HOMOPOLYMER-RICH REGIONS IN MAMMALIAN DNA

by

ALAN SHENKIN

A thesis presented for the degree of

Doctor of Philosophy,

The University of Glasgow

February, 1974.

Now some parts are the special instrument of an action; others have been so constituted that it seems as if the action could not take place without them; others have been laid down in order that the action may be better performed; and still others have been created for the protection and preservation of the whole.

(Fabricius, 1621)

ACKNOWLEDGEMENTS

I am indebted to a number of persons for their help with various aspects of this work. In particular, I would like to express my gratitude to:-

Professor R.M.S. Smellie and the late Professor J.N. Davidson for their advice and encouragement, and for providing the necessary facilities for this work in the Department of Biochemistry, University of Glasgow.

Dr. R.H. Burdon, for his astute and stimulating guidance throughout the performance and presentation of this study.

Dr. R. Eason for assistance with sedimentation velocity analyses.

Dr. Anna Hell, The Beatson Institute for Cancer Research, for advice regarding preparation and shearing of DNA.

Drs. D. Bell, A.M. Campbell, R. Eason, D. Jolly and N. Wilkie, for donation of specimens of viral DNA.

Dr. R. Elton, The Institute of Virology, University of Glasgow, for his advice regarding the statistical interpretation of results.

The staff of the Wellcome Cell Culture Unit for assistance with the preparation of cells.

Mrs. E. Hanton for her painstaking yet rapid typing of this thesis.

And to my wife, Leonna, for her unfailing support, encouragement and understanding throughout the period of this work.

ABBREVIATIONS

The standard abbreviations, as recommended by the Editors of the Biochemical Journal (Biochem. J. (1973) 131, 1-20) are used throughout this thesis with the following additions:-

dA-rich region dC-rich region dG-rich region dT-rich region	A region within a DNA molecule where one of the DNA strands is considerably enriched for a particular nucleotide, as specified.
dA·dT-rich region	A region within native DNA which is considerably enriched for dA·dT base-pairs
A-rich region C-rich region G-rich region U-rich region	A region within an RNA molecule which is considerably enriched for a particular nucleotide, as specified.

Deoxyhomopolymers are abbreviated so as to specify the sugar residue e.g. poly(dA), poly(dG), whereas the sugar residue of ribohomopolymers is not specified e.g. poly(A), poly(G).

BSS	balanced salt solution
HAP	hydroxylapatite
HnRNA	heterogeneous nuclear RNA
PPLO	pleuro-pneumonia like organisms
rDNA	the genes for ribosomal RNA
RNase	ribonuclease (E.C.2.7.7.16)

SDS	sodium dodecyl sulphate
SSC	0.15 M NaCl - 0.015 M trisodium citrate pH 7.0
0.1 x SSC, 2 x SSC etc.	multiples of the above composition of SSC

CONTENTS

List of Figures

List of Tables

SUMMARY

The presence of deoxyadenylate-rich (dA-rich) and deoxyguanylate-rich (dG-rich) regions in mammalian DNA has been demonstrated by hybridisation with [^{3}H] poly(U) and [^{3}H] poly(C).

For BHK-21/C13 cells, the maximum levels of these homopolymer-rich regions, as detected by the hybridisation technique, are about 0.41% of the DNA for dA-rich, and 0.1% of the DNA for dG-rich regions. Since the hybrids are largely sensitive to digestion with RNase, it is probable that the majority of hybrid molecules contain mismatched base-pairs. 0.13% of the DNA consists of dA-rich regions ranging in size from 25-130 nucleotides long and containing about 2-6% of bases other than adenine. On the other hand, dG-rich regions less than 40 nucleotides long comprise 0.07% of the DNA and contain 10-30% of bases other than guanine.

Exhaustive RNase digestion of the hybrids enables detection of pure deoxyhomopolymeric regions in the DNA. Pure sequences of poly(dA) of average size about 31 nucleotides long account for 0.008% of BHK-21/C13 DNA, whereas poly(dG) sequences about 17 nucleotides long comprise 0.0016% of the DNA.

The organisation of these sequences within the genome has been investigated. Both dA-rich and dG-rich regions are present within DNA sequences of widely varying base composition. Extensive shearing of the DNA is required to produce some enrichment for dA-rich sequences in the A + T rich fraction, although dG-rich sequences are slightly

enriched in the G + C rich fraction of even unsheared DNA. The buoyant density of hybrid molecules was found to be significantly greater than that of the unhybridised DNA only when highly sheared DNA was used. Furthermore, dA-rich and dG-rich regions were shown to be associated with rapidly, intermediately, and slowly renaturing DNA sequences. These findings all suggest that the dA-rich and dG-rich regions have a widespread distribution throughout DNA molecules. The pure poly(dA) and poly(dG) sequences also appear to be scattered throughout DNA molecules. In situ hybridisation studies with [^{3}H] poly(U) further suggest that the dA-rich regions are not localised to any particular chromosome or to any specific region of the chromosomes.

Analysis of DNA from a number of different species has shown that, in general, the dA-rich and dG-rich regions are present to a much higher level in mammalian DNA than in bacterial, bacteriophage, and mammalian viral DNAs.

Evidence for the existence of A-rich, U-rich and G-rich RNA species in BHK-21/C13 cells has also been obtained.

The characteristics of the homopolymer-rich regions in DNA suggest that the dA-rich and pure poly(dA) sequences are more likely to have a significant role than dG-rich and poly(dG) sequences. The possible functions of these unusual deoxynucleotide sequences are reviewed.

SECTION I INTRODUCTION

A. General

The structure of DNA has intrigued scientists for many years. Despite the considerable advances which have been made in the past 20 years, little is still known about the organisation of nucleotide sequences within each DNA molecule, especially in higher organisms. This is primarily due to the enormous complexity of DNA, there being about 6.5×10^9 base-pairs per diploid mammalian genome (Davidson, 1972). Besides those DNA sequences which give rise to particular proteins, there exist other types of sequence in mammalian DNA, especially regulatory sequences and possibly non-functional sequences (Comings, 1972). The present work represents a study of a particular class of DNA sequences in the genome of animal cells - those which consist very largely of just one type of nucleotide. It is hoped that studies of this type will lead to a greater understanding of the structural organisation of the mammalian genome.

B. The structure of DNA

The discovery of the double-stranded nature of DNA by Watson & Crick (1953) has been of fundamental importance in the development of our knowledge of the mechanism of DNA replication and transcription. They proposed that the DNA molecule is double-stranded and in the form of a right-handed helix. The two polynucleotide chains, which are of opposite polarity, are wound around the same axis and held together by hydrogen bonds between the bases, adenine in one strand always pairing with thymine in the other, and

guanine always pairing with cytosine. These specific base-pairing mechanisms readily suggest methods by which the sequence of nucleotides in the DNA can be faithfully replicated prior to cell division, and by which appropriate regions can be transcribed into RNA (uridine in RNA base-pairs with adenine in DNA). In their original model, Watson & Crick further suggested that the base-pair is horizontal i.e. at 90° to the axis of the helix. It is, however, now realised that the secondary structure of DNA, as measured by X-ray fibre diffraction, depends to a large extent upon the salt concentration and humidity at which the studies are performed (Table I.1).

The B structure of DNA in fibres of 92% humidity is believed to correspond most closely to the structure in solutions of low ionic strength (Tunis-Schneider & Maestre, 1970). This structure is very similar to the original Watson-Crick model and is shown in Fig. I.1.

Not only does the secondary structure of the DNA depend upon the salt concentration and the humidity, but it also depends upon the base composition of the DNA. Thus X-ray scattering studies on DNA in solution demonstrate that although DNA of average base composition is normally in the B form, very A + T rich DNA does not seem to adopt this form (Bram, 1971). Moreover, analysis of DNA structure by fibre diffraction (Bram & Tougard, 1972) and infra-red spectroscopy (Pilet & Brahms, 1972) has demonstrated that whereas DNA of average base composition readily adopts the A form at low salt concentration or low humidity, very A + T rich DNA only very rarely takes up the A configuration.

Table I.1. Dimensions of the different forms of DNA

DNA	Pitch	Residues per turn	Angle between perpendicular to helix axis and bases
A form, Na salt, 75% humidity	28.15	11	20°
B form, Na salt, 92% humidity	34.6	10	
B form, Li salt, 66% humidity	33.7	10	2°
C form, Li salt, 66% humidity	31.0	9.3	6°
DNA-RNA hybrid, Na salt, 75% humidity	28.8	11	~ 20°
Poly[d(A-T)], Li salt, 66% humidity	33.4	10	

From Davies (1967)

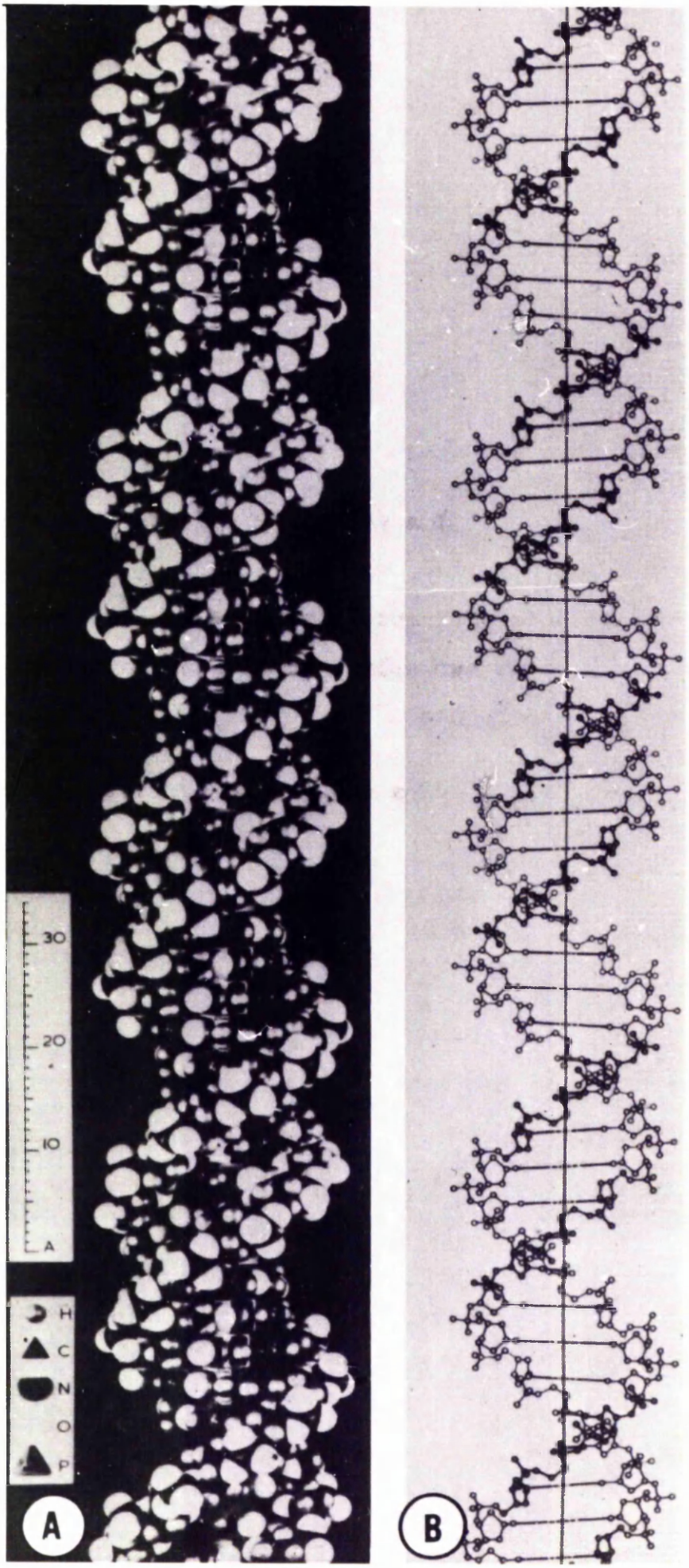
Fig. I.1.

The secondary structure of DNA

(A) Molecular model of a DNA double helix, showing the two spiral grooves on the surface of the molecule.

(B) Diagram of a DNA double helix in the B form.

(From Dupraw, 1970)


30
20
10
A
H
C
N
O
P
A
B

The ability of a region of DNA to take up the A configuration may be important since RNA does not adopt the B configuration (Davies, 1967) and DNA-RNA hybrids exist in the A form (Milman et al., 1967). Thus, when a DNA sequence is transcribed into RNA, the DNA molecule must presumably adopt the A form (Milman et al., 1967; Bram, 1972). If it cannot take up the A configuration then it may never be transcribed (Bram, 1972).

C. Chromosome structure

Besides the highly organised double-helical secondary structure of DNA, it is also evident that the overall packaging of the DNA into the nuclei of mammalian cells involves considerable folding of the DNA molecule. In this way, the total length of DNA per diploid human nucleus, about 174 cm (Dupraw, 1970) can be organised inside a nucleus only 5-10 μ in diameter. The basic unit of this further level of organisation is the chromosome, there being 23 pairs of chromosomes in the diploid human cell. Chromosomes show considerable heterogeneity of size and shape (Dupraw, 1970). Viewed by light or electron microscopy they are seen to be composed of fibres which are folded in very complex ways (Fig. I.2).

1. Components of chromosome fibres

a) DNA

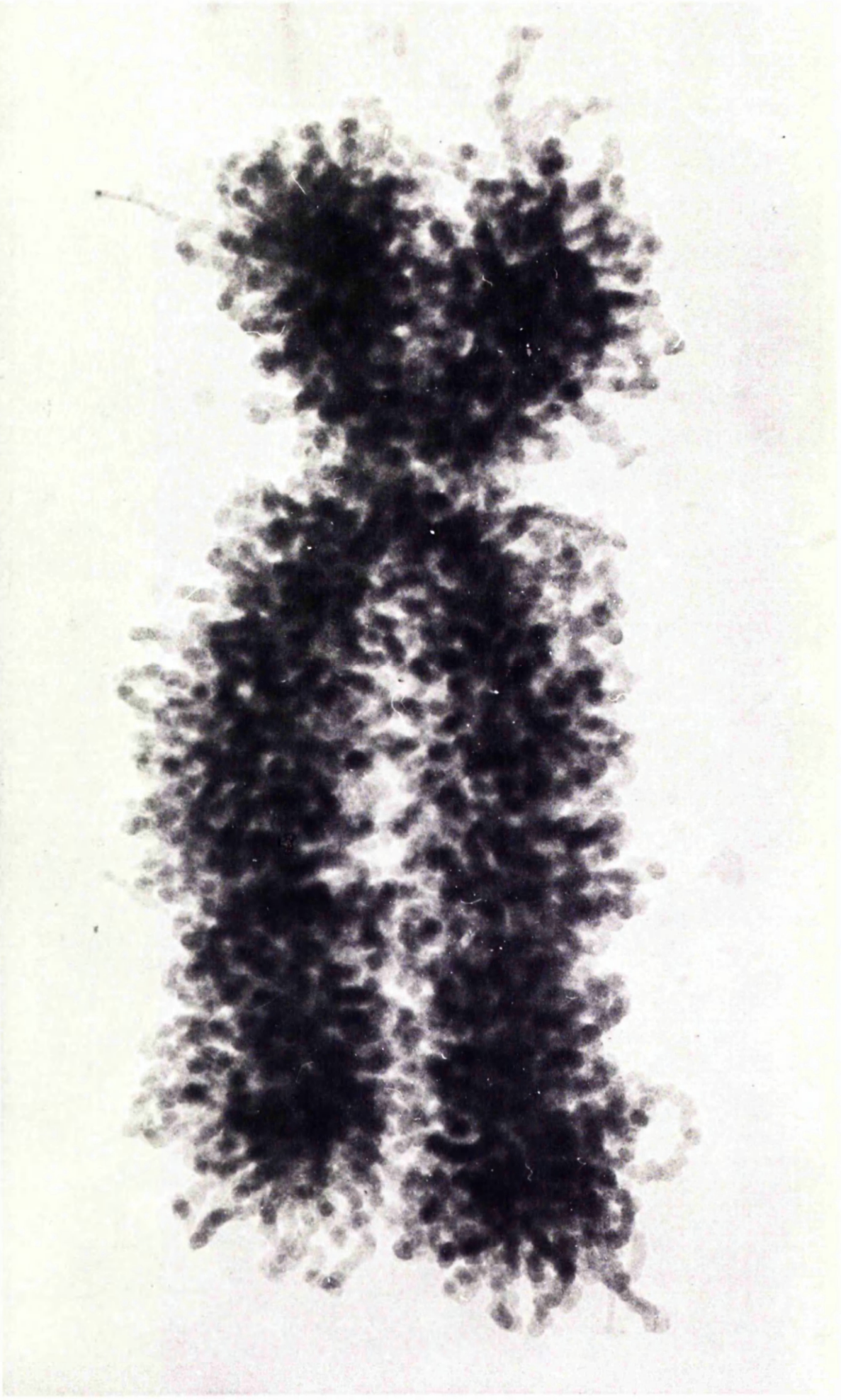
DNA forms the essential backbone of the chromosome fibre (Dupraw, 1970; Huberman, 1973). Thus, hydrolysis with proteolytic agents does not affect the size of the DNA recovered, so there are probably no non-nucleotide "linkers" in chromosomal DNA (Huberman & Riggs, 1966). Each fibre may

Fig. I.2.

Electron micrograph of a human chromosome

Human chromosome 12, about 100,000 x magnified.

(From Dupraw, 1970)



contain just a single DNA molecule which is folded in a particular way. Evidence for this comes from attempts to isolate DNA of very high molecular weight. For example, DNA molecules have been isolated from Drosophila melanogaster which are at least half as long as would be expected if each chromosome contained only a single DNA molecule (Kavenoff & Zimm, 1973).

b) Histones

Histones are basic proteins bound to the DNA of the chromosomes in most eukaryotic organisms. Histones have been classified into five main groups depending largely upon the amino-acid composition. A number of different nomenclatures have been proposed for these histone fractions (Huberman, 1973). The complete or partial amino-acid sequence of a representative of each class is known (Iwai et al., 1970; De Lange et al., 1969, 1972; Ogawa et al., 1969; Rall & Cole, 1971; Sugano et al., 1972). One important feature of these sequences is the non-uniform distribution of basic amino-acids, nearly all of them being crowded into one half of the molecule. Such a region therefore has a high concentration of positive charges and may be the region which interacts with negatively charged phosphate groups on DNA.

c) Non-histone proteins

Those proteins which are associated with chromosomal material but are not histones are called non-histone proteins. They are a much more heterogeneous group than the histones. Analysis by polyacrylamide gel electrophoresis reveals some 10-30 major species together with very many minor species (Elgin et al., 1971; Teng et al., 1971). Unlike the

histones, the amount of non-histone protein can vary considerably, both from one cell type to another (Mirsky & Ris, 1951) and at different stages in the cell cycle (Huberman & Attardi, 1966). The bulk of the non-histone protein fraction probably has enzymic activity e.g. DNA polymerase (Patel et al., 1967), RNA polymerase (Weiss, 1960; Huang et al., 1960), protein methylase (Sekeris et al., 1967; Burdon & Garven, 1971), protein phosphokinase (Kleinsmith & Allfrey, 1969) and many more (Elgin et al., 1971). As discussed later, the minor species of non-histone protein probably have a regulatory role (Introduction, Section I.F.2.b.ii).

2. The dimensions of chromosome fibres

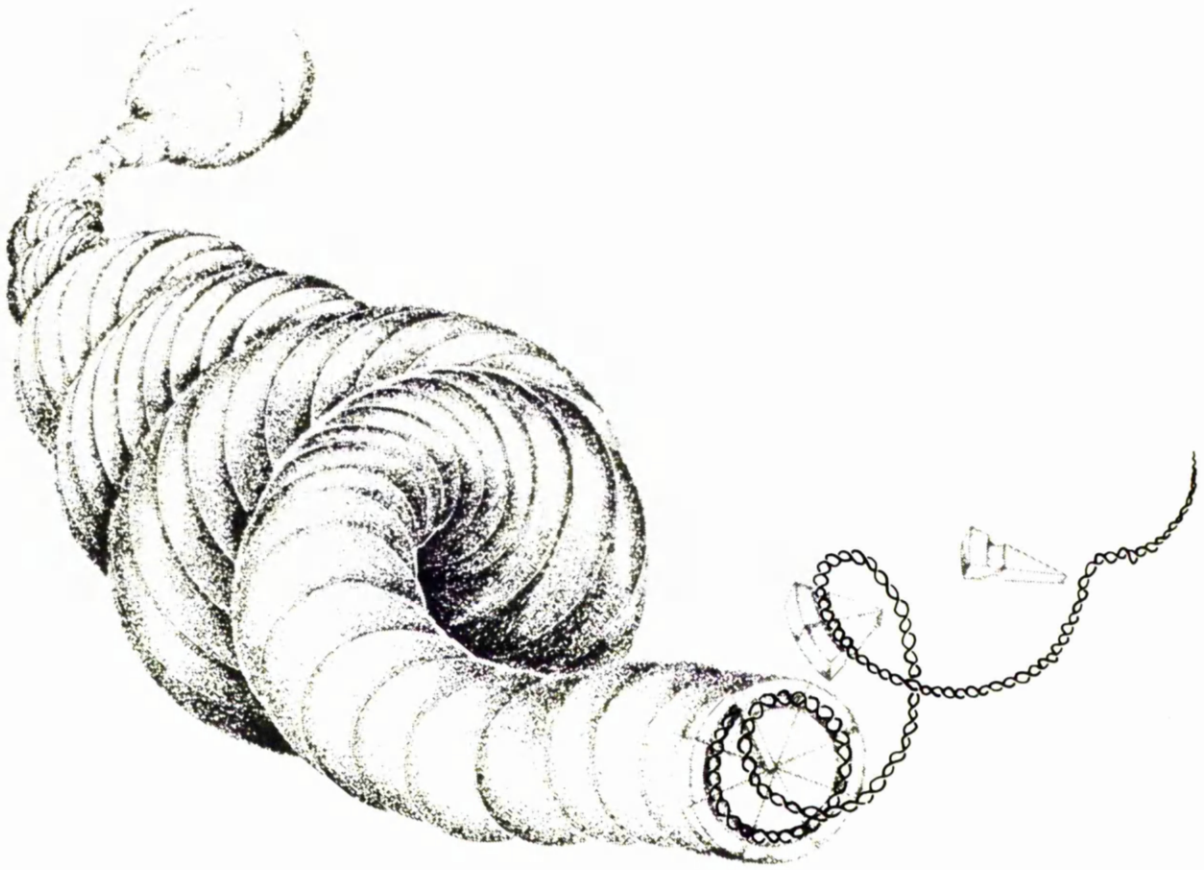
Since the structure of chromosome fibres varies depending upon the method of preparation, there is considerable disagreement about the specific dimensions of these fibres (Huberman, 1973). Measurements of diameter range from 30 Å (Zubay & Doty, 1959; Solari, 1971) to as much as 500 Å (Dupraw, 1970). Which value (or values) corresponds to the in vivo structure of the chromosomes is still uncertain. One study which should be mentioned is that of Davies & Small (1968) who suggested, on the basis of electron micrographs of thin tissue sections, that certain regions of the chromosomes consist of fibres about 130 Å in diameter, each fibre being composed of a coil of thinner material 20-40 Å thick. These conclusions are supported by X-ray diffraction data, from which it was interpreted that the chromosome consists of a supercoil of the DNA duplex with a pitch of 120 Å and a diameter of about 100 Å (Richards & Pardon, 1970; Pardon & Wilkins, 1972).

Fig. I.3.

Supercoiling within chromosomal fibres

A model for the mechanism by which DNA is tightly packed within chromosomal fibres, first by supercoiling to form a thin fibre, and then by supercoiling again to form a thicker fibre. The DNA molecule is held in its supercoiled configuration by histones and perhaps non-histone proteins, which are shown hypothetically as wedge-shaped molecules with two binding sites each.

(After Dupraw, 1970)



This would condense the DNA by a factor of 2.8:1 along the fibre.

This model may not be sufficiently ordered to produce the degree of condensation necessary to package the DNA, since Dupraw & Bahr (1969) have calculated that the DNA packing ratio (DNA length : fibre length) in interphase chromosomes is 56:1. They propose that this degree of packing can be most easily obtained if the chromosome consists of a supercoiled supercoil of DNA (Fig. I.3). To account for the differences in fibre diameter observed using the same method of preparation (Dupraw, 1970), the extent of supercoiling may be different at different regions of the chromosome.

D. Heterogeneity of DNA sequences

There is about 700 times more DNA in the haploid genome of mammals than there is in the genome of Escherichia coli. The DNA of E.coli consists of a fairly homogeneous collection of about 4000 genes, most of which are present as single copies (Britten & Kohne, 1968; Laird, 1971). However, the much larger genome of higher organisms is not a comparable homogeneous mixture of many more genes, but rather is composed of a very heterogeneous collection of different families of DNA sequences. This heterogeneity of DNA sequences has been observed in three main ways.

1. Differences in mean base composition

In 1961, Kit observed that when DNA from a number of different animals was centrifuged to equilibrium in CsCl, satellite peaks of varying buoyant density and size were obtained. These satellite peaks were therefore composed of

DNA with a mean base composition different from main band DNA (Schildkraut et al., 1962). One of the most extensively studied satellites is mouse satellite DNA which comprises about 10% of the genome (Kit, 1961) and has a base composition of about 34% G + C, in comparison to 42% G + C for main band DNA (Flamm et al., 1967). When this satellite was isolated by preparative ultracentrifugation, it was observed to renature very rapidly (Walker & McLaren, 1965). On the basis of the kinetics of reassociation, it has been estimated that mouse satellite DNA is composed of a basic unit of 300-400 nucleotides, repeated about 10^6 times (Waring & Britten, 1966; Flamm et al., 1969). However, sequencing studies suggest that the repeating unit is only 8-13 bases long (Southern, 1970). This overestimation of sequence length by renaturation data is probably the result of mismatching of bases, since the repeating units, although very similar, are probably not identical (Southern, 1970, 1971; Sutton & McCallum, 1971).

For a family of sequences to occur as a distinct satellite, not only must they have a base composition distinct from that of the main band DNA, but the sequences must also be clustered together. If they were interspersed within larger blocks of average base composition, they would not separate out as a unique band.

2. Differences in the degree of repetition of sequences

Heterogeneity in the eukaryotic genome can also be classified on the basis of the degree of repetition of the sequences. Britten & Kohne (1968) were first to demonstrate that DNA of higher organisms is composed of three main classes - highly repetitious, moderately repetitious and non-repetitious sequences. This conclusion was based upon

studies of the rate of reassociation of DNA which had been sheared to short pieces of about 500 nucleotides long and then denatured. The rate of renaturation is expressed in terms of "C_0t" values, in which "C_0" is the initial concentration of DNA in moles of nucleotides per litre and "t" is the time in seconds. C_0t values are expressed on a logarithmic scale ranging from 10^{-4} to about 10^4. DNA renaturing at low C_0t values (10^{-4} to 10^{-1}) is defined as being composed of highly repetitious sequences (usually about 15 - 40% of the total), DNA renaturing at high C_0t values (greater than 10^3) is minimally or non-repetitious (usually more than 50% of the total), whereas DNA renaturing at intermediate C_0t values (10^0 to 10^2) is moderately repetitious (Fig. I.4) (Britten & Kohne, 1968; Laird, 1971). The proportion of DNA in each of these classes varies considerably in different species.

Types of rapidly renaturing DNA

a) Satellite DNA sequences

As discussed above, these are clustered, highly repetitive sequences which, by virtue of their distinct base composition, can be separated from the rest of the DNA by centrifugation.

b) Hidden satellite sequences

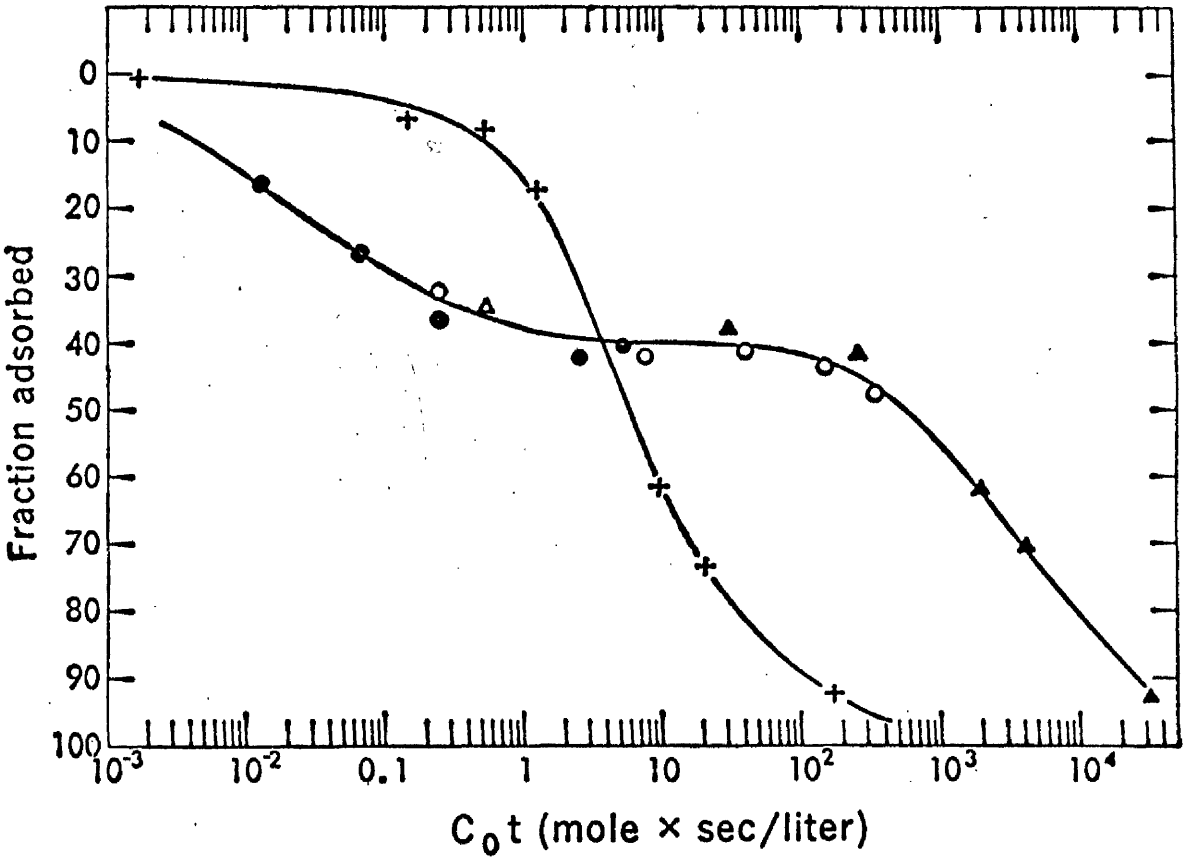
Some families of clustered repetitious sequences may be hidden within the main band of DNA after centrifugation in CsCl because they have the same mean base composition as main band DNA. Such "hidden satellites" may be detected by centrifugation in caesium sulphate containing silver or mercury salts (Corneo et al., 1970, 1972).

Fig. I.4.

The kinetics of reassociation of DNA

Denatured DNA samples of average size 500 nucleotides long were incubated at 60°C for various periods in 0.12 M phosphate buffer and then were passed over a hydroxylapatite column. Crosses refer to E.coli DNA. Other points were obtained with calf-thymus DNA.

(From Britten & Kohne, 1968)


Fraction adsorbed
0
10
20
30
40
50
60
70
80
90
100
10^{-3}
10^{-2}
0.1
1
10
10^{2}
10^{3}
10^{4}
$C_0 t$ (mole × sec/liter)

c) Other repetitious sequences

About 30-40% of the DNA of higher organisms consists of repeated sequences. At least some of these sequences are closely clustered (see above). The exact arrangement of the remaining repetitious sequences is a matter of some controversy. Davidson and coworkers (1973) have analysed DNA of Xenopus laevis by measurement of reassociation rate of sheared DNA fragments of different sizes, using hydroxyl-apatite to distinguish between denatured and partially reassociated sequences. On the other hand, Thomas and coworkers have studied DNA fragments (mainly from Drosophila polytene chromosomes, but also from mouse liver and Necturus erythrocytes) for their ability to form ring structures after treatment with exonuclease III, an enzyme which degrades DNA sequences from both 3' ends (Richardson, 1966; Lee & Thomas, 1973; Pyeritz & Thomas, 1973; Bick et al., 1973; Thomas et al., 1973). Both groups of workers agree that repetitious sequences are widely distributed in DNA molecules. The essential difference in their models can be summarised as follows:- Davidson et al. (1973) consider that 50% of DNA consists of closely interspersed repetitious (300 ± 100 nucleotides long) and non-repetitious (800 ± 200 nucleotides long) sequences, so that the repetitious sequences comprise 30-40% of each unit length of repetitious plus non-repetitious sequence i.e. the repetitious sequences are not clustered (less than 8% of Xenopus DNA was thought to be closely clustered). Thomas and coworkers, however, suggest that 50% of the DNA consists of regions in which about 80% of the total length consists of tandemly repeating sequences i.e. all the repetitious sequences of one type are densely

clustered into one region. At the present time, it is difficult to reconcile the differences in these models, and further work is clearly required to resolve this problem. As is discussed later, which of these models is correct may be of importance in understanding the mechanism of control of transcription (Introduction, Sections I.F.2.b.v, and I.G).

d) Foldback DNA

On some occasions, almost instantaneous renaturation occurs. This suggests that some sequences may not have to find each other in solution, because they are already attached perhaps by a single strand folding back on itself and renaturing with a similar but reversed repeat. This has been termed "foldback DNA" (Britten & Smith, 1970).

3. Differences in timing of replication of DNA sequences

When mammalian cells are synchronised in tissue culture and pulse-labelled at different times during the S period, DNA which replicates early is relatively G + C rich (about 43.6% G + C), whereas DNA which replicates late is relatively A + T rich (about 38.7% G + C) (Tobia et al., 1970; Bostock & Prescott, 1971; Flamm et al., 1971). Furthermore, when synchronised Chinese hamster cells are pulse-labelled at short intervals through S phase, the pulse-labelled DNA frequently shows one or more subcomponents whilst control DNA from unsynchronised cells does not (Comings, 1972). This suggests that within main band DNA there are several families, each family having a similar base composition and replicating at similar times.

E. RNA of eukaryotic cells

Eukaryotic cells contain a wide variety of RNA species of sizes ranging from about 2.5×10^4 to at least 10^7 daltons. These RNA molecules arise by transcription of particular regions of DNA by the enzyme RNA polymerase (E.C.2.7.7.6), of which there are at least three main types in mammalian cells (Roeder & Rutter, 1970; Horgen & Griffin, 1971). Extensive work has recently been directed towards establishing the nature of the genes for these various RNA species, and how the immediate transcription products are modified to produce mature RNA (Burdon, 1971b).

1. Ribosomal RNA

Ribosomes, which are small ribonucleoprotein particles found in the cytoplasm of eukaryotic cells, are an integral component of the protein synthesising machinery of the cell. They consist of a large and a small subunit, the chemical composition of each being approximately 50% RNA and 50% protein. The RNA component of the large subunit has a sedimentation value of about 28S (about 1.4×10^6 daltons) and that of the small subunit is 18S (about 0.7×10^6 daltons) (Darnell, 1968). Associated with the large subunit is a 5S RNA species 120 nucleotides long (Brownlee et al., 1968). Moreover, a 5.8S component (previously known as 7S RNA) may be associated with 28S RNA by hydrogen bonding (Pene et al., 1968; Rubin, 1973).

The base composition of 28S RNA is 63-67% G + C, and of 18S RNA is 53-58% G + C (Birnstiel et al., 1968; Maden, 1971), whereas the average G + C content of most mammalian cells is about 40-42% G + C. Since the genes for 18S and 28S ribosomal RNA are clustered at the nucleolar organiser

region of the chromosomes in Xenopus laevis, this abnormal base composition has allowed the isolation and analysis of these genes (Birnstiel et al., 1968; Brown et al., 1972). There are about 400-500 repeating units of ribosomal DNA per haploid set of chromosomes (Birnstiel et al., 1968; Brown & Weber, 1968), the 28S and 18S genes being separated from one another by a spacer region of even higher G + C content (69-73% G + C), which is transcribed at the same time as the 28S and 18S ribosomal DNA (Dawid et al., 1970). Each transcription unit (28S rDNA + spacer + 18S rDNA) is separated from the next by a further spacer region which is not transcribed (Reeder & Brown, 1970; Brown et al., 1972). The immediate transcription product is a large RNA species, 45S in HeLa cells (Weinberg et al., 1967) which is then modified by a series of cleavages of the polynucleotide chain, and methylation of particular nucleotides, to produce the mature 28S, 18S and 5.8S RNA species which are transferred to the cytoplasm (Weinberg & Penman, 1970; Maden, 1971).

5S RNA is also G + C rich (about 58% G + C), but the genes for 5S RNA are not associated with the nucleolar organiser (Brown & Weber, 1968; Brown et al., 1971). There are about 24,000 genes for 5S RNA per haploid genome of Xenopus, these genes being clustered together, but separated from one another by non-transcribed spacer sequences which are very A + T rich (about 30% G + C) (Brown et al., 1971).

