sodium chloride at* 25°. (—) sodium ribonucleate in neutral solution (R. A. Cox, *J. Polymer Sci.*, 1960, **47**, 441); (○) sodium deoxyribonucleate, data of Conway and Butler, *J. Polymer Sci.*, 1953, **11**, 277; 1954, **12**, 199; (- -) values calculated (R. A. Cox, *loc. cit.*) for RNA if simple polyelectrolyte behaviour were followed as in equation (1) assuming $m = 0.6$

Figure 5b *Temperature-dependence of reduced viscosity of high-polymer RNA preparations of* E. Coli *RNA* (2·3 *mg./ml.*) *in phosphate buffer*, pH 7·3, *of ionic strength* 0·1, *in the presence of* 0·01M-versene (after Spirin, Colloq. int. Cent. nat. Rech. Sci. No. 106, Acides Ribonucleique et Polyphosphates; Structure, Synthèse, et Fonctions, Strasbourg, p. 81)

on ionic strength since equation (2) is no longer obeyed and the observed viscosity is about one-quarter of the predicted value. When the temperature is increased, the viscosity of the polynucleotide (*e.g.*, in 0·1M-NaCl) increases about four-fold until the polynucleotide becomes entirely single-stranded. Judged by many criteria the transition between the entirely single-stranded and partly double-helical forms takes place reversibly. The extent to which double-helical regions are formed depends on the precise nucleotide sequence. Random

[71] A. Maeda, *J. Biochem.* (*Tokyo*), 1961, **50**, 377; A. S. Spirin, *Biochemistry* (*U.S.S.R.*), 1961, **26**, 454; C. G. Kurland, *J. Mol. Biol.*, 1960, **2**, 83; R. F. Gesteland and H. Boedtker, *ibid.*, 1964, **8**, 496.

[72] (*a*) R. A. Cox and U. Z. Littauer, *Biochim. Biophys. Acta*, 1962, **61**, 197; U. Z. Littauer and H. Eisenberg, *Biochim. Biophys. Acta*, 1959 **32**, 320; (*b*) H. Boedtker, *J. Mol. Biol.*, 1960, **2**, 171.

copolymers of AMP and UMP or GMP and CMP readily form such structures. The transition between one form and another can be followed by measuring the changes in extinction,[65,67] optical rotatory dispersion,[67b] circular dichroism,[73] ^{1}H n.m.r. spectra,[74] mass per unit length,[75] or hydro-dynamic volume[65a,72] or by the techniques of polarography,[76] infrared spectroscopy,[77] and electron microscopy.[62,78]

6 The Macromolecular Structure and Properties of the Principal Species of RNA

A. Messenger. RNA.—The isolation of pure messenger RNA has proved difficult so that the macromolecular properties of this species are not known in detail. The size of messenger is determined by the size of the protein or proteins for which it codes. Messenger RNA may be polycistronic, *i.e.*, code for more than one protein, so that a wide range of molecular weights may be encountered.[79] There is increasing evidence that the messenger for haemoglobin has been isolated[80] and its sedimentation coefficient is about 8—10*S* in 0·1M-KCl in accord with a chain length of *ca.* 450 residues, the appropriate size of messenger coding for one of the polypeptide chains (about 150 amino-acid residues) of hæmoglobin. Messenger from rat liver appears to be heterogeneous[81] and some species may be as large as 10^6 daltons or more, judged by their sedimentation properties.

There is no information about the secondary structure of messenger RNA.[82] However, it has been shown that the presence of stable secondary structure inhibits the translation of the genetic message. The synthetic messenger polyuridylic acid, which has little or no organised secondary structure due either to stacking or hydrogen bonding, very efficiently directs the synthesis of polyphenylalanine.[83] If poly-U is first allowed to react with poly-A to form the double-helical complex poly(A + U) there is no synthesis of polyphenylalanine. The nucleotide sequence of 'messenger' RNA has not been studied.

[73] J. Brahms, *Proc. Roy. Soc.*, 1967, A, **297**, 152.
[74] C. C. Macdonald, W. D. Phillips, and M. Penswick, *Biopolymers*, 1965, **3**, 609.
[75] V. Luzzati, J. Witz, and A. Mathis, 'Genetic Elements: Properties and Function', ed. D. Shugar, Academic Press, New York, 1967, p. 41; S. N. Timasheff, J. Witz, and V. Luzzatti, *Biophys. J.*, 1961, **1**, 525.
[76] E. Palaček, *J. Mol. Biol.*, 1966, **20**, 263.
[77] H. T. Miles, *Biochim. Biophys. Acta*, 1958, **30**, 324; F. B. Howard, J. Frazier, M. F. Singer, and H. T. Miles, *J. Mol. Biol.*, 1966, **16**, 415.
[78] D. Danon, Y. Marikovsky, and U. Z. Littauer, *J. Biophys. Biochem. Cytol.*, 1961, **9**, 253; N. A. Kisselev, L. P. Gavrilova, and A. S. Spirin, *J. Mol. Biol.*, 1961, **3**, 778.
[79] B. H. Hoyer, B. J. McCarthy, and E. T. Bolton, *Science*, 1963, **140**, 1408; K. Asano, *J. Mol. Biol.*, 1965, **14**, 71; G. Attardi, H. Parnas, M. I. Hivany and B. Attardi, *ibid.*, 1966, **20**, 145; K. Scherrer, K. Marcaud, F. Zajdela, I. M. London, and F. Gros, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **56**, 1571; D. D. Brown and E. Littner, *J. Mol. Biol.*, 1966, **20**, 81, 95.
[80] G. Marbaix and A. Burny, *Biochem. Biophys. Res. Comm.*, 1964, **16**, 522; A. Burny and G. Marbaix, *Biochim. Biophys. Acta*, 1965, **103**, 409; G. Huez, A. Burny, G. Marbaix, and E. Schramm, *European J. Biochem.*, 1967, **1**, 179.
[81] T. Staehelin, F. O. Wettstein, H. Oura, and H. Noll, *Nature*, 1964, **201**, 264.
[82] E. P. Guiduschek, J. W. Moohr, and S. B. Weiss, *Proc. Nat. Acad. Sci. U.S.A.*, 1962, **48**, 1078.
[83] J. H. Matthaei and M. W. Nirenberg, *Proc. Nat. Acad. Sci. U.S.A.*, 1961, **47**, 1580, 1588.