2. Transfer RNA

The RNA species which provides the adaptor mechanism by which the four letter alphabet of the nucleic acids is translated into the twenty letter alphabet of proteins, is transfer RNA (tRNA) (Crick, 1958). tRNA has a complex

secondary structure and contains a high proportion of modified bases. Many tRNA molecules have now been sequenced (Holmquist et al., 1973).

Transfer RNA is synthesised by transcription of tRNA genes of which there are about 8000 per haploid mammalian genome (Birnstiel et al., 1972; Clarkson et al., 1973; Shenkin, unpublished results). There are about 40 species of tRNA and hence an average of 200 genes for each individual tRNA species (Birnstiel et al., 1972). These genes together with spacer DNA, also are clustered within the DNA (Clarkson et al., 1973). The transcription product of the tRNA genes is, as for rRNA, a precursor of the final mature tRNA species (Burdon et al., 1967; Bernhardt & Darnell, 1969). The conversion of the precursor species to mature tRNA takes place in the cytoplasm and involves shortening of the polynucleotide chains and modification of bases (Burdon & Clason, 1969; Mowshowitz, 1970; Smillie & Burdon, 1970).

3. Messenger RNA and heterogeneous nuclear RNA

Certain regions of DNA specify, by their base sequence, the amino acid sequence of particular proteins. These regions are known as structural genes. The mechanism by which the nucleotide sequence of the DNA is decoded in this way, involves the synthesis of an intermediary RNA species, known as messenger RNA (mRNA) (Crick, 1958).

The demonstration of mRNA in eukaryotes has proved much more difficult than in prokaryotes. Analysis of RNA synthesised in eukaryotic cells during very short incubations with radioactive precursors (less than 30 min) shows that most of the radioactivity is found in the nucleus in a collection of RNA molecules of variable size, ranging from 20-100 S by

sedimentation value or about 1-10 x 10^6 daltons in molecular weight (Houssais & Attardi, 1966; Warner et al., 1966; Darnell, 1968; Holmes & Bonner, 1973). This RNA species is known as heterogeneous (or heterodisperse) nuclear RNA (HnRNA). However, a small proportion of the radioactivity is found in the cytoplasm, associated with ribosomes (Darnell, 1968). It is now clear that this ribosome associated RNA is, indeed, the mRNA of eukaryotes (Darnell, 1968; Williamson, 1972; Darnell et al., 1973).

One current problem is whether HnRNA is a precursor of mRNA. About 90% of HnRNA molecules turn over in the nucleus and thus cannot be precursors of mRNA (Scherrer & Marcaud, 1965; Soeiro et al., 1966). Moreover, mRNA molecules, being of average sedimentation value about 16-18S are very much smaller than HnRNA molecules (20-100S) (Darnell, 1968; Holmes & Bonner, 1973). On the other hand, a number of lines of evidence suggest that a small fraction of HnRNA molecules are precursors of mRNA:-

(i) The base composition of mRNA is approximately the same as that of HnRNA (about 45% G + C) (Darnell, 1968).

(ii) In cells transformed by small DNA viruses, virus specific RNA is found in the nucleus as part of high molecular weight HnRNA molecules which also contain cellular sequences (Wall & Darnell, 1971), but in the cytoplasm, the virus specific mRNA molecules are smaller and are uniform in size (Lindberg & Darnell, 1970; Darnell et al., 1973).

(iii) Hybridisation of HnRNA to DNA can be partially inhibited by cytoplasmic mRNA, whereas hybridisation of mRNA can be completely prevented by HnRNA (Arion & Georgiev, 1967; Georgiev et al., 1972). Other studies, in which HnRNA

from reticulocytes was hybridised against a complementary DNA copy of haemoglobin mRNA, and in which HnRNA isolated from erythroblasts or myeloma cells was injected into Xenopus oocytes, also suggest that these HnRNA species contain the mRNA sequences for haemoglobin and immunoglobulin (Melli & Pemberton, 1972; Stevens & Williamson, 1972, 1973; Williamson et al., 1973). These conclusions, however, depend upon the HnRNA preparations being absolutely free from contamination with mRNA, a criterion which has not yet been established (Lane et al., 1973).

(iv) The 5' terminal triphosphate group which is present on large HnRNA molecules is lost during the production of small HnRNA species (Georgiev et al., 1972).

(v) A sequence of poly(A), about 150-200 nucleotides long, is present at the 3' end of certain HnRNA molecules (Edmonds & Caramela, 1969; Edmonds et al., 1971; Darnell et al., 1971a, b, 1973; Mendecki et al., 1972; Greenberg & Perry, 1972; Jelinek et al., 1973; Nakazato et al., 1973). The larger the HnRNA molecule, the higher the proportion of molecules containing a poly(A) segment (Jelinek et al., 1973). However, only about 20-40% of even very large HnRNA molecules contain poly(A) (Greenberg & Perry, 1972; Jelinek et al., 1973). By contrast, almost all mRNA species (with the exception of histone mRNA) have a poly(A) sequence, also 150-200 nucleotides long, at the 3' end of the molecule (Lim & Canellakis, 1970; Lee et al., 1971; Edmonds et al., 1971; Darnell et al., 1971a, b, 1973; Adesnick et al., 1972; Molloy et al., 1972a; Greenberg & Perry, 1972; Molloy & Darnell, 1973; Nakazato et al., 1973; Jelinek et al., 1973). It has therefore been suggested that a portion of certain

HnRNA molecules, containing the 3' end (and therefore the poly(A) sequence) may be transferred to the cytoplasm as mRNA (Darnell et al., 1971a, 1973).

Thus the overall evidence at present strongly suggests that a proportion of HnRNA molecules do serve as precursors for mRNA. The role of those HnRNA molecules which are not precursors for mRNA, and of the very large portion of each precursor molecule which is lost during maturation to mRNA is still uncertain. It is possible that such regions may play some role in the control of transcription (see Section I.F.2.b.v).

4. Other RNA species

In 1965, Huang & Bonner identified a species of RNA which was associated with chromosomes. This RNA was thought to be covalently associated with histones, was on average 30-40 nucleotides long, and had a very high content of modified bases (Huang & Bonner, 1965; Holmes et al., 1972). Many workers, however, have now suggested that this chromosomal RNA may not be a separate class of cellular RNA, but rather it is derived by degradation of tRNA, rRNA or HnRNA (evidence reviewed by Weinberg, 1973). Conclusive evidence for the exact nature of chromosomal RNA is still therefore required.

Nuclei of eukaryotic cells also contain a small amount of low molecular weight stable RNA species of largely unknown function, and mitochondria contain a small subsystem of tRNA, rRNA and mRNA capable of synthesising a limited number of proteins. Reviews of this nuclear RNA (Weinberg, 1973) and mitochondrial RNA (Borst & Grivel, 1971; Hirsch & Penman, 1973) have been recently published and these species will not be

further considered here.

F. Control of protein synthesis

Differentiated cells have a characteristic phenotype, which reflects a particular pattern of gene activity. The differential genetic expression which produces these various protein patterns at different stages of development and in different tissues may involve the regulation of the synthesis of the particular proteins at any of a number of stages.

(1) Differences in the number of gene copies present
(2) Differential RNA transcription
(3) Non-random stabilisation and maturation of potential mRNA species within the pool of HnRNA molecules e.g. poly(A) addition (see Section I.E.3)
(4) Regulation of the transport of "mature" potential mRNA from the nucleus to the cytoplasm
(5) Differential stability of mRNA within the cytoplasm
(6) Modulation of the translational efficiency of mRNA-ribosome complexes involved in protein synthesis
(7) Differential stability of the protein products.

Of particular relevance to the current work are headings (1) - (3) above. The other topics will not be reviewed here (see Wainwright, 1972).

1. Differences in the number of gene copies

Alteration in the amount of a particular type of RNA could clearly be obtained if the number of copies of the genes involved was able to vary. It has, however, been believed for many years that as a general rule, all the somatic cells of an individual organism contain identical genomes (e.g. Sturtevant, 1951; Davidson, 1972). Some

evidence for this comes from the constancy of DNA content and base composition (Davidson, 1972), and from the similarity in hybridisation properties of repetitious DNA sequences from different mouse tissues (McCarthy & Hoyer, 1964; Tobler, 1972). However, the most convincing evidence for constancy of DNA is the demonstration that nuclei from both developing and mature tissues can support normal development of enucleate eggs (Gurdon & Uehlinger, 1966; Laskey & Gurdon, 1970). These nuclei must therefore contain all the essential genes present in the original fertilised egg from which the donor organism developed. Furthermore, if alteration in the number of gene copies were responsible for changes in the amount of a protein synthesised, then it would be expected that tissues producing large amounts of a particular protein would have many copies of the appropriate gene. The evidence currently available suggests that this is not the case - for example, reticulocytes probably contain only one gene for haemoglobin per haploid genome (Bishop et al., 1972; Harrison et al., 1972; Bishop & Rosbach, 1973), and silkworms probably have only 1-3 copies of the gene for silk fibroin per haploid genome (Suzuki et al., 1972).

It should, however, be pointed out that certain rare exceptions to this general rule are known. In particular, a marked reversible alteration in the genome in certain differentiating systems occurs, resulting in selective gene amplification. One example of this is the amplification of genes for rRNA in Xenopus laevis. Somatic cells contain only one nucleolar organiser, comprising about 400-500 genes for 28S and 18S RNA per haploid genome (Birnstiel et al.,

1968; Brown & Weber, 1968), whereas oocytes contain about 600-1600 copies of the nucleolar organiser (Brown & Dawid, 1968; Perkowska et al., 1968; Gall, 1969). These extra copies of rDNA are non-chromosomal and form a cytogenetically distinct nuclear cap.

2. Differential RNA transcription

a) Control of transcription by the functional state of RNA polymerase

Control of transcription by RNA polymerase has been particularly actively studied in bacterial cells, where it has been shown that the RNA polymerase consists of several subunits - $\alpha_2\beta\beta'\omega\sigma$ (Travers & Burgess, 1969; Travers, 1971). The β subunits are involved in binding to DNA (Travers, 1971) whereas σ factor is required for initiation of transcription (Travers & Burgess, 1969). Moreover, other factors e.g. rho (Roberts, 1969) and psi (Travers, 1971) have been identified and shown to modify RNA polymerase activity in E.coli. Relatively little is still known about the exact subunit structure and the roles of subunits of eukaryotic RNA polymerases (Kedinger et al., 1972).

b) Control of transcription by the functional state of the DNA

Although the DNA content of all cells in the mammalian organism is probably identical (see above), each cell synthesises its own specific pattern of proteins, and in some cases particular proteins are synthesised only in certain cell types e.g. trypsinogen and chymotrypsinogen in the pancreas, or peptide hormones in the anterior pituitary gland. It is becoming increasingly clear that this control is achieved by "switching-off" the genes responsible, in

cells not synthesising particular proteins (Dupraw, 1970; Comings, 1972). This switching-off mechanism is believed to involve a process of masking of the appropriate region of DNA, whereas "switching-on" requires a reversal of this process. How, in molecular terms, might this be obtained?

In nuclei of eukaryotic cells, long DNA molecules occur complexed with histone proteins, non-histone proteins and a small amount of RNA. This complex is known as chromatin. Whether or not a region of DNA is transcribed into RNA probably depends to a large extent upon the interaction of these various elements in chromatin.

(i) The role of histones

That histones might have a role in regulating gene expression was suggested more than 20 years ago (Stedman & Stedman, 1950). Evidence for this has come from the ability of histones to inhibit transcription of DNA (Huang & Bonner, 1962; Paul & Gilmour, 1968; Bonner et al., 1968; Spelsberg & Hnilica, 1971). Some results have implied that lysine-rich (Barr & Butler, 1963; Huang et al., 1964) or arginine-rich (Allfrey et al., 1963; Holoubek, 1966) histones may be more effective as template inhibitors. These conflicting reports may well represent differences in experimental technique (Allfrey & Mirsky, 1963; Smart & Bonner, 1971). Two interesting observations regarding lysine-rich histones should however be made. First, removal of lysine-rich histone (together with some acidic protein) leads to stimulation in transcription of repetitious DNA sequences (Georgiev et al., 1966). Secondly, Koslov & Georgiev (1970) observed that the RNA synthesised on chromatin was of relatively short chain length, but after

removal of lysine-rich histones, the chains were as long as those synthesised on DNA. Lysine-rich histones may, therefore, prevent the movement of RNA polymerase along DNA molecules, repetitious DNA sèquences being particularly affected.

It is most unlikely that there is sufficient heterogeneity within the histones to allow for specific repression of individual genes. There are only 5 main classes of histones. These may display some microheterogeneity due to acetylation, methylation, and phosphorylation (Elgin et al., 1971; Comings, 1972). However, acetylation does not seem to alter the association of the histones with DNA (Allfrey et al., 1964; Clarke & Byvoet, 1969), or the degree of suppression of RNA synthesis (Clarke & Byvoet, 1969). The role of acetylation is therefore not clear. The role of methylation is also unresolved, although it is of interest that a chromatin preparation which catalyses methylation of histones will also bring about the methylation of certain cytosine residues in DNA (Burdon, 1971a; Burdon & Douglas, 1974).

Phosphorylation may well, however, play some role in regulating the effect of histones. Histone kinases whose activity can be modulated by cAMP are present (Langan, 1969) and phosphorylation of lysine-rich histones modifies the binding of the histone to DNA (Adler et al., 1971).

In summary, the relative lack of heterogeneity of the histones, together with the infrequency of amino-acid substitution with evolution (De Lange et al., 1969) and the fact that histones of genetically active and inactive chromatin (see below) are very similar, lead to the

conclusion that histones are non-specific repressors of gene activity rather than specific repressor molecules.

It should perhaps also be mentioned that chromosomal RNA has been implicated as a type of control element (Huang & Huang, 1969; Dahmus & Bonner, 1970). However, as discussed earlier, the very existence of chromosomal RNA is still very much in doubt. Understanding the possible role of any chromosomal RNA must await its more accurate characterisation.

(ii) The role of non-histone proteins

It now appears much more likely that specific modulators of gene activity are to be found in the non-histone protein fraction of chromatin (Paul, 1971). The regulatory proteins in the non-histone protein fraction are probably individually present in much smaller amounts than those with enzymic or structural functions but they show much greater variability (Elgin et al., 1971; Comings, 1972). Evidence which supports this concept comes from the heterogeneity of the non-histone proteins (Elgin et al., 1971), from the species and organ specificity of their electrophoretic pattern (Platz et al., 1970), from the metabolic activity of the fraction, and from the fact that non-histone proteins respond to hormone administration (Comings, 1972). Reconstitution experiments have provided much of the fine evidence for the role of non-histone protein in stimulating gene expression (Paul & Gilmour, 1968; Gilmour & Paul, 1969, 1970). The specificity of the interaction of non-histone protein with DNA has been demonstrated by hybridisation studies on the RNA produced from reconstituted chromatin (Gilmour & Paul, 1969, 1973). Early experiments of this type

only detected the synthesis of RNA from repetitious sequences of DNA (Gilmour & Paul, 1969). However, recently acidic proteins from foetal mouse liver have been shown to interact with liver or brain chromatin so as to permit synthesis of mRNA for globin, a messenger species which is known to be transcribed from unique sequences in DNA (Gilmour & Paul, 1973; Harrison et al., 1972; Bishop & Rosbach, 1973).

It would therefore appear that a non-specific inhibition of template activity is produced by the histones. This can, however, be modified by particular non-histone proteins so that specific RNA sequences are synthesised.

(iii) Euchromatin and heterochromatin

It is of interest to consider how these effects might be brought about. Thin-sectioned interphase nuclei generally reveal the chromatin in two distinct states (a) relatively diffuse, finely filamentous euchromatin which is active in incorporating DNA or RNA precursors (Hay & Revel, 1963; Littau et al., 1964) and (b) compact, granular heavily-staining heterochromatin which is synthetically almost inert (Heitz, 1933; Frenster et al., 1963; Dupraw, 1970). The relative amounts of euchromatin and heterochromatin varies considerably in different species but on average 10-20% of the chromatin is heterochromatin.

DNA content of euchromatin and heterochromatin

(1) Satellite DNA is highly enriched in constitutive heterochromatin i.e. heterochromatin occurring in the homologous portions of both homologous chromosomes (Pardue & Gall, 1970; Jones & Robertson, 1970; Mattoccia & Comings, 1971; Yasmineh & Yunis, 1971). However, by no means all of the repetitious DNA is localised to the heterochromatin.

Thus, RNA synthesised from repetitious DNA sequences has been shown to hybridise *in situ* with euchromatic portions of the chromosomes (Arrighi *et al.*, 1970; Hennig *et al.*, 1970; Jones & Robertson, 1970). This observation is of considerable importance as regards the possibility that repetitious sequences might serve as control elements in transcription, and would therefore be expected to be present in euchromatin (Section I.F.2.b.v).

(2) Main band heterochromatic DNA is relatively A + T rich whereas euchromatic DNA is relatively G + C rich (Mattoccia & Comings, 1971). The significance of this is uncertain but it may be related to the late replication of heterochromatin (Comings, 1972).

The differences in genetic activity of euchromatin and heterochromatin are probably related to the compactness of the chromatin fibres. Heterochromatin has a greater fibre diameter than euchromatin (Dupraw, 1970). In this respect, it is significant that there is considerably more non-histone protein in euchromatin than in heterochromatin (Dupraw, 1970). The condensed heterochromatin is likely to be associated with very tight packing of the DNA, such as in a supercoiled supercoil configuration (Fig. I.3) (Dupraw, 1970). Such a structure would provide little room for the activity of large globular polymerase molecules (60-100 Å in diameter).

(iv) Models of chromosome structure and their relationship to gene activity

Although euchromatin as a whole is genetically active, only particular regions of it are transcribed at any one time. How is activity of euchromatin controlled? It seems

likely that one can extend the argument presented above for the inactivity of heterochromatin. Many workers have now suggested that the DNA of euchromatin consists of alternating regions of tightly compacted and loosely compacted nucleoprotein (Dupraw, 1970; Crick, 1971; Paul, 1972). This has been concluded from studies on the characteristic banding patterns of the polytene chromosomes of diptera, the band regions containing high concentrations of DNA and histones. Histones reduce the pitch of the DNA helix (Bram, 1971) and therefore cause tightening of the helix. This may well induce supercoiling of the nucleoprotein complex. For transcription of a segment of DNA to take place, some modification of this compact structure would be necessary to allow access of an RNA polymerase molecule. Two recent models of how this might be achieved will be briefly discussed.

Crick (1971) has suggested that the coding sequences of the DNA are in the interband regions, the compact region providing the recognition sites for control elements. He speculates that such a binding site might be produced in the compact region if hairpin regions of supercoiled DNA could be partially unwound, perhaps by the binding of histones to neighbouring regions in the DNA, so that some single-stranded DNA is produced at the tip of the loop. Evidence for the existence of such single-stranded regions in DNA is currently lacking (Levy & Simpson, 1973) although there have been two recent reports of the isolation of single-stranded DNA from eukaryotes (Tapiero et al., 1972; Amalric et al., 1973).

On the other hand, Paul (1972) has suggested that the

binding site for RNA polymerase is near the band-interband junction, and that transcription moves into the band region from the initiation site. Binding of some polyanionic substance, perhaps a non-histone protein, to a site closely linked to the promoter would be expected to cause a local reduction in supercoiling of the DNA by competing with DNA for some of the positive charges on the histones. RNA polymerase would then be able to approach the promoter, RNA synthesis could proceed, and the nascent RNA (also a polyanion) would cause further unwinding of the adjacent compact nucleohistone.

(v) The structure of the operon in eukaryotic cells

The organisation of the transcriptional unit and the associated control sequences in eukaryotic DNA can now be considered. The concept of the operon (Jacob & Monod, 1961) provided a unified interpretation of regulatory phenomena in prokaryotes. They proposed that each operon, containing one or more structural genes, is controlled by an operator sequence in the adjacent DNA. When a protein repressor molecule (synthesised from a regulator gene) is bound to the operator, RNA polymerase is prevented from transcribing the structural genes. Transcription of the structural genes may be induced by the binding of some activator molecule to the repressor, which prevents it from binding to the operator. This model is so simple and so elegant that it has been considered as the basis of most models of the operon in eukaryotes. However, the model in eukaryotes should also provide a role for repetitious and unique DNA sequences, for the existence of large HnRNA molecules which are degraded in the nucleus, and should provide a mechanism for integrated control of

several different genes in response to a single stimulus e.g. a hormone.

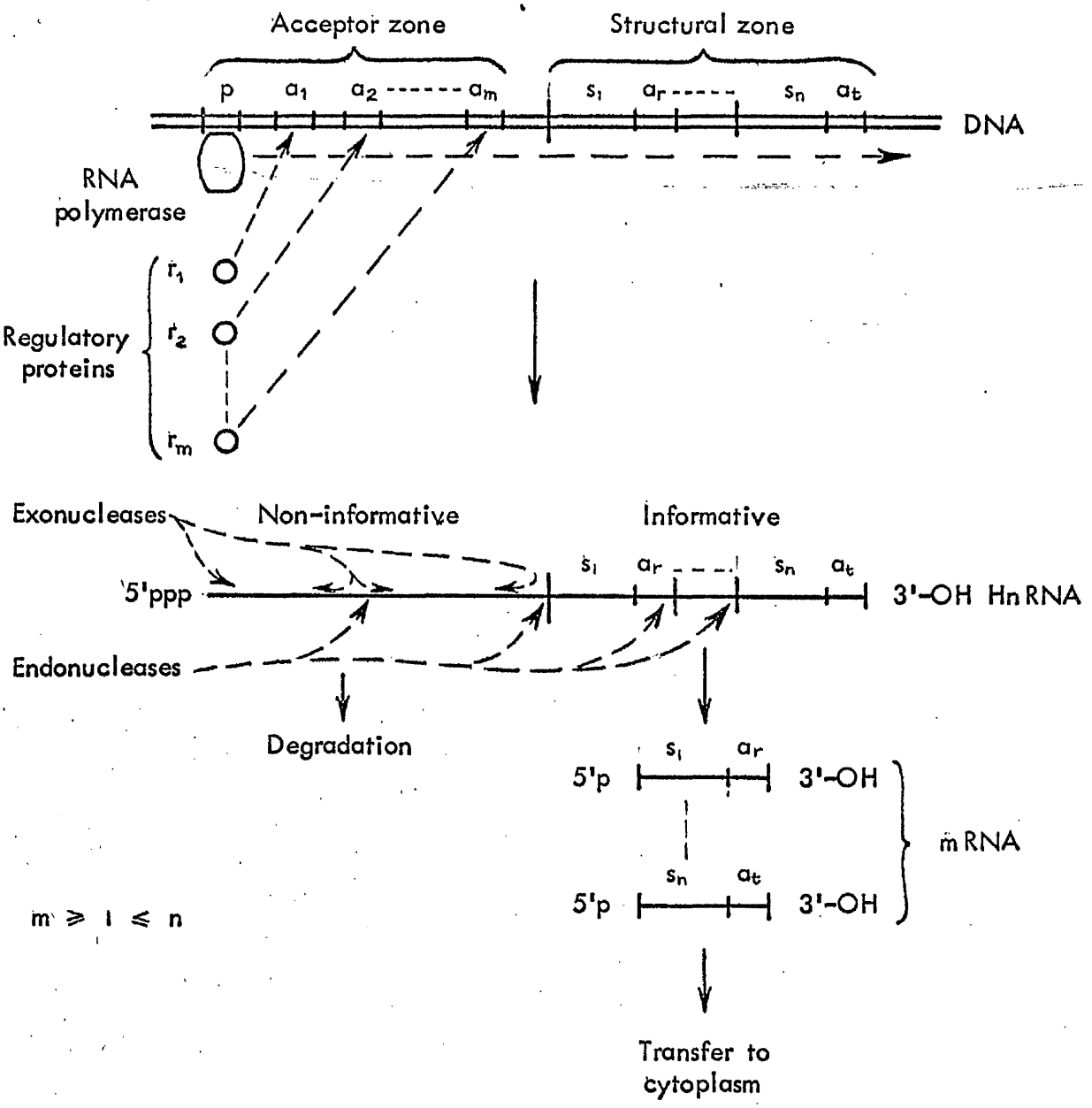
(1) The model of Britten & Davidson

The first model to include a role for the repetitious sequences in DNA was proposed by Britten & Davidson (1969). They suggested that producer genes (i.e. structural genes) are controlled by a number of receptor genes (i.e. operators). Similar receptor genes may be found associated with a number of different producer genes. The receptor gene, which is normally switched-off by binding of histones, can be switched-on by combination with an appropriate activator molecule. They proposed that the activator molecule may be HnRNA or chromosomal RNA, but their model could also accommodate its being non-histone protein (see above). The activator species is produced from integrator genes, which themselves are under control of a sensor gene, which responds to the inducing stimulus. Several integrator genes may be associated with a single sensor gene and similar integrator genes may be controlled by different sensor genes. Thus by stimulation of a large number of integrator genes, a single stimulus can lead to integrated activation at a large number of receptor genes and hence producer genes. The elements of the integrator genes and receptor genes are therefore repetitive sequences of DNA. This model is clearly consistent with the data of Davidson et al. (1973) who suggested that repetitious DNA sequences (i.e. control sequences) are interspersed among unique sequences (i.e. structural genes). However, this model alone does not provide a convincing role for large HnRNA molecules which are degraded within the nucleus.

Fig. I.5.

The structure of the transcriptional unit in eukaryotes (after Georgiev, 1969; Georgiev et al., 1972)

p, promoter; a_1----a_m, acceptor sites; a_r----a_t, other service sequences; s_1----s_n, structural genes. It should be noted that the post-transcriptional addition of poly(A) to HnRNA and mRNA molecules is not included in this model.


Acceptor zone
Structural zone
p
a_1
a_2
a_m
s_1
a_r
s_n
a_t
DNA
RNA polymerase
Regulatory proteins
r_1
r_2
r_m
Exonucleases
Non-informative
Informative
5'ppp
3'-OH Hn RNA
Endonucleases
Degradation
5'p
3'-OH
m RNA
$m \geqslant 1 \leqslant n$
Transfer to cytoplasm

(2) The model of Georgiev

Georgiev (1969) has proposed a mechanism for the regulation of gene transcription based upon a model of a giant compound operon. He suggested that the giant HnRNA species are polycistronic transcripts from compound operons (Fig. I.5). Each operon is composed of an "acceptor" region containing a series of regulator gene loci, and an informative zone of contiguous structural genes. RNA polymerase, which binds to a promoter adjacent to the acceptor regions, is only free to transcribe the operon if no regulatory proteins have bound to any of the acceptor loci. Transcription of any of the structural genes would require prior transcription of the entire sequence of regulator gene loci. Once completed, the giant HnRNA molecule is released, the non-informative portion is degraded and individual mRNA species are excised and transported to the cytoplasm. This model certainly would account for the turn-over of high molecular weight HnRNA. Moreover, the demonstration that the 5' ends of giant HnRNA species are transcribed from repetitive regions in the DNA (Georgiev et al., 1972) would be consistent with Britten & Davidson's proposal (1969) that repetitive sequences are important in integrating control of various operons. Georgiev et al. (1972) further suggested that regions at the 3' end of mRNA molecules are also transcribed from repeated sequences which may represent more control elements. One aspect of Georgiev's model for which there is still little experimental evidence is the concept that HnRNA is polycistronic (Darnell et al., 1973). With this proviso, Georgiev's model, when considered in the light of Britten & Davidson's proposals for sensor,

integrator, and receptor genes, represents a useful working model for further consideration of the mammalian genome.

G. Functional gene number in eukaryotes

It is clear from the above discussion that by no means all of the DNA is informational, some of the DNA being required for control purposes. The exact amount of informational DNA in eukaryotes is, however, unknown. A reasonable estimate could be obtained if the number of genes were known.

1. Estimation of gene number

The number of genes in the *Drosophila* genome can be estimated from the number of bands in the polytene chromosomes, since genetic studies strongly suggest that each band + interband region corresponds to a single gene (Muller, 1967; Shannon *et al.*, 1970; Thomas, 1971). This gives a value of about 5000 genes, and if there were on average 1000 nucleotide-pairs per gene, this would correspond to about 5% of the *Drosophila* genome being informational (Thomas, 1971).

Another method of estimating the number of genes is by consideration of the number of mutations per generation. On average, each structural gene sustains a deleterious mutation in 10^5 generations (Muller, 1967). On this basis, it has been calculated that the maximum number of functional genes in the human genome is about $3\text{-}4 \times 10^4$ otherwise the genetic load would become intolerable (Muller, 1967; Ohno, 1971; Ohta & Kimura, 1971). If the average gene size is 1000 nucleotides long, then only $3\text{-}4 \times 10^7$ nucleotides would be informational out of a total of 3×10^9 nucleotide

pairs per haploid genome i.e. about 1% of the total. By comparison of mutation rates of nucleotides and amino acids, Ohta & Kimura (1971) have further argued that only about 6% of the sequences in mammalian DNA can be under the intensive selection that has characterised the evolution of cytochrome c, globin chains, and histones. These estimates of gene number may be low if different regions in DNA are subject to different mutation rates (Cox, 1972), or if only a proportion of genes when mutated are capable of producing a lethal morphological phenotype (O'Brien, 1973). Thus, a minimum of about 1% but probably somewhat more than 10% of the DNA is informational.

It should be pointed out that the proportion of informational DNA clearly differs in different species. For example the salamander Amphiuma contains about 26 times as much DNA per haploid genome as man, but does not show more genetic polymorphism (Comings & Berger, 1969). It is therefore probable that most of this extra DNA is non-informational.

2. The master-slave hypothesis

One model of the eukaryotic genome which would account for the apparently excessive amount of DNA is the master-slave hypothesis (Callan & Lloyd, 1960; Callan, 1967; Thomas, 1971). In this model, genes are thought to be arranged as a series of tandemly repeating copies, each set of genes consisting of a master and a number of slaves. An important feature is that the nucleotide sequence in the slave genes should be kept constant by periodic checking of the sequence against that of the master. One implication of the model is that it could provide a mechanism by which the repetitious genes for

ribosomal, transfer and 5S RNA species (at least) could maintain similar sequences.

Some evidence for this model has been obtained by Thomas and coworkers (Lee & Thomas, 1973; Pyeritz & Thomas, 1973; Thomas et al., 1973) (see Section I.D.2.c). They have concluded that about 50% of each chromomere of Drosophila consists of repetitious sequences arranged in tandem i.e. about 15,000 nucleotides which would correspond to 10-20 copies of a particular gene. The important corollary to this conclusion is that the essential genes of Drosophila should be found in slightly repetitious DNA. Results obtained by measurements of reassociation kinetics have, however, suggested that most essential genes are found in unique sequences of DNA (Laird, 1971; Harrison et al., 1972; Suzuki et al., 1972; Bishop & Rosbach, 1973; Davidson et al., 1973). Further work must therefore be directed towards establishing the exact number of copies of individual genes.

3. Possible roles of non-informational DNA

If one disregards the master-slave hypothesis for want of definite evidence, it can be concluded from the above discussion that more than 80% of the DNA of eukaryotes may be non-informational. If so, a number of roles can be postulated for this DNA.

(a) It may serve as control sequences in DNA, in regulating transcription of informational DNA. As Paul (1972) has pointed out, quite large stretches of DNA might be required to act as binding sites for regulatory proteins, if a large destabilising effect is required to open up the superhelix. However, such control sequences may well be under similar selection pressures as informational sequences (Thomas, 1971;

Cox, 1972), and their level will therefore be included in the content of essential DNA sequences as estimated by genetic studies.

(b) They might be transcribed into HnRNA sequences which are turned over in the nucleus. The transcript might include copies of regulatory regions in DNA (Georgiev, 1969; Georgiev et al., 1972), or alternatively, copies of some DNA sequences of no known function (i.e. transcribed "junk" DNA) simply by their proximity to the informational DNA (Comings, 1972). In either event, the proportion of each HnRNA molecule which is informational is still not known. However, assuming the initial transcription product may be a giant HnRNA molecule of about 30,000 nucleotides long (Georgiev et al., 1972; Holmes & Bonner, 1973; Darnell et al., 1973), if it were monocistronic only about 3% (1000 nucleotides) would be informational.

(c) At least some of the non-informational DNA is probably not transcribed and is of no known function. Heterochromatic DNA, including most satellite DNAs, would be an example of this type. However, heterochromatin rarely constitutes more than 20% of the genome (Comings, 1972). Another example would be non-transcribed spacer DNA (Brown et al., 1971, 1972). Such DNA sequences may be conserved simply because they are interspersed between useful genes.

4. Differences in base composition between informational and non-informational DNA

Utilising the amino-acid sequences of a number of known mammalian proteins, and assuming a random use of the various codons for a particular amino-acid, it can be calculated that the base composition of that portion of the genome

which is informational is about 44-48% G + C (King & Jukes, 1969; Comings, 1972). It should be pointed out that this calculation may be invalid since certain codons may be under-represented (Elton, 1973). However, these values are in good agreement with the base composition of mRNA and HnRNA (Darnell, 1968). By contrast, the G + C content of the DNA as a whole is 40-42%. These results suggest that non-informational DNA may be relatively A + T rich. The reason for this is currently unknown. One possibility is that methylation of cytosine residues might be an important characteristic of essential, transcribed and translated DNA (Comings, 1972). If deamination were to occur, thymidine would result. Such a transition in duplicated, non-essential DNA would not be selected against, and could result in the gradual shift of the base composition of non-essential DNA to A + T richness. This conclusion is, of course, consistent with the suggestion of Bram (1972) that very A + T rich DNA may not be transcribed (Introduction, Section I.B).

H. Aims of the present work

The above discussion has attempted to highlight some of the areas of uncertainty in our knowledge of the processes involved in transcription of DNA. Some of the major problems are:- What is the nature of the regulatory sequences in DNA? How do regulatory elements recognise and interact with the regulatory sequences? What is the structure and function of repetitious sequences of DNA? What are the relative amounts of informational and non-informational DNA? To what extent are non-informational sequences transcribed? And how are all these various sequences organised in the genome?

Studies of the organisation of nucleotide sequences in mammalian DNA have been handicapped by the considerable difficulty involved both in performing genetic studies in eukaryotes, and in detecting particular polynucleotide sequences. One type of sequence which might serve as a suitable marker, if it were present, would be a homopolymeric or homopolymer-rich region in DNA. The existence of such sequences was suggested several years ago by Szybalski and coworkers. They demonstrated that if certain mammalian DNAs are denatured in the presence of a vast excess of poly(U) or poly(G), then the buoyant density of the DNA increases (Opara-Kubinska et al., 1964; Kubinski et al., 1966). They concluded that these DNA species contain dA-rich and dC-rich sequences, capable of forming hydrogen-bonded complexes with these ribopolymers. However, the nature of the complexes formed was never fully established, and since no steps were taken to remove mismatched regions from the complex, it is uncertain to what extent specific base-pairing had, in fact, taken place. Despite this limitation, the general technique of ribohomopolymer interaction with DNA, followed by CsCl density gradient centrifugation has found considerable application in the study of bacteriophage (Kubinski et al., 1966; Hradecna & Szybalski, 1967; Guha & Szybalski, 1968; Summers & Szybalski, 1968), mammalian viral (Kubinski & Rose, 1967; Rust, 1970), mitochondrial (Borst & Ruttenberg, 1972) plant (Quetier & Vedel, 1973) and satellite (Maio, 1971; Kurnit et al., 1972) DNAs. The ability of all DNAs tested to form complexes with either poly(U) or poly(G) led to the suggestion that pyrimidine-rich clusters may serve as markers for the binding of RNA polymerase (Szybalski et al., 1966).

Indeed, Becker (1972) has further speculated that deoxyribonucleotide homopolymeric regions of different types may serve as binding sites for polymerases with different functions.

It therefore seemed important to clarify whether there are regions in the mammalian genome which consist largely of clusters of a particular nucleotide. A further reason for investigating such regions in DNA, has been the recent upsurge of interest in homopolymeric regions in RNA. At the outset of this work, only the long (150-200 nucleotides long) poly(A) sequences in HnRNA and mRNA had been characterised to any extent. It was, however, not clear whether these sequences might arise by transcription of a homopolymeric region in DNA or by some post-transcriptional process. The same question can be asked of the short A-rich and U-rich sequences in HnRNA which have been more recently discovered. Although demonstration of homopolymeric regions in DNA would not prove that these regions are necessarily transcribed, it would be helpful to know if the genetic potential exists in DNA to produce homopolymeric regions in RNA.

The availability of commercial preparations of ^{3}H-labelled ribohomopolymers suggested that the existence of true homopolymeric or homopolymer-rich regions in DNA could be precisely demonstrated, and further that the characteristics of these regions could be investigated, using standard techniques of nucleic acid hybridisation (Gillespie, 1968). Thus, [^{3}H] poly(U) would hybridise with dA-rich regions in DNA, and [^{3}H] poly(C) would hybridise with dG-rich regions. These two ribohomopolymers were chosen for most studies because of the ease with which unhybridised ribopolymer could be removed with pancreatic RNase. Moreover, the

organisation of homopolymeric sequences within DNA might be investigated by studying their distribution in different fractions of DNA.

The present study was therefore undertaken to ascertain whether homopolymer-rich regions exist in mammalian DNA, and if so, to investigate the characteristics of these regions particularly with respect to their size, purity and organisation through the genome.

Table II.1.

Constituents of Eagle's Minimal Essential Medium (MEM), as used in Department of Biochemistry, University of Glasgow.

MEM amino acids	mg/litre	MEM vitamins	mg/litre
L-arginine	126.4	D-calcium pantothenate	2.0
L-cystine	24.0	choline chloride	2.0
L-glutamine	292.0	folic acid	2.0
L-histidine HCl	38.3	i-inositol	4.0
L-isoleucine	52.5	nicotinamide	2.0
L-leucine	52.5	pyridoxal HCl	2.0
L-lysine	73.1	riboflavin	0.2
L-methionine	14.9	thiamine HCl	2.0
L-phenylalanine	33.0		
L-threonine	47.6		
L-tryptophan	10.2		
L-tyrosine	36.2		
L-valine	46.9		

Inorganic salts and other components

	mg/litre
$CaCl_2.6H_2O$	393.0
KCl	400.0
$MgSO_4.7H_2O$	200.0
NaCl	6800.0
$NaH_2PO_4.2H_2O$	140.0
glucose	4500.0
$NaHCO_3$	2240.0
phenol red	15.3
streptomycin	100.0
penicillin	100,000 units/litre

Section II MATERIALS AND METHODS

IIA MATERIALS

(1) Biological

(a) BHK-21/C13 cells.

BHK-21/C13 cells, an established line of hamster fibroblasts (Macpherson & Stoker, 1962) were subcultured by the staff of the Wellcome Cell Culture unit of this department.

Tissue culture media and materials.

BHK-21/C13 cells were grown in Eagle's Minimal Essential Medium (based on the Glasgow Modification of Busby et al., 1964, see Table II.1.), to which had been added calf serum to 10% (v/v) and tryptose phosphate broth to 10% (v/v). This medium is designated as ETC_{10}.

Calf serum, amino-acids and vitamins were obtained from Bio-Cult Laboratories Ltd., Paisley, Scotland. Penicillin and streptomycin were supplied by Glaxo Pharmaceuticals, London. Tryptose phosphate broth was a 2.95% (w/v) solution of Difco Bacto tryptose phosphate broth (Difco Laboratories, Michigan, U.S.A.) in distilled water.

Balanced Salt Solution (BSS - Earle, 1943).

This contained 0.116 M NaCl, 5.4 mM KCl, 1 mM $MgSO_4$, 1 mM NaH_2PO_4, 1.8 mM $CaCl_2$ and 0.002% (w/v) phenol red. The pH of this solution was adjusted to 7.0 with 8.4% (w/v) $NaHCO_3$.

Phosphate Buffered Saline (PBS).

This consisted of 0.17 M NaCl, 3.4 mM KCl, 10 mM Na_2HPO_4 and 2.4 mM KH_2PO_4, pH 7.4.

Versene solution was 0.6 M disodium EDTA in PBS, to

which was added 0.002% (w/v) phenol red.

Trypsin/versene was 0.05% (w/v) trypsin (from Difco Laboratories, Michigan, U.S.A.) in the above versene solution.

(b) Krebs II mouse ascites tumour cells.

These were propagated by serial intraperitoneal transplantation in mice of the departmental colony and were harvested routinely after 10-14 days of growth. Cells were washed twice by suspension in ice-cold BSS followed by centrifugation at 800 g for 2 min at 4°C.

(2) Chemical

(a) General

All chemicals were, whenever possible, of Analar or comparable grade. CsCl was obtained from Hopkin and Williams Ltd., Chadwell Heath, Essex, England. Hydroxylapatite (Biogel HTP) was purchased from Bio-Rad Laboratories, Richmond, California. Acrylamide was obtained from Eastman Organic Chemicals, New York. Methylene bisacrylamide, ammonium persulphate and hydrogen peroxide were supplied by BDH Biochemicals Ltd., Dorset. N,N,N',N'-tetramethyl ethylene diamine (Temed) was obtained from Koch-Light Laboratories, Colnbrook, Bucks. Baycovin (diethylpyrocarbonate) was obtained from Bayer Chemical Co., Germany. Colcemid was purchased from Ciba Laboratories, Horsham, Sussex.

(b) Nucleic acids

(i) Naturally occurring.

Calf thymus, salmon sperm, E.coli, Clostridium perfringens and Micrococcus lysodeikticus DNAs were all purchased from Worthington Biochemical Corporation, New Jersey.

Generous gifts of other naturally occurring nucleic

acids are gratefully acknowledged as follows:-
Adenovirus 5 DNA, herpes simplex Type I (strain α) DNA and bacteriophage T4 DNA from Dr. N. Wilkie, Institute of Virology, University of Glasgow.
Replicating form (RF) of ØX174 DNA and bacteriophage λ and T7 DNAs from Dr. A. Campbell and Dr. D. Jolly, Department of Biochemistry, University of Glasgow.
SV40 DNA from Dr. R. Eason, Department of Biochemistry, University of Glasgow.
Cytoplasmic RNA from BHK-21/C13 cells, labelled with ^{14}C-uridine, from Dr. J. Abrahams, Department of Biochemistry, University of Glasgow.

(ii) Synthetic

1. Radioactively labelled.

[^{3}H] poly(U) (specific activity 4.3-5.1 µg/µc), [^{3}H] poly(C) (specific activity 8.5-8.9 µg/µc), [^{3}H] poly(A) (specific activity 11.1 µg/µc) and [^{3}H] poly(G) (specific activity 13.2 µg/µc) were all purchased from Miles Laboratories Inc., Stoke Poges, England. [^{3}H] poly(U) (specific activity 28.9 µg/µc) and [^{3}H] poly(C) (specific activity 36.7 µg/µc) were also obtained from Schwartz-Mann Orangeburg, New York.

2. Non-radioactive.

Poly(U), poly(C), poly(A) and poly(G) were obtained from Miles Laboratories, Stoke Poges, England. Poly(dA), poly(dT), poly(dG), poly(dC), poly(dA).poly(dT) and poly[d(A-T)] were all supplied by P-L Biochemicals Inc., Wisconsin, U.S.A.

(c) Enzymes

Pancreatic ribonuclease (RNase A, protease free; E.C.2.7.7.16) was purchased from the Sigma Chemical Co. Ltd.,

London. RNase T_1 and RNase T_2 were manufactured by Sankyo Co. Ltd., Tokyo, Japan and were marketed by Calbiochem Ltd., Los Angeles, U.S.A. RNase preparations were dissolved in 0.1 x SSC and any traces of deoxyribonuclease (DNase) were destroyed by heating pancreatic RNase to 100°C for 10 min, or RNase T_1 or T_2 to 80°C for 5 min. Pronase, B grade, was also obtained from Calbiochem Ltd. Before use, pronase was autodigested at 37°C for 2 hours to remove contaminating nucleases.

(d) Nitrocellulose filters.

Filters of diameter 25 mm were Sartorius Membranfilters, 0.45 μ pore size, and were supplied by V.A. Howe and Co. Ltd., London. Filters of diameter 13 mm (0.45 μ pore size) and Swinnex 13 mm filter holders, were purchased from Millipore (U.K.) Ltd., Middlesex.

(e) Scintillation spectroscopy.

For most experiments, a toluene based scintillation fluid, consisting of 5 g 2,5 diphenyloxazole (PPO) (Koch-Light Laboratories, Colnbrook, Bucks.) per litre Analar toluene was used.

For assay of radioactivity in hydrolysates of gel slices a mixture of toluene scintillator (6 parts) with 2-methoxyethanol (4 parts) was employed.

(f) Autoradiographic and photographic materials.

AR10, fine grain autoradiographic stripping film, Kecordak Micro-phile film, and D19B and D8 developer fluids were obtained from Kodak. Amphix fixative was supplied by May & Baker Ltd., Dagenham, Essex. Giemsa stain (BDH Biochemicals Ltd.) contained 0.75% Giemsa in glycerol-methanol (1:1, v/v).

IIB METHODS

(1) Cell culture techniques

(a) Growth and harvesting of cells

BHK-21/C13 cells were cultured routinely as monolayers in rotating 80 oz bottles, according to the method of House & Wildy (1968). 25 x 10^6 cells were seeded into 200 ml of ETC_{10} and incubated at 37°C for 3 days in an atmosphere of 5% CO_2 (to maintain the buffering capacity of the medium), by which time the cell density had reached 4 x 10^8 to 10^9 cells per bottle. For serial passaging, cells were removed from the bottle by treatment with 25 ml trypsin/versene, after which 2.5 ml serum was added to inhibit tryptic activity. Cells were then pelleted by centrifugation at 800 g for 2 min at 4°C, washed twice with 50 ml ice-cold BSS, resuspended in ETC_{10} and dispensed in aliquots of 25 x 10^6 into sterile 80 oz bottles containing 200 ml ETC_{10}. Alternatively, harvesting of cells was performed by scraping cells off the glass with a rubber wiper into 50 ml BSS. Cells were pelleted by centrifugation at 800 g for 2 min at 4°C, and washed twice with 50 ml cold BSS.

(b) Contamination checks

All sterile media and passaged cells were checked regularly for bacterial, fungal or PPLO (mycoplasma) infection, as follows:-

Bacterial contamination.

Aliquots were grown on blood agar plates and heart infusion broth at 37°C. Results were considered negative if no growth was seen after 7 days.

Fungal contamination.

Aliquots were added to Sabouraud's medium and grown at

32°C. Again, no visible growth after 7 days was taken to indicate the absence of mycotic contamination.

PPLO infection

PPLO agar plates were seeded with passaged cells by piercing the agar surface with a charged pasteur pipette. Plates were incubated in an atmosphere of 5% CO_2 in N_2 at 37°C. PPLO infected cultures were also prepared as controls. After 7 days incubation, the plates were examined microscopically for the characteristic "fried egg" appearance of PPLO colonies.

(2) Preparation of DNA

The procedure used for preparation of DNA varied depending upon the source.

(a) Modified method of Marmur.

The standard method for preparation of DNA from BHK-21/C13, Krebs II ascites tumour, and mouse spleen cells was a modification of the method of Marmur (1961). Mouse spleen cells were first dispersed by brief (30 sec) homogenisation of spleen in ice-cold BSS. Cells were washed twice by suspension in cold BSS followed by centrifugation at 800 g for 2 min. Cells were then suspended in BSS, made 2% with respect to sodium dodecyl sulphate and heated to 60°C for 10 min. Sodium perchlorate (5 M) was added to a final concentration of 1 M, and the lysate was extracted with an equal volume of chloroform-isoamyl alcohol (24:1, v:v) for 30 min at room temperature. The aqueous phase was recovered by centrifugation at 10,000 g for 10 min and an equal volume of absolute alcohol was layered over the surface. DNA was spooled out on to a glass rod and was redissolved in SSC. This deproteinisation was repeated

twice. Pancreatic RNase was added to 50 µg/ml and the solution was incubated at 37°C for 1 hour. 500 µg pronase/ml was then added and the incubation was continued for a further 2 hours. Sodium perchlorate was again added to 1M, and chloroform-isoamyl alcohol extraction was repeated until no precipitate was visible at the interphase (3-4 extractions). After precipitation with absolute alcohol, DNA was dissolved in SSC, and one-ninth volume of 3.0 M sodium acetate-0.001 M EDTA, pH 7.0 was added. DNA was then precipitated with 0.54 volumes isopropyl alcohol, spooled on to a glass rod, redissolved in 0.1 x SSC and the DNA concentration was estimated assuming that a concentration of 50 µg/ml has an absorbance at 260 nm approximately equal to 1 (Hirschmann & Felsenfeld, 1966). 100-200 µg DNA was added to a concentrated CsCl solution, the density was adjusted to 1.710 g/cm^3 and the solution was centrifuged at 33,000 rpm for 48 hours in the fixed angle 40 rotor of a Beckman Model L ultracentrifuge. Fractions were collected dropwise, the absorbance at 260 nm was measured and those fractions containing DNA were pooled and dialysed extensively against 0.1 x SSC. This treatment was necessary to remove undigested RNA fragments, which are mainly poly(A) and poly(G) (Results, Section III.I), from the DNA preparations. The average single-stranded molecular weight of this DNA was about 20×10^6. DNA prepared in this way was used for all experiments where unsheared DNA was required.

(b) Hydroxylapatite absorption chromatography method.

For experiments using sheared DNA, on some occasions DNA was prepared by the hydroxylapatite method of Britten et al., (1970). This method is based upon the observation

that in the presence of 8 M urea - 0.24 M potassium phosphate (pH 6.8), DNA is absorbed to hydroxylapatite (HAP) whereas RNA and protein do not bind (Britten et al., 1970). The tissue was first suspended in 10 volumes 8 M urea - 0.24 M potassium phosphate - 1% sodium dodecyl sulphate - 0.01 M EDTA, pH 6.8 and was then sheared in a sealed filled vessel with a Waring Blendor at full speed for 6 periods of 15 sec, alternating with 45 sec cooling in ice between each period. The lysate was then added to a thick slurry of HAP (5 g per g of tissue used) suspended in 8M urea - 0.24 M potassium phosphate, and was left for one hour at room temperature, with occasional mixing. The suspension was then poured into a sintered glass funnel (Gallenkamp Sinta No. 3, approximately 30 g HAP per funnel of diameter 9.5 cm) and the liquid drawn off under low suction. The HAP was washed with several volumes of 8 M urea - 0.24 M potassium phosphate, the mixture being stirred frequently to prevent channelling, until no further material was eluted, as judged by A_{260} and A_{280} measurements (about 2 l./30 g HAP). Washing was continued with 0.014 M phosphate buffer (about 500 ml per 30 g HAP), the removal of urea being monitored by measurement of refractive index. DNA was then eluted with 200-300 ml 0.4 M phosphate buffer, and was dialysed against 10-20 volumes of distilled water. After concentration on the rotary evaporator, DNA was finally dialysed against 0.1 x SSC. The DNA product had an A_{260}:A_{280} ratio of 1.8-2.0 and an A_{260}:A_{230} ratio of 2.3-2.4, values very similar to those of the DNA products of Method (a) above. DNA prepared by the HAP method is completely free of contamination with poly(A) or poly(G) (see Results, Section III.I) and has an average

single-stranded molecular weight of 1.1×10^6.

Comparison of DNA samples prepared by the two methods described above ((a) and (b)) revealed no significant difference in their ability to hybridise with [^{3}H] poly(U) or [^{3}H] poly(C).

(c)' Preparation of DNA from viruses.

Viral DNA was kindly donated by several workers (see Materials Section). The methods used in purification of these DNA samples are briefly summarised here. DNA was prepared from purified bacteriophage by the phenol procedure of Mandell & Hershey (1960). Adenovirus 5 DNA was prepared by phenol-SDS extraction of virus (Wilkie et al., 1973), and DNA from herpes simplex (Type 1, strain α) was extracted with phenol-SDS and separated from host DNA by CsCl gradient centrifugation (Bell et al., 1971). Standard techniques were used to prepare the relaxed circular forms of ØX174-RF DNA (Rush et al., 1967) and SV40 DNA (Rush et al., 1971).

(3) Shearing of DNA and estimation of molecular weight

DNA preparations (100 µg/ml in 0.1 x SSC) were sheared by sonication in a Dawe Soniprobe. The glass vial containing 5 ml DNA solution was held in a pre-cooled metal container, surrounded by melting ice. Successive cycles of sonication for 15 sec followed by cooling for 45 sec were performed until the desired size of DNA was achieved. The average single-stranded molecular weight was determined by sedimentation velocity through 0.9 M NaCl - 0.1 N NaOH in the Model E analytical ultracentrifuge. The observed S value was corrected to $S^{o}_{20,w}$ (the value expected with DNA at infinite dilution in a solution having the density and viscosity of water at 20°C) by multiplying by 1.16 (Studier,

1965). The molecular weight (M) can then be estimated using the relationship of Studier (1965):-

$$S^{o}_{20,w} = 0.0528\ M^{0.400}$$

It should be noted that this relationship was derived from studies with relatively long DNA fragments (Studier, 1965). The validity of this relationship at molecular weights less than 10^6 has not been proven. The smallest DNA fragments obtained, estimated to have an average single-stranded molecular weight of about 60,000, was produced by sonication at 7 mA for 3 min.

(4) Fractionation of DNA

(a) On CsCl gradients

The mole per cent guanine plus cytosine (G + C) content of DNA is linearly related to the buoyant density of the DNA in CsCl (Schildkraut et al., 1962). Thus, when DNA is centrifuged to equilibrium in a CsCl density gradient, fractions at different regions of the gradient correspond to DNA molecules of different G + C content. 100-300 µg DNA in 0.1 x SSC (0.5 ml) was added to 4.5 ml CsCl solution (density 1.78 g/cm^3) and the final density adjusted to 1.700 g/cm^3. Density was calculated, using the relationship of Ifft et al. (1961), from the refractive index, which was measured using an Abbe refractometer. The solution was overlaid with paraffin oil, and the mixture was centrifuged at 33,000 rev/min at 20°C in the fixed angle 40 rotor of a Beckman model L ultracentrifuge, for 64 hours. 6 drop fractions were collected from the bottom of the tube and the refractive index of every fifth fraction was measured. 0.5 ml 0.1 x SSC was added and the absorbance at 260 nm was measured.

(b) By hydroxylapatite chromatography

Mixtures of denatured and reassociated DNA can be conveniently separated by chromatography on hydroxylapatite (HAP) at 60°C in 0.12 M potassium phosphate buffer pH 6.8 (PB). Under these conditions native DNA, but not denatured DNA, is absorbed to HAP (Bernardi, 1965; Walker & McLaren, 1965; Miyazawa & Thomas, 1965; Britten & Kohne, 1968).

HAP was suspended in 0.12 M PB and the slurry was poured into a water-jacketed Whatman chromatography column (diameter 25 mm) at 60°C. The HAP was allowed to settle and it was then washed with 50 ml 0.12 M PB under gravity.

Samples of DNA in 0.12 M PB were denatured at 100°C for 15 min and were reannealed at 60°C to various C_0t values.

C_0t = DNA concentration (moles nucleotides/l.) x time (secs). (Britten & Kohne, 1968).

An approximate value for C_0t can be easily calculated as follows:-

$$C_0t = \frac{A_{260} \text{ x time (hour)}}{2} \quad \text{(Kohne, 1968)}$$

Thus, incubation of a reassociation mixture with an optical density of 2 for 1 hour results in a C_0t of about 1.

After reassociation, the DNA was applied immediately to the HAP column at 60°C. Columns were operated at 100 µg DNA/g HAP, well below the measured capacity of the HAP. Single-stranded DNA was washed out in 30 ml of 0.12 M PB. The fragments of DNA containing some duplex regions were eluted with 15 ml 0.3 M PB at 60°C. 5 ml fractions were collected throughout, and A_{260} and A_{320} were measured. To correct for any optical density due to suspended particles of HAP the A_{320} was subtracted from the A_{260}. Column recoveries were in the range of 95-100% with respect to

optical density.

Fractions of DNA which were to be investigated in hybridisation reactions were dialysed against 0.1 x SSC. However, when further fractionation on HAP was to be carried out, samples were adjusted to a suitable concentration by rotary evaporation and then dialysed against 0.12 M PB.

(5) Hybridisation Procedures

(a) With [^{3}H] poly(U) and [^{3}H] poly(C)

Contamination of solutions or glassware with pancreatic RNase would lead to rapid degradation of [^{3}H] poly(U) and [^{3}H] poly(C) (Markham & Smith, 1952; Volkin & Cohn, 1953). To destroy traces of RNase, all solutions used were, wherever possible, autoclaved at 15 p.s.i. for 30 min, and pipettes and test tubes were held in a flame for several seconds prior to use (Burdon, 1967).

(i) With DNA in solution.

Unless otherwise specified, hybridisation experiments were carried out with both DNA and ribohomopolymer in solution. DNA was denatured by heating to 100°C for 15 min in 0.1 x SSC, followed by rapid cooling in an ice-bath. Varying amounts of denatured DNA in 1 ml 2 x SSC were incubated with [^{3}H] poly(U) for 4 hr at 20°C or for 30 min at 62°C with [^{3}H] poly(C) (Results Section III.A). In most experiments, the incubation mixtures were cooled in ice, 5 ml cold 2 x SSC were added, and hybrids were collected on 2.5 cm nitrocellulose filters (Sartorius Membranfilters 0.45 μ pore diameter) (Gillespie, 1968), which had been presoaked for at least 30 min in 2 x SSC. After washing with 50 ml 2 x SSC, filters were transferred to scintillation vials and incubated at 20°C for 30 min with 5 ml 2 x SSC containing pancreatic

RNase (5 µg/ml for [^{3}H] poly(U), 10 µg/ml for [^{3}H] poly(C)). Under these conditions, poly(dA).[^{3}H] poly(U) and poly(dG).[^{3}H] poly(C) hybrids are stable to digestion with RNase, whereas unhybridised [^{3}H] poly(U) and [^{3}H] poly(C) are completely degraded (Results, Section III.B). Vials were cooled for 1 min in ice, and both sides of the filters were then washed with 50 ml 2 X SSC. Filters were oven dried, covered with toluene based-scintillator fluid and assayed for radioactivity by scintillation spectrometry.

For experiments involving the subsequent analysis of the hybrid molecules in CsCl density gradients, hybrids were not collected directly on to filters, but were treated in solution with RNase. Precipitation of the RNase resistant radioactivity with cold 5% trichloracetic acid demonstrated that the levels of hybrid detected with these two methods were not significantly different.

(ii) With DNA bound to filters.

This technique was used mainly for analysis of fractions of DNA from CsCl density gradients and for determination of thermal melting profiles. DNA, usually in 0.7 ml of 0.1 x SSC, was denatured by addition of an equal volume of 1 N NaOH, which was neutralised 15-20 min. later by addition of 3 volumes of a solution containing 3 M NaCl, 1 N HCl, and 1 M Tris (pH 8.0) in a ratio of 2:1:1 (Birnstiel et al., 1968). The solution was then passed under gravity through 13 mm HAWP Millipore filters, presoaked in 2 x SSC and held in Swinnex-13 filtration units which were fitted with 10 ml Perspex tubes. Filters were then washed with 5 ml 6 x SSC, dried at 80°C for 2 hours under vacuum, numbered with a soft lead pencil and stored at -20°C (Gillespie & Spiegelman, 1965). When required, the

filters were soaked in 2 x SSC, covered with 2 x SSC containing ^{3}H-labelled ribohomopolymer, the volume being sufficient to allow the filters to move freely in solution. The temperature and time of incubation were as for hybridisation in solution. Filters were then washed by a batch method (Birnstiel et al., 1968) by placing them in a 2 l. beaker with 1 l. 2 x SSC at room temperature and keeping them under continuous agitation for 30 min. This washing procedure was repeated twice. The filters were then incubated with pancreatic RNase (5 µg/ml for [^{3}H] poly(U) and 10 µg/ml for [^{3}H] poly(C)) for 30 min at 20°C, rinsed twice with 2 x SSC and oven dried, before counting in the scintillation counter.

(b) With [^{3}H] poly(A)

Hybridisation experiments with [^{3}H] poly(A) were performed with DNA in solution, since it proved difficult to consistently degrade unhybridised [^{3}H] poly(A) which had bound to nitrocellulose filters.

Denatured DNA was incubated with [^{3}H] poly(A) at 45°C for 4 hours in 1 ml 2 x SSC (Results, Section III.A). Incubation mixtures were cooled in ice and 4 ml of a solution of 0.625 M trisodium citrate pH 7.0 containing 6.25 µg RNase T_2/ml was added. The final solution, which effectively contained 5 µg RNase T_2/ml in 0.5 M trisodium citrate was incubated at 30°C for 30 min. Under these conditions [^{3}H] poly(A) is completely digested to acid soluble fragments whereas hybrid molecules of poly(dT).[^{3}H] poly(A) are stable (Fig. II.1). It should be noted that with different salt concentrations (Fig. II.1), pure hybrids of poly(dT).[^{3}H] poly(A) are not totally resistant to digestion with RNase T_2.

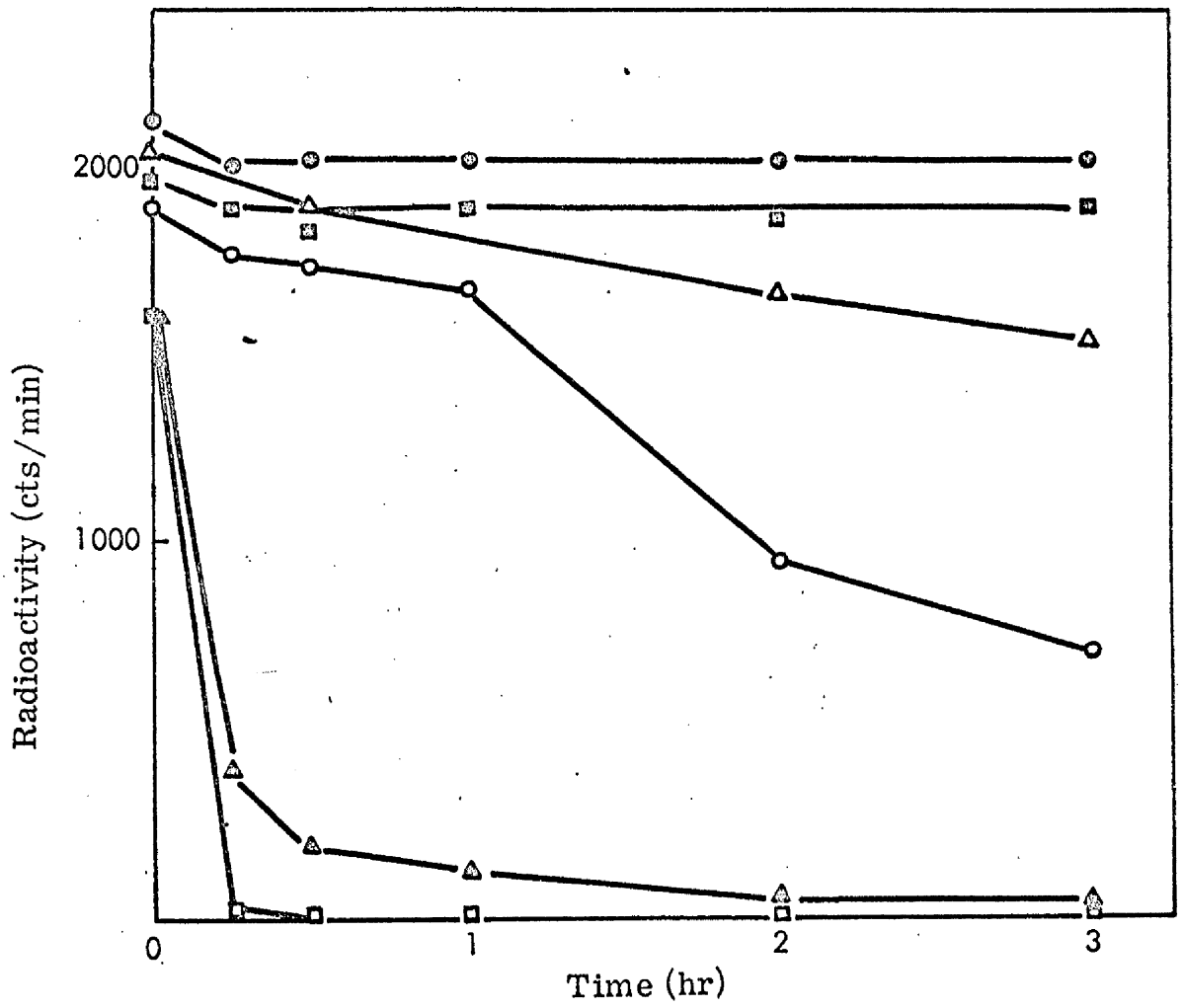
Fig. II.1.

Stability of poly(dT)·[^{3}H] poly(A) hybrids to digestion with RNase T_2

1 µg poly(dT) was incubated with 0.25 µg [^{3}H] poly(A) in 1 ml 2 x SSC at 45°C for 4 hours. 4 ml of citrate buffer containing RNase T_2 was added so that the final concentration of RNase T_2 was 5 µg/ml, and tubes were incubated at 30°C. At various times, 0.8 ml was removed and trichloracetic acid precipitable radioactivity was estimated.

Final concentration of buffer in each incubation was:- ●——● 0.5 M trisodium citrate pH 7.0; ▩——▩ 0.5 M trisodium citrate pH 6.0; △——△ 0.35 M trisodium citrate pH 6.0; o——o 0.25 M trisodium citrate pH 4.5.

Control incubations in the absence of any added poly(dT) were also performed:- □——□, trisodium citrate 0.5 M pH 7.0, 0.5 M pH 6.0, and 0.35 M pH 6.0; ▲——▲ trisodium citrate, 0.25 M pH 4.5.


2000
1000
0
1
2
3
Radioactivity (cts/min)
Time (hr)

Ice-cold trichloracetic acid was added to a final concentration of 5%, the tubes were kept in ice for 15 min, and acid insoluble radioactivity was collected on nitrocellulose filters and assayed by scintillation spectrometry.

(6) Thermal dissociation of hybrids

The thermal dissociation profile of hybrids can be studied using three techniques. With the DNA, and therefore the hybrid, bound to filters, either the amount of ^{3}H-labelled ribopolymer dissociated from the hybrid, or the amount still present in the hybrid can be measured. Where hybrid molecules are in solution, only the amount of ribopolymer remaining in the hybrid can readily be estimated.

(a) With DNA filter bound

(i) Measurement of [^{3}H] poly(U) and [^{3}H] poly(C) eluted from hybrids.

This method was used to study the melting profiles of hybrids of [^{3}H] poly(U) and [^{3}H] poly(C) with dA-rich and dG-rich regions, respectively. Nitrocellulose filters were loaded with 20 µg BHK-21/C13 DNA as described above (Section II.B.5.a.ii) and poly(U)-DNA and poly(C)-DNA hybrids were formed under optimal conditions (Section II.B.5.a.i). Filters were washed three times with 1 l. 2 x SSC, and treated with pancreatic RNase (5 µg/ml for [^{3}H] poly(U), 10 µg/ml for [^{3}H] poly(C)) for 30 min at 20°C. After washing in 2 x SSC, any remaining nuclease activity was inactivated by incubating filters in 2 x SSC containing 0.1% diethylpyrocarbonate (Baycovin) for 30 min at 20°C. Filters were then rinsed in 2 x SSC and dried. A filter was then placed in 1 ml 1 x SSC (for poly(U)) or 1 ml 2 x SSC (for poly(C)) at a given temperature, and then successively transferred to fresh 1 ml

portions of buffer, which were heated in temperature increments of about 5°C. After 5 min at each temperature, the released RNA was recovered by trichloracetic acid precipitation and counted. For comparison, the melting profile of such DNA.poly(U) and DNA.poly(C) hybrids were also analysed using the other techniques (see below, II.B.6.a.ii and II.B.6.b), the results obtained with the different methods being virtually identical.

(ii) Measurement of [^{3}H] poly(U) and [^{3}H] poly(C) remaining in hybrid.

This method was used mainly to study the melting profiles of hybrids of [^{3}H] poly(U) and [^{3}H] poly(C) with DNA, after exhaustive digestion with pancreatic RNase. It was found necessary to employ this method because of difficulty in consistently precipitating the eluted ribohomopolymer, possibly because of the relatively small size of these fragments (Results, Section III.D).

Formation of hybrids and inactivation of RNase was as in Section II.B.6.a.(i) above, the main difference being that RNase treatment of hybrids was carried out with 20 µg/ml at 20°C for 24 hours. Two or three filters were then placed in 1 ml of 1 x SSC containing 50 µg unlabelled poly(U) (for DNA.poly(U) hybrids) or 1 ml 2 x SSC containing 50 µg unlabelled poly(C) (for DNA.poly(C) hybrids), and tubes were held at a particular temperature for 5 min. Filters were then rinsed briefly in 50 ml ice-cold 2 x SSC, were dried, and radioactivity remaining on the filters was estimated.

(b) With DNA in solution

In the cases of the poly(dA).[^{3}H] poly(U) hybrid and poly(A).[^{3}H] poly(U) complex, dissociation was performed in

solution since poly(dA) does not bind to nitrocellulose filters under the conditions used (see Results Section, Table III.2). After the formation of the hybrid in solution, 50 µg cold poly(U) was added to prevent any rapid reassociation of [^{3}H] poly(U) to poly(dA) or poly(A). The salt concentration was adjusted to 1 x SSC and aliquots of the mixture were heated in temperature increments of 5°C. After 5 min at each temperature, RNase was added to 5 µg/ml, and the mixture incubated at 20°C for 30 min. The RNase resistant radioactivity which remained was then assayed by trichloracetic acid precipitation.

(7) Analysis of size of hybridised [^{3}H] poly(U) and [^{3}H] poly(C)

(a) Preparation of samples

Hybrids of BHK-21/C13 DNA with [^{3}H] poly(U) and [^{3}H] poly(C) were formed and collected on nitrocellulose filters as described above. The filters were then incubated at 20°C in 5 ml 2 x SSC containing 5 µg RNase/ml for 30 min, or 20 µg RNase/ml for 1 hr or 24 hours. After washing extensively with 2 x SSC, filters were incubated with 0.1% diethylpyrocarbonate as described above. 50 µg of either unlabelled poly(U) or poly(C) were added (the same ribopolymer as had been hybridised) in 2 ml water and filters were heated to 50°C for 5 min to ensure complete dissociation of the hybrid. The eluate was dialysed overnight against 5 l. water, lyophilised, and dissolved in 100 µl half-strength electrophoresis buffer (see below) containing 10% (w/v) sucrose and bromophenol blue 0.01% (w/v) (Loening, 1969). Samples containing cytoplasmic RNA labelled with ^{14}C-uridine, from BHK-21/C13 cells, were also prepared so that 5S and 4S RNA

species could be used for size reference. However, since poly(U) for example, is known to have negligible secondary structure at room temperature (Michelson et al., 1967), it was important to ensure that the mobility of the 5S and 4S RNAs corresponded respectively to polynucleotides of 120 (Brownlee et al., 1968) and approximately 80 (Holmquist et al., 1973) nucleotides long with minimal secondary structure. For this purpose, ^{14}C-cytoplasmic RNA, [^{3}H] poly(U) and [^{3}H] poly(C) were pre-treated with 1 M formaldehyde-0.01% SDS (w/v) at 63°C for 15 min (Boedtker, 1968). Urea was added to 8 M and samples were heated to 50°C for 1 min (Jelinek et al., 1973). Sucrose (in 8 M urea) was then added to 20% (w/v) and bromophenol blue to 0.01% (w/v).

(b) Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis was carried out as described by Loening (1967, 1969). The basic gel buffer used was 36 mM tris-HCl, 30 mM Na_2HPO_4, 1 mM EDTA, pH 7.8. Gels containing 10% (w/v) acrylamide, 0.25% (w/v) methylene bisacrylamide, 0.33% (v/v) Temed, and 0.025% (w/v) ammonium persulphate were prepared. When samples had been pre-treated with urea, both gels and gel buffer also contained 8 M urea.

Gels were run in Perspex tubes, 10 cm long and 6.0 mm internal diameter, 2 ml of gel mixture producing a gel 6 cm long. Gels were subjected, at first, to electrophoresis at 2 mA per tube for 15 min. Electrophoresis was continued at 5 mA per tube until the marker dye had migrated 3 cm from the top of the gel in studies of poly(dA) and poly(dG) regions, or 4 cm from the gel surface in studies of dA-rich and dG-rich regions. Gels were then frozen in solid CO_2 (Drikold, ICI), sliced at approximately 1 mm intervals,

and the resulting slices were dried at 60°C for 1 hour in scintillation vials. 0.3 ml hydrogen peroxide (100 volumes) was added, vials were tightly capped, and were then heated at 60°C until gel slices had completely dissolved (about 16 hours). 10 ml toluene/methoxyethanol scintillator was added and samples were counted in the liquid scintillation counter. This method for assay of radioactivity in gel slices is a modification of the method of Tischler & Epstein (1968).

(8) Chromosomal localisation of dA-rich regions

In general, the procedure described by Gall & Pardue (1972) was employed.

(a) Preparation of metaphase chromosomes

To a culture of BHK-21/C13 cells, growing exponentially on glass cover slips in a 10 cm Petri dish, was added colcemid to a final concentration of 5 μg/ml. Incubation was continued for a further 4 hr after which the medium was removed and 5 ml hypotonic (0.95% w/v) sodium citrate was added. After 20 min, 5 ml ice-cold ethanol-acetic acid, in a ratio of 3:1, was added very slowly. The mixture was removed carefully, and replaced with a further 5 ml fixative for 10 min, after which the cells were washed with 45% acetic acid. The cover slips were then removed and allowed to dry in air at room temperature.

(b) Removal of RNA and residual protein

Coverslips were placed in 0.2 N HCl at room temperature for 30 min to remove any remaining basic proteins, and were then rinsed in several changes of distilled water and dried in air. They were then incubated in a solution of 2 x SSC containing 100 μg RNase/ml at 20°C for 2 hours, and were

rinsed in 3 changes of 2 x SSC. Cover slips were then passed through solutions of 70% and 95% ethanol and were dried in air.

(c) Denaturation of DNA

Cover slips were placed in 0.07 N NaOH at room temperature for 2 min. They were then transferred to 70% ethanol for 10 min, changing the ethanol twice to remove alkali. The specimens were rinsed in 3 changes of 95% ethanol over a period of 10 min and were dried in air.