B. Hybridisation of RNA with DNA.—The formation of a double-helix between a polyribonucleotide strand and a complementary polydeoxyribonucleotide strand is predicted on the basis of the theory of protein biosynthesis.[84] The first demonstration that a hybrid helix could be formed was achieved by use of model polynucleotides. The interaction *in vivo*[85] and *in vitro* between RNA and denatured DNA was developed as a means of isolating messenger RNA:[86] t-RNA and ribosomal RNA interact with only a small fraction of the total DNA.

C. Ribosomal RNA.—Ribosomes are nucleoprotein particles containing 50—60% of RNA which can be dissociated reversibly into two unequal subparticles having the same RNA:protein ratio. The structural RNA from each subparticle appears to be a continuous polynucleotide chain.[65a] The molecular weight of RNA from the smaller subparticle is about $0{\cdot}5 \times 10^6$ daltons irrespective of its origin. Two species of RNA have been isolated from the larger subparticle. The larger species has a molecular weight[9,65a] of about $(1—1{\cdot}6) \times 10^6$ daltons (depending on the source of the ribosomes). The smaller species[87] is known as 5S-RNA and has a molecular weight of about 35,000 daltons. The RNA species differ in their nucleotide composition (see Table 3) and appear to have few sequences in common since each species hybridises with its own DNA cistron.[88] The nucleotide composition of bacterial ribosomal RNA does not vary to the same extent as the base composition[89] of bacterial DNA. Presumably, the ribosomal RNA cistron which accounts for no more than 3% of the total DNA is copied very many times.

Higher organisms (unlike bacteria which seem to have only one species of ribosome) contain ribosomes of different sizes. The ribosomes of the cytoplasm have a molecular weight of about 4×10^6 daltons whereas ribosomes of the mitochondrion (or of chloroplasts) are about 3×10^6 daltons.[90] The larger species of RNA from cytoplasmic ribosomes has a molecular weight of about $1{\cdot}6 \times 10^6$ daltons compared with about $1{\cdot}0 \times 10^6$ daltons for RNA isolated from mitochondrial ribosomes.

[84] A. Rich, *Proc. Nat. Acad. Sci. U.S.A.*, 1960, **46**, 1044; C. L. Schildkraut, J. Marmur, J. Fresco, and P. Doty, *J. Biol. Chem.*, 1961, **236**, PC4.
[85] G. Richter and H. Sanger, *Biochim. Biophys. Acta*, 1965, **95**, 362; M. N. Hayashi and M. Hayashi, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 635; Y. Miura, H. Itoh, K. Sunaga, T. Nishizawa, and I. Ohki, *Biochim. Biophys. Acta*, 1967, **134**, 258.
[86] H. F. Lodish and N. D. Zinder, *Biochem. Biophys, Res. Comm.*, 1965, **19**, 269; D. Gillespie and S. Spiegelman, *J. Mol. Biol.*, 1965; **12**, 830; B. D. Hall and S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.*, 1961, **47**, 137.
[87] R. J. Bachvaroff and V. Tongur, *Nature*, 1966, **211**, 248; R. Rosset and R. Monier, *Biochim. Biophys. Acta*, 1963, **68**, 653.
[88] S. A. Yankofsky and S. Spiegelman, *Proc. Nat. Acad. Sci. U.S.A.*, 1963, **48**, 1069; 1963, **49**, 538; G. Attardi, P. Huang, and S. Kabat, *ibid.*, 1965, **54**, 185; H. Wallace and M. Birnstiel, *Biochim. Biophys. Acta*, 1966, **114**, 296; I. Merts, W. Schulze, and L. R. Overby, *Arch. Biochem. Biophys.*, 1966, **115**, 197; D. Apirion, *J. Mol. Biol.*, 1967, **30**, 255.
[89] A. N. Belozersky and A. S. Spirin, 'The Nucleic Acids', ed. E. Chargaff and J. N. Davidson, Academic Press, New York, 1960, vol. 3, p. 147.
[90] (*a*) G. Brawerman, *Biochim. Biophys. Acta*, 1963, **72**, 317; J. L. Chen and S. G. Wildman, *Science*, 1967, **155**, 1271; T. W. O'Brien and G. F. Kalf, *J. Biol. Chem.*, 1967, **242**, 2180; (*b*) P. J. Rogers, B. N. Preston, E. B. Titchener, and A. W. Linnane, *Biochem. Biophys. Res. Comm.*, 1967, **27**, 405.

Table 3 *Base composition of ribosomal RNA isolated from animals, plants, and bacteria**

Organism	S-*value*	*Moles/100 moles of nucleotides*				*Molar ratios*	
		CMP	*AMP*	*GMP*	*UMP*	$(A + U)/(C + G)$	*Ref.*
Krebs ascites cells	30	30·6	16·0	36·1	17·1	0·50	*a*
	18	27·9	21·2	29·4	21·4	0·74	
Rabbit reticulocytes	28	31·6	16·4	35·3	16·6	0·49	*b*
	18	28·8	20·5	30·7	20·0	0·68	
Xenopus laevis	28	30	17	37	16	0·49	*c*
	18	29	22	31	18	0·67	
Potato tuber	25	22·0	25·1	31·7	21·2	0·86	
	16	22·2	25·4	27·2	25·2	1·02	
Pea seedlings	25	22·6	23·6	32·1	21·6	0·82	*d*
	16	20·1	23·7	31·1	25·1	0·96	
Cauliflower	28	22·6	25·7	33·1	18·4	0·79	*e*
	18	20·6	25·6	32·9	21·0	0·87	
Escherichia coli	23	21·5	25·4	33·5	19·6	0·82	*f*
	16	22·7	24·8	31·0	21·5	0·86	
Bacillus subtilis	23	22·5	26·5	32·0	19·3	0·84	*f*
	16	22·3	26·5	29·6	21·6	0·93	
Pseudomanas aeruginosa	23	21·2	26·3	31·2	21·3	0·91	*f*
	16	21·6	25·1	32·8	20·5	0·84	
Drosophila melanogaster	28	19·6	30·8	22·5	27·1	1·38	*g*
	16	20·3	28·8	23·5	27·4	1·28	
Hyalophora cecropia	(total RNA)	23·4	23·2	31·2	22·2	0·83	*h*