(d) Hybrid formation with [^{3}H] poly(U)

10 µl drops of [^{3}H] poly(U) (Miles Laboratories) (14 µg/ml) in 2 x SSC were placed on siliconised slides, the coverslips were applied to the drop, and the preparations were incubated at 20°C overnight. The coverslips were then removed, washed thoroughly in 2 x SSC and incubated with 20 µg RNase/ml for 30 min at 20°C. Further exhaustive washing in several changes of 2 x SSC was followed by passing the coverslips through 70% and 95% ethanol, after which they were dried in air.

(e) Autoradiography

Coverslips were mounted on glass slides and covered with fine grain stripping film (AR10, Kodak). Specimens were left for 4 months and were then developed for 5 min in Kodak D19b developer, rinsed in water, and fixed with Amphix. After rinsing in several changes of distilled water, preparations were lightly stained with freshly diluted (1:20 v/v) Giemsa (about 30 sec). Slides were rinsed in water and dried in air. A drop of DePex mounting medium (Gurr Ltd., London) was added and sections were covered with a 13 mm coverslip.

(f) Photography

Preparations were examined with a Leitz microscope, using a green filter, and photographs of suitable areas were taken using Recordak Micro-phile film (Kodak). Negatives were developed for 2 min in Kodak D8 developer, and fixed in Amphix. Prints were produced by the Department of Medical Illustrations, University of Glasgow.

Background grains were estimated by cutting out from the prints both the chromosomes and also random areas outwith the chromosomes and weighing. It was calculated that a background of 6 grains was present over the same area as was occupied by the chromosomes.

Section III RESULTS

(A) Conditions required for formation of hybrids between ribohomopolymers and BHK-21/C13 DNA

The ribohomopolymers poly(U), poly(C) and poly(A) were all found to hybridise readily to denatured BHK-21/C13 DNA. However, the accuracy of base-pairing within a nucleic acid hybrid depends, to a large extent, upon the conditions used in its preparation (McCarthy & Church, 1969). It was therefore of prime importance to establish standard conditions for formation of DNA·ribohomopolymer hybrids. One method of carrying out such an investigation is to incubate the RNA species with DNA under varying conditions, the input ratio of RNA:DNA being on the rising part of the saturation curve (Results, section III.C). The optimal temperature for hybridisation is then the temperature at which the reaction rate is maximal, the duration of the incubation being the time required for complete hybridisation at the optimal temperature (Birnstiel et al., 1972).

With [^{3}H] poly(U) the reaction proceeds most rapidly at 18-22°C, and is complete in 4 hours (Fig. III.1). For all experiments involving poly(U)-DNA hybrid formation, an incubation at 20°C for 4 hours was therefore performed.

Hybrid formation with [^{3}H] poly(C) takes place optimally at 54-62°C (Fig. III.2a). To provide relatively stringent conditions for specific base-pairing, further experiments were performed at 62°C. The hybridisation reaction with poly(C) is very rapid and is complete in 5 min at 62°C. However, since the hybrids are stable to prolonged incubation (Fig. III.2b), for convenience of handling a 30 min

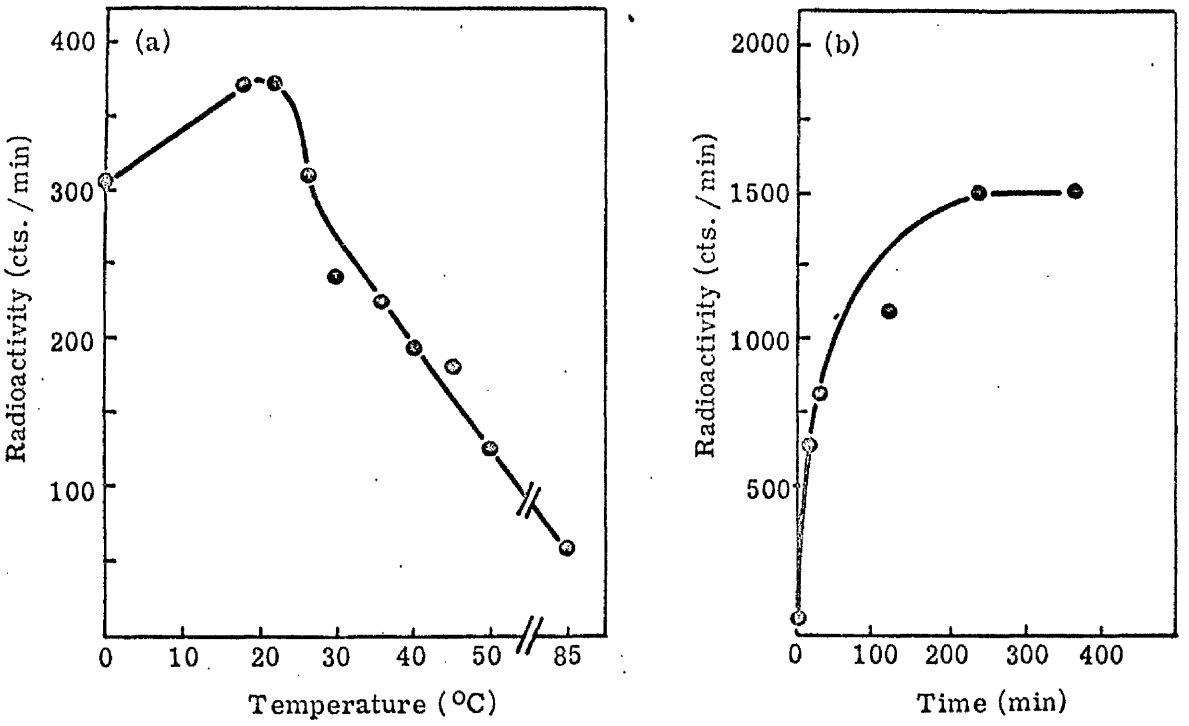


Fig. III.1.

Hybridisation of [^{3}H] poly(U) to BHK-21/C13 DNA

5 µg denatured BHK-21/C13 DNA was incubated with 0.07 µg [^{3}H] poly(U) in 0.55 ml 2 x SSC (a) for 20 min at different temperatures, (b) at 20°C for different periods. Hybrids were collected on nitrocellulose filters and processed as in METHODS.

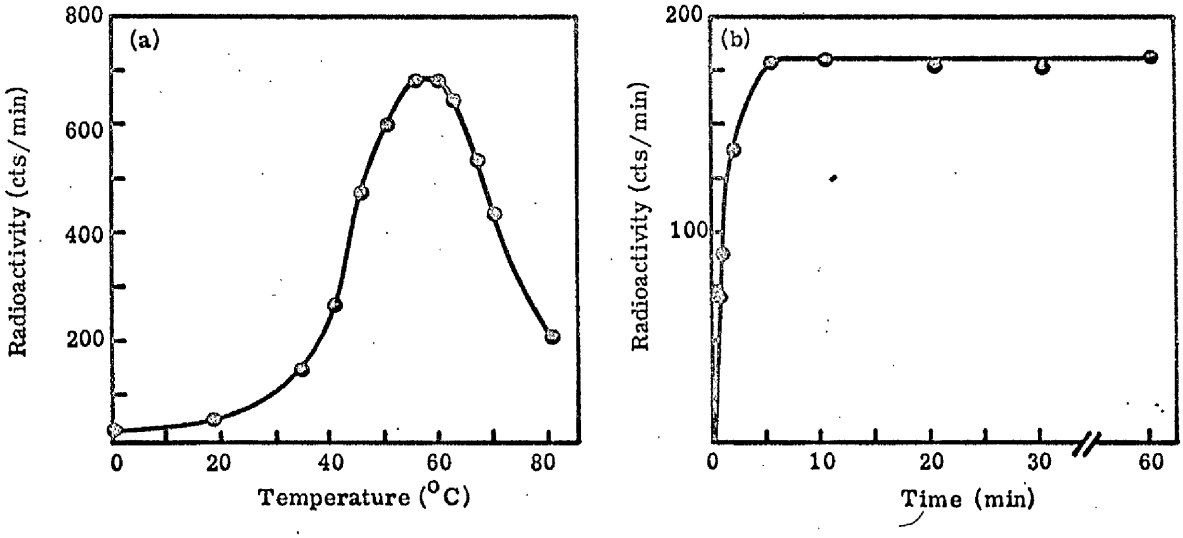


Fig. III.2.

Hybridisation of [^{3}H] poly(C) to BHK-21/C13 DNA

10 µg denatured BHK-21/C13 DNA in 1 ml 2 x SSC was incubated with (a) 0.1 µg [^{3}H] poly(C) for 20 min at different temperatures or (b) 0.02 µg [^{3}H] poly(C) at 62°C for different periods. Hybrids were collected on nitrocellulose filters and processed as in METHODS.

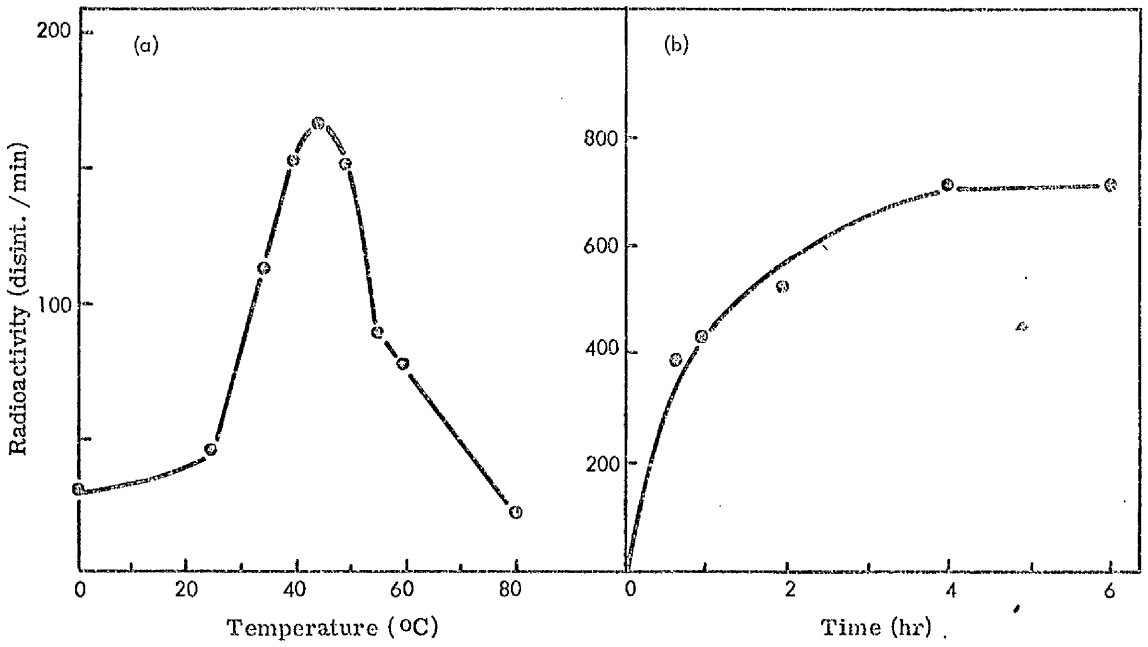


Fig. III.3.

Hybridisation of $[^3H]$ poly(A) to BHK-21/C13 DNA

10 µg denatured BHK-21/C13 DNA was incubated in 1 ml 2 x SSC with 0.1 µg $[^3H]$ poly(A) (a) for 20 min at different temperatures, (b) at 45°C for different periods. Hybrids were incubated with 5 µg RNase T_2/ml in 0.5 M trisodium citrate pH 7.0 at 30°C for 30 min, and the acid insoluble radioactivity was estimated, as described in METHODS.

incubation at 62°C was chosen as the standard method for poly(C)-DNA hybrid formation.

Similarly, a 4 hour incubation at 45 C was found to be optimal for production of [^{3}H] poly(A)-DNA hybrids (Fig. III.3). Although attempts were also made to form hybrids of BHK-21/C13 DNA with [^{3}H] poly(G) it proved difficult to find conditions for treatment with RNase T_1 in which hybrid molecules are resistant to digestion whereas unhybridised [^{3}H] poly(G) is digested. Thus, no results of [^{3}H] poly(G) hybridisation are included in this study.

These results provided the initial demonstration that sequences at least partially complementary to these ribohomopolymers are present in BHK-21/C13 DNA:- deoxyadenylate-rich (dA-rich) sequences which hybridise to poly(U); the complementary deoxythymidylate-rich (dT-rich) sequences which hybridise to poly(A); and deoxyguanylate-rich (dG-rich) sequences which hybridise to poly(C). On the basis of these observations, further experiments were designed to elucidate the nature of these regions which hybridise to ribohomopolymers.

(B) Stability of the DNA.[^{3}H] poly(U) and DNA.[^{3}H] poly(C) hybrids to digestion with RNase

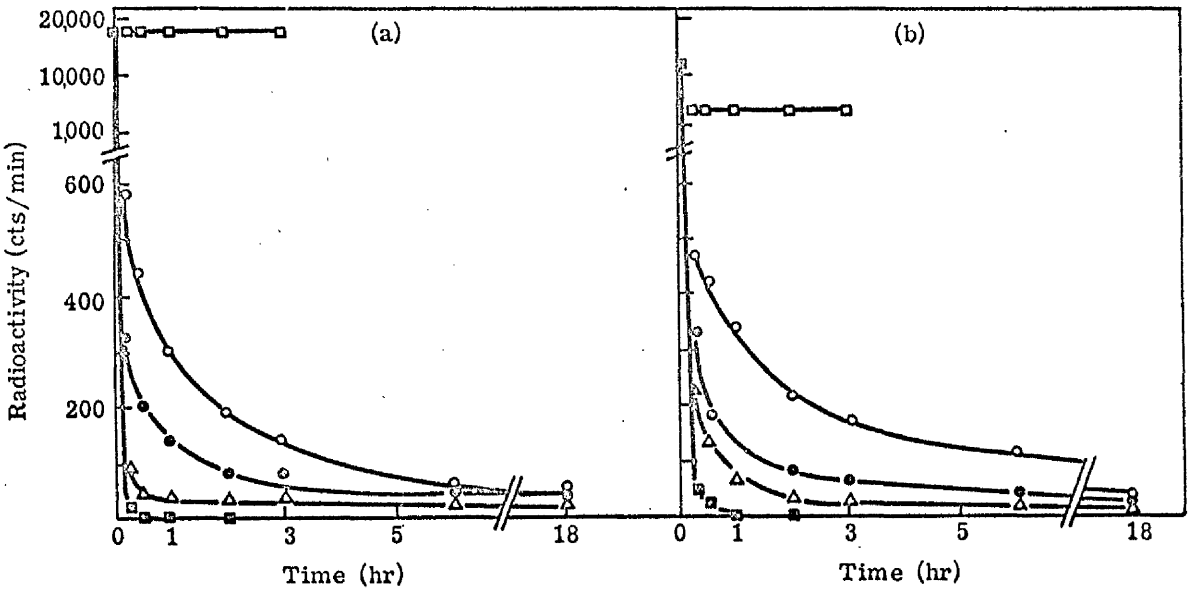
For an incubation period of 30 min at 20°C, it was found that the pancreatic RNase concentration required to completely degrade the unhybridised ^{3}H-labelled ribohomopolymers to acid-soluble fragments was 5 µg/ml for poly(U) and 10 µg/ml for poly(C). Using these conditions for the digestion of DNA-ribohomopolymer complexes with RNase, any [^{3}H] poly(U) or [^{3}H] poly(C) which is not digested, is presumably protected through hybrid formation with dA-rich or dG-rich regions, respectively, in the DNA (Gillespie, 1968). The

stability of the hybrids to digestion with pancreatic RNase is shown in Fig. III.4. Increasing the concentration of RNase, or the duration or temperature of the incubation, led to a marked reduction in the level of [^{3}H] poly(U) or [^{3}H] poly(C) remaining in the hybrids. Since the accurately matched hybrids, poly(dA)·poly(U) and poly(dG)·poly(C), are very resistant to RNase (Fig. III.4), it is probable that the reduced stability of the DNA-ribohomopolymer hybrids to RNase is due to the progressive digestion of the ribohomopolymer from partially mismatched regions within the hybrid. The milder the RNase digestion the more mismatched would be the hybrid detected i.e. the less dA-rich or dG-rich is the region binding ribohomopolymer. Therefore, for investigations of the dA-rich and dG-rich regions, only the minimal conditions, as described above, which are required for complete digestion of ribohomopolymer alone, were employed. On the other hand, a small proportion of the hybridised [^{3}H] poly(U) and [^{3}H] poly(C) was found to be resistant to vigorous RNase treatment (20 µg/ml at 20°C for 24 hours). Such hybrids will provide an estimate of the level of pure sequences of poly(dA) and poly(dG) in the DNA. It is possible that such hybrids may still contain a very small proportion of mismatched base-pairs which for some unknown reason are resistant to RNase digestion. Proof of this would, however, require DNA sequence analysis of purified hybrids. In further discussion of the hybrid complexes resistant to exhaustive RNase treatment, it is assumed that the DNA sequence hybridised is a pure deoxyhomopolymer.

Fig. III.4.

Stability of DNA·[^{3}H] poly(U) and DNA·[^{3}H] poly(C) hybrids to RNase

Incubation mixtures contained (a) 2 μg BHK-21/C13 DNA with 0.2 μg [^{3}H] poly(U); or (b) 10 μg BHK-21/C13 DNA with 0.2 μg [^{3}H] poly(C), in 1 ml 2 x SSC. Hybrids were collected on to nitrocellulose filters, treated with RNase for different periods and washed as in METHODS; —o—o—, 5 μg RNase/ml at 20°C; —●—●—, 20 μg RNase/ml at 20°C; —△—△—, 20 μg RNase/ml at 30°C. The hybrids formed by incubation of 0.2 μg poly(dA) in (a) with 0.2 μg [^{3}H] poly(U), and of 0.2 μg poly(dG) in (b) with 0.2 μg [^{3}H] poly(C), were treated in solution with RNase under all the above conditions, and acid-precipitable radioactivity measured, —□—□—. Digestion of (a) 0.2 μg [^{3}H] poly(U), and (b) 0.2 μg [^{3}H] poly(C) with 5 μg RNase/ml at 20°C, —■—■—.


(a)
(b)
20,000
10,000
1,000
600
400
200
Radioactivity (cts/min)
0
1
3
5
18
Time (hr)

(C) Estimation of the level of dA-rich and dG-rich sequences in BHK-21/C13 DNA

The level of dA-rich and dG-rich regions in BHK-21/C13 DNA were estimated by construction of standard saturation curves for ribohomopolymer hybridisation. These yielded values of 0.13% of the DNA for the dA-rich regions, and 0.072% of the DNA for dG-rich regions (Fig. III.5). These estimates are based upon the level of hybrid present after 30 min RNase digestion at 20°C, a point on the steeply descending portion of the RNase digestion curve (Fig. III.4). Slight modifications in the concentration of RNase, or in the temperature or time of digestion, therefore lead to changes in the saturation values obtained. It is therefore possible to obtain a range of different saturation values, which in turn reflects the extent of mismatched base-pairs within the hybrid i.e. the proportion of heterologous bases present within the dA-rich or dG-rich regions (Table III.1). In the case of the dA-rich regions the greatest amount of poly(U) which could be hybridised corresponded to 0.41% of the DNA i.e. dA-rich regions with a relatively large proportion of heterologous bases, whereas 0.008% of the DNA could form complexes with poly(U) which are stable to exhaustive digestion with RNase (Table III.1). On the basis of the above discussion, 0.008% of the DNA can be considered to consist of regions of poly(dA).

It is possible that the values for the levels of dA-rich and poly(dA) sequences may be overestimates, since poly(U) probably forms triple-stranded complexes with pure poly(dA) of the form poly(dA)·2 poly(U) (Riley *et al.*, 1966).

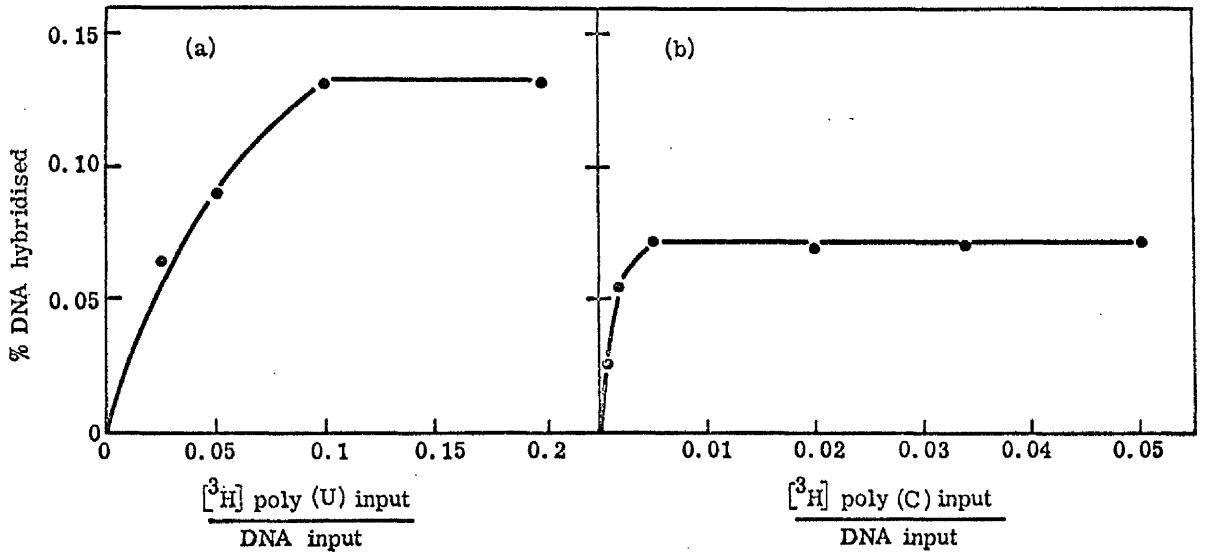


Fig. III.5.

Saturation curves for hybridisation of [^{3}H] poly(U) and [^{3}H] poly(C) with BHK-21/C13 DNA

Increasing amounts of ^{3}H-labelled ribohomopolymer were hybridised to denatured DNA under standard conditions (a) [^{3}H] poly(U) with 2 µg DNA, (b) [^{3}H] poly(C) with 10 µg DNA. Hybrids were collected on nitrocellulose filters and treated with pancreatic RNase (5 µg/ml for [^{3}H] poly(U), 10 µg/ml for [^{3}H] poly(C)) for 30 min at 20°C. Filters were washed and hybrids were estimated. Technical details were as described in METHODS.

Table III.1.

The effect of pancreatic RNase on the level of dA-rich and dG-rich sequences detected

Hybrids were formed between denatured BHK-21/C13 DNA and [^{3}H] poly(U) or [^{3}H] poly(C) under optimal conditions, and were collected on nitrocellulose filters, as described in METHODS. Filters were treated with pancreatic RNase as shown. The figures of disint/min represent the net radioactivity after subtraction of the value found in the absence of DNA. Virtually identical results were obtained when hybrids were estimated using RNase treatment in solution followed by precipitation with trichloracetic acid.

Specific activity [^{3}H] poly(U) = 435,000 disint/min/µg

Specific activity [^{3}H] poly(C) = 247,000 disint/min/µg

*Specific activity [^{3}H] poly(C) = 257,000 disint/min/µg

Table III.1. The effect of pancreatic RNase on the level of dA-rich and dG-rich sequences detected

RNase treatment at 20°C	DNA input (μg)	[^{3}H] poly(U) in hybrid (disint/min)	% of DNA in hybrid	DNA input (μg)	[^{3}H] poly(C) in hybrid (disint/min)	% of DNA in hybrid
1 μg/ml 5 min	2	3570	0.41	10	2535	0.103
5 μg/ml 5 min	2	2700	0.31	10	2410	0.098
5 μg/ml 30 min	2	1130	0.13	-	-	-
10 μg/ml 30 min	-	-	-	10	1770	0.072
20 μg/ml 30 min at 30°C	2	112	0.0126	10	385	0.0156
20 μg/ml 6 hr	2	111	0.0126	10	81	0.0033
20 μg/ml 18 hr	20	691	0.0080	100	492*	0.0019
20 μg/ml 24 hr	20	724	0.0083	100	395	0.0016

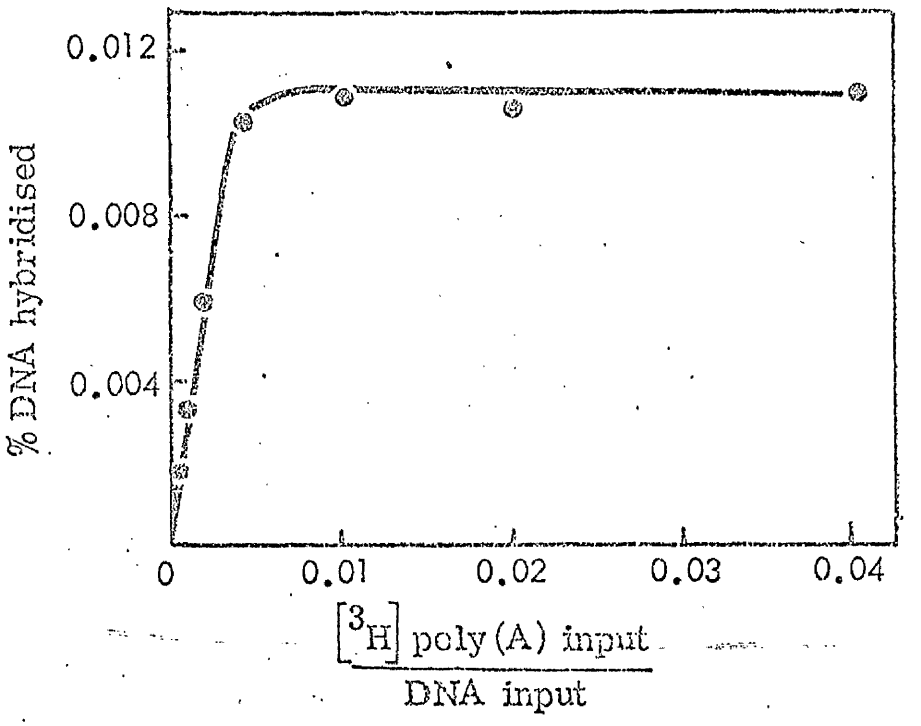


Fig. III.6.

Saturation curve for hybridisation of [^{3}H] poly(A) with BHK-21/C13 DNA

Increasing amounts of [^{3}H] poly(A) were hybridised with 10 µg denatured BHK-21/C13 DNA at 45°C for 4 hours. Hybrids were incubated in 0.5 M trisodium citrate pH 7.0 with 5 µg/ml RNase T_2 at 30°C for 30 min and the acid insoluble radioactivity was estimated as in METHODS.

However, the extent to which such triple-stranded complexes can form with natural DNA is not known. One line of evidence which suggests that at least some of the hybridised poly(U) is in two-stranded complexes comes from hybridisation studies with [^{3}H] poly(A). Three stranded complexes of the form poly(dT).2 poly(A) are not produced (Riley et al., 1966). The saturation level for the DNA·[^{3}H] poly(A) hybrid, using a 30 min digestion at 30°C with RNase T_2 (5 µg/ml) is 0.011%, a value in close agreement with that found for similar digestion conditions of DNA·[^{3}H] poly(U) hybrids with pancreatic RNase (Fig. III.6 and Table III.1). Exact comparisons cannot be made due to uncertain differences in specific activities of the enzymes. RNase T_2 activity was about 0.3 - 0.5 units/µg, as assayed by the method of Uchida & Egami (1967), whereas the pancreatic RNase activity was about 0.1 unit/µg, as assayed by the method of Kunitz (1940). These assays are very different and units of activity cannot readily be compared. The above results do, however, suggest that the level of poly(A) hybridisation is unlikely to be only 50% of the level of poly(U) hybridisation. Marshall & Gillespie (1972) have also recently found that poly(A) and poly(U) hybridise to the same extent (0.005%) with poly(dT) and poly(dA) sequences in human DNA. These studies therefore support the contention that at least some of the natural DNA·poly(U) complexes may be two-stranded.

The corresponding levels of hybrids of BHK-21/C13 DNA with [^{3}H] poly(C) are up to 0.1% of the DNA consisting of dG-rich regions whereas 0.0016% of the DNA is poly(dG) (Table III.1). There is probably a direct 1:1 relationship

between poly(C) hybridised and the amount of dG-rich sequences in the DNA, since three-stranded complexes of poly(dG) and poly(C) are not produced (Chamberlin, 1965).

(D) Size of homopolymer-rich regions in BHK-21/C13 DNA

(1) Size of the dA-rich and dG-rich regions

Direct measurement of the size of the dA-rich and dG-rich regions in the hybrids is difficult since it would require the purification of the hybrid molecules from an incubation mixture which also contained single-stranded and rapidly reassociated DNA. An indirect way of estimating the size of the regions is by measuring the size of the [^{3}H] poly(U) and [^{3}H] poly(C) which is present in hybrids with DNA. The [^{3}H] poly(U) eluted from the BHK-21/C13 DNA·[^{3}H] poly(U) hybrid has a wide spectrum of sizes as estimated by electrophoretic mobility in 10% polyacrylamide gels (Fig. III.7). In the phosphate buffer suggested by Loening (1969) for polyacryl - amide gel electrophoresis, the marker 5S and 4S RNA species both have a secondary structure which renders the molecules more compact, and therefore causes them to move faster through the gel than would be expected for molecules of the same molecular weight but with minimal secondary structure. This is clearly demonstrated by treating cytoplasmic RNA with formaldehyde to eliminate hydrogen-bonded double-helical segments (Boedtker, 1968) and then subjecting the RNA to electrophoresis in buffer containing 8 M urea which minimises 'base-stacking' interactions. Under these conditions, 4S RNA migrates about 49% of the distance of the bromophenol blue marker, whereas 5S RNA migrates to about 30% of the dye marker (Fig. III.8b). When compared to the mobilities

Fig. III.7.

Estimation of size of dA-rich and dG-rich regions

Hybrids were formed between 20 μg BHK-21/C13 DNA and 2 μg [^{3}H] poly(U) ((a) and (b)), or 50 μg BHK-21/C13 DNA and 1 μg [^{3}H] poly(C) ((c) and (d)) and collected on nitrocellulose filters. RNase digestion was with 5 μg/ml for 30 min in (a) and (c), or 20 μg/ml for 1 hr in (b) and (d). The eluted ribohomopolymer was analysed in 10% polyacrylamide gels as in METHODS. The mobility of formaldehyde treated 5S and 4S RNA species as determined in parallel gels containing 8 M urea is marked by the arrows. The mobility of the input [^{3}H] poly(U) and [^{3}H] poly(C) is also indicated.

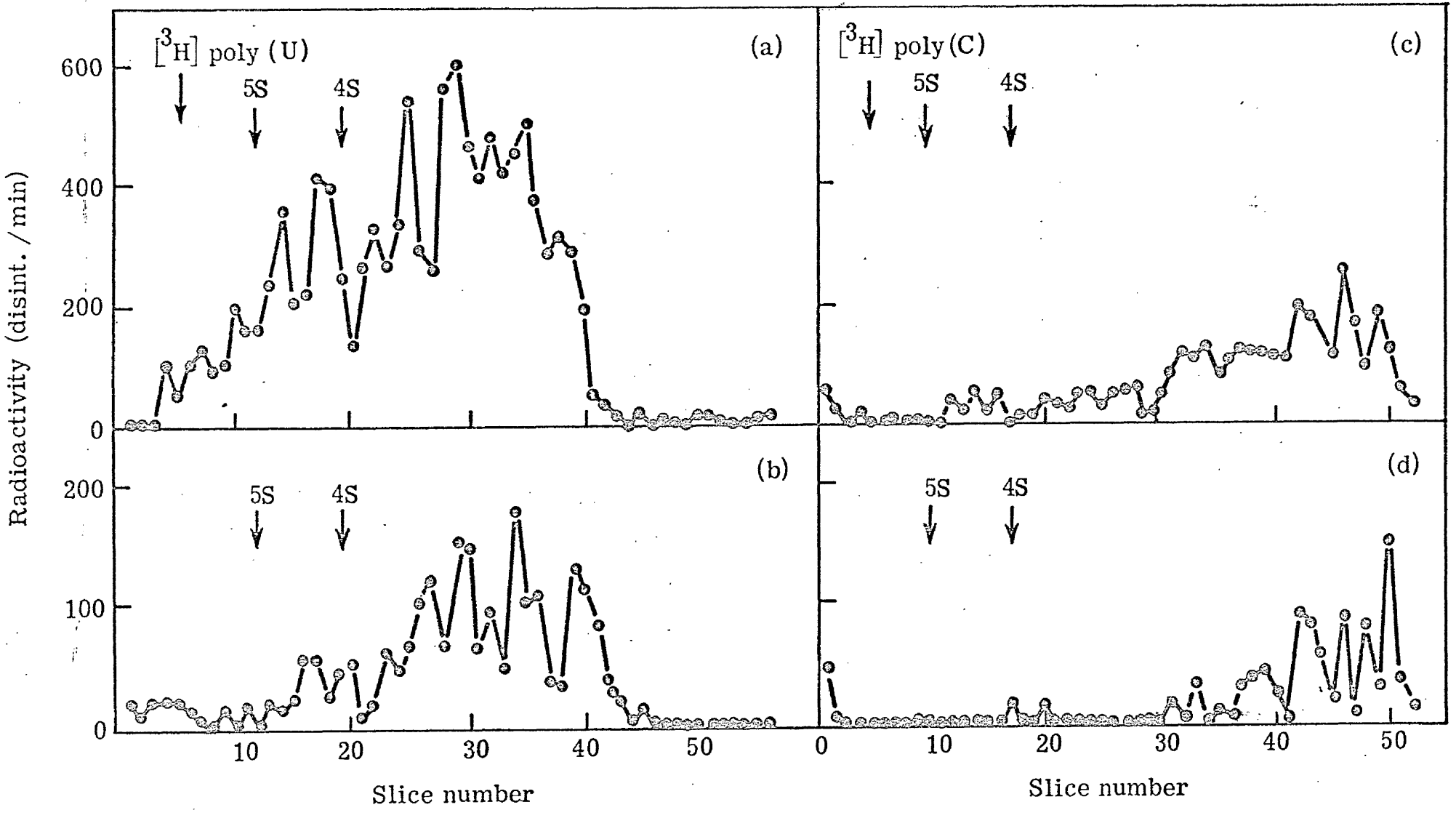

[3H] poly (U)
(a)
5S
4S
[3H] poly (C)
(c)
(b)
(d)
600
400
200
0
100
10
20
30
40
50
Radioactivity (disint. /min)
Slice number
Slice number

Fig. III.8.

Effect of treatment with formaldehyde and urea on the electrophoretic mobility of 5S and 4S RNA

^{14}C-labelled cytoplasmic RNA was treated with 1 M formaldehyde/0.01% SDS at 63°C for 15 min. Urea was added to 8 M, and the solution was incubated at 50°C for 1 min. Electrophoresis was performed (a) of ^{14}C-RNA which had not been treated with formaldehyde/urea, in 10% polyacrylamide gels, (b) of formaldehyde/urea treated ^{14}C-RNA in 10% polyacrylamide gels containing 8 M urea. The centre of the band of bromophenol blue (BPB) marker is indicated. Technical details are described in METHODS.

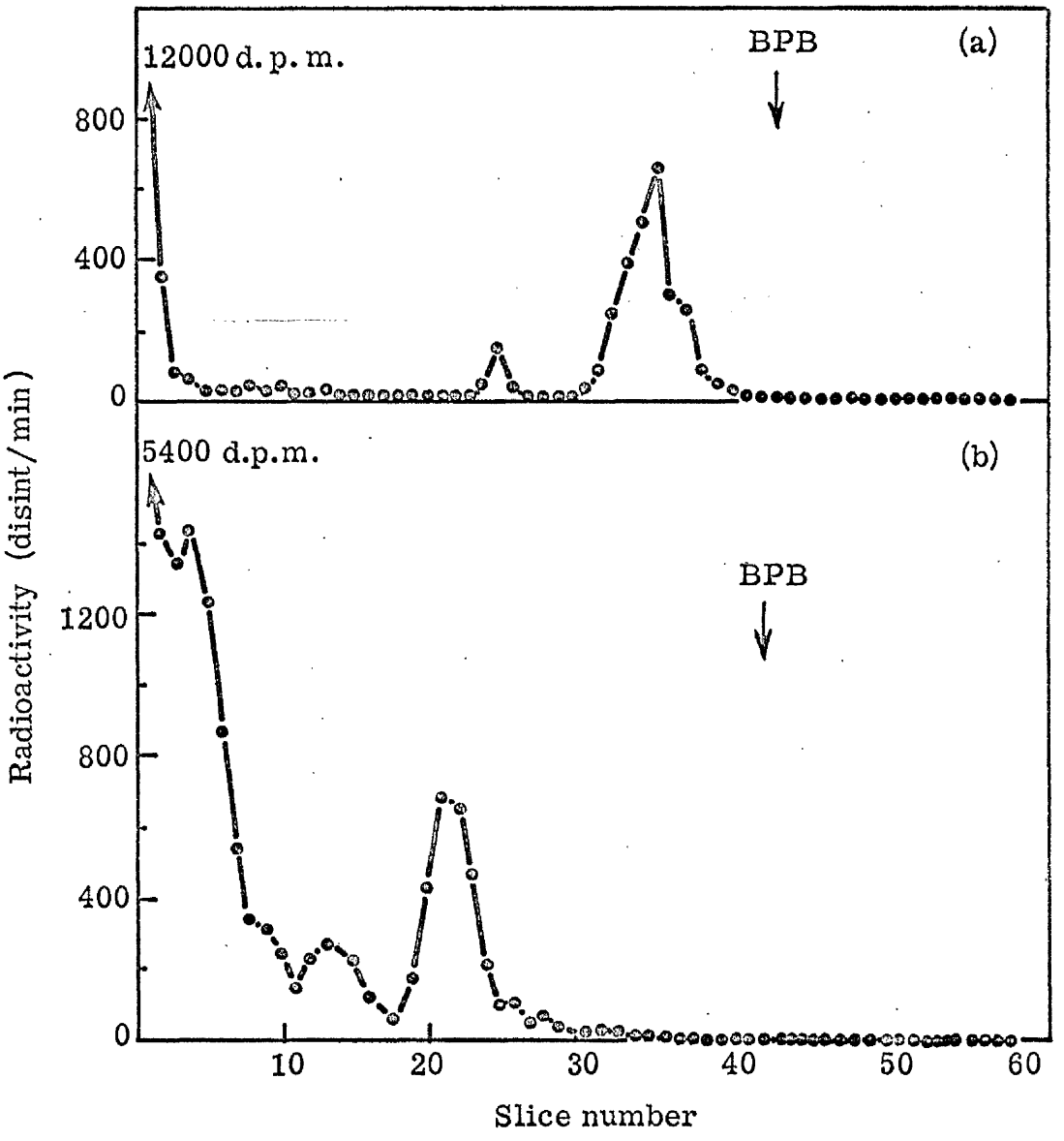
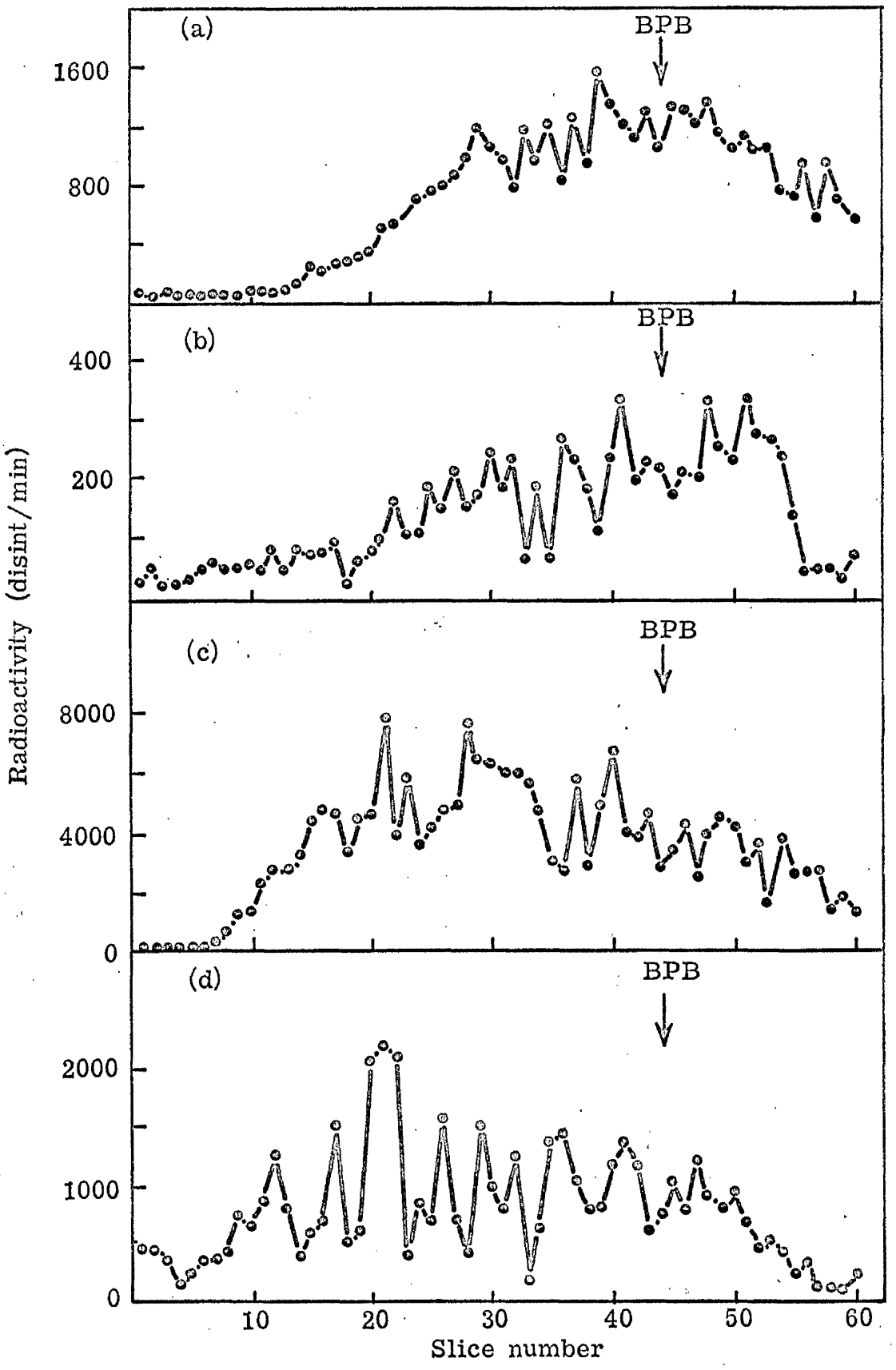

12000 d.p.m.
BPB
(a)
800
400
0
5400 d.p.m.
(b)
BPB
1200
800
400
0
10
20
30
40
50
60
Radioactivity (disint/min)
Slice number

Fig. III.9.

Effect of treatment with formaldehyde and urea on the electrophoretic mobility of poly(U) and poly(C)

Electrophoresis was performed of specimens (different batches from those used in Fig. III.7 and Fig. III.10) of:- (a) [^{3}H] poly(U) and (c) [^{3}H] poly(C) which had not been treated with formaldehyde/urea, in 10% polyacrylamide gels; (b) [^{3}H] poly(U) and (d) [^{3}H] poly(C) which had been treated with formaldehyde/urea as in the legend to Fig. III.8, in 10% polyacrylamide gels containing 8 M urea. The centre of the band of bromophenol blue (BPB) marker is indicated. Technical details were as described in METHODS.


(a)
BPB
1600
800
(b)
BPB
400
200
(c)
BPB
8000
4000
0
(d)
BPB
2000
1000
0
Radioactivity (disint/min)
10
20
30
40
50
60
Slice number

of these RNA species prior to disruption of secondary structure (Fig. III.8.a), these values constitute a reduction in mobility of about 40% for 4S RNA and about 47% for 5S RNA. In contrast, the electrophoretic mobility of neither [^{3}H] poly(U) nor [^{3}H] poly(C) in 10% polyacrylamide gels is significantly altered by the formaldehyde/urea treatment (Fig. III.9), confirming the lack of secondary structure of these polymers at room temperature (Michelson et al., 1967). It is therefore necessary to take the electrophoretic mobility of fully extended 4S and 5S RNA species, as standards for estimation of size of the poly(U) and poly(C) fragments eluted from hybrids (Fig. III.7). The average size of 4S RNA is about 80 nucleotides long (Holmquist et al., 1973) and 5S RNA is 120 nucleotides long (Brownlee et al., 1968). On the assumption that electrophoretic mobility is proportional to the log (molecular weight) (Loening, 1969), it can be calculated that poly(U) eluted from a hybrid treated with 5 µg RNase/ml for 30 min ranges in size from 130 to 25 nucleotides long, the average size being about 50 nucleotides long. After a more vigorous RNase digestion (20 µg/ml for 1 hour) the size of the [^{3}H] poly(U) eluted is reduced to about 66 to 20 nucleotides long.

The [^{3}H] poly(C) eluted from hybrids with BHK-21/C13 DNA is considerably smaller than the [^{3}H] poly(U) (Fig. III.7). Again, the more vigorous the digestion with RNase, the smaller the fragments produced. The estimated range of sizes of eluted [^{3}H] poly(C) is 39 to 13 nucleotides long, falling to less than 28 nucleotides long after more vigorous RNase treatment.

The results are consistent with the hypothesis that increased RNase concentration leads to digestion of [^{3}H] poly(U) and [^{3}H] poly(C) at certain mismatched regions within the hybrids. The overall size of the dA-rich and dG-rich regions cannot therefore be estimated in this way. The range of sizes of poly(U) and poly(C) eluted, corresponding to sequence lengths of up to about 130 nucleotides long and up to about 40 nucleotides long respectively, therefore only represent the minimum lengths of dA-rich and dG-rich sequences in the DNA.

(2) Size of poly(dA) and poly(dG) regions

Similarly, the size of the [^{3}H] poly(U) and [^{3}H] poly(C) fragments which remain hybridised after exhaustive RNase digestion (20 µg/ml at 20°C for 24 hours) gives a measure of the size of the poly(dA) and poly(dG) sequences in the DNA (Fig. III.10). As expected, these sequences are shorter than the corresponding dA-rich and dG-rich regions, the mean size of the poly(dA) sequences being about 31 nucleotides long (range 23 to 60), the poly(dG) sequences averaging about 17 nucleotides long (range 11 to 29). It should perhaps be noted that a shoulder is consistently found in the trailing edge of the profile of [^{3}H] poly(U) eluted from the hybrids with poly(dA) sequences (Fig. III.10 and III.21). This suggests that there may be two fairly homogeneous populations of poly(dA) sequences in BHK-21/C13 DNA, the major population having a mean size of about 30 nucleotides long, the minor population being slightly larger with a mean size of about 50 nucleotides long. Alternatively, this minor peak might represent poly(U) eluted from hybrids

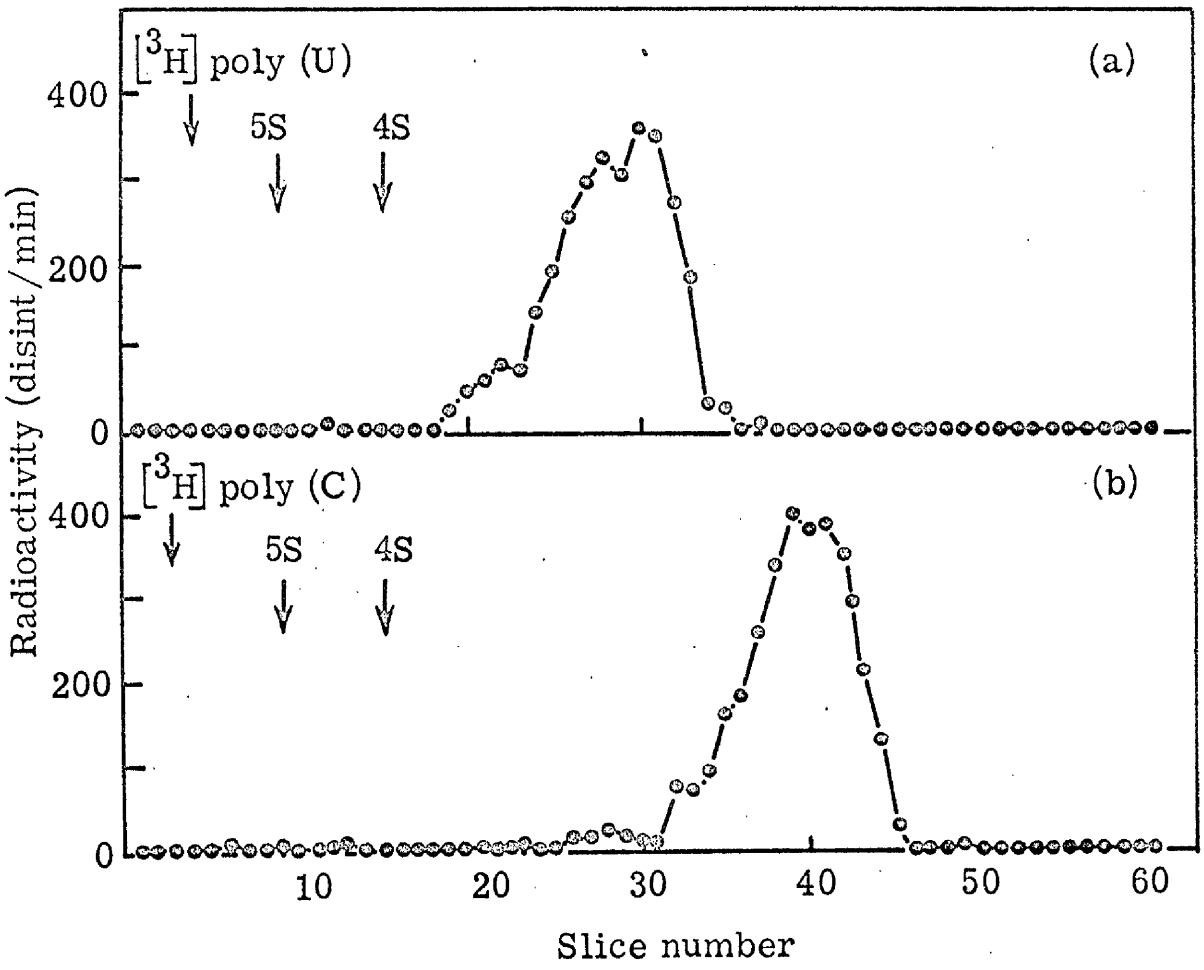


Fig. III.10.

Estimation of size of poly(dA) and poly(dG) sequences

Hybrids were formed between (a) 100 µg filter-bound DNA and 10 µg [^{3}H] poly(U) or (b) 1 mg filter-bound DNA and 20 µg [^{3}H] poly(C), treated with 20 µg RNase/ml for 24 hr at 20°C, and the size of the ribohomopolymers eluted from the filters was estimated by gel electrophoresis as in METHODS and in the legend to Fig. III.7.

in which for some unknown reason not all mismatched base-pairs have been attacked by RNase.

(E) Investigation of the accuracy of base-pairing within the hybrids

(1) Melting temperatures

In view of these results, an estimate of the degree of mismatching within the DNA·poly(U) and DNA·poly(C) hybrids would be valuable. One of the most useful criteria in assessing the extent of non-complementary base pairs within a hybrid is its Tm, or mean thermal dissociation (McCarthy & Church, 1969). The Tm for the BHK-21/C13 DNA.poly(U) hybrid after gentle RNase treatment is 34-35°C (Fig. III.11.a). This would appear to be considerably lower than the Tm of the pure poly(dA).poly(U) hybrid. However, as Thomas & Dancis (1973) have pointed out, the Tm of a hybrid can be related to its length, by the expression $b\Delta Tm = 820°C$ where b is the length of the duplex. Our data suggests that this relationship holds for the deoxyhomopolymer·ribohomopolymer hybrids. Thus poly(dA).poly(U) hybrids, about 600 nucleotides long, melt at 53°C in 0.2 M Na^+ ion (Riley et al., 1966). The expected Tm of poly(dA).poly(U) of infinite length would therefore be 54°C. In our studies, poly(dA) about 100 nucleotides long was used. From the above relationship, the expected fall in Tm would be 8°C, i.e. a Tm of 46°C. This was, indeed, the value which was found (Fig. III.11.a). Further confirmation that this relationship holds for short hybrids comes from our inability to detect significant amounts of hybrid shorter than 25 nucleotides long (Figs. III.7 and III.10). Such hybrids would be expected to be unstable at 20°C. The average size

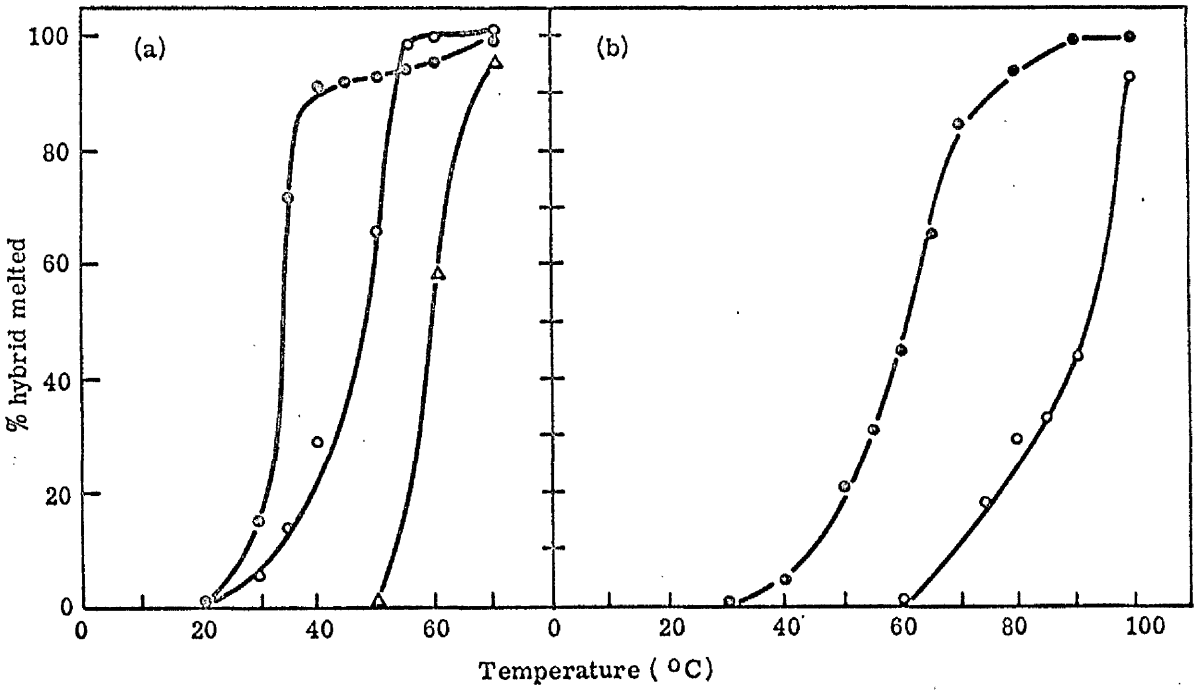


Fig. III.11.

Melting curves of DNA·poly(U) and DNA·poly(C) hybrids after gentle RNase treatment

Hybrids of DNA or homopolymer and saturating amounts of [^{3}H] poly(U) or [^{3}H] poly(C) were formed under standard conditions and dissociated as described in the METHODS section. Hybrids were treated at 20°C for 30 min with (a) 5 µg RNase/ml, (b) 10 µg RNase/ml.

a) Dissociation in 1 x SSC of BHK-21/C13 DNA·[^{3}H] poly(U), ●——●; poly dA·[^{3}H] poly(U), o——o; poly(rA)·[^{3}H] poly(U), Δ——Δ.

b) Dissociation in 2 x SSC of BHK-21/C13 DNA·[^{3}H] poly(C), ●——●; poly(dG)·[^{3}H] poly(C), o——o.

of poly(U) fragments eluted from hybrids after gentle RNase digestion is about 50 nucleotides long (Fig. III.7). Using the above relationship, a pure poly(dA).poly(U) hybrid 50 nucleotides long would be expected to melt at 38°C. Our value of 34-35°C implies a $\triangle$Tm of about 4°C. The exact effect of mismatched base pairs on the Tm of the hybrid is still uncertain, but current evidence suggests that each 1% of mismatching leads to a fall in Tm of between 0.7°C and 2.2°C (Ullman & McCarthy, 1973). The DNA.poly(U) hybrids, stable to digestion with 5 µg RNase/ml for 30 min, are therefore probably on average between 2-6% mismatched. The exhaustively digested DNA.poly(U) hybrid is about 31 nucleotides long, and if pure, would be expected to melt at about 28°C. In fact, the Tm was found to be 29°C (Fig. III.12.a). This result confirms that after exhaustive RNase treatment, poly(U) is hybridised to apparently pure sequences of poly(dA) in DNA.

Furthermore, the Tm of the BHK-21/C13 DNA.poly(C) hybrid, which averages 25 nucleotides long after mild RNase digestion is about 61°C (Fig. III.11.b). The anticipated Tm for a pure poly(dG).poly(C) hybrid of this length is 82°C in 2 x SSC, as calculated from the data of Chamberlin (1965), correcting for salt concentration (Tm = 16.6 log M + 0.41 [% G + C] + 81.5, for M $<$ 0.5 molar, (Schildkraut & Lifson, 1965)) and size (Thomas & Dancis, 1973). Depending upon the $\triangle$Tm for 1% mismatching, it can be calculated that these hybrids are highly mismatched, to an extent of 10-30%. On the other hand, pure poly(dG).poly(C) hybrids of average size 17 nucleotides long would be expected to have a Tm of

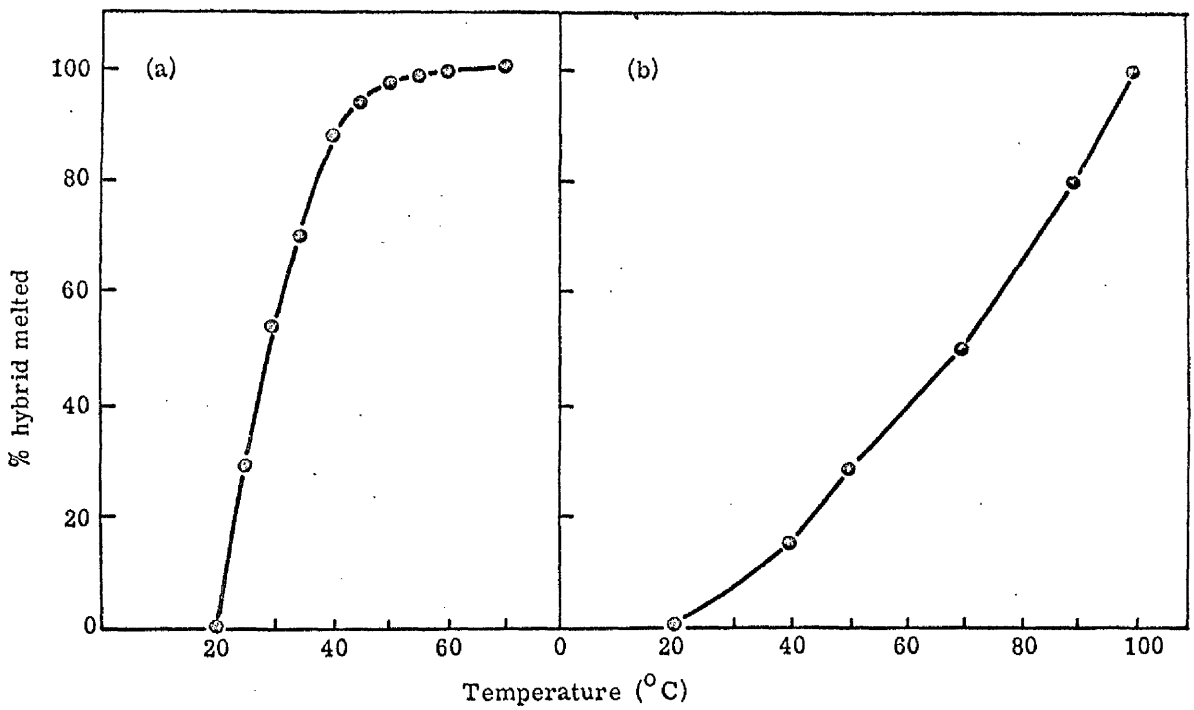


Fig. III.12.

Melting curves of DNA·poly(U) and DNA·poly(C) hybrids after exhaustive RNase treatment

Hybrids were formed between 400 µg of DNA, fixed on nitrocellulose filters, and 20 µg of [^{3}H] poly(U) or [^{3}H] poly(C) under optimal conditions. Filters were treated with RNase 20 µg/ml at 20°C for 24 hours. Dissociation of hybrids on (a) 2 filters, (b) 3 filters was performed at various temperatures, as described in METHODS and the radioactivity remaining on the filter was estimated in each case.

a) Dissociation in 1 x SSC of BHK-21/C13 DNA·[^{3}H] poly(U) hybrid

b) Dissociation in 2 x SSC of BHK-21/C13 DNA·[^{3}H] poly(C) hybrid.

67°C, using the above relationships. The exhaustively digested DNA.poly(C) hybrids have a Tm of about 69°C (Fig. III.12.b), confirming the presence of pure poly(dG) sequences in the DNA.

These results are therefore consistent with the findings already presented (Results, Sections III.B and III.D) which suggested that when a gentle digestion with RNase is performed, a significant degree of mismatching remains in the hybrid, whereas exhaustive treatment with RNase leads to the removal of mismatched base-pairs.

(2) Hybridisation of poly(U) and poly(C) with synthetic deoxyhomopolymers

A further test of whether the uridine residues of poly(U) are hybridised only to deoxyadenosine in DNA is to investigate whether poly(U) can form a stable complex with any deoxyhomopolymer other than poly(dA) (Table III.2).

It is interesting to note that the hybrid formed between poly(dA) and [^{3}H] poly(U) is not assayed by the standard nitrocellulose filter binding technique, although it can be readily detected as an acid insoluble, RNase resistant complex (Table III.2). The failure of the poly(dA).poly(U) hybrid to bind to nitrocellulose filters is due to the fact that poly(dA) itself does not bind to such filters (Table III.3). Poly(dA) is therefore an unusual example of a single-stranded DNA molecule which does not bind to nitrocellulose in the presence of a high salt concentration. On the other hand, the duplex poly(dA).poly(dT) is capable of binding to nitrocellulose filters and the triplet complex (poly(dA).poly(dT)).poly(U)

Table III.2. Hybridisation of [^{3}H] poly(U) to synthetic deoxypolymers

	[^{3}H] poly(U) cts/min on filter	
	Filtration through nitrocellulose membranes then RNase treatment	Treatment with RNase in solution then acid precipitation
poly(dA)	65	5040
poly(dA).poly(dT)	1927	2293
poly(dT)	0	250
poly [d(A-T)] (denatured)	0	19
poly(dC)	0	9
poly(dG)	7	35

0.2 µg [^{3}H] poly(U) was incubated with 0.2 µg deoxypolymer at 20°C for 4 hours in 1 ml 2 x SSC. Poly[d(A-T)] was heated to 100°C for 5 min and [^{3}H] poly(U) was added at 100°C, prior to incubation at 20°C. Hybrids were estimated in two different ways as shown. RNase treatment was with 5 µg/ml for 30 min at 20°C.

can be quantitatively measured in this way. It would seem likely, therefore, that hybrids between DNA and poly(U) can only be detected by the nitrocellulose filter technique, by virtue of the binding to the filters of DNA polynucleotide sequences adjacent to the dA-rich regions, rather than direct binding of the dA-rich regions themselves.

Table III.3. Binding of nucleic acid homopolymers to nitrocellulose filters

	Initial A_{260}	A_{260} after passing through nitrocellulose filters	% binding to filter
poly(dA)	0.12	0.12	0
poly(dA).poly(dT)	0.16	0.00	100
poly(dT)	0.115	0.00	100
poly(dG)	0.15	0.02	87
poly(dC)	0.105	0.03	71
poly(A)	0.21	0.02	91
poly(U)	0.18	0.12	33
poly(G)	0.23	0.14	39
poly(C)	0.21	0.18	14

2 ml aliquots of solutions of nucleic acid polymers, at a concentration of 5-10 μg/ml in 2 x SSC, were passed through 2.5 cm nitrocellulose filters, at a rate of 2 ml/min. The presence of 0.005 M Mg^{++} had no effect on binding of the polymers.

However, the other single-stranded homopolymeric deoxynucleotides tested, poly(dT), poly(dG) and poly(dC), all of which can bind to nitrocellulose filters (Table III.3), gave

no detectable hybridisation with poly(U) as assayed by the nitrocellulose method, and very little or no significant hybridisation as assayed by the formation of RNase resistant complexes (Table III.2). The copolymer poly[d(A-T)] also did not hybridise to [^{3}H] poly(U). Thus, under the hybridisation conditions used, stable complexes between [^{3}H] poly(U) and deoxypolymers other than poly(dA) were not detected, testifying to the specificity of hydrogen bonding between uridine and deoxyadenosine.

Similarly, significant hybrid formation between poly(C) and deoxyhomopolymers is only observed with poly(dG), although a small amount of RNase resistant complex is formed with poly(dC) (Table III.4). Thus stable hydrogen-bonded base-pairs are probably only formed between cytosine and deoxyguanosine under the conditions used.

Table III.4. Hybridisation of [^{3}H] poly(C) to deoxyhomopolymers

	[^{3}H] poly(C) cts/min on filter	
	Filtration through nitrocellulose membranes then RNase treatment	Treatment with RNase in solution, then acid precipitation
poly(dA)	0	0
poly(dT)	0	19
poly(dC)	23	86
poly(dG)	607	865

0.2 µg [^{3}H] poly(C) was incubated with 0.2 µg deoxyhomopolymer at 62°C for 30 min in 1 ml 2 x SSC. Hybrids were estimated in two different ways as shown. RNase treatment was with 10 µg/ml for 30 min at 20°C.

(3) Competition of hybridisation of [^{3}H] poly(U) and [^{3}H] poly(C) by other ribohomopolymers

Another way of looking for non-specific hybridisation of poly(U) or poly(C) to sequences within DNA other than poly(dA) or poly(dG) respectively, is to determine if the hybridisation of the ribohomopolymer can be affected by the presence of other ribohomopolymers. For example, if poly(U) could complex to poly(dG) or dG-rich sequences in DNA, then it would be expected that poly(C) would greatly reduce the extent of hybridisation of poly(U) to DNA. It was observed

Table III.5. Effects of ribohomopolymers on hybridisation of [^{3}H] poly(U) and [^{3}H] poly(C) to BHK-21/C13 DNA

Addition	[^{3}H] poly(U) cts/min	[^{3}H] poly(C) cts/min
nil	152	427
poly(U)	22	448
poly(C)	160	30
poly(G)	157	855
poly(A)	3320	474

Standard hybridisation reactions were carried out on mixtures containing 1 µg unlabelled ribohomopolymer plus (a) 1 µg denatured BHK-21/C13 DNA with 0.2 µg [^{3}H] poly(U), or (b) 10 µg denatured BHK-21/C13 DNA with 0.2 µg [^{3}H] poly(C). Hybrids were collected and RNase treated on nitrocellulose filters as in METHODS.

(Table III.5) that although unlabelled poly(U) effectively competes out the hybridisation of [^{3}H] poly(U) to DNA, neither poly(C) nor poly(G) had any significant effect.

The greatly increased level of [^{3}H] poly(U) collected on nitrocellulose filters after addition of poly(A) is due to the ability of poly(A) alone, and also of poly(A).poly(U) complexes to bind to such filters (Table III.3, and Burdon & Shenkin, 1972). It is therefore conceivable that some of the [^{3}H] poly(U) could have hybridised to dT-rich sequences in the DNA, since a competition experiment with poly(A) is clearly impracticable. This, however, is unlikely since poly(U) does not hybridise directly to pure poly(dT) (Table III.2).

Similarly, hybridisation of [^{3}H] poly(C) to DNA is not significantly affected by the presence of excess poly(U) or poly(A), although addition of poly(G) leads to an apparent increase in [^{3}H] poly(C) hybridised (Table III.5). This is probably due to the somewhat limited affinity of poly(G) itself to bind to nitrocellulose filters (Table III.3). Thus no further conclusion can be drawn regarding the extent of hybridisation of [^{3}H] poly(C) to dC-rich regions within the DNA.

In summary, these experiments (Results, Sections III.E.(2) and (3)) indicate that under the conditions used for the formation and assay of hybrids, stable complexes are only detected of poly(U) with dA-rich or poly(dA) regions, or of poly(C) with dG-rich or poly(dG) regions (and possibly to a slight extent to dC-rich regions). It is most improbable that within the dA-rich sequences hybridised by poly(U), there should also exist regions which are dT-rich, dG-rich or dC-rich. Furthermore, the dG-rich regions hybridised by poly(C) probably contain no dA-rich or dT-rich sequences;

although a small level of dC-rich sequences may be present. These results, of course, do not preclude the possibility that short stretches of mismatching with two or three different bases may occur within regions of otherwise accurately matched base-pairs of dA.U or dG.C.

(F) Distribution of dA-rich and dG-rich regions within DNA molecules

(1) dA-rich and dG-rich regions in DNA molecules of different base composition

Since both dA-rich and dG-rich regions represent quite a significant proportion of the total DNA, it is important to determine whether these regions have a widespread distribution throughout DNA molecules or whether they are localised to only a few molecules. One way of attempting to answer this question is to determine whether any DNA satellite of abnormal base composition exists, or can be produced by shearing, which is enriched in the ability to form hybrids with [^{3}H] poly(U) or [^{3}H] poly(C). After banding the native DNA to equilibrium in a CsCl density gradient, fractions from the gradient were denatured, bound to nitrocellulose filters and assayed for ability to hybridise with ribohomopolymer by the batch method of Birnstiel et al. (1968). Both [^{3}H] poly(U) and [^{3}H] poly(C) were found to hybridise to DNA fractions from all parts of the gradient (Fig. III.13). Even after extensive shearing of the DNA, the hybridisation profile closely follows the DNA absorbance profile, suggesting that both the dA-rich and the dG-rich regions are distributed throughout DNA molecules of all base compositions. In the case of the dG-rich regions,

Fig. III.13

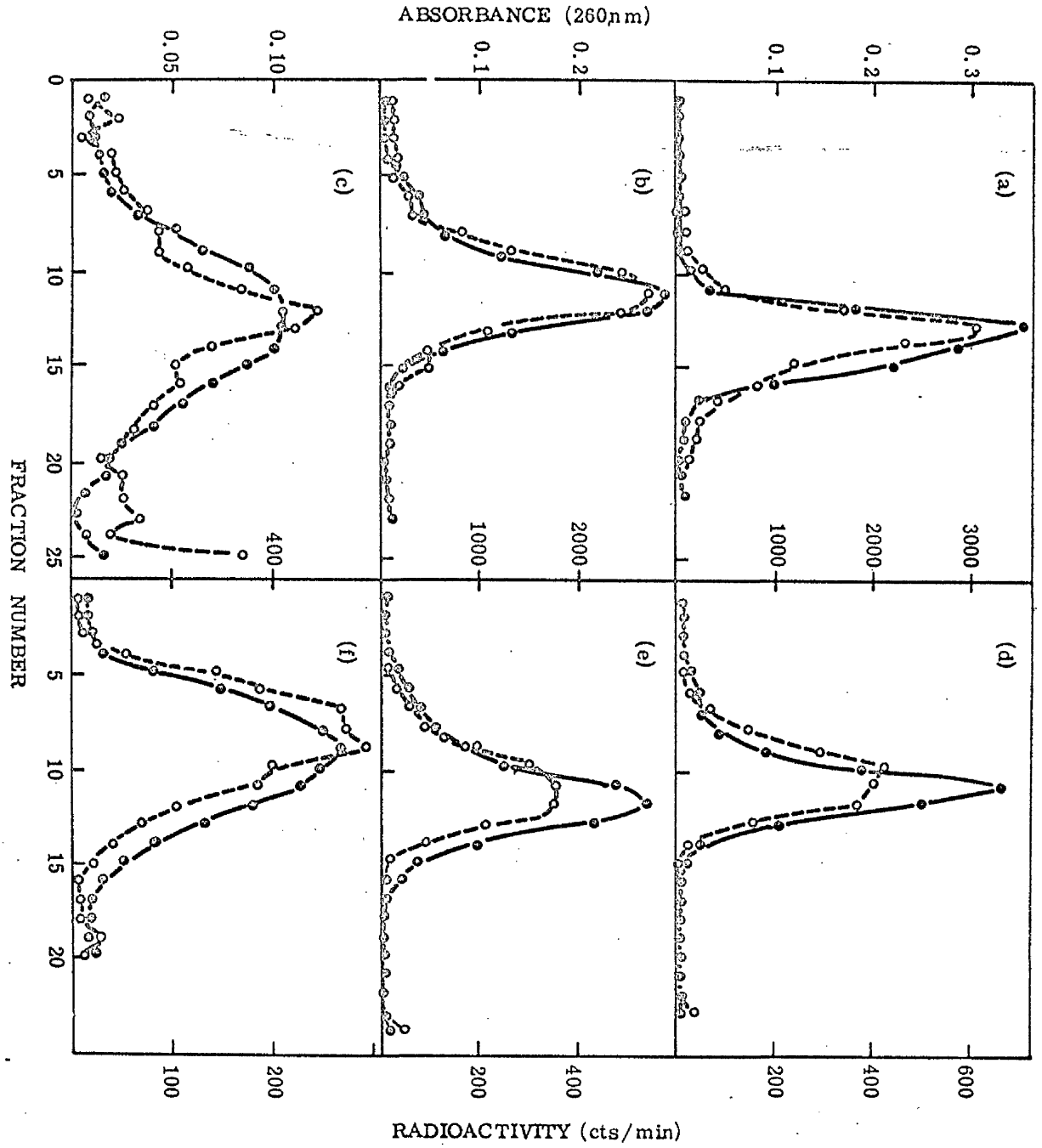
Distribution of dA-rich and dG-rich regions within CsCl density gradient fractions of BHK-21/C13 DNA

100 μg DNA in 0.1 x SSC (0.5 ml) was added to 4.5 ml CsCl solution (density 1.78 g/cm^3) and the final density was adjusted to 1.700 g/cm^3. The mixture was centrifuged at 33000 rev/min at 20°C in a Spinco angle 40 rotor for 64 hours. Six-drop fractions were collected into 0.5 ml 0.1 x SSC, the absorbance (260 nm) was measured, and the DNA was denatured and bound to nitrocellulose filters. All the filters from one gradient were then incubated with either 2 μg [^{3}H] poly(U) ((a), (b) and (c)) or 2 μg [^{3}H] poly(C) ((d), (e) and (f)), under optimal conditions for hybrid formation, and filters were washed and RNase treated by the batch method ((a), (b) and (c), 5 μg RNase/ml; (d), (e) and (f), 10 μg RNase/ml). Technical details are described in METHODS.