* Data taken from (*a*) L. Montagnier and A. D. Bellamy, *Biochim. Biophys. Acta*, 1964, **80**, 157; (*b*) R. H. DeBellis, N. Gluck, and P. A. Marks, *J. Clin. Invest.*, 1964, **43**, 1329; (*c*) D. D. Brown and J. B. Gurdon, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **51**, 139; (*e*) C. J. Pollard, *Biochem. Biophys. Res. Comm.*, 1964, **17**, 171; (*f*) J. E. M. Midgeley, *Biochim. Biophys. Acta*, 1962, **61**, 513; (*g*) G. Hastings and K. Kirby, *Biochem. J.*, 1966, **100**, 532; (*h*) G. R. Wyatt and B. Lenzin, *Biochim. Biophys. Acta*, 1965, **103**, 588. Ribosomal RNA from different organs of the same species have the same nucleotide composition (Hirsch, *Biochim. Biophys. Acta*, 1966, **123**, 246) and the same sedimentation coefficient (R. E. Click and B. L. Tint, *J. Mol. Biol.*, 1967, **25**, 111).

Comparatively little is known about the secondary structure of ribosomal RNA. Apparently structural RNA of both the smaller and larger subparticles have a hairpin-loop structure within the ribosome which is retained in solution after the protein moiety is removed.[91] About 50—70% of the base-residues form base-pairs.[65,92] The average size of the hairpin loops was estimated, on the basis of the statistics of degradation of linear polymers, to be 25 ± 5 residues and 35 ± 7 residues respectively for RNA from the smaller and larger ribosomal subparticles of rabbit reticulocytes.[66] These values are towards the lower limit obtained from measurements of mass per unit length.[51,75]

The secondary structure of ribosomal RNA 'melts' over a broad temperature range, suggesting that the double-helical segments differ appreciably either in the number or composition of the base-pairs.[92] Analysis of the nearest neighbours of the nucleotides of rat liver ribosomal RNA shows that long runs of a particular sequence can be excluded.[93] Partial enzymic hydrolysis leads to the production of discrete fragments which differ in nucleotide composition.[53a,94] The stepwise hydrolysis of rat liver ribosomal RNA suggests that uracil residues are unevenly distributed along the chain.[95] The heterogeneity of the hairpin loops is matched by the heterogeneity of ribosomal proteins.[96]

Ribosomal RNA exhibits hysteresis[93,97] on titration over the range pH 3—7, since curve I (Figure 6) is followed on titration from pH 7 to pH 3 whereas curve II (Figure 6) is followed on titration with alkali from pH 3 to pH 7. Curves I and II are accurately reproduced on successive titration cycles and the hysteresis loop can be scanned by titration from intermediate pH values. The observations suggest that more than one conformation of ribosomal RNA is stable over the range pH 3—7. Fragments of ribosomal RNA having a molecular weight of about 13,000 daltons, obtained by hydrolysis in 0·3N-KOH, form crystallites which have an *X*-ray diffraction pattern that is characteristic of double-helical RNA.[98] The relation of these fragments to the intact molecule has yet to be established but they may be single intact or damaged hairpin loops.

[91] C. E. Hall and H. S. Slater, *J. Mol. Biol.*, 1959, **1**, 329; H. E. Huxley and G. Zubay, *ibid.*, 1960, **2**, 10; G. Zubay and M. H. F. Wilkins, p. 105; F. Bonhoeffer and H. K. Schachman, *Biochem. Biophys. Res. Comm.*, 1960, **2**, 366; D. Schlessinger, *J. Mol. Biol.*, 1970, **2**, 92; A. Klug, K. C. Holmes, and J. T. Finch, *J. Mol. Biol.*, 1961, **3**, 87; A. Blake and A. R. Peacocke, *Nature*, 1965, **208**, 1319; P. McPhie and W. Gratzer, *Biochemistry*, 1966, **5**, 1310; I. R. Cotter, P. McPhie, and W. Gratzer, *Nature*, 1967, **216**, 864.

[92] R. A. Cox, *Biochem.*, 1966, **98**, 841.

[93] A. A. Hadjiolov, P. V. Venkov, and L. B. Dolopchiev, *Biochim. Biophys. Acta*, 1965, **108**, 220.

[94] P. McPhie, J. Hounsell, and W. B. Gratzer, *Biochemistry*, 1965, **5**, 988; H. Gould, *ibid.*, p. 1103; H. Gould, *Biochim. Biophys. Acta*, 1966, **123**, 441; N. Delihas and J. Bertman, *J. Mol. Biol.*, 1966, **21**, 391; N. Delihas, *Biochemistry*, 1967, **6**, 3356.

[95] A. A. Hadjiolov, P. V. Venkov, L. B. Dolapochiev, and D. D. Genchev, *Biochim. Biophys. Acta*, 1967, **142**, 111.

[96] M. G. Hamilton and M. E. Ruth, *Biochemistry*, 1967, **6**, 2585; W. Möller and A. Chrambach, *J. Mol. Biol.*, 1967, **23**, 377; P. B. Moore, R. R. Traut, H. Noller, P. Pearson, and H. Delius, *ibid.*, 1968, **31**, 441.