(a) and (d) - unsheared DNA (average single-stranded mol. wt. = 20 x 10^6).

(b) and (e) - DNA prepared by the hydroxylapatite method (average single-stranded mol. wt. = 1.1 x 10^6).

(c) and (f) - sonicated DNA (average single-stranded mol. wt. = 1.2 x 10^5). Density increases from right to left. Absorbance 260 nm, o——o; radioactivity, o----o.


ABSORBANCE (260nm)
RADIOACTIVITY (cts/min)
FRACTION NUMBER
(a)
(b)
(c)
(d)
(e)
(f)

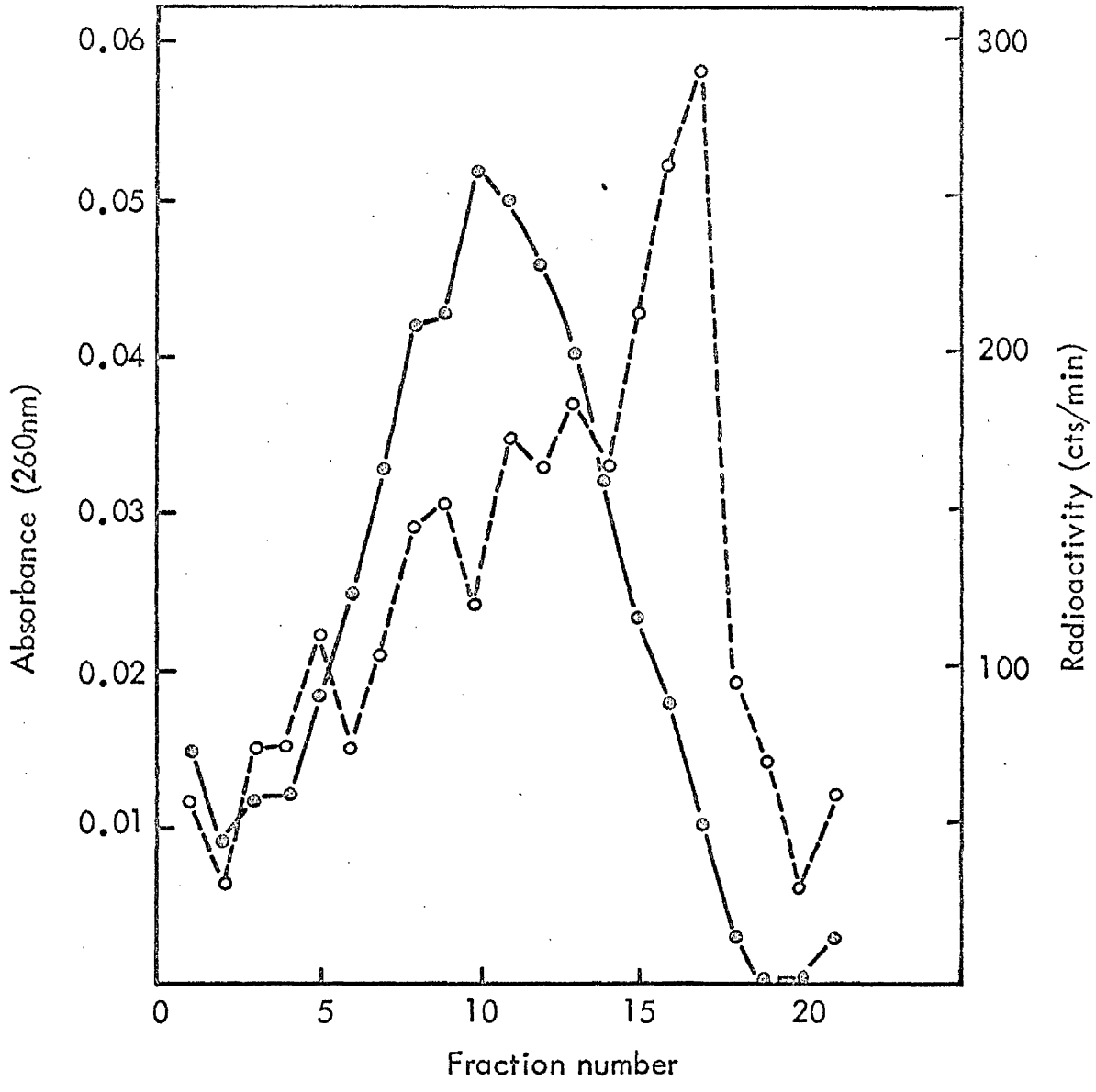
a small degree of skewing of the hybridisation profile towards the G + C rich region of the gradient was observed, this being slightly accentuated in sheared DNA preparations. This result would be consistent with the occurrence of dG-rich sequences in DNA molecules of all base compositions, those molecules which are G + C rich being only slightly enriched for the dG-rich regions.

The apparent absence of a large dA-rich satellite in the A + T rich fraction of even highly sheared DNA is surprising. Only a very small satellite would seem to be present, but since the dA-rich regions are up to 130 nucleotides long (Fig. III.7) it might be expected that in DNA molecules only 400 nucleotides long, the bulk of the DNA molecules with a long dA-rich region within their polynucleotide sequence, would be found in the A + T rich region on the gradient. The probable reason for this paradox is observed in the properties of the hybrid formed between poly(dA) and [^{3}H] poly(U). Such a hybrid binds only to a very small extent to nitrocellulose filters, although it is readily detectable by treatment in solution with RNase followed by acid precipitation (Table III.2). As already pointed out (Results, Section III.E.2), this is due to the inability of poly(dA) itself to bind to nitrocellulose filters. If DNA is extensively sheared, some fragments may be produced which are so dA-rich that they fail to bind to nitrocellulose filters. The ability of these fragments to hybridise to poly(U) would therefore be undetected. This possibility was tested by hybridising the CsCl density gradient fractions of sonicated DNA in solution with [^{3}H] poly(U) and collecting the RNase resistant radioactivity by

Fig. III.14.

Hybridisation of [^{3}H] poly(U) to CsCl density gradient fractions of sonicated BHK-21/C13 DNA

25 μg BHK-21/C13 DNA, sonicated to average single-stranded mol. wt. 1.2 x 10^5, was banded to equilibrium in 5 ml CsCl and fractions collected into 0.1 x SSC as in Fig. III.13. After measurement of absorbance, fractions were dialysed against three changes of 5 l. 2 x SSC. 1 μg [^{3}H] poly(U) (Schwartz-Mann) in 0.1 ml 2 x SSC was added to each dialysate and the final volume was adjusted to 1 ml 2 x SSC. Samples were heated to 100°C for 2 min then rapidly cooled and incubated at 20°C for 4 hours. Samples were digested with 5 μg RNase/ml for 30 min at 20°C, and then precipitated with 5% trichloracetic acid. ●——●, absorbance; o----o, radioactivity. Density increases from right to left.


0.06
0.05
0.04
0.03
0.02
0.01
Absorbance (260nm)
300
200
100
Radioactivity (cts/min)
0
5
10
15
20
Fraction number

acid-precipitation (Fig. III.14). By this assay procedure it was demonstrated that although the dA-rich regions are found in DNA molecules of all base compositions, extensive shearing of the DNA does, indeed, lead to the production of a very A + T rich DNA satellite which is highly enriched for the dA-rich regions.

Poly(dG) and poly(dG).poly(C) hybrids do bind to nitrocellulose filters (Tables III.3 and III.4), therefore the comparable experiment with sonicated DNA and [^{3}H] poly(C) was not performed.

It can therefore be seen that both dA-rich and dG-rich regions are distributed throughout DNA molecules of a wide variety of base compositions, even after extensive shearing of the DNA. Such shearing does, however, lead to the production of DNA fragments, some of which are enriched for dA-rich regions. On the other hand, some enrichment for dG-rich regions is found in the G + C rich fraction of both extensively sheared and relatively unsheared DNA.

(2) Distribution of poly(dA) and poly(dG) sequences within DNA molecules of different base composition

In a similar way, the distribution of poly(dA) and poly(dG) sequences in DNA molecules can be investigated. Both the poly(dA) and poly(dG) sequences can be detected in DNA molecules of a wide variety of base composition (Fig. III.15). With the DNA preparation used, which had been sheared to a single-stranded molecular weight of 1.1×10^6, it is possible that some enrichment for poly(dA) sequences is present in A + T rich DNA. However, the low level of radioactivity present in the hybrid prevents a definite conclusion being reached on this point. As was observed with

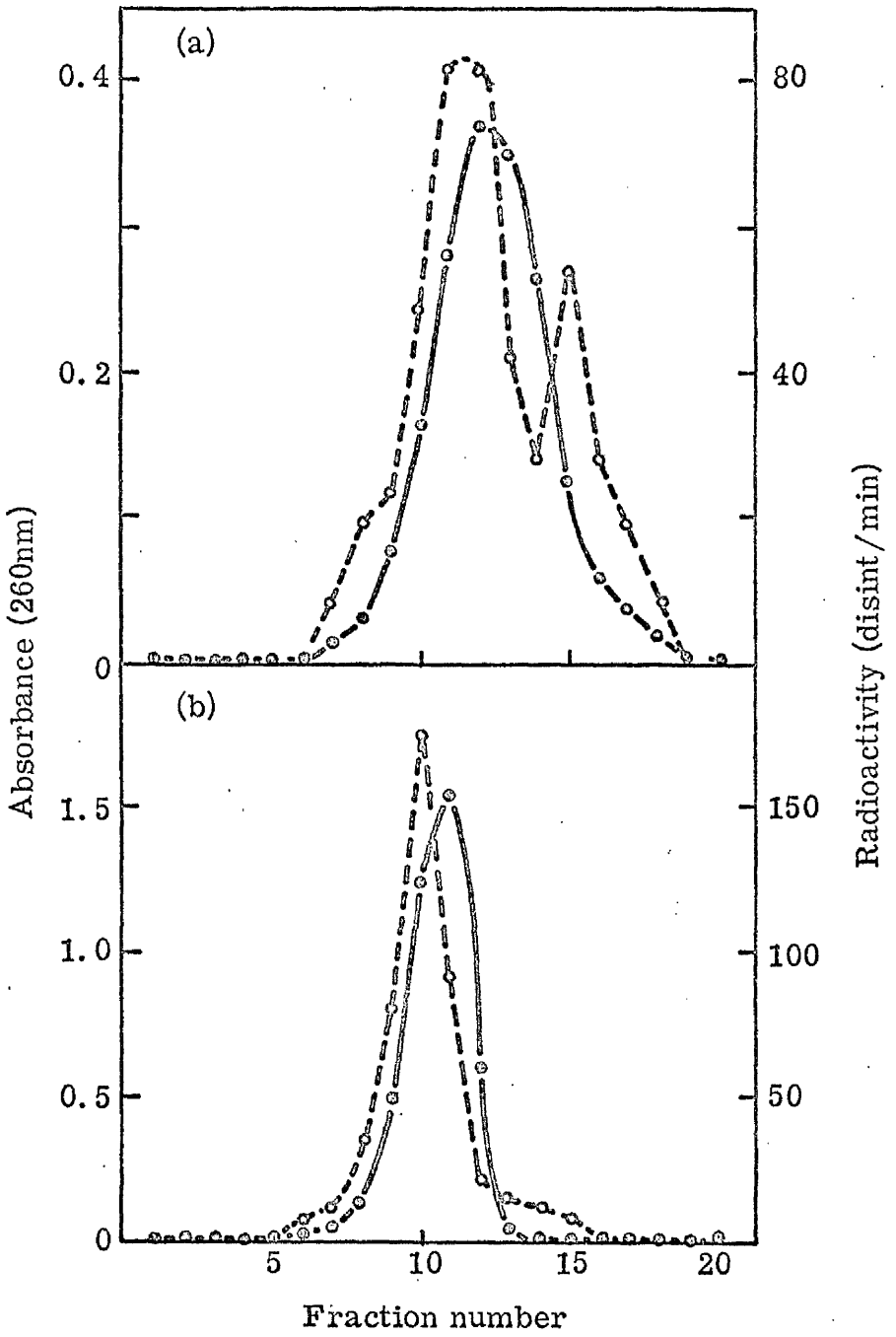
Fig. III.15.

Distribution of poly(dA) and poly(dG) regions within CsCl gradient fractions of BHK-21/C13 DNA

(a) 100 μg DNA and (b) 400 μg DNA, of average single-stranded mol. wt. 1.1×10^6 was centrifuged to equilibrium in CsCl, average density 1.700 g/cm^3, and fractionated as in the legend to Fig. III.13. After measurement of absorbance (260 nm), fractions were bound to nitrocellulose filters, fractions from (b) being further divided so that no more than 20 μg DNA was applied to any one filter. All the filters from one gradient were then incubated with either (a) 20 μg [^{3}H] poly(U) or (b) 10 μg [^{3}H] poly(C) under optimal conditions for hybrid formation, and filters were washed and RNase treated (20 μg/ml for 24 hours at 20°C) by the batch method. Technical details are described in METHODS.

Absorbance 260 nm, ●——●; radioactivity, o----o.

Density increases from right to left.


(a)
(b)
0.4
0.2
0
1.5
1.0
0.5
0
80
40
150
100
50
5
10
15
20
Absorbance (260nm)
Radioactivity (disint/min)
Fraction number

the dG-rich regions, the poly(dG) sequences were found to be slightly enriched in the G + C rich fractions of the DNA.

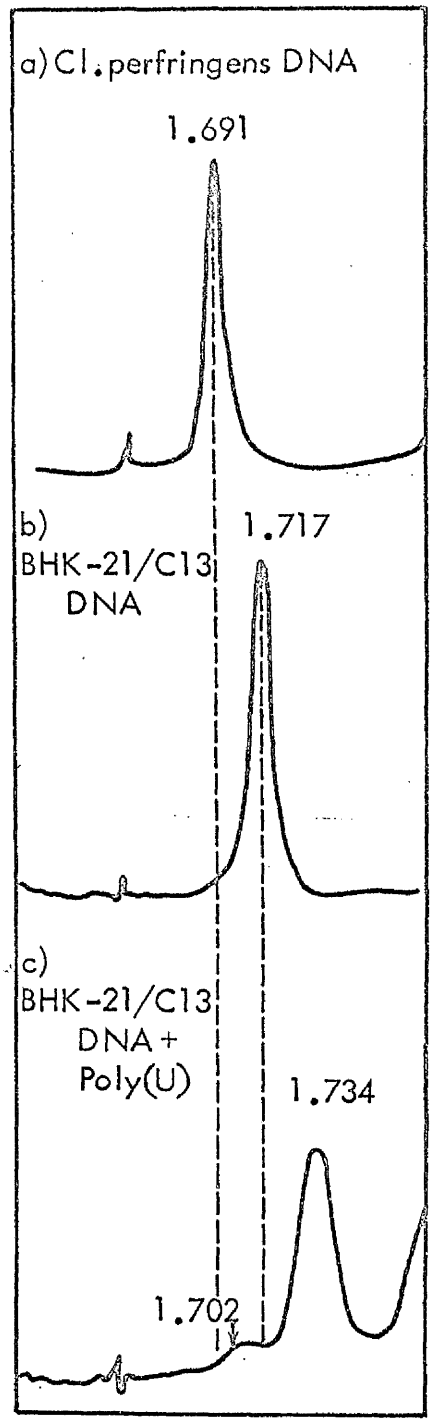
(3) Attempts to separate hybrid molecules from unhybridised DNA

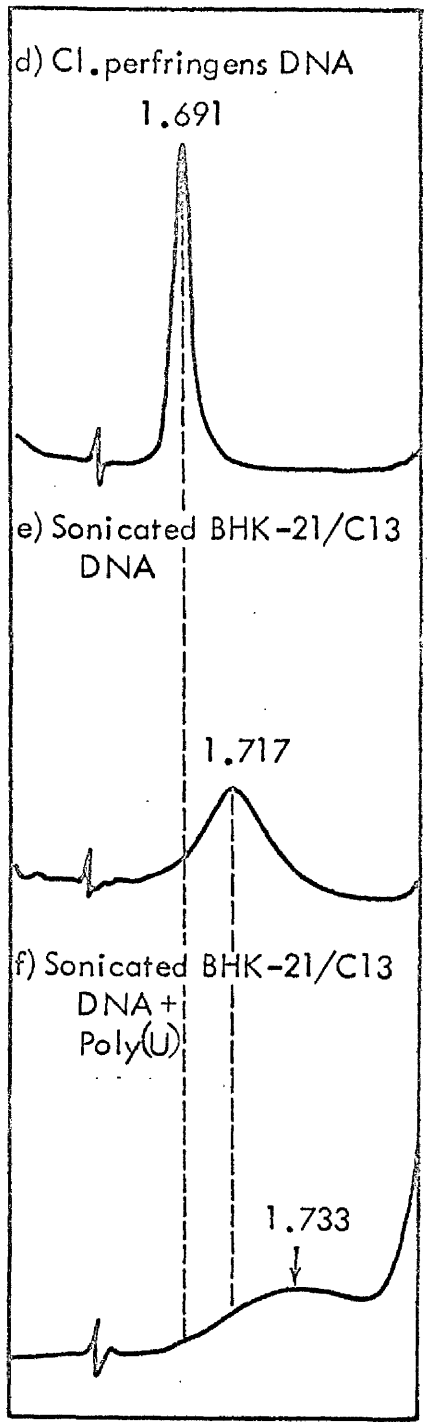
Although both the dA-rich and dG-rich regions are distributed throughout DNA molecules with a wide range of base compositions, it is uncertain what proportion of DNA molecules of any given base composition contain such a region. For example, if the dA-rich regions were found in all DNA molecules, approximately 0.13% of each molecule would be dA-rich. On the other hand, only a small number of DNA molecules may contain dA-rich regions. If this were the case, then the dA-rich regions would comprise a relatively large proportion of those DNA molecules. It might then be possible to separate those DNA molecules which had hybridised poly(U) from those which had not, by centrifugation in CsCl. Initially, an attempt was made to visualise such a separation in the analytical ultracentrifuge (Fig. III.16). In order to maximise the density increment due to the binding of poly(U), no RNase treatment was performed. The bulk of the unsheared denatured DNA was observed to have bound poly(U), a new symmetrical peak being observed at density 1.734 g/cm^3 (Fig. III.16.c). The small peak at density 1.702 g/cm^3 probably represents rapidly reassociated DNA. However, by this procedure, even after shearing the DNA to an average single-stranded length of 400 nucleotides, the bulk of the DNA would still seem to have complexed poly(U) (Fig. III.16.f). Since a maximum of 0.4% of the DNA consists of dA-rich regions, the complexes

Fig. III.16.

Effect of poly(U) on buoyant density profile of denatured BHK-21/C13 DNA

50 µg denatured DNA, either unsheared (average single-stranded mol. wt. 20 x 10^6) or sonicated (average single-stranded mol. wt. 1.2 x 10^5) was incubated with 50 µg poly(U) in 1 ml 2 x SSC under standard conditions (see METHODS). 0.1 ml aliquots of incubation mixtures or of control DNAs were added to 0.8 ml saturated CsCl and the average density was adjusted to 1.720 g/cm^3. Samples were centrifuged at 44,000 rev/min for 16 hours at 25°C in the model E analytical ultracentrifuge. Cells were scanned automatically at 265 nm, and the absorbance profile was recorded. Cl.perfringens DNA was centrifuged separately since even in its native form it interacts with poly(U) leading to a broadening of the density profile.


a) Cl. perfringens DNA
1.691
b)
BHK-21/C13
DNA
1.717
c)
BHK-21/C13
DNA +
Poly(U)
1.734
1.702


d) Cl. perfringens DNA
1.691
e) Sonicated BHK-21/C13
DNA
1.717
f) Sonicated BHK-21/C13
DNA +
Poly(U)
1.733

observed in the absence of RNase digestion must therefore be extremely non-specific. Indeed, digestion of the complex with RNase (5 µg/ml for 30 min at 20°C) prior to centrifugation led to the complete absence of the new bands, no significant increase in the density of the denatured DNA being observed (results as for Fig. III.16.b and e).

Although it was not possible to visualise a separation of true hybrids from non-hybridised DNA in this way, it seemed likely that a minor degree of separation might be demonstrated using the ^{3}H-labelled polyribonucleotides, and determining the relationship of the radioactivity profile to the DNA absorbance profile. For BHK-21/C13 DNA·[^{3}H] poly(U) hybrids, it was indeed observed that the progressive decrease in size of the DNA led to a shift in the hybrid molecules away from the bulk of the DNA towards heavier buoyant densities (Fig. III.17). The comparable experiment with BHK-21/C13 DNA·[^{3}H] poly(C) hybrids demonstrated that even with unsheared DNA the buoyant density profile of the hybrid molecules was to the heavy side of the DNA profile (Fig. III.18). In contrast to the result with [^{3}H] poly(U), shearing of the DNA prior to hybridisation with [^{3}H] poly(C) did not lead to the majority of hybrid molecules becoming significantly more dense with respect to the unhybridised DNA (Fig. III.18). However, with the most highly sheared DNA preparation (Fig. III.18.e) some skewing of the hybridisation profile is observed, indicating the formation of some hybrid molecules of particularly high density.

These results suggest that dA-rich and dG-rich regions are not clustered within short regions in particular DNA

Fig. III.17.

Isopycnic centrifugation of BHK-21/C13 DNA·[^{3}H] poly(U) hybrids in CsCl .

10 µg [^{3}H] poly(U) (Schwartz-Mann) was hybridised with 100 µg denatured DNA of five different mean molecular weights in 2.2 ml 2 x SSC. An equal volume of 2 x SSC containing 20 µg RNase/ml was added, and incubation at 20°C continued for 1 hour. Solid CsCl was added until the final density was 1.730 g/cm^3. After centrifugation at 33,000 rev/min for 64 hours, 6 drop fractions were collected. 10 µl of each sixth fraction was removed for measurement of refractive index. 0.5 ml 0.1 x SSC was added, the absorbance of each fraction was measured and fractions were incubated with ice-cold 5% trichloracetic acid for 30 min. Acid precipitable radioactivity was then assayed. Mean single-stranded molecular weights (a) 20 x 10^6, (b) 1.1 x 10^6, (c) 3 x 10^5, (d) 2 x 10^5, (e) 6 x 10^4. ●——●, absorbance 260 nm; o----o, [^{3}H] poly(U) acid precipitable radioactivity; △——△, density.

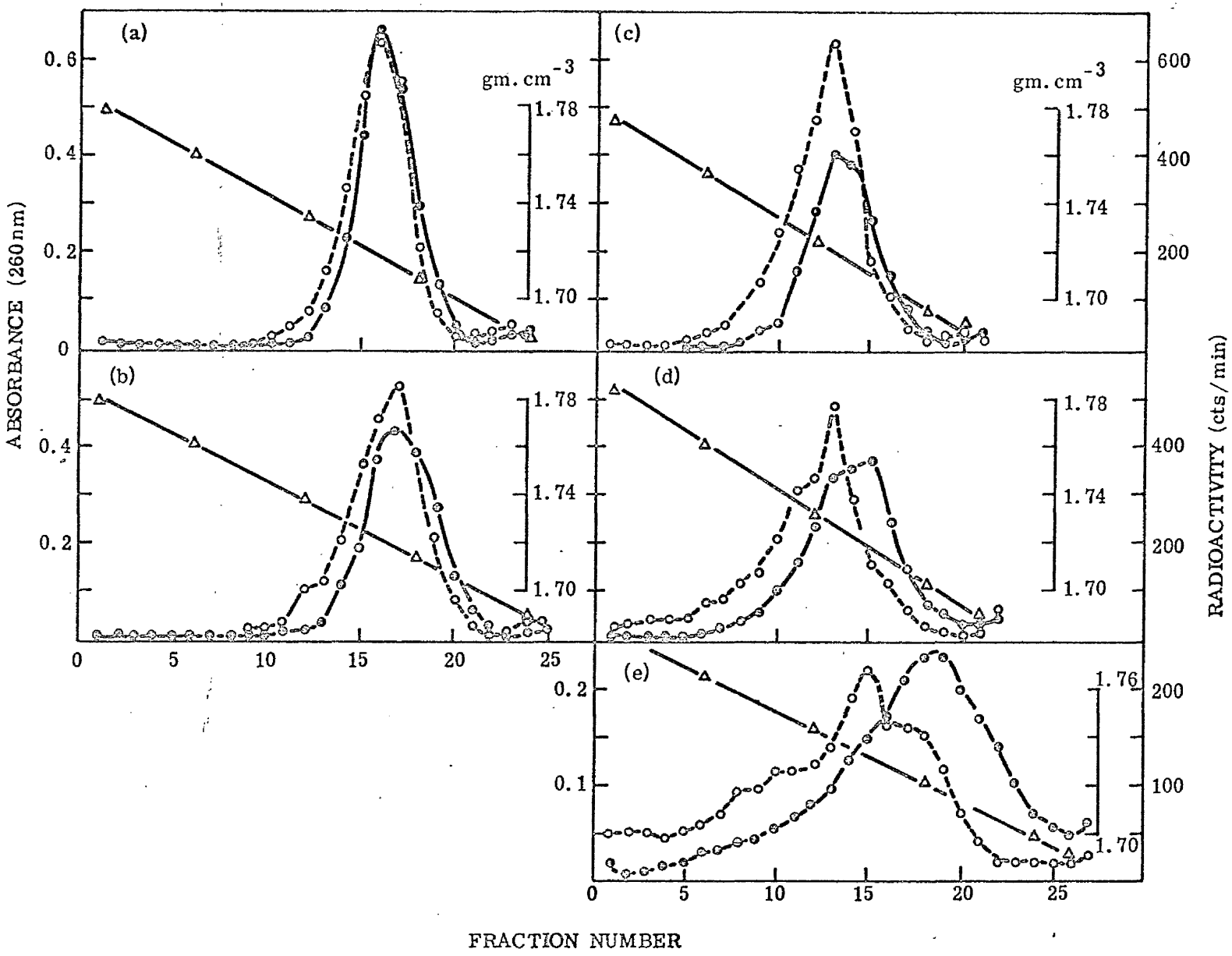
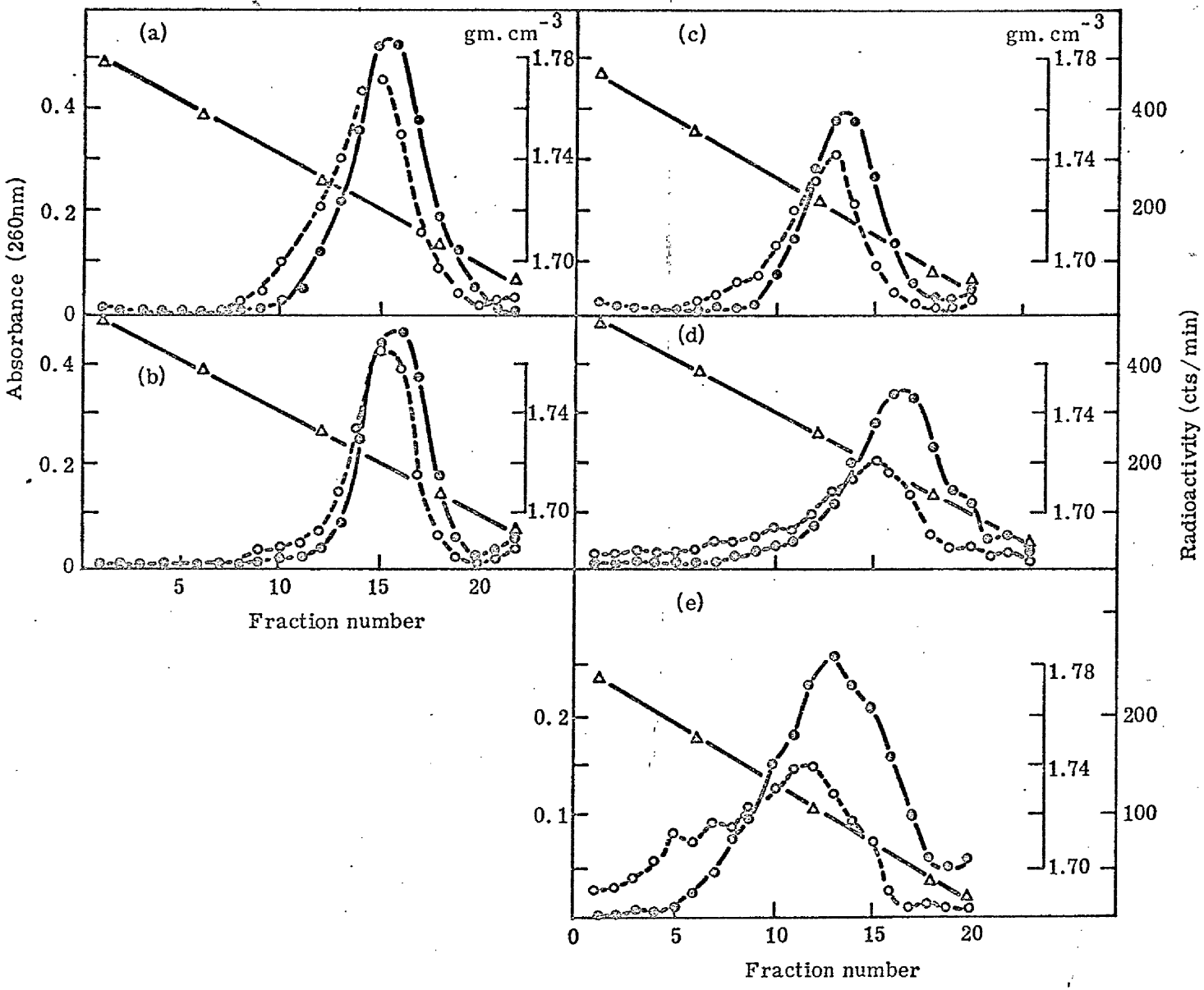

(a)
(b)
(c)
(d)
(e)
ABSORBANCE (260 nm)
RADIOACTIVITY (cts/min)
gm. cm^{-3}
1.78
1.74
1.70
1.76
FRACTION NUMBER

Fig. III.18.

Isopycnic centrifugation of BHK-21/C13 DNA·[^{3}H] poly(C) hybrids in CsCl

2 μg [^{3}H] poly(C) was hybridised with 100 μg of DNA samples (Fig. III.17) under standard conditions. Hybrids were RNase treated and processed as in the legend to Fig. III.17.

o——o, absorbance 260 nm; o----o, [^{3}H] poly(C) acid precipitable radioactivity; △——△, density.


(a)
(b)
(c)
(d)
(e)
gm. cm^{-3}
1.78
1.74
1.70
Absorbance (260nm)
0.4
0.2
0
0.1
400
200
100
5
10
15
20
Fraction number
Radioactivity (cts/min)

molecules. As is discussed later (see Discussion Section), it can be estimated that there is probably a minimum of 1000 nucleotides between successive dA-rich and dG-rich regions, and a maximum of 38,500 nucleotides between dA-rich and 36,000 nucleotides between dG-rich regions.

(4) Investigation of dA-rich and dG-rich regions within repetitious DNA sequences

DNA of eukaryotes can be considered to be composed of two main types of sequences, those sequences present only once in the genome i.e. unique sequences, and sequences of which many copies are present i.e. repetitious or reiterated sequences (Britten & Kohne, 1968). The repetitious sequences can be further classified depending upon the number of copies of each sequence. Up to 1,000,000 copies of a single sequence may exist in calf thymus DNA (Britten & Kohne, 1968). As discussed earlier (Introduction, Section I.D.2), recent investigations of the arrangement of repetitious sequences within the genome, have suggested that at least some of the repetitious sequences are distributed between the tracts of unique sequences (Davidson et al., 1973). It was of considerable interest, therefore, to determine the distribution of dA-rich and dG-rich sequences in DNA fragments of different degrees of reiteration. BHK-21/C13 DNA was fractionated into highly repetitious sequences (reassociated by C_0t = 0.05), moderately repetitious sequences (reassociated between C_0t = 0.05 - C_0t = 10), slightly repetitious sequences (reassociated between C_0t = 10 - C_0t = 100) and sequences which are probably unique (not reassociated by C_0t = 100) (Britten & Kohne, 1968) (Fig. III.19). These fractions were tested for their ability to hybridise with [^{3}H] poly(U) and

Fig. III.19.

Fractionation of BHK-21/C13 DNA by rate of reassociation

2 mg of sonicated BHK-21/C13 DNA was reassociated at 60°C to C_0t = 10. Separation of single-stranded (S.S.) DNA from double-stranded reassociated DNA (D.S.) was carried out on hydroxylapatite at 60°C as described in the METHODS section. Further fractionation of the C_0t = 10 S.S. DNA at C_0t = 100, and of the C_0t = 10 D.S. DNA at C_0t = 0.05 was performed. Samples were dialysed against 0.1 x SSC prior to hybridisation. In calculating the % DNA obtained in each fraction, allowance has been made for the yield from the hydroxylapatite column (95-100%) and the samples of DNA taken for analysis (Fractions I and II).

---o---

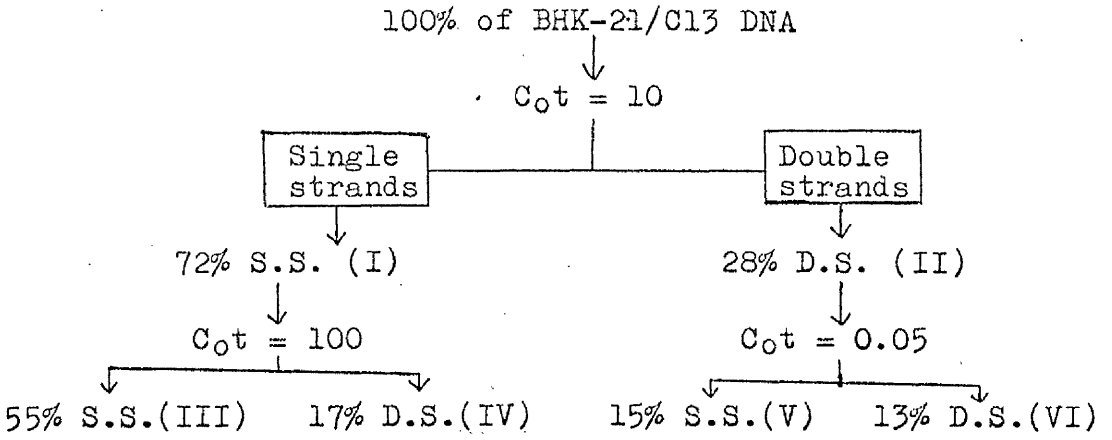
Table III.6.

Hybridisation of [^{3}H] poly(U) and [^{3}H] poly(C) to DNA fractions with different rates of reassociation

DNA fractions were adjusted to 10 µg/ml in 0.1 x SSC. (a) 1 µg DNA in 0.9 ml 2.2 x SSC, and (b) 8 µg DNA in 0.9 ml 2.2 x SSC were heated to 100°C for 15 min. 0.1 ml (0.2 µg) of (a) [^{3}H] poly(U), (b) [^{3}H] poly(C) was added immediately and incubations were continued (a) at 20°C for 4 hours, (b) at 62°C for 30 min. Mixtures were digested for 30 min at 20°C (a) with 5 µg RNase/ml, (b) with 10 µg RNase/ml. Acid insoluble radioactivity was estimated as described in METHODS. Control values in the absence of added DNA have been subtracted (42 disint/min for [^{3}H] poly(U), 10 disint/min for [^{3}H] poly(C)).

Fig. III.19

Fractionation of BHK-21/C13 DNA by rate of reassociation



---o---

Table III.6.

Hybridisation of [^{3}H] poly(U) and [^{3}H] poly(C) to DNA fractions with different rates of reassociation

	[^{3}H] poly(U) hybridised (disint/min)	[^{3}H] poly(C) hybridised (disint/min)
Total DNA	1339	816
C_0t = 10 S.S. (I)	1320	674
C_0t = 10 D.S. (II)	1173	457
C_0t = 100 S.S. (III)	1050	412
C_0t = 100 D.S., C_0t = 10 S.S. (IV)	881	291
C_0t = 0.05 S.S. C_0t = 10 D.S. (V)	1290	592
C_0t = 0.05 D.S. (VI)	330	432

[^{3}H] poly(C). It can be clearly seen (Table III.6) that both dA-rich and dG-rich sequences are present in all fractions of the DNA tested.

It can be calculated from the levels of dA-rich and dG-rich sequences (Fig. III.5) that there are more than 300,000 dA-rich and dG-rich regions per diploid genome (see Discussion). It might therefore be anticipated that both the dA-rich and the dG-rich regions would be largely found in the highly repetitive fraction of the DNA. This was not found to be the case. Indeed, there would appear to be a particularly low level of dA-rich regions in the very rapidly reassociating fraction of BHK-21/C13 DNA (Table III.6). The probable explanation for this rather anomalous finding is that the dA-rich and dG-rich regions themselves may not affect the reassociation of DNA strands under the conditions used. Thus, since poly(dA).poly(dT) of high molecular weight melts at about 73°C in 0.18 M Na^{+} (Riley et al., 1966), poly(dA).poly(dT) 50 base-pairs long (the average size of the dA-rich regions in DNA (Fig. III.7)) would be expected to melt at 57°C (using the relationship of Thomas & Dancis, 1973). Since much of the dA.dT-rich regions in DNA are not pure poly(dA).poly(dT) (Results, Sections III.B,D, and E.1), these regions will be expected to melt a few degrees below 57°C. Hence most of the dA-rich regions in the DNA will not base-pair with the complementary dT-rich regions at 60°C in 0.12 M phosphate buffer, the conditions used for the reassociation.

Similarly, high molecular weight poly(dG).poly(dC) which melts at 89°C in 0.03 M Na^{+} (Chamberlin, 1965) would melt at about 102°C in 0.18 M Na^{+} (using the relationship of

Schildkraut & Lifson, 1965). Poly(dG).poly(dC) of average size 25 base pairs long (the average size of dG-rich regions in DNA, (Fig. III.7)) would therefore be expected to melt at 69°C in 0.18 M Na^+ (Thomas & Dancis, 1973). Since the Tm of hybrids of poly(C) to dG-rich regions is about 20°C below the Tm for the pure poly(dG).poly(C) hybrid (Results, Section III.E.1), the dG.dC-rich regions in DNA will probably dissociate at about 49°C in 0.18 M Na^+.

Thus neither dA.dT-rich nor dG.dC-rich regions would be expected to reassociate at 60°C in 0.12 M phosphate buffer. The results described in Table III.6 therefore demonstrate the distribution of dA-rich and dG-rich sequences which are covalently linked to repetitive and non-repetitive sequences of a more "normal" base composition.

It is important, however, to note that the levels of hybridisation observed with the individual fractions does not, in every case, correspond to the level observed in the DNA sample from which the fractions were derived. In particular, the hybridisation of both [^{3}H] poly(U) and [^{3}H] poly(C) to C_0t = 10 single-stranded DNA (fraction I) is considerably greater than to both the C_0t = 100 single-stranded DNA (fraction III) and the C_0t = 100 double-stranded DNA (fraction IV). Further deductions regarding the relative extent of dA-rich and dG-rich regions within the DNA should therefore not be made. It can, however, be reasonably concluded that the dA-rich and dG-rich regions are not confined to a single class of repetitious sequences, but rather that they are present in many families of DNA sequences of widely varying degrees of repetition.

(5) Summary

The investigations presented in this section have demonstrated by a number of means that the dA-rich and dG-rich regions are widely distributed throughout DNA molecules. Not only are the dA-rich and dG-rich regions found in DNA molecules of a wide variety of base composition and of varying degrees of reiteration, but also the hybridisation studies using sheared DNA have further indicated that clustering of sequences within short segments of any one molecule does not take place. As is discussed later, this type of distribution of the dA-rich and dG-rich sequences enables hypotheses to be made regarding the possible functions of these sequences.

(G) Chromosomal localisation of dA-rich regions

By the technique of in situ hybridisation, the location of dA-rich regions in BHK-21/C13 interphase and metaphase nuclei was examined. After hybridisation to interphase nuclei, [^{3}H] poly(U) is distributed fairly uniformly over the whole nucleus (Plate 1). The background grains are very low, demonstrating the specific uptake of the [^{3}H] poly(U) into the nucleus. Similar specificity of labelling of chromosomes in metaphase nuclei is shown in Plate 2. Despite autoradiographic exposure for four months, individual chromosomes are still only very lightly labelled, averaging only 2-3 grains per chromosome. However, by comparison of a number of metaphase preparations (e.g. Plate 3) it can be concluded that the grains are widely distributed throughout the chromosomes, no one chromosome being particularly heavily labelled. Moreover it would

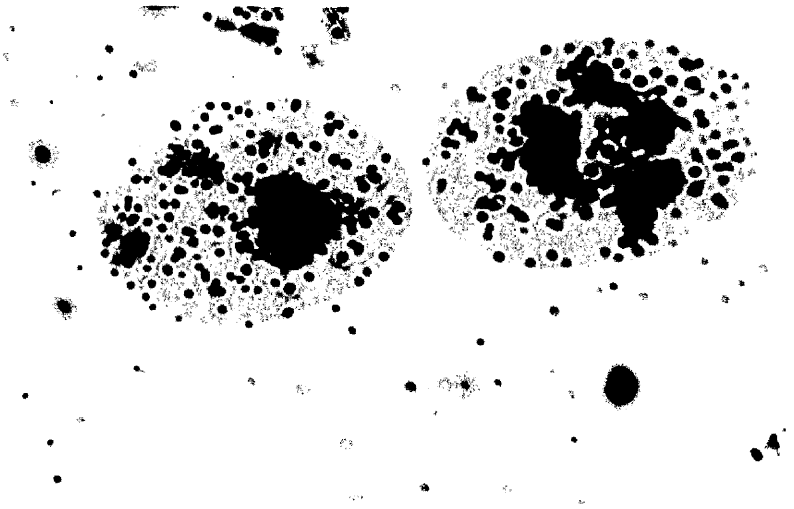


Plate 1 Autoradiograph of *in situ* hybrids of [^{3}H] poly(U) with interphase nuclei from BHK-21/C13 cells.

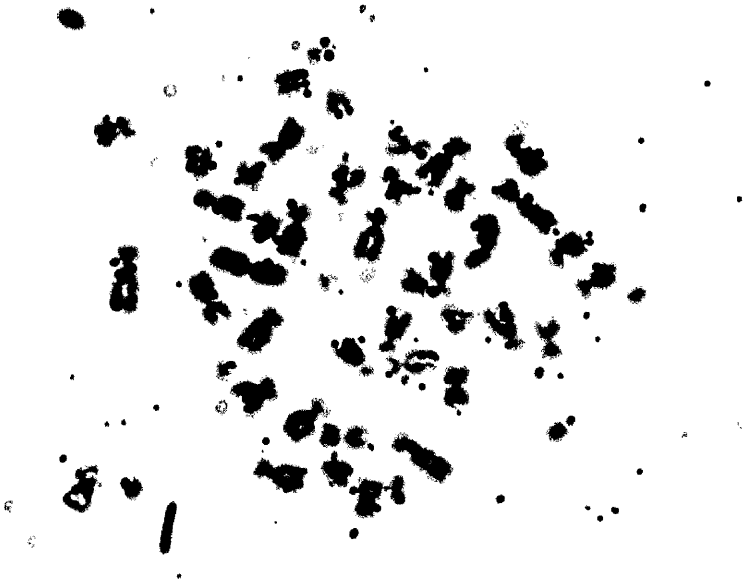
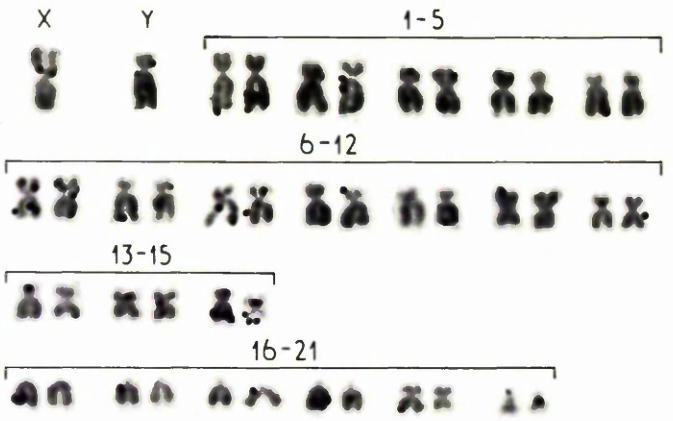
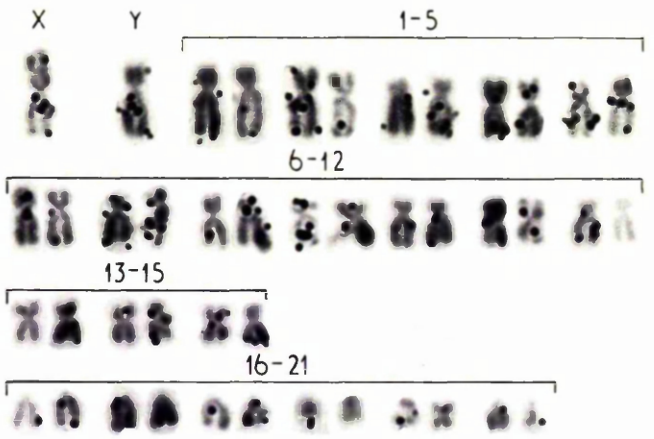
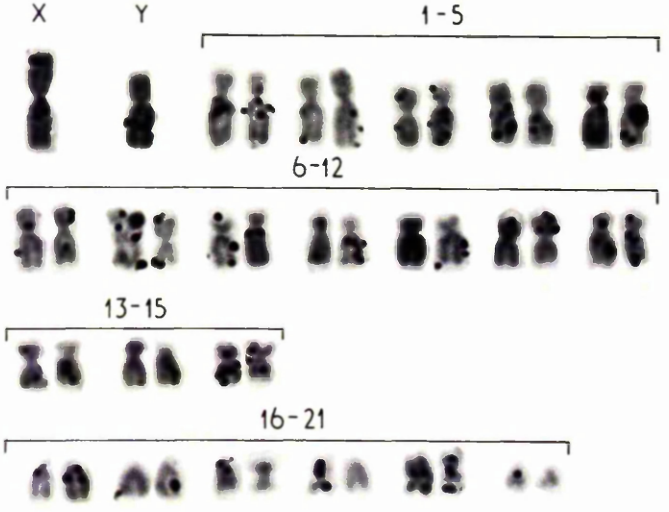


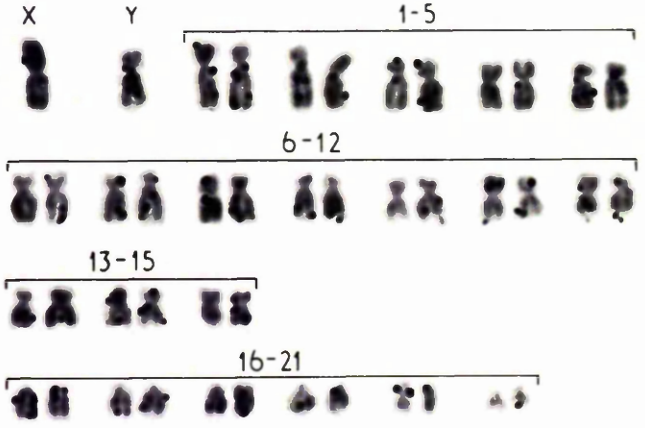
Plate 2 Autoradiograph of *in situ* hybrids of [^{3}H] poly(U) with metaphase chromosomes from BHK-21/C13 cells.

Plate 3 Autoradiographs of *in situ* hybrids of [^{3}H] poly(U) with metaphase chromosomes from BHK-21/C13 cells. Chromosomes from four nuclei were arranged roughly into groups on the basis of size, and by reference to the ratios of lengths of long:short arms (Marshall, 1972).


X
Y
1-5
6-12
13-15
16-21


X
Y
1-5
6-12
13-15
16-21


X
Y
1-5
6-12
13-15
16-21


X
Y
1-5
6-12
13-15
16-21

seem unlikely that the dA-rich regions are consistently localised to only one region of the chromosomes, on different chromosomes the label being found at the centromeres, telomeres, on the long or on the short arms. The differences in labelling pattern between the two assigned members of each pair of chromosomes, and between the same chromosomes in different nuclei, are probably due to the low level of label incorporated, and also to the difficulty in accurately karyotyping BHK-21/C13 chromosomes without the use of quinacrine fluorescence (Marshall, 1972). The average level of labelling of metaphase chromosomes is 92 grains per nucleus (after deduction of the background of 6 grains over a comparable area). Assuming 10% efficiency of autoradiographic exposure (Gall & Pardue, 1971), and the DNA content of each nucleus to be 6.6 picogram (Davidson, 1972), for a [^{3}H] poly(U) preparation of specific activity 550,000 d.p.m. per μg it can be calculated that 0.147% of the DNA has been bound by [^{3}H] poly(U). For the interphase nuclei, the average net number of grains is 146 per nucleus, corresponding to 0.232% of the DNA. The reason for the difference between these two values is uncertain, but may be related to differences in the efficiency of RNase digestion in such morphologically different nuclei. It is encouraging, however, that both values are in reasonable agreement with the saturation value for purified DNA (Fig. III.5).

Similar <u>in situ</u> hybridisation studies with [^{3}H] poly(C) showed no significant retention of radioactivity within the nucleus. The reason for this is not known, but it may be due, at least partly, to the lower level of dG-rich than

Table III.7.

The level of dA-rich sequences and dG-rich sequences in different DNA species

DNA	% of DNA dA-rich regions	% of DNA dG-rich regions
BHK-21/C13	0.13	0.072
Krebs II mouse ascites tumour	0.24	0.076
Mouse spleen	0.27	0.097
Calf thymus	0.07	0.028
Salmon sperm	0.04	0.035
E.coli	N.D.	0.001
M.lysodeikticus	0.01	N.D.
C.perfringens	0.037	N.D.
SV40	N.D.	0.008
Adenovirus 5	N.D.	0.004
Herpes Simplex	N.D.	0.056
ØX 174 (RF)	N.D.	N.D.
λ	0.024	0.002
T7	N.D.	N.D.
T4	N.D.	0.001

Hybridisation mixtures included 1 μg of each DNA with 0.2 μg [^{3}H] poly(U) or 8 μg of each DNA with 0.2 μg [^{3}H] poly(C). Hybrids were prepared and estimated under standard conditions as described in METHODS.

N.D. = not detectable.

dA-rich regions (Fig. III.5) as well as to the relatively low specific activity of the [^{3}H] poly(C) compared to the [^{3}H] poly(U).

(H) The level of dA-rich and dG-rich regions in DNA of different species

In an attempt to shed some light upon the possible functions of the dA-rich and dG-rich regions, DNA from a number of different species was prepared, and the saturation levels with [^{3}H] poly(U) and [^{3}H] poly(C) were ascertained in each case (Table III.7). For the DNAs tested, DNA of mammalian origin was found to have a higher level of dA-rich and dG-rich regions than DNA from bacteria, mammalian viruses, or bacteriophage. Also, in general the dA-rich regions are more plentiful than dG-rich regions, but this finding is by no means universal, salmon sperm DNA and herpes simplex DNA being examples of the antithesis of this.

(I) Evidence that A-rich, U-rich and G-rich RNA species occur in BHK-21/C13 cells and their significance with regard to DNA preparation procedure

The function of the homopolymer-rich regions in DNA is uncertain. One possibility is that they may be transcribed into RNA. If this is the case, then the transcription product will be RNA species with regions rich in a particular nucleotide. Over the past few years, many laboratories have reported the detection of such RNA species (see Discussion). In the course of some of the preliminary experiments in this study, certain lines of evidence were obtained which suggested the existence of, in particular, A-rich, U-rich and G-rich regions in RNA of BHK-21/C13 cells.

(1) Evidence for A-rich sequences in RNA

Some early BHK-21/C13 DNA specimens, prepared by the method of Marmur (1961), were not further purified by centrifugation in CsCl gradients. Since the only step taken to remove RNA from such specimens involved digestion with pancreatic RNase, any RNA species very rich in A or G would be largely resistant to digestion. Three observations were made which suggested the presence of A-rich or poly(A) sequences in such DNA preparations.

(a) A greater amount of [^{3}H] poly(U) complexed to DNA from such preparations than to DNA further purified either by centrifugation in CsCl or by treatment with 0.07 N NaOH at 37°C for 18 hours (which totally degrades RNA (Schmidt & Thannhauser, 1945)), or to DNA prepared by the hydroxylapatite (HAP) method (Britten et al., 1970) (Table III.8). This result supports the contention that DNA prepared by the HAP method does not contain poly(A).

Table III.8. Hybridisation of [^{3}H] poly(U) to BHK-21/C13 DNA prepared in different ways

	[^{3}H] poly(U) hybridised cts/min
M-DNA	252
M-DNA, purified in CsCl gradients	159
M-DNA treated at 37°C overnight with 0.07 N NaOH	162
DNA prepared by hydroxylapatite method	153

1 µg DNA was hybridised with 0.2 µg [^{3}H] poly(U) under standard conditions, and hybrids were assayed as in METHODS. M-DNA = DNA prepared by the method of Marmur (1961).

(b) During the experiment in which an attempt was made to separate DNA.[^{3}H] poly(U) hybrids from unhybridised DNA (Results, Section III.F.3), on one occasion DNA which had not been further purified was used. After centrifugation through CsCl, a considerable amount of the RNase resistant [^{3}H] poly(U) was found at the bottom of the gradient (Fig. III.20). This result was not obtained when DNA already purified on CsCl or hydroxylapatite was used (Fig. III.17). This result again is consistent with the presence of poly(A) in the DNA preparation which could form a high density poly(A).[^{3}H] poly(U) complex. It is possible that some of the poly(U) at the bottom of the gradient is not complexed to poly(A). This is, however, unlikely since control experiments in the absence of any DNA input demonstrated that no acid precipitable [^{3}H] poly(U) remained after digestion with RNase.

(c) A further study involved an estimation of size of the poly(dA) sequences in BHK-21/C13 DNA by measurement of the size of poly(U) eluted from hybrids (Results, Section III.D). On one occasion impure DNA, thought to contain poly(A), was used. It was observed on this occasion that some sequences of poly(U) of up to about 175 nucleotides long, were eluted from the hybrids (Fig. III.21). This result is consistent with the interpretation that such impure DNA preparations contain poly(A), since it is known that long sequences of poly(A) of 150-200 nucleotides long are present in HnRNA and mRNA of mammalian cells (Edmonds & Caramela, 1969; Darnell et al., 1971a). On the other hand, when DNA was purified in CsCl gradients or by the hydroxylapatite method, the longest fragments of poly(U) observed in such a study

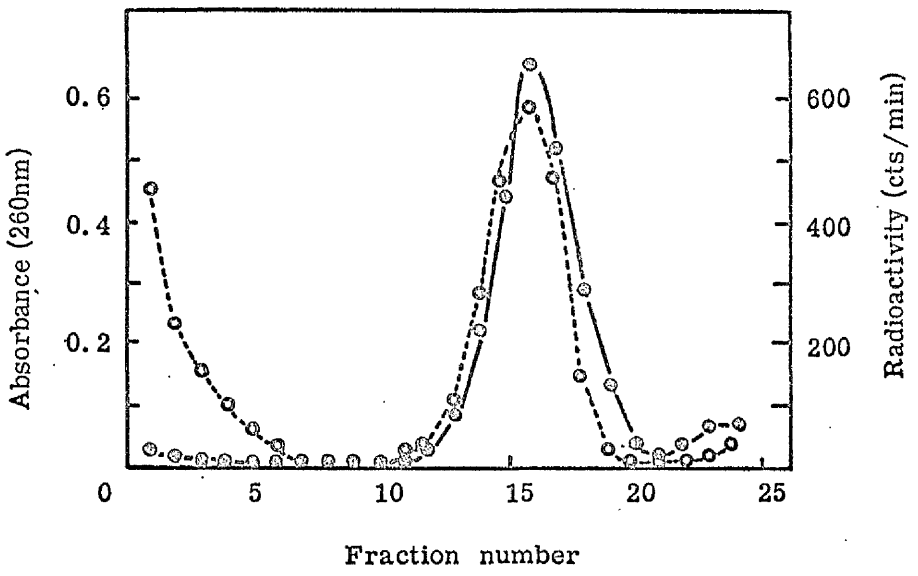


Fig. III.20.

CsCl gradient centrifugation of complexes formed between [^{3}H] poly(U) and BHK-21/C13 DNA which had been prepared by the method of Marmur.

10 µg [^{3}H] poly(U) (Schwartz-Mann) was hybridised under standard conditions with 100 µg denatured BHK-21/C13 DNA, prepared by the method of Marmur but not further purified. Hybrids were treated with 10 µg RNase/ml at 20°C for 1 hour, and were then centrifuged for 64 hours at 33,000 rev/min in a CsCl gradient of average density 1.730 g/cm^3. Fractions from the gradient were collected and absorbance and acid-insoluble radioactivity were measured as in METHODS. Density increases from right to left. Absorbance, ⊙——⊙; Radioactivity o----o.

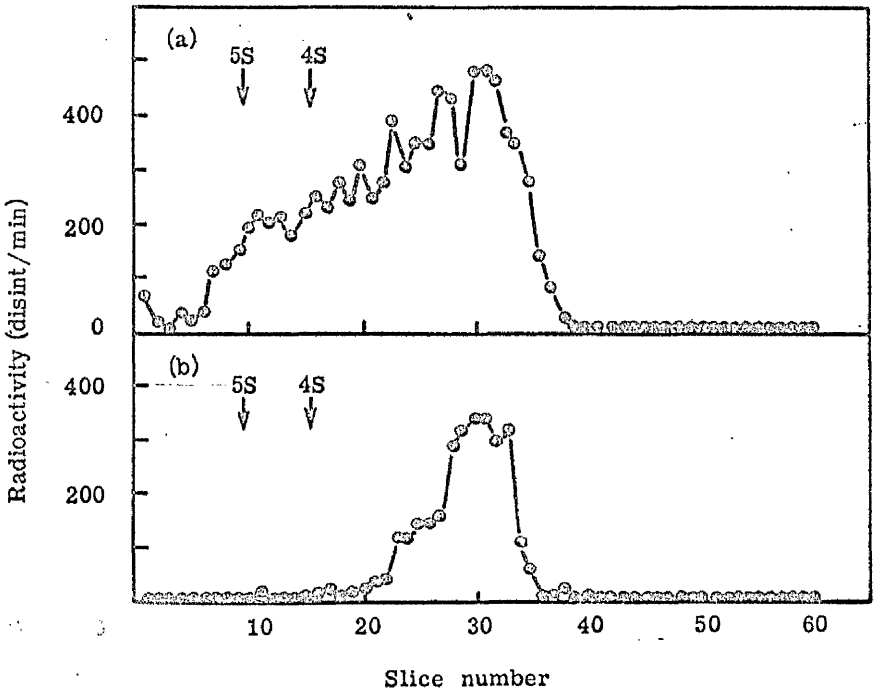


Fig. III.21.

Size of poly(A) sequences present in preparations of BHK-21/C13 DNA

(a) 11 nitrocellulose filters each containing 20 µg BHK-21/C13 DNA prepared by the method of Marmur but not further purified, and (b) 9 nitrocellulose filters each containing 20 µg BHK-21/C13 DNA prepared by the hydroxylapatite method, were incubated with 20 µg [^{3}H] poly(U) in 2 ml 2 x SSC at 20°C for 4 hours. Filters were washed and treated with RNase, 20 µg/ml at 20°C for 24 hours. Hybridised poly(U) was then eluted and analysed in 10% polyacrylamide gels as detailed in METHODS. The position of marker 5S and 4S RNAs, with minimal secondary structure (see Results, Section III.D), is shown by the arrows.

are 60 nucleotides long (Fig. III.10 and III.21).

In summary, these results demonstrate that A-rich or poly(A) sequences of up to 175 nucleotides long are indeed present in BHK-21/C13 cells. Because such sequences are both resistant to RNase and also bind to nitrocellulose filters (Lee et al., 1971) if they contaminate DNA preparations, complexes of poly(A).[^{3}H] poly(U) as well as of DNA.[^{3}H] poly(U) will be detected. It is therefore of fundamental importance to ensure that DNA preparations are free of poly(A), either by further purification of DNA in CsCl gradients, or by preparing the DNA by the hydroxyl-apatite method.

(2) Evidence for U-rich sequences in RNA of BHK-21/C13 cells

U-rich sequences can be detected in pulse-labelled high molecular weight RNA of BHK-21/C13 cells (Burdon & Shenkin, 1972). One method of detecting such sequences involves hybridisation of the RNA to poly(A) and then collecting both the poly(A) and any RNA molecules complexed to the poly(A) (presumably by U-rich regions) on nitrocellulose filters (Burdon & Shenkin, 1972). A further method of detecting such sequences might make use of the fact that (poly(dA).poly(dT)).poly(U) hybrids bind to nitrocellulose filters (Table III.2). When this was attempted, using RNA pulse-labelled with ^{3}H-uridine, some of the RNA does complex with poly(dA).poly(dT) (Table III.9). As expected, only a very small amount of the ^{3}H-U label collected on the nitrocellulose filter is RNase resistant i.e. only the portion which is complexed to poly(dA).poly(dT). The level of ^{3}H-radioactivity binding to the filter in the

absence of poly(dA).poly(dT) is a measure of the level of RNA molecules containing poly(A) sequences. These results therefore provide further proof of the existence of U-rich sequences in high molecular weight RNA of BHK-21/C13 cells.

Table III.9. Hybridisation of [^{3}H]-U-pulse labelled RNA to poly(dA).poly(dT)

[^{3}H]-U-RNA (μg)	Poly(dA).Poly(dT) (μg)	RNase treatment	Radioactivity on filter cts/min
55	2.8	-	172
55	2.8	5 μg/ml 30 min	11
55	-	-	41
55	-	5 μg/ml 30 min	9

BHK-21/C13 cells were exposed to ^{3}H-uridine for 12 min and the labelled RNA was extracted by the hot phenol-SDS technique at pH 5.1 (Burdon & Clason, 1969; Burdon & Shenkin, 1972). This RNA was a gift of Dr. R.H. Burdon. RNA was incubated with or without poly(dA).poly(dT) at 20°C for 4 hours in 2 x SSC and was then passed slowly through nitrocellulose filters. RNase treatment was performed as shown, and filters were washed with 2 x SSC and counted as in METHODS.

(3) Evidence for G-rich sequences in RNA of BHK-21/C13 cells

It has been pointed out above, that DNA prepared by the method of Marmur (1961) and not further purified may contain not only A-rich RNA sequences but also G-rich RNA sequences. Two lines of evidence support this.

(a) The level of hybridisation of poly(C) to DNA which has not been further purified, is slightly greater than the

level when DNA is used which has been purified on CsCl gradients or hydroxylapatite (Table III.10). This method of detection of G-rich sequences will, however, be less sensitive than for the detection of poly(A) sequences, since poly(G) has less affinity than poly(A) for nitrocellulose filters (Table III.3).

Table III.10. Hybridisation of [^{3}H] poly(C) with BHK-21/C13 DNA prepared in different ways

	[^{3}H] poly(C) hybridised cts/min
M-DNA	560
M-DNA, purified in CsCl gradients	475
DNA prepared by hydroxylapatite method	470

10 µg denatured BHK-21/C13 DNA was hybridised with 0.2 µg [^{3}H] poly(C) under standard conditions, and hybrids were assayed as in METHODS. M-DNA = DNA prepared by the method of Marmur (1961).

(b) Centrifugation through CsCl gradients of hybrids of [^{3}H] poly(C) with "impure" DNA, demonstrates that some of the poly(C) is held in a very dense RNase resistant complex (Fig. III.22). This high density complex is not observed when CsCl purified DNA or DNA prepared by the hydroxylapatite method is used (Fig. III.18). This is very suggestive that G-rich or poly(G) sequences are present in the RNA of BHK-21/C13 cells. It can be argued that some of the high density [^{3}H] poly(C) may not be complexed to G-rich RNA species, but is detected due to inadequate

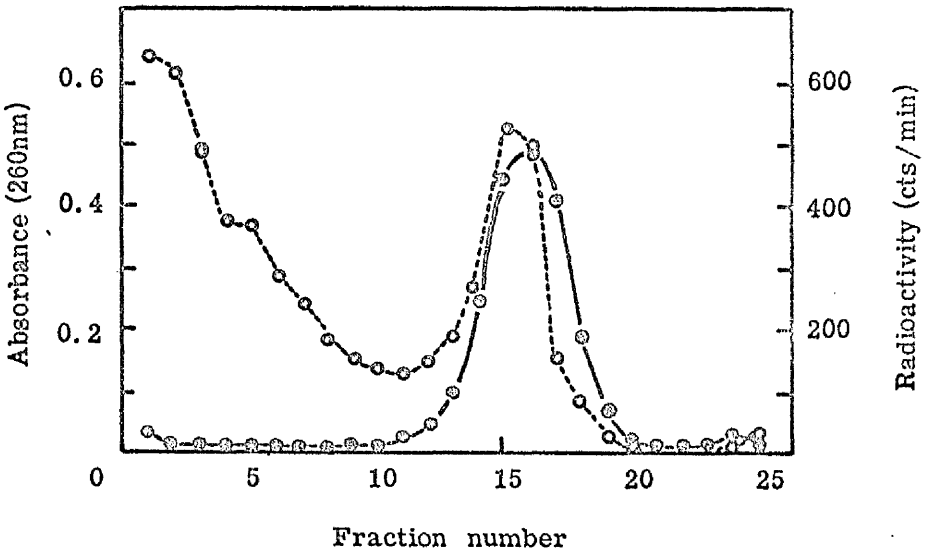


Fig. III.22.

CsCl gradient centrifugation of complexes formed between [^{3}H] poly(C) and BHK-21/C13 DNA which had been prepared by the method of Marmur

2 µg [^{3}H] poly(C) was hybridised under standard conditions with 100 µg BHK-21/C13 DNA which had been prepared by the method of Marmur but had not been further purified. Hybrids were RNase treated and processed as in the legend to Fig. III.20. Density increases from right to left.

Absorbance, ●——●; Radioactivity, o----o.

digestion with RNase. This is unlikely since control experiments with no added DNA demonstrated that RNase completely degraded all the [^{3}H] poly(C) to acid soluble fragments. It is, therefore, probable that G-rich sequences are present in BHK-21/C13 RNA. As for the poly(A) sequences mentioned above, it is clearly very important to use a purification procedure which will remove such G-rich RNA sequences from the DNA preparation.

In summary, these experiments have demonstrated that A-rich, U-rich and G-rich sequences are probably present in RNA of BHK-21/C13 cells. No attempt has been made in this study to further characterise them. However, the recent interest in homopolymeric sequences in RNA has led to their being intensively investigated in many other laboratories. The properties of these ribohomopolymer-rich regions, and their relation to the homopolymer-rich regions in DNA reported in this study, will be discussed in the next section.

SECTION IV DISCUSSION

A. Techniques used to investigate homopolymer-rich regions in DNA

Previous attempts to identify regions in DNA composed largely of only one type of nucleotide have employed basically two types of technique.

1. Depurination

Treatment of DNA with acid, either with or without diphenylamine, leads to the digestion of purine residues and the production of pyrimidine oligonucleotides of the form Py_nP_{n+1} (Burton & Petersen, 1970; Shapiro, 1967). The pyrimidine tracts produced can then be separated by chromatography or electrophoresis.

The largest deoxyhomopolymeric runs characterised by these treatments have been $(T)_{11}$ in Bacillus subtilis DNA (Rudner et al., 1972) and $(C)_9$ in chick embryo orphan virus (Bellet et al., 1972). In mammalian DNA, $(T)_7$ and $(C)_4$ have been detected (Sneider, 1971). The interpretation of these results, particularly with mammalian DNAs, has been complicated by the failure, in most cases, to account for all the DNA applied to the chromatography columns. In general, only 90-99% of the pyrimidine tracts was recovered and characterised. Since the pure deoxyhomopolymeric runs, of average size $(A)_{31}$ (and therefore $(T)_{31}$) and $(G)_{17}$ (and therefore $(C)_{17}$), detected in the present study amount to less than 0.01% of the DNA, such tracts would probably be missed in most studies using pyrimidine tract analysis. However, one recent investigation of pyrimidine tracts in HeLa DNA, labelled to high specific activity with ^{32}P,

demonstrated that long polypyrimidine tracts are indeed present, although the exact composition of these tracts was not established (Birnboim et al., 1973).

Another limitation of these pyrimidine tract studies is that no conclusions can be reached regarding the relationship of these tracts to one another along the DNA molecule (Mushynski & Spencer, 1970a, b). Thus long regions which are almost pure polypyrimidine may have just a few purine residues scattered within the region, so that only short pyrimidine tracts are observed.

Despite these problems, it should be emphasised that pyrimidine tract analysis has been of considerable value in providing information regarding short tracts, usually of less than 8 nucleotides long. One interesting feature which emerges from such studies is that in general, long isostichs containing T are more common than those containing C (Spencer & Chargaff, 1963; Sneider, 1971), results which are consistent with those presented in Table III.7.

2. Ribohomopolymer binding

The technique of ribohomopolymer binding followed by centrifugation in CsCl (Kubinski et al., 1966) is probably very much less precise than pyrimidine tract analysis from the point of view of detecting deoxyhomopolymeric regions, since the exact nature of the complexes observed in such studies is uncertain (Kubinski et al., 1966; Summers & Szybalski, 1968; Borst & Ruttenberg, 1972). A number of experimental results lead to the conclusion that there is only a limited amount of true "Watson-Crick" hydrogen-bond formation between the various ribohomopolymers and DNA.

(a) The marked concentration dependence of complex

formation strongly suggests that much of the complex is due to aggregation of the ribohomopolymers (Kubinski et al., 1966; Sheldrick & Szybalski, 1967).

(b) Most, if not all, of the density increment brought about by binding ribohomopolymer can be abolished by treatment with RNase (Results, Section III.F.3; Kubinski et al., 1966).

(c) In many cases there is only a very poor correlation between the effect of poly(G) binding and the distribution of oligo(C) tracts (Mushynski & Spencer, 1970a, b; Kizer & Saunders, 1972).

(d) The melting profile of the bacteriophage T7 DNA-poly(G) complex obtained by Summers & Szybalski (1968) shows a Tm of 52°C in 3 x SSC. Using the data of Chamberlin (1965), and the relationships of Schildkraut & Lifson (1965) and Thomas & Dancis (1973), it can be calculated that a pure hybrid of poly(dC).poly(G), 20 base pairs long (the length of the dC-rich regions in T7 DNA) would melt at about 110°C. The relatively very low Tm actually found therefore suggests a large amount of mismatching within the complex.

(e) The difficulty in forming complexes of DNA with poly(A) although poly(U) binds easily, suggests that mismatched regions may exist which produce different degrees of destabilisation with different ribohomopolymers (Borst & Rüttenberg, 1972).

(f) The results of a study of the complex formation of poly(U) with sheared fragments of mammalian DNA, further suggested that if no RNase treatment were used, then the complexes detected would be very non-specific (Fig. III.16).

It seems likely, therefore, that the DNA sequences

detected by the ribohomopolymer binding technique are enriched in a particular nucleotide but probably also contain a significant proportion of other nucleotides. As pointed out already, this technique has proved very useful, particularly in studies involving the separation of DNA strands (Introduction, Section I.H). However, for the purpose of studying the homopolymer-rich regions themselves it has several disadvantages as compared to molecular hybridisation (see below).

3. Molecular hybridisation

The availability of ^{3}H-labelled ribohomopolymers of high molecular weight has enabled the use of molecular hybridisation in this study to examine certain DNAs for regions consisting largely of the complementary deoxyribonucleotide. This technique offers a number of advantages over pyrimidine tract analysis and ribohomopolymer binding. First, the level of even very small amounts of deoxyhomopolymer can be estimated. Secondly, studies on the distribution of sequences both within and between different DNA fractions are possible. Thirdly, the type of complexes which are formed can be characterised. Moreover, these various studies are facilitated by the ability to perform a number of analyses simultaneously. Thus, comparisons of different specimens, and studies on the reproducibility of data, can be relatively easily made.

B. Evaluation of the hybridisation technique

The technique of DNA-RNA hybridisation was introduced in 1961 by Hall & Spiegelman who separated hybrid molecules from unhybridised molecules, by density gradient

centrifugation. The procedure was made more specific and also considerably less tedious, by the introduction of RNase treatment (Yanofsky & Spiegelman, 1962) and the use of DNA immobilised on columns (Bautz & Hall, 1962; Bolton & McCarthy, 1962). Much more convenient techniques using nitrocellulose filters to bind denatured DNA and hybrid molecules were developed by Nygaard & Hall (1964) and Gillespie & Spiegelman (1965). The amount of information which can be obtained by the rapid analysis of a large number of samples in this way, has made this one of the most powerful tools available for investigation of DNA sequence composition and organisation.

It is now clear that a critical evaluation both of the hybridisation techniques used and the quality of the hybrid molecules produced should be made in all studies (McCarthy & Church, 1969; Kennell, 1970). Depending upon the conditions employed during hybridisation, the accuracy of base-pairing within the hybrid can be varied. The higher the temperature of the incubation, the less mismatching is found in the hybrid (McCarthy & Church, 1970; Kennell, 1970). One standard method of carrying out hybridisation studies is to use the temperature at which the hybridisation reaction proceeds most rapidly (Nygaard & Hall, 1964; Gillespie, 1968; Birnstiel et al., 1972). This temperature is usually 20-25°C below the Tm of the hybrid. Thus, by performing hybridisation at $(T_m - 25)^{\circ}C$, some mismatching of base-pairs may well occur within the hybrid. The exact nature of the hybrids produced under these conditions must therefore be investigated. To this end, three main methods were used in this study.