[97] R. A. Cox, *Biochim. Biophys. Acta*, 1963, *A*, **68**, 401; R. A. Cox, *ibid.*, 1963, **72**, 203.

[98] M. Spencer and F. Poole, *J. Mol. Biol.*, 1965, **11**, 314; W. J. Pigram, J. Littlechild, and M. Spencer, *J. Mol. Biol.*, 1968, in the press.

Studies of nucleotide sequence have been confined[99] to the identification of the base-residues at the C(3′)-OH and C(5′)-OH termini.

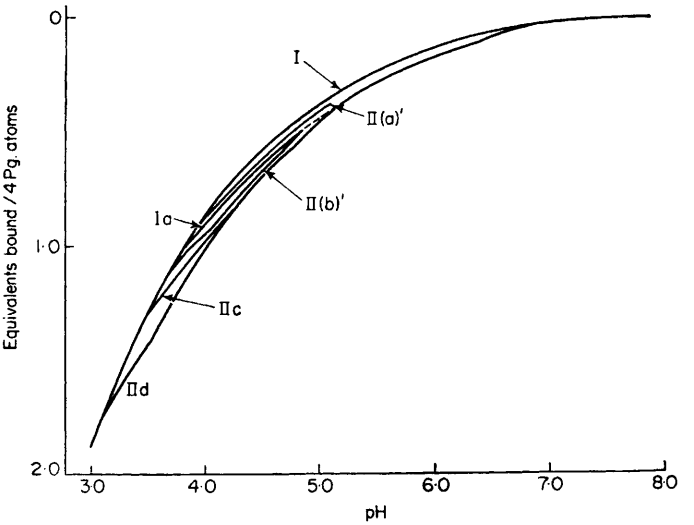

Figure 6 *Scanning curves of ribosomal RNA from* E. coli
Solvent: 0·1M-NaCl at 0·1°. Curve I: Titration with acid from pH 8. Curves I and II*a*: Titration from pH 8·00 to pH 3·80 to pH 5·2 to pH 3·80. Curves I, II*b*, and I*a*: respectively: Titration from pH 8·0 to pH 3·58, pH 3·58 to pH 4·88 to pH 3·58. Curves I and II*c*: Titration from pH 8·0 to pH 3·37 to pH 8·0. Curves I and II*d*: Titration from pH 8·0 to pH 2·80 to pH 8·0

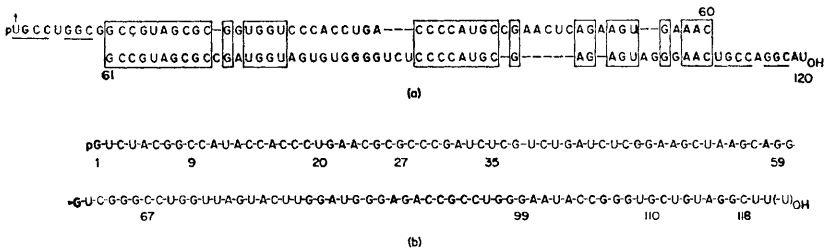

Figure 7a *Homologies between the two halves of the sequence of* 5S-*RNA* The residues are numbered as in Figure 1. Homologies are shown by the boxed areas. Dashes are where gaps have to be left in the sequence in order to maximise these homologies. The underlining shows similarities between the two ends of the molecule (G. G. Brownlee, F. Sanger, and B. G. Barrell, *Nature*, 1967, **215**, 735).

Figure 7b *Sequence of KB cell 5S RNA* (B. G. Forget and S. M. Weissman, *Science*, 1967, **158**, 1695)

[99] B. G. Lane, *Biochemistry*, 1965, **4**, 212; J. A. Hunt, *Biochem. J.*, 1965, **95**, 541; J. E. M. Midgley, *Biochim. biophys. Acta*, 1965, **95**, 232; J. E. M. Midgley, *ibid.*, 1966, **123**, 210; J. E. M. Midgley and D. J. McIlreavy, *ibid.*, 1967, **145**, 512; D. J. McIlreavy, and J. E. M. Midgeley, *ibid.*, 1967, **142**, 47; M. Takanami, *J. Mol. Biol.*, 1967, **23**, 135.

D. 5S-RNA.—The function of 5S-RNA associated with the larger ribosomal subparticle is unknown but the nucleotide sequence of 5S-RNA from *Escherichia coli* (V) and from KB carcinoma cells (VI) has been reported (Figure 7). In both cases the molecules may have a clover-leaf conformation[100] like that suggested for t-RNA (Figure 8) although this notion is far from established. Residues 31—37 (VI) and 41—47 are identical and so are the sequences 66—71 and 93—98: in (V) the sequences 10—19, and 61—70 are identical and so are the sequences 35—42 and 90—97 (see Figure 7). 5S-RNA may be able to combine specifically with t-RNA.

E. Transfer RNA.—The role of transfer RNA in protein biosynthesis was described above. This adaptor molecule is capable of combining chemically with the particular amino-acid specified by the anticodon. The nucleotide sequences of 5 t-RNA molecules from yeast have been determined, four of which are given in Figure 8. Each species appears to be capable of forming a clover-leaf structure. The possible conformation of the anticodon loop deduced from model building is given in Figure 9. The general features of the clover-leaf structure include a high concentration of dihydrouracil in one loop, and the anticodon is flanked by uncommon base-residues: the anticodon itself may include[101] a 'wobble', *i.e.*, a base-residue that is capable of forming base-pairs other than A:U and G:C. The regulation of secondary structure by means of the large number of modified base-residues is of particular interest.

t-RNA possesses a specific tertiary structure.[102] When certain species of t-RNA are heated to 70° in the presence of ethylenediaminetetra-acetic acid and cooled, their biological activity and chromatographic behaviour are altered. The original active form may be regained by heating to 70° in the presence of Mg^{2+}.

The secondary structure of t-RNA has been intensively studied by optical[103] and physical [104] methods as well as by chemical methods[11b] based on measurements of the rate of reaction with reagents that are specific for particular base-residues (see Table 2). In general, the results are in agreement with the clover-leaf structure.

F. Viral RNA.—Each species of viral RNA has unique biological properties[105] which arise from a unique primary structure (the nucleotide composition and

[100] H. Boedtker and D. G. Kelling, *Biochem. Biophys. Res. Comm.*, 1967, **29**, 758; C. R. Cantor, *Nature*, 1967, **216**, 513.
[101] F. H. C. Crick, *J. Mol. Biol.*, 1966, **19**, 548.
[102] W. Gartland and N. Sueoka, *Proc. Nat. Acad. Sci. U.S.A.*, 1966, **55**, 948; T. Lindahl, A. Adams, and J. R. Fresco, *ibid.*, p. 941; A. Adams, T. Lindahl, and J. R. Fresco, *ibid.*, 1967, **57**, 1684.
[103] C. R. Cantor, S. R. Jaskunas, and I. Tinoco, *J. Mol. Biol.*, 1966, **20**, 39; J. N. Vournakis and H. A. Scheraga, *Biochemistry*, 1966, **5**, 2997.
[104] J. A. Lake and W. W. Beeman, *J. Mol. Biol.*, 1968, **31**, 115; V. G. Tumangan, N. G. Esipova, and L. L. Kiselev, *Doklady Biochemistry*, 1966, **168**, 180; C. C. MacDonald, W. D. Phillips, and J. Penswick, *Biopolymers*, 1965, **3**, 609; D. Bell and G. J. Russell, *Biochemistry*, 1967, **6**, 3363.
[105] A. J. D. Bellet, *J. Virology*, 1967, **1**, 245.

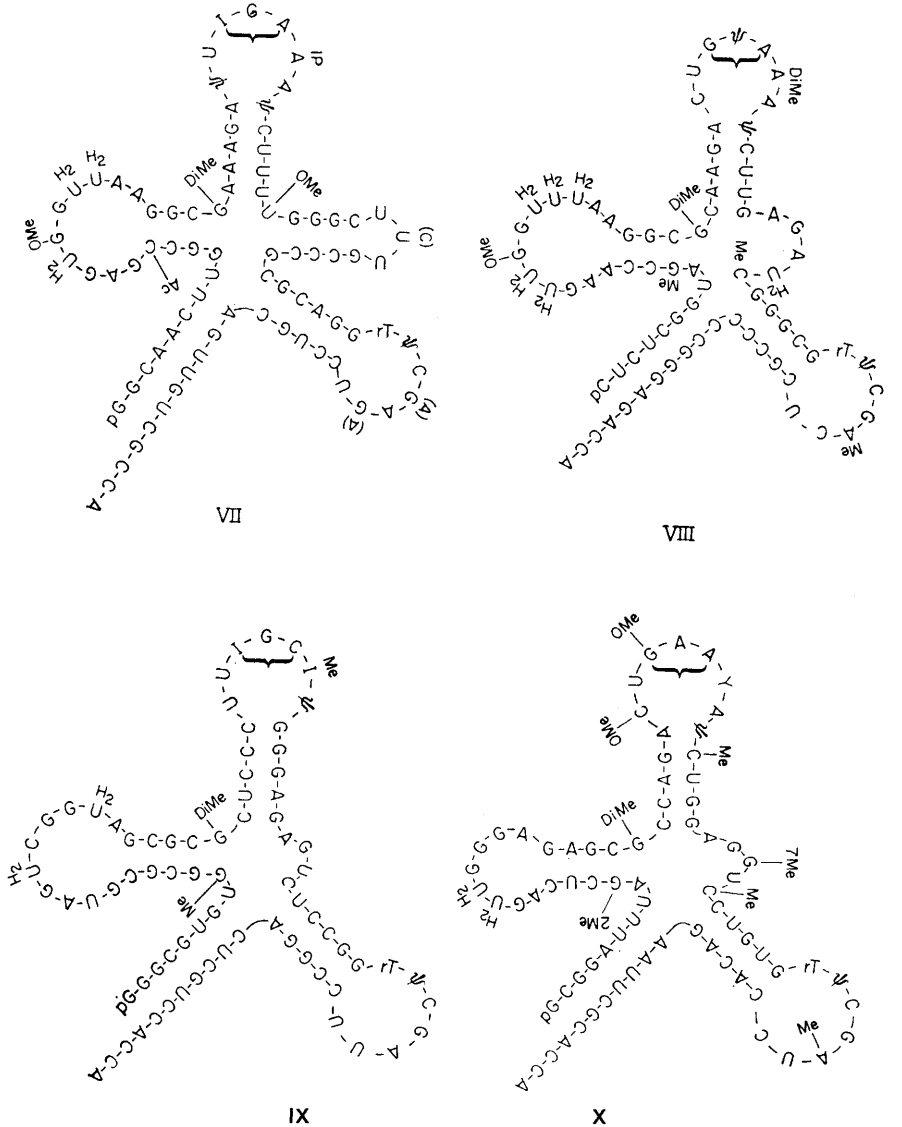

Figure 8 *Sequences and possible secondary structure of t-RNAs* (VII) Two serine t-RNAs [The three nucleotides in serine t-RNA (II) which differ from those in serine t-RNA (I) are indicated in parentheses] (H. G. Zachau, D. Dütting, and H. Feldmann, *Angew. Chem. Internat. Edn.*, 1966, **5**, 422; *Z. physiol. Chem.*, 1966, **347**, 212; H. Feldmann, D. Dütting, and H. G. Zachau, *ibid.*, p. 236; D. Dütting, H. Feldman, and H. G. Zachau, *ibid.*, p. 249). (VIII) Tyrosine t-RNA (J. T. Madison, G. A. Everett, and H. Kung, *Science*, 1966, **153**, 531). (IX) Alanine t-RNA (R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir, *Science*, 1964, **147**, 1462), (X) Phenylalanine t-RNA (U. L. Raj Bhandary, S. M. Chang, A. Stuart, R. D. Faulkner. R. H. Hoskinsau, and H. G. Khorana, *Proc. Nat. Acad. Sci., U.S.A.*, 1967, **57**, 571)

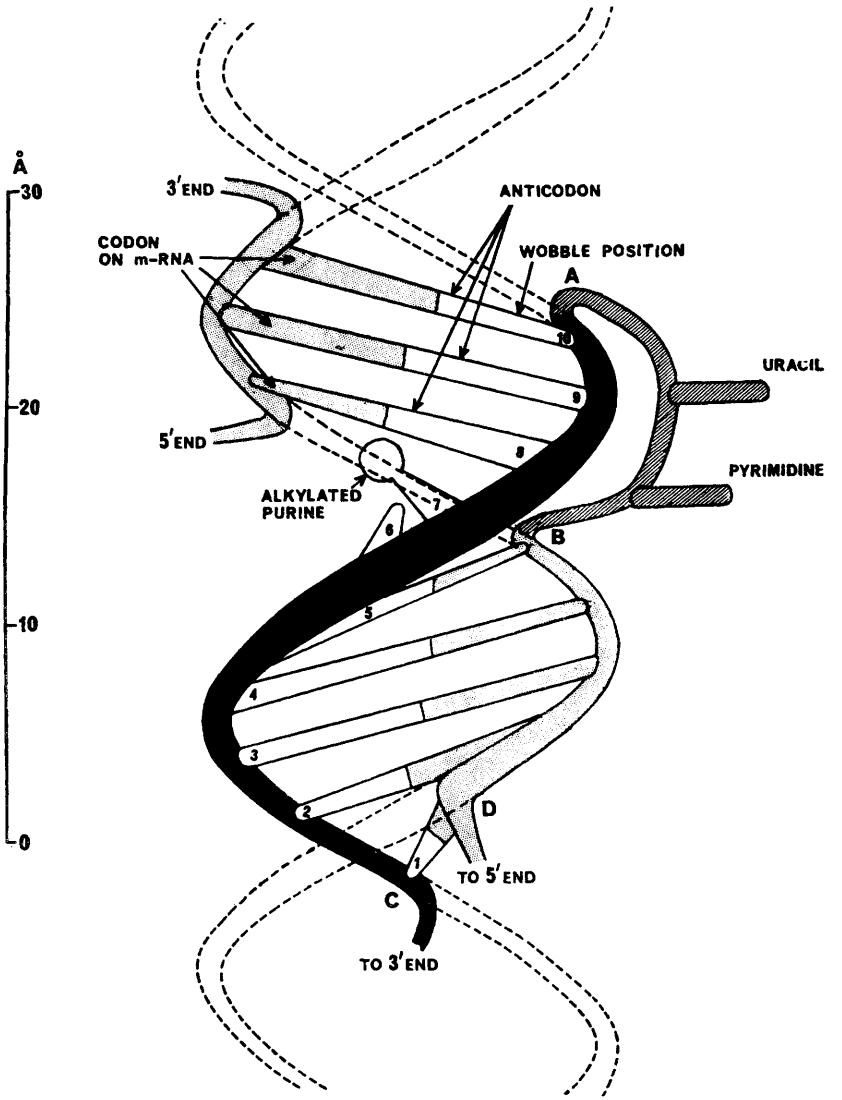

Figure 9 *Schematic diagram of the t-RNA anticodon loop illustrating its relationship to the codon and the helical character of the structure* The letters *A*, *B*, *C*, and *D* identify the same points on the structure as in Figure 3. The bases of nucleotides 1 to 10 are stacked on one another and follow the regular helix which is shown black. The chain of the anticodon double-helix between *D* and *B* is shaded like the codon to indicate that they follow the same helix. This helix is complementary to the black one. The two nucleotides not in the standard conformation are represented by dark line shading. The representation of their conformation is very schematic because they lie behind nucleotides 8, 9, and 10 in the black chain. The dotted lines indicate the generic helix from which the structure can be imagined to be derived (W. Fuller and A. Hodgson, *Nature*, 1967, **215**, 817)

molecular weights of RNA from viruses are compared in Tables 4 and 5). A few of the viruses so far examined have a double-stranded RNA component,[106] *e.g.*, reovirus and wound tumour virus. The molecular architecture of double-stranded RNA has been studied.[58]

The smallest known viruses are called satellites because they can replicate only when another virus, *e.g.*, tobacco mosaic virus or alfalfa mosaic virus, is present. The RNA component is about 10^5 daltons in size [107] and is sufficiently long to code for one or two small proteins.[108]

Next in size come the spherical bacterial viruses[109] such as f2, R17, and M.S.2, whose RNA component is about 1×10^6 daltons. There are a few plant viruses in this class such as bromegrass mosaic virus.[110]

There are two types of virus containing RNA of about 2×10^6 daltons: the rod-shaped plant viruses,[111,112] such as tobacco mosaic virus and spherical viruses such as picornaviruses[113] (*e.g.*, poliovirus) and some spherical plant viruses such as turnip yellow mosaic virus.

The picornaviruses are the smallest RNA-containing animal viruses and a great deal more is known about their biochemistry than about their structure.

The arboviruses (*e.g.*, Sindbis and Semliki Forest virus) are more complex than picornaviruses since they have a thick lipoprotein membrane which surrounds a central core that resembles a picornavirus. Little is known about the RNA moiety of these viruses. Similarly, our knowledge of the RNA of the myxoviruses[113] (*e.g.*, influenza virus or Newcastle disease virus) is limited because of the difficulties in obtaining quantities of the virus and in isolating undegraded RNA.[114] In the majority of animal, plant, and bacterial RNA viruses so far studied, the RNA is single-stranded. It is probable that the RNA moiety of these viruses is a single molecule. The molecular weight of the RNA moiety varies from about 1 to 3×10^6 daltons. The RNA moiety is a genetic element and the amount of genetic information is proportional to the length of the polynucleotide chain. The introduction of viral RNA into its host cells may lead to the synthesis of complete virus. It was first shown in 1956 that when RNA from

[106] P. J. Gomatos and I. Tamm, *Proc. Nat. Acad. Sci. U.S.A.*, 1963, **50**, 878.
[107] M. E. Reichmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 1009.
[108] J. M. Clark, A. Y. Chang, S. Speigelman, and M. E. Reichmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **54**, 1193.
[109] N. D. Zinder, *Ann. Rev. Microbiol.*, 1965, **19**, 455; H. Hoffmann-Berling, H. C. Kaerner, and R. Knippers, *Adv. Virus Res.* 1966 **12**, 329.
[110] L. E. Boekstahler and P. Kaesberg, *Nature*, 1961 **190**, 192; L. E. Boekstahler and P. Kaesberg, *Biophys. J.*, 1962, **2** 1.
[111] R. Markham, 'Progress in Nucleic Acid Research', ed. J. N. Davidson and W. E. Cohn, Academic Press, New York, 1963, vol. 2, p. 61.
[112] H. Schuster, 'The Nucleic Acids', ed. E. Chargaff and J. N. Davidson, Academic Press, New York, 1960, vol. 3, p. 245.
[113] C. H. Andrewes and H. G. Pereira, 'Viruses of Vertebrates', Ballière, Tindall and Cassell, London, 1964, 2nd ed.; F. M. Burnet and W. M. Stanley, 'The Viruses', Academic Press, New York, 1959, vol. 3; F. L. Schaffer and C. E. Schwerdt, 'Viral and Rickettsial Infections of Man', F. L. Horsfall, and I. Tamm, Pitman Medical Pub. Corp., London, 1965, 4th ed., p. 94.
[114] P. H. Duesberg and W. S. Robinson, *Proc. Nat. Acad. Sci. U.S.A.*, 1965, **54**, 794; M. W. Pons, *Virology*, 1967, **31**, 523.

Table 4 *Some RNA viruses of plants and bacteria*

Group	*Virus*	*Particle weight*	*M.W. of RNA*	*Isolated RNA shown to be infective*	*Nucleotide composition (moles/100 mmoles nucleotide)*			
					CMP	*AMP*	*GMP*	*UMP*
	Turnip yellow mosaic	$5{\cdot}0 \times 10^6$	$1{\cdot}7 \times 10^6$	YES	38·1	22·6	17·2	22·1
	Tomato bushy stunt	$10{\cdot}7 \times 10^6$	$1{\cdot}7 \times 10^6$		20·8	25·7	27·9	25·7
Spherical	Southern bean mosaic	$6{\cdot}6 \times 10^6$	$1{\cdot}4 \times 10^6$		23·0	25·8	26·0	25·3
(plant)	Tobacco necrosis	$8{\cdot}0 \times 10^6$	$1{\cdot}5 \times 10^6$		22·0	27·9	24·4	25·7
	Tobacco ingot	$3{\cdot}4 \times 10^6$	$1{\cdot}5 \times 10^6$		23·2	23·9	24·7	28·2
	bromegrass	$4{\cdot}6 \times 10^6$	$1{\cdot}0 \times 10^6$					
Satellite (plant)	Satellite tobacco necrosis	$2{\cdot}0 \times 10^6$	$3{\cdot}94 \times 10^5$		22·1	28·0	24·0	24·9
Rod-shaped	Tobacco mosaic	$40{\cdot}0 \times 10^6$	$2{\cdot}2 \times 10^6$	YES	18·5	29·8	25·3	26·3
(plant)	Potato X	$36—39 \times 10^6$	$2{\cdot}0 \times 10^6$		22·8	34·2	21·8	21·3
Double-helical plant RNA	wound tumour	$70{\cdot}0 \times 10^6$	$10{\cdot}0 \times 10^6$		19·1	31·1	18·6	31·3
Spherical	QB	—	$1{\cdot}1 \times 10^6$		24·7	22·3	23·7	29·4
(bacterial)	f_2	—	$1{\cdot}8 \times 10^6$	YES	25·9	22·1	26·8	25·1
	R17	$3{\cdot}6\text{-}4{\cdot}19 \times 10^6$	$1{\cdot}1 \times 10^6$		24·9	23·1	26·3	25·7
	M.S.2	$3{\cdot}6 \times 10^6$	$1{\cdot}1 \times 10^6$	YES	24·9	22·8	27·1	25·2
	fr	$4{\cdot}1 \times 10^6$	$1{\cdot}2 \times 10^6$		24·9	24·3	27·1	23·7

Table 5 *RNA viruses of vertebrates*

Group	*Example*	*Particle weight or diameter*	*M.W. of RNA*	*Whether isolated RNA has been shown to be infective*	*Nucleotide composition* (moles/100 moles nucleotides)*			
					CMP	*AMP*	*GMP*	*UMP*
	Poliomyelitis			YES	22·0	29·0	24·0	25·0
Picornaviruses	Foot-and-mouth disease			YES	28·0	26·0	24·0	22·0
	Encephalomyo-carditis	22—27 mμ	$\sim 2 \times 10^6$	YES	23·0	27·0	24·0	26·0
	Coxsakie			YES	23·0	29·0	24·0	24·0
	Rhinoviruses			YES				
Reovirus	Reovirus†	70×10^6, 60–90 mμ	$\sim 10 \times 10^6$		21·0	28·0‡	22·3	29·0
	Eastern equine encephalomyo-carditis	25—50 mμ	2×10^6	YES				
	Sindbis	40—48 mμ			24·9	29·6	25·8	25·5
Arboviruses	Semliki forest	50 mμ		YES				
	West Nile	20—30 mμ		YES				
Myxoviruses	Influenza	80—120 mμ	2×10^6	?	24·0	23·1	20·1	32·8
Paramyxo-viruses	Newcastle disease virus	100—200 mμ	$3—8 \times 10^6$		27·0	26·1	24·9	22·0

* Data from F. L. Schaeffer and C. E. Schwerdt, 'Viral and Rickettsial Infections in Man', ed. F. H. Horsfall and I. Tamm, Pitman Publishing Corporation, London, 4th edn., 1965, p. 94.

† Reovirus RNA may exist on more than one molecular species within the virus (A. R. Bellamy, L. Shapiro, J. T. August, and W. J. Joklik, *J. Mol. Biol.*, 1967, **29**, 1).

‡ The high ratio of AMP is due to the presence of low molecular weight RNA rich in adenine residues (A. R. Bellamy, L. Shapiro, J. T. August, and W. J. Joklik, *J. Mol. Biol.*, 1967, **29**, 1).

tobacco mosaic virus was rubbed into tobacco leaves, the cells became infected with tobacco mosaic virus.[115] More recently, RNA from a bacterial virus R17 was added to a cell-free system of *Escherichia coli* ribosomes and supernatant enzymes and the synthesis of viral coat protein was demonstrated.[116] The viral RNA not only codes for its coat protein but also for its own RNA polymerase.[117]

Comparatively little is known about the secondary structure of viral ribonucleic acids[118] although they adopt a hairpin-loop form in neutral salt solution.[65,72b] The structure in solution is not necessarily the same as in the virus. For example, the RNA component of tobacco mosaic virus is entirely single-stranded within the virus[119] but has a well defined partly double-helical form in solution.[65,72b,120] In the case of spherical viruses, such as turnip yellow mosaic virus, the secondary structure of the RNA moiety may be similar within the virus[121] and also after isolation.[122] The determination of the nucleotide sequence in the vicinity of the C(5′)-OH[123] and C(3′)-OH[124] termini has been attempted for RNA from tobacco mosaic virus and for RNA from the bacterial viruses f2, R17, and M.S.2. The replicative form of viral RNA has been isolated in many cases and was shown to be double-helical by its resistance to degradation by ribonuclease, its sharp melting properties, and its buoyant density.[125]

7 Conclusions

The ribonucleic acids have fascinating physical and chemical properties in addition to their important biological role. The correlation between structure and function remains a challenge, even in the case of t-RNA. There is a need for nucleotides of known sequence that also form hairpin loops so that physical chemistry of short double-helical segments can be established. The secondary

[115] A. Gierer and G. Schramm, *Nature*, 1956, **177**, 702; A. Gierer and G. Schramm, *Z. Naturforsch.*, 1956, **11**b, 138; H. Fraenkel-Conrat, B. Singer, and R. C. Williams, *Biochim. Biophys. Acta*, 1957, **25**, 87.

[116] D. Nathans, G. Notani, J. H. Schwartz, and N. Zinder, *Proc. Nat. Acad. Sci. U.S.A.*, 1962, **48**, 1424; Y. Ohtaka and S. Spiegelman, *Science*, 1963, **142**, 493; D. Nathans, *J. Mol. Biol.*, 1965, **13**, 521; M. R. Capecchi and D. Gussin, *Science*, 1965; **149**, 417; M. R. Capecchi, *J. Mol. Biol.*, 1966, **21**, 173.

[117] G. N. Gussin, *J. Mol. Biol.*, 1966, **21**, 435.

[118] S. Sprecher-Goldberger, *Archiv. für gesamte Virusforschung*, **1967, 20, 225.**

[119] W. Ginoza, *Nature*, 1958, **181**, 958.

[120] A. Gierer, *Z. Naturforsch.*, 1958, **13**b, 477.

[121] J. T. Finch and A. Klug, *J. Mol. Biol.*, 1966, **15**, 344; A. Klug, W. **Longley, and R.** Leberman, *ibid.*, p. 315; G. Zubay and M. H. F. Wilkins, *ibid.*, 1960, **2**, 105.

[122] R. E. F. Matthews and R. K. Ralph, *Adv. Virus Res.*, 1966, **12**, 273.

[123] R. Roblin, *J. Mol. Biol.*, 1968, **31**, 51.

[124] B. Singer, M. Sherwood, and H. Fraenkel-Conrat, *Biochim. Biophys. Acta*, 1965, **108**, 306; P. R. Whitfield, *ibid.*, p. 202; H. L. Weith and P. T. Gilham, *J. Amer. Chem. Soc.*, 1967, **89**, 21, 5473; R. de Wachter and W. Fiers, *J. Mol. Biol.*, 1967, **30**, 507.

[125] L. Montagnier and F. K. Sanders, *Nature*, 1963, **199**, 664; R. Langridge, M. A. Billeter, P. Borst, R. H. Burdon, and C. Weissmann, *Proc. Nat. Acad. Sci. U.S.A.*, 1964, **52**, 114; Y. Watanabe, *Biochim. Biophys. Acta*, 1965, **95**, 515; D. Baltimore, *J. Mol. Biol.*, 1966, **18**, 421; R. K. Ralph, R. E. F. Matthews, A. I. Matus, and H. G. Mandel, *ibid.*, 1965, **11**, 202; B. Francke and P. H. Hofschneider, *ibid.*, 1966, **16**, 544; G. Feix, R. Pollet, and C. Weissman, *Proc. Nat. Acad. Sci. U.S.A.*, 1968, **59**, 145; N. R. Pace, D. H. L. Bishop, and S. Spiegelman, *ibid.*, p. 139.

structure of ribosomal and viral RNA and other RNA species having a hairpin-loop conformation may then be better understood. The structure of ribonucleoproteins remains ill-defined except possibly for small viruses. Future developments may depend on the extension of sequencing technique to larger molecules, on improved methods for the synthesis of polynucleotides of known sequence, and on the production of crystals of nucleic acids and nucleoproteins for *X*-ray crystallographic studies.