(a) First, the resistance of hybrid molecules to digestion with pancreatic RNase was investigated. The relative instability of the hybrids of poly(U) and poly(C) with DNA to prolonged RNase digestion, when compared with the marked stability of pure poly(dA).poly(U) and poly(dG).poly(C) hybrids (Fig. III.4), strongly suggests that the hybrids with DNA are significantly mismatched. This conclusion is supported by the finding that the hybridised poly(U) and poly(C) become smaller with prolonged RNase digestion of hybrids(Fig. III.7 and III.10). It seems likely, therefore, that RNase progressively hydrolyses the ribohomopolymer at points corresponding to mismatched base-pairs, leading to the production of only short stretches of ribopolymer hybridised to DNA. The shorter the hybrid molecule, the more unstable it is, this being particularly evident with poly(dA).poly(U) hybrids (Walker, 1969; Thomas & Dancis, 1973). A small proportion of mismatched base-pairs within the hybrid might therefore have a very substantial effect upon the amount of ribohomopolymer which can remain in stable complex with DNA, after RNase digestion. A corollary to this is that estimates of the levels of, in particular, the dA-rich regions may be low, since short dA-rich sequences (less than 25 nucleotides long, see Results, section III.E.1) will not be detected in the assay used.

The small proportion of the initial complex which remains resistant to exhaustive RNase digestion probably consists of hybrids of pure sequences of deoxyhomopolymer with ribohomopolymer. It should again be emphasised that the level of the poly(dA) sequences detected are almost certainly low, since sequences of pure poly(dA) less than

25 nucleotides long will not form stable complexes at the incubation temperatures used.

(b) Secondly, the inability to form significant amounts of hybrids of poly(U) and poly(C) with deoxypolymers other than poly(dA) and poly(dG) respectively, and the effects of competition with ribohomopolymers, further suggested that major degrees of mismatching were unlikely (Results, Section III.E.2 and 3).

(c) However, since mismatching of base-pairs is clearly present within most of the hybrid molecules it would be of great value to be able to estimate the extent of this mismatching. The most useful technique available for obtaining such an estimate is by studying the thermal dissociation profile of the hybrids. As already discussed (Results, Section III.E.1), analysis of the dissociation curves of DNA.poly(U) and DNA.poly(C) hybrids are complicated by the short nature of the hybrids. A reasonable interpretation of the results can be obtained using the relationship of Thomas & Dancis (1973), $b\Delta T_m = 820^{o}C$, where b is the length of the duplex. This leads to the conclusion that after mild RNase digestion the hybrids of poly(U) with DNA are probably about 2-6% mismatched, and hybrids of poly(C) with DNA are 10-30% mismatched, whereas after exhaustive RNase treatment, as expected, the hybrids probably contain no mismatched base-pairs.

The validity of the relationship of Thomas & Dancis (1973) has not yet been rigorously demonstrated, but results obtained both in this study (Results, Section III.E.1) and by Riley et al. (1966) strongly suggest that the relationship holds for deoxyhomopolymer·ribohomopolymer hybrids.

A definitive answer to the validity of the relationship could clearly be obtained if homopolymers of a specific length were purified. In the meantime, it can reasonably be concluded that the hybridisation conditions used in this study lead to the production of hybrids containing only a small proportion of mismatched base-pairs. Digestion with RNase is essential to remove both unhybridised ribohomopolymer and also regions of mismatching within the hybrid. However, this treatment leads to a marked reduction in the amount of hybrid which remains stable.

C. The organisation of homopolymer-rich regions in DNA molecules

It is of interest to consider the possible distribution of the dA-rich and dG-rich regions throughout the genome. If there are 6.5×10^9 base-pairs present per nucleus (Davidson, 1972), 0.13% would represent approximately 338,000 tracts of dA-rich regions, averaging 50 nucleotides long. Similarly 0.07% would correspond to about 364,000 tracts of dG-rich regions having an average length of 25 nucleotides. It could be envisaged that such a large number of tracts may be found clustered together in a specific region of the DNA, or alternatively, may be scattered at random throughout the DNA. The data presented demonstrate that both types of region are to be found throughout fractions of the DNA of widely differing G + C content. Moreover, it has also been shown that when BHK-21/C13 DNA is fractionated on the basis of its reassociation kinetics (Fig. III.19), both dA-rich and dG-rich regions are present in the highly repetitious, moderately repetitious and also in single copy DNA sequences

(Table III.6). It would thus seem likely that both types of region are widely distributed throughout the DNA. If they are situated at regular intervals along the whole length of the DNA, then a dA-rich region would be found every 38,500 nucleotides, and a dG-rich region every 36,000 nucleotides. These values are therefore the maximum distance between two similar deoxynucleotide-rich regions.

What is likely to be the minimum distance between two similar regions? This can be inferred from studies of the distribution of densities of hybrid molecules using DNA sheared to different sizes (Figs. III.17 and III.18). The density of DNA fragments after hybridisation with RNA is approximately proportional to the amount of DNA within such fragments which is in an RNA-DNA form (Clarkson et al., 1973). Consider the binding of [^{3}H] poly(U) to DNA fragments of average size 1000 nucleotides (Fig. III.17(c)). Some of the DNA molecules have clearly bound poly(U) since there is a very slight shift in the density of hybrid molecules in relation to the bulk DNA. Binding of one tract of poly(U) of average size 50 nucleotides long (that is, 5% of the DNA in a fragment hybridised) would be expected to increase the density of the DNA from 1.717 g/cm^3 to 1.723 g/cm^3 (Clarkson et al., 1973). However, if any of the DNA fragments 1000 nucleotides long were to bind two tracts of poly(U), the density of such hybrid molecules would be about 1.729 g/cm^3. This should be clearly visible as a distinct shoulder in the radioactivity profile. Such a shoulder was not found (Fig. III.17(c)). It can therefore be concluded that the dA-rich regions in DNA must be more than 1000 nucleotides apart. Because the dG-rich regions are, on average, about half the

size of the dA-rich regions (Fig. III.7), the density increment on hybridising one homopolymer-rich tract will be less on binding poly(C) than on binding poly(U). The presence of a shoulder in the curve of the distribution of densities of hybrid molecules will therefore be more difficult to detect. Nonetheless, from the results obtained with DNA sheared to an average size of 1000 and 650 nucleotides long (Fig. III.18(c) and (d)), it would seem unlikely that a large fraction of the dG-rich regions could be so clustered that two or more regions are present in stretches of DNA, 1000 nucleotides long or less.

The point should however be made that when hybrids were formed between DNA fragments of average size 200 nucleotides long and both poly(U) and poly(C), a small fraction of the hybrid molecules were found to have densities as high as 1.76 g/cm^3 or more (Fig. III.17(e) and III.18(e)). This fraction is difficult to quantitate but is unlikely to be more than 10% of the total hybrid in either case. This finding implies that some of the DNA molecules have bound more than 35% ribohomopolymer (Clarkson et al., 1973), that is, an average of 70 nucleotides of RNA. In the case of the dA-rich regions, these may well represent examples of poly(U) binding to the longer regions which are known to be present (Fig. III.7(a)). Thus only one tract of dA-rich region would be present per DNA fragment. On the other hand, no dG-rich regions longer than 40 nucleotides long were observed in BHK-21/C13 DNA (Fig. III.7(c)). It is likely therefore that there do exist certain regions in which 2-3 dG-rich regions are clustered on a sequence of DNA only about 200 nucleotides long. However, since the density profile was

not significantly altered by shearing the DNA from 65,000 - 1,000 nucleotides long (Fig. III.18(a-c)), there is probably no large scale clustering of regions in DNA molecules of 1000 nucleotides or more.

The vast majority of dA-rich and dG-rich regions are, therefore, probably widely distributed throughout the DNA, at least 1000 nucleotides separating similar tracts. In this context, it should be noted that regions of very low melting temperature have recently been visualised with the electron microscope in chick fibroblast and Chinese hamster DNA (Evenson et al., 1972). At least some of these regions will be very rich in A·T base-pairs. Of particular interest is the observation that these regions are found throughout the DNA with a probable periodicity of around 6000 nucleotides (Evenson et al., 1972).

Considering the pure homopolymeric sequences in BHK-21/C13 DNA, the level of 0.008% corresponds to about 34,000 sequences of pure poly(dA) of average size 31 nucleotides long, whereas 0.0016% corresponds to about 12,200 sequences of pure poly(dG) of average size 17 nucleotides long per diploid genome. Studies on the distribution of these pure homopolymeric sequences proved more difficult than for the homopolymer-rich regions, because of the much larger amount of DNA which must be used to incorporate a sufficient level of radioactivity into the hybrid. However, it was possible to show that these sequences are present in DNA molecules of a wide variety of base compositions (Fig. III.15). The pure sequences of deoxyhomopolymers may also, therefore, have a widespread distribution in BHK-21/C13 DNA, similar to that of the overall families of dA-rich and dG-rich regions.

D. The organisation of homopolymer-rich regions in chromosomes

The pattern of folding of chromatin fibres within the chromosomes is still not understood (Huberman, 1973). Since the homopolymer-rich regions seem to be widely distributed throughout DNA molecules, they might serve as markers to assist in the interpretation of chromosome structure. It would be of particular interest to know if they are scattered throughout the chromosomes or if they are clustered to particular regions of chromosomes. In situ hybridisation studies were performed in an attempt to answer this question.

The exact nature of the nuclear material to which the [^{3}H] poly(U) bound is uncertain. Very little protein is likely to have remained after the extensive acid and salt washes which were used. There is no direct evidence concerning whether DNA or RNA is the molecule in situ which is complexed. However, binding was observed after denaturation with alkali, a technique which will tend to destroy RNA. Moreover, there is relatively little binding of poly(U) to cytoplasmic material, although the cytoplasm would be expected to have considerably more poly(A) (in mRNA) than would be present in the nucleus (in HnRNA)(Darnell, 1968; Jelinek et al., 1973; Greenberg & Perry, 1972). These findings strongly suggest that DNA is the substance bound by poly(U). This conclusion is in agreement with the findings of Jones et al. (1973) who showed that [^{3}H] poly(U) will hybridise in situ with Rhynchosciara angelae chromosomes.

The in situ hybrids obtained with BHK-21/C13 cells demonstrate that the dA-rich regions are, indeed, widely distributed throughout interphase and metaphase chromosomes.

No particular region appears to be enriched or depleted in dA-rich regions. The relatively low level of grains observed over the chromosomes precludes a definite conclusion on this point. Jones et al. (1973) however found that only specific regions of the polytene chromosomes of R.angelae bound [^{3}H] poly(U). They classified these regions into three types:- (a) The heterochromatic terminal and subterminal regions of certain chromosomes which bound [^{3}H] poly(U) intensely. These regions are known to be very rich in satellite DNA in R.hollandei and probably also in R.angelae (Gall et al., 1971). (b) Certain heterochromatic regions which bound [^{3}H] poly(U) diffusely. (c) Certain euchromatic regions which also bound [^{3}H] poly(U) intensely. In polytene chromosomes, at least, it seems likely that dA-rich regions are found at particular regions of both heterochromatin and euchromatin. It is uncertain to what extent the dA-rich regions detected in BHK-21/C13 chromosomes reflect euchromatic or heterochromatic DNA.

With reference to the possibility that particular regions of eukaryotic metaphase chromosomes might be particularly enriched for dA-rich sequences, recent studies on banding patterns in mammalian DNA should be considered. New staining techniques have been developed which lead to the production of heavily stained bands across both chromatids of metaphase chromosomes at specific sites (Fig. IV.1). The banding techniques have been classified as follows:-

(i) Q-banding. Treatment of chromosomes with acridine dyes, especially quinacrine and quinacrine mustard, leads to the development of a specific fluorescence pattern (Caspersson et al., 1968; Marshall, 1972). This pattern is thought to be

Fig. IV.1.

G-banding patterns of human chromosomes. In each set of four chromosomes, the first two are homologues from one cell, while the third and fourth are examples from two other cells.

(From Drets & Shaw, 1971)


1
2
3
1a
4
5
6
7
8
9
1b
10
11
12
X
13
14
15
16
17
18
1c
19
20
21
22
Y

largely due to the greater fluorescence of the dyes when bound to A + T rich DNA (Weisblum & de Haseth, 1972; Ellison & Barr, 1972; de la Chapelle et al., 1973). It seems likely that the type and extent of base sequence repetition within A + T rich regions also influences these Q-banding patterns (Selander & de la Chapelle, 1973). The role of proteins, if any, in Q-banding is uncertain (Selander & de la Chapelle, 1973).

(ii) C-banding. Denaturation of the DNA in situ in the chromosomes, followed by partial renaturation and staining with Giemsa, leads to specific staining of centromeric heterochromatin (Arrighi & Hsu, 1971). The pattern produced in this way probably reflects the degree of repetition of nucleotide sequences.

(iii) G-banding. Modifications of the C-banding technique have allowed differential staining of the arms of the chromosomes with Giemsa (Schnedl, 1971; Sumner et al., 1971; Drets & Shaw, 1971). With certain exceptions, these G-banding patterns are identical to Q-banding (Selander & de la Chapelle, 1973), suggesting a similar mechanism of production of bands. The role of protein in G-banding is obscure, but it has been suggested that non-histone proteins may play some part (Comings et al., 1973).

(iv) R-banding. By heating the chromosome preparation to about 88°C followed by Giemsa staining, a pattern (R-banding) is obtained which is the reverse of G-banding (Dutrillaux & Lejeune, 1971).

(v) Use of fluorescent anti-nucleoside antibodies. Fluorescent anti-adenosine antibodies have been shown to bind to DNA and give rise to a fluorescence pattern very

similar to that obtained by Q-banding (Dev et al., 1973). On the other hand, anti-cytosine antibodies preferentially bind to areas which stain poorly with quinacrine (Schreck et al., 1973).

Although the exact mechanism of band production is uncertain, it seems clear that regions which stain have specific base compositions and nucleotide sequences (de la Chapelle et al., 1973; Huberman, 1973). Since the bands always cross the whole chromatid perpendicular to the long axis of the chromosomes, similar sequences must be found near one another in the chromosome (Sutton, 1972; Huberman, 1973). Thus the structure of metaphase chromosomes is not random, as it would appear from electron and light microscopy (Fig. I.2) but it is highly ordered. This degree of order is made even more remarkable if there is just a single DNA molecule in each chromosome (Kavenoff & Zimm, 1973; Petes & Fangman, 1972).

To what extent the dA-rich regions detected by in situ hybridisation of BHK-21/C13 chromosomes correspond to these chromosome banding patterns is uncertain. A much greater level of radioactivity would have to be incorporated into hybrids to enable comparative studies.

In this study, I was unable to obtain satisfactory binding of [^{3}H] poly(C) to cytological preparations of BHK-21/C13 cells. Jones et al. (1973) were also unable to bind [^{3}H] poly(C) to Rhynchosciara chromosomes. It is therefore possible that there are particular technical problems involved in the in situ hybridisation of poly(C). For example, Kubinski et al. (1966) suggested that poly(C) binding will be promoted by acid conditions. It should be

noted that staining studies suggest that G + C rich regions are widely distributed throughout the chromosomes (Schreck et al., 1973). The relation of such regions to the dG-rich sequences observed in mammalian DNA is, of course, not known.

E. The levels of homopolymer-rich regions in different species

One of the striking features of this study was the observation that, in general, DNA from prokaryotes contains very low levels, if any, of the homopolymer-rich regions (Table III.7). This finding has a number of implications.

1. Since the prokaryotic DNAs studied ranged from those which were more A + T rich (e.g. C.perfringens) to those which were more G + C rich (e.g. M.lysodeikticus) than the eukaryotic DNAs, the binding of ribohomopolymer does not depend on the overall base composition but rather on the nucleotide sequence. This, therefore, is further evidence for the specificity of base-pairing within the hybrids which were formed.

The main exception to this general rule was that Herpes simplex DNA, which is about 67% G + C, was found to hybridise well with poly(C). The significance of this observation is not known.

2. The considerable increase in buoyant density after binding of poly(U) or poly(IG) to denatured C.perfringens, λ, T4 and T7 DNA (Kubinski et al., 1966), contrasts with the low levels of hybrids formed between poly(U) or poly(C) and these DNAs (Table III.7). This poor correlation supports the conclusions already reached (Discussion, Section IV,A.2) that the type of complex detected by the ribohomopolymer

binding technique of Kubinski et al. (1966) is not a true nucleic acid hybrid. The results obtained in such studies should not, therefore, be compared with the results obtained in the present work.

3. The similarity in nearest neighbour frequency patterns of eukaryotic DNAs and small DNA viruses has led to the hypothesis that certain small DNA viruses may have evolved from the eukaryotic genome (Subak-Sharpe, 1967). However, the levels of dA-rich and dG-rich regions in SV40 DNA is very much less than in eukaryotic DNAs (Table III.7). The absence of dT-rich regions has been confirmed in other studies on SV40 DNA (Weinberg et al., 1972; Aloni, 1973). These differences would not be expected to be demonstrable by nearest neighbour analysis, since the dA-rich and dG-rich regions even in eukaryotes account for such a small fraction of the total genome. Such differences suggest that these DNA viruses might not have evolved from the eukaryotic host. However, up to about 36,000 nucleotides might exist between two dA-rich or dG-rich regions (Discussion, Section IV.C) and SV40 DNA is only about 5,000 base-pairs long (Rush et al., 1971). It is, therefore, still possible that the SV40 DNA could have been derived from an area in the eukaryotic genome which did not contain any dA-rich or dG-rich regions.

F. Possible functions of homopolymer-rich sequences in DNA

The functions of the dA-rich, dG-rich, and the pure poly(dA) and poly(dG) sequences is, as yet, unknown. It is, however, interesting to speculate on the possible roles of these regions in mammalian DNA.

1. Do they have any function?

One possibility which must first be considered is that these sequences may have no definite function. For example, it might be argued that since the levels of dA-rich and dG-rich regions vary considerably in different eukaryotic species (Table III.7) they cannot have some essential function. As has already been discussed, certain regions of DNA may well consist of relatively useless "junk" DNA (Introduction, Section I.G). Such regions include most satellite DNAs and types of spacer DNA. However, satellite DNA sequences must be closely clustered to enable them to form a satellite of atypical base composition. The evidence presented in this study has demonstrated that neither dA-rich nor dG-rich regions are clustered to any significant extent (Discussion, Section IV.C), and that they certainly do not form satellite bands with what would be a most characteristic base composition (Fig. III.13).

The possibility that these regions may be found in a type of spacer DNA is very real. Although little is known about spacer DNA, the experimental evidence for the spacer DNA between ribosomal RNA genes and transfer RNA genes, indicates that it has a most atypical base composition. Thus, the spacer DNA between genes for 18S and 28S ribosomal RNA is about 69-73% G + C for Xenopus species (Brown et al., 1972), and about 77% G + C in HeLa cells (Jeanteur et al., 1968), whereas the spacer DNA between 5S RNA genes in Xenopus is about 70% A + T (Brown et al., 1971). The spacer DNA between transfer RNA genes in Xenopus is also A + T rich (Clarkson et al., 1973). The distribution of bases between the two strands of such spacer DNAs is not known, but it

would be most unlikely if more than 50% of any one strand consisted of just one nucleotide. One example of the base composition of individual strands of DNA molecules of such an atypical overall base composition is mouse satellite DNA (Flamm et al., 1969). Although the overall base composition is 67% A + T, one strand contains 46% A, the other 21% A.

If the homopolymer-rich sequences have no real function other than to contribute towards filling in gaps between sequences of informational DNA, then there would be little evolutionary constraint upon their sequences. It might therefore be expected that the observed frequency of occurrence of sequences of a particular size and base composition would reflect the estimated random occurrence of such sequences. Considering a random sequence of deoxynucleotides - if base X has probability p, then the number r of bases X in a random sequence of length n is binomial (n,p)

$$\text{i.e. } Pr(r) = \binom{n}{r} p^r (1 - p)^{n-r}$$

Table IV.1. shows the probability of the chance occurrence of sequences with different proportions of a single base (i.e. corresponding to different levels of mismatching of hybrids of ribohomopolymers) in particular lengths of polynucleotide. Two overall base compositions have been considered - one strand of the DNA being comprised of a single base to a level of 50% (very unlikely, but just possible in view of the atypical base composition of spacer DNAs (see above)), or a level of 30% (perhaps a more reasonable average). A stretch of DNA 3×10^9 nucleotides long, i.e. about 50% of one strand of mammalian DNA has been chosen arbitrarily as the maximum amount of spacer DNA which

Table IV.1.

The probability of the random occurrence of homopolymer-rich sequences of different sizes and different degrees of purity. For explanation see text.

Base composition of DNA	Length of sequence (n)	Probability of X = n	No. of sequences of $(X)_n$ expected in 3×10^9 bases	Probability of $X > 0.9n$ i.e. 10% mismatched	No. of sequences with $X > 0.9n$ expected in 3×10^9 bases	Probability of $X > 0.8n$ i.e. 20% mismatched	No. of sequences with $X > 0.8n$ expected in 3×10^9 bases	Probability of $X \geq 0.7n$ i.e. 30% mismatched	No. of sequences with $X \geq 0.7n$ expected in 3×10^9 bases
50% X	50	9.0×10^{-16}	0	2.1×10^{-9}	6	1.2×10^{-5}	3.6×10^{4}	3.3×10^{-3}	9.9×10^{6}
i.e. p=0.5	30	9.3×10^{-10}	3	4.2×10^{-6}	1.3×10^{4}	7.1×10^{-4}	2.1×10^{6}	2.1×10^{-2}	6.2×10^{7}
	20	9.7×10^{-7}	2.9×10^{3}	2.0×10^{-4}	6×10^{5}	5.9×10^{-3}	1.8×10^{7}	5.8×10^{-2}	1.8×10^{8}
	15	3.2×10^{-5}	9.6×10^{4}	4.9×10^{-4}	1.5×10^{6}	1.8×10^{-2}	5.4×10^{7}	1.5×10^{-1}	4.5×10^{8}
30% X	50	7.1×10^{-27}	0	1.1×10^{-18}	0	3.1×10^{-13}	0	6.4×10^{-9}	19
i.e. p=0.3	30	2×10^{-16}	0	1.1×10^{-11}	0	2.2×10^{-8}	66	7.3×10^{-6}	2.2×10^{4}
	20	3.5×10^{-11}	0	3.7×10^{-8}	1.1×10^{2}	5.5×10^{-6}	1.7×10^{4}	2.6×10^{-4}	7.8×10^{5}
	15	1.4×10^{-8}	42	5.2×10^{-7}	1.6×10^{3}	9.2×10^{-5}	2.8×10^{5}	3.6×10^{-3}	1.1×10^{7}

Experimental values for BHK-21/C13 DNA (6.5×10^9 base-pairs)

poly(dA) - average size 31 nucleotides long - 34,000 sequences

poly(dG) - average size 17 nucleotides long - 12,200 sequences

dA-rich - average size 50 nucleotides long - about 2-6% mismatched - 338,000 sequences

dG-rich - average size 25 nucleotides long - about 10-30% mismatched - 364,000 sequences.

might have such an atypical base composition. On this basis, only 3 pure poly(dA) sequences 30 nucleotides long would be expected. On the other hand, 338,000 dA-rich (70 - 80% dA) regions 50 nucleotides long would be found by chance in 3×10^9 nucleotides of random sequence and overall base composition 50% dA. However, the DNA·poly(U) hybrids from which the number of dA-rich regions was calculated (Fig. III.5) are probably only about 4% mismatched (Results, Section III.E). It is therefore most unlikely that the pure poly(dA) and dA-rich regions are solely a chance phenomenon. By contrast, 12,200 poly(dG) sequences 17 nucleotides long and 364,000 dG-rich regions (80% dG) some 20-30 nucleotides long would arise by chance if 50% (but not 30%) of the 3×10^9 bases were dG. It is therefore considerably more likely that the dG-rich and poly(dG) regions could have occurred by chance.

Two further lines of experimental evidence should also be considered. First, if dA-rich and dG-rich regions were produced by random mutations of some ancestral sequence, it would be expected that short dA-rich and dG-rich regions would be considerably more frequent than long regions. Moreover, a skewed distribution of dA-rich and dG-rich regions in the A + T rich fraction and G + C rich fraction respectively would be found. Neither result was the case for the dA-rich regions (Figs. III.7 and III.13). The dG-rich regions do, however, tend to follow this expected pattern.

It can be concluded that the dA-rich and poly(dA) regions at least are so plentiful and of such a size and distribution that they may well have some biological functions.

A proportion of the dG-rich and pure poly(dG) sequences may also play some significant role, but it is possible that the majority of such sequences do not. The above arguments, of course, do not explain the widely different levels of these regions found in different eukaryotes. One possible reason for this is that small variations in the size of the regions, or the interspersion of just 2-3 heterologous bases in a region say 50 nucleotides long would considerably affect the stability of hybrids. Thus, although the overall amount of dA-rich and dG-rich regions in the different species may be similar, under the hybridisation conditions used very different levels would be observed.

If one assumes that at least some of these sequences have a useful role, what might this role be? The widespread distribution of the homopolymer-rich sequences would be consistent both with the possibility that these sequences serve as regulatory regions which bind modulating elements so as to control transcription or replication of DNA sequences, and also with the possibility that these sequences might themselves be transcribed into RNA molecules. As pointed out by Georgiev (1969) these functions are not necessarily mutually exclusive, since the regulatory sequences might be transcribed into non-informational RNA. It is convenient, however, to consider the evidence for these two possibilities separately.

2. Binding of regulatory elements

The recent studies of Bram (1971) and Bram & Tougard (1972) have demonstrated that very A + T rich DNA has a different secondary structure from DNA of average base composition. Such an unusual structure would clearly

facilitate the recognition of such sites by incoming regulatory elements. In this context, it is of interest to note that some dA·dT-rich DNA in Rhynchosciara chromosomes may be made more available for complexing by being extended out from the chromosomal axis (Jones et al., 1973). Furthermore, Crick (1971) has postulated the presence of single-stranded DNA which is involved in binding regulatory substances. He suggested that unwinding of the DNA to produce such single-strands might be localised by the presence of dA·dT-rich regions.

(a) Binding of RNA polymerase

The hypothesis that pyrimidine-rich regions in DNA might serve as binding sites for RNA polymerase (i.e. promoters) was first proposed by Szybalski et al. (1966). This was based upon the observation that pyrimidine-rich clusters could be observed in a wide variety of DNAs, and also on the fact that the pyrimidine-rich clusters were present on the DNA strand transcribed into mRNA (Szybalski et al., 1966). This hypothesis was extended by Becker (1972) to include different types of homopolymer-rich regions. However, the evidence presented by these workers that RNA polymerase might bind to such sites is, at best, circumstantial. More concrete evidence comes from the demonstration that poly[d(A-T)] and A + T rich DNA are more effective templates for RNA synthesis using E.coli RNA polymerase, than natural DNA, G + C rich DNA, or poly(dG)·poly(dC) (Maitra et al., 1966; Chamberlin et al., 1970; Shishido & Ikeda, 1972). Moreover, E.coli RNA polymerase binds preferentially to A + T rich DNA sequences in bacteriophage λ (Le Talaer & Jeanteur, 1971), T5 and T7

(Le Talaer et al., 1973), f1 (Shishido & Ikeda, 1972) and fd DNAs (Heyden et al., 1972). However, the base composition of these promoter regions ranges from 57% to only 67% A + T, and pyrimidine tract analysis of the isolated promoter of fd DNA reveals the longest homopolymer to be only $(T)_5$ (Heyden et al., 1972). Moreover, bacteriophage λ DNA (at least) has very low levels of dA-rich regions (Table III.7). It seems likely therefore that the promoters in these prokaryotic organisms are moderately A + T rich DNA sequences, rather than being very dA-rich or very dT-rich. No promoters have yet been isolated from eukaryotic DNA. Hence, firm conclusions regarding the presence of dA-rich or dT-rich regions in the promoters of eukaryotic organisms cannot be made.

(b) Binding of DNA polymerase

G + C rich DNA has been shown to be a particularly good primer for DNA polymerase (Loeb, 1969). However, the nature of the DNA polymerase initiation sites in vivo are unknown.

(c) Binding of histones

Certain histones have a marked preference for binding to particular types of DNA. Thus, lysine-rich histones have a high affinity for A + T rich regions in DNA (Ohba, 1966; Mazen & Champagne, 1968; Combard & Vendrely, 1970), whereas arginine-rich histones have a higher affinity for G + C rich regions in DNA (Kleiman & Huang, 1971; Clark & Felsenfeld, 1972; Smythies et al., 1972). However, neither the exact base composition nor the nucleotide sequences of these binding sites is currently known. Hence the relationship of these sites to the deoxyhomopolymer-rich regions cannot currently be assessed.

(d) Binding of repressors

Both very A + T rich DNA and poly[d(A-T)] have a stronger affinity for the lac repressor in vitro than DNA of low or moderate A + T content (Lin & Riggs, 1970). However, recent sequence studies have demonstrated that although the lac operator is indeed A + T rich, homopolymeric regions are not present (Gilbert & Maxam, 1973).

(e) Binding of RNA species

It has been suggested that certain RNA species might be involved as elements in the control of transcription, by binding to double-stranded DNA (Bekhor et al., 1969; Britten & Davidson, 1969). There is, at present, little conclusive evidence either for the existence of such RNA species or for the three-stranded nucleic acid complexes postulated. It should, however, be mentioned that dA·dT-rich regions would probably be able to form a three-stranded complex of the form poly(dA).poly(dT).poly(U) (Riley et al., 1966; Table III.2 and III.9). It can therefore be speculated that RNA species containing a poly(U)-rich region might be able to combine with the dA·dT-rich regions in DNA and thus in some way modify the transcriptional activity at neighbouring regions in the DNA.

3. Transcription into homopolymer-rich RNA species

Over the past few years, the presence of sequences within RNA which are rich in a single type of nucleotide has been demonstrated. Two basic mechanisms for the biosynthesis of such RNA sequences can be envisaged. First, these sequences might arise by transcription of the complementary homopolymer-rich regions in DNA. Alternatively, a post-transcriptional mechanism might be employed, by which a

number of identical nucleotides are added to the end of newly transcribed RNA sequences. In this latter mechanism, there is no requirement for the involvement of homopolymer-rich regions in the DNA (Darnell et al., 1971a).

(a) Long sequences of poly(A)

Long tracts of poly(A), some 150-200 nucleotides long, are found at the 3'-end of most mRNA and some HnRNA molecules (Introduction, Section I.E.3). These sequences have been shown to consist of pure poly(A) (Molloy & Darnell, 1973; Brownlee et al., 1973). The mode of synthesis of these sequences has been studied in a number of ways.

(i) Kinetic studies using Actinomycin D have shown that HnRNA synthesis is inhibited to a much greater extent than poly(A) synthesis (Darnell et al., 1971a).

(ii) No pure poly(dT) sequences 150-200 nucleotides long exist in the mammalian genome (Fig. III.10; Birnboim et al., 1973). Comparable studies on the slime mould Dictyostelium have shown that the poly(A) tracts in its mRNA are about 100 nucleotides long, but no poly(dT) tracts longer than 25 nucleotides can be detected (Firtel & Lodish, 1973). It is also noteworthy that mRNA of SV40 (Weinberg et al., 1972), Herpes simplex (Silverstein et al., 1973), and adenovirus 2 (Philipson et al., 1971) all contain long tracts of poly(A), whereas the DNA of these species do not contain detectable dA-rich or dT-rich regions (Table III.7; Philipson et al., 1971; Weinberg et al., 1972). One possible exception to this general rule is vaccinia viral mRNA, the poly(A) segment of which might be produced by transcription from poly(dT) in the vaccinia DNA (Kates, 1970; Kates & Beeson, 1970).

(iii) Enzymes are present in the mammalian nucleus which will add a poly(A) tract to the 3' end of RNA molecules (Edmonds & Abrams, 1960; Burdon, 1963).

This evidence strongly suggests that the long tracts of poly(A) found in eukaryotic HnRNA and mRNA are synthesised by a post-transcriptional mechanism. The addition of the poly(A) region may therefore be a further step in modulating gene expression (Darnell et al., 1971a; 1973).

(b) Short sequences of poly(A)

Short sequences of poly(A) (or oligo(A)), some 20-30 nucleotides long, are found at some internal region of HnRNA molecules (Nakazato et al., 1973). These regions are present in amounts equimolar to the long poly(A) tracts. It is, however, not yet clear if both types of poly(A) tract are found in a single HnRNA molecule. Similar short poly(A) sequences have been observed in Dictyostelium precursor mRNA molecules (Lodish, 1973). Although no definite details of their mode of synthesis are yet available, it is significant that pure sequences of poly(dA) (and therefore presumably also poly(dT)) about 30 nucleotides long are present in the mammalian genome (Fig. III.10), and poly(dT) tracts 20-25 long are present in Dictyostelium DNA (Firtel & Lodish, 1973). These regions may well, therefore, be formed by a transcriptional mechanism. If these short poly(A) tracts are situated near the longer poly(A) sequences in the HnRNA molecules, it is possible that the short tracts may act as some type of primer for the post-transcriptional addition of long poly(A) sequences.

(c) Other A-rich regions

The size and purity of the poly(A) regions in haemoglobin

mRNA has been a matter of some dispute. Although Lim & Canellakis (1970) originally demonstrated that the poly(A) in rabbit haemoglobin mRNA is about 50-70 nucleotides long, and Mansbridge et al. (1973) found that two classes of pure poly(A) sequences of about 75 and 50 residues are present in mouse haemoglobin mRNA, duck haemoglobin mRNA has been shown to have a poly(A) tract 150 nucleotides long (Pemberton & Baglioni, 1972). On the other hand, Burr & Lingrel (1971) and Hunt (1973a) have found that only a very short oligo(A) segment, about 6-7 nucleotides long is found at the 3' end of the rabbit haemoglobin mRNA molecule. These differences may reflect, to some extent, different modes of preparation of mRNA, since the poly(A) tract becomes shorter with age (Sheiness & Darnell, 1973). Alternatively, the poly(A) tracts of haemoglobin mRNA may not be pure poly(A), but rather very A-rich (Hunt, 1973b).

If short regions of poly(A) (less than about 100 nucleotides long) are indeed found in newly synthesised haemoglobin mRNA, a transcriptional origin could be postulated for them (Fig. III.7 and III.10). A transcriptional mechanism would also seem more likely to account for the introduction of "non-A" bases into an A-rich region in RNA. Further sequence studies on haemoglobin mRNA and other specific mRNA species are clearly required to resolve these problems.

(d) U-rich regions

U-rich regions have been demonstrated in HnRNA of mammalian cells (Burdon & Shenkin, 1972; Molloy et al., 1972b). These regions are about 30 nucleotides long and may contain up to 20% of bases other than U (Molloy et al., 1972b). Studies on the kinetics of their synthesis, and their position

within the HnRNA molecule suggests a transcriptional origin (Burdon, personal communication). Such regions could, of course, be transcribed from the short dA-rich regions which have been demonstrated in this study.

With regard to a transcriptional mechanism for the synthesis of these short A-rich and U-rich regions in RNA, the speculation that very A + T rich DNA may not be transcribed should be borne in mind (Bram, 1972) (Introduction, Section I.B).

If these 30 nucleotide long regions of poly(A) and U-rich sequences are indeed formed by a transcriptional mechanism, further information regarding the proportion and size of HnRNA molecules carrying such sequences may permit calculation of the proportion of the genome which is transcribed into HnRNA. For example, if there is one short poly(A) sequence in each HnRNA molecule, and these sequences are pure poly(A), then the diploid BHK-21/C13 genome can code for a maximum of 34,000 HnRNA molecules. If nascent HnRNA molecules are, on average, 30,000 nucleotides long, 34,000 molecules would account for about 16% of the total genome. The figure of 34,000 is clearly less than the approximate value of 60,000 - 80,000 genes per diploid genome suggested by studies of mutation rate (Muller, 1967; Ohno, 1971; Ohta & Kimura, 1971). There are, however, very many more dA-rich than pure poly(dA) tracts (Discussion, Section IV.C), and hence, depending upon the exact structure of the homopolymer-rich regions in HnRNA, a larger proportion of the genome could be calculated to be transcribed into HnRNA.

(e) G-rich regions

Lim and coworkers (1970) demonstrated that polyribosomal

RNA contains, besides poly(A) tracts, segments which are very G-rich. These G-rich regions have not yet been fully characterised, but it is possible that they could arise by transcription of short dC-rich regions in the DNA.

(f) C-rich regions

It has recently been observed that oligo-dG is a successful primer for the synthesis of a DNA copy of immunoglobulin mRNA using reverse transcriptase (Diggleman et al., 1973). This suggests that a C-rich region might exist in the mRNA. If so, it almost certainly arises by transcription of a dG-rich region in the DNA (Diggleman et al., 1973).

4. Functions of homopolymer-rich regions in RNA

It seems most unlikely that long or short poly(A) tracts, or the U-rich or G-rich sequences in RNA are translated into protein, since this would imply the production of a protein which contains sequences of largely a single amino acid. Such sequences have not been detected (Kretsinger & Garmany, 1970). The C-rich region in immunoglobulin mRNA might well be translated, however, since the proline residues at positions 119-120 of the amino acid sequence, together with the neighbouring amino acids, could account for a short C-rich cluster 6-12 nucleotides long (Diggleman et al., 1973).

The role of the long sequences of poly(A) in HnRNA and mRNA is still uncertain. Since such poly(A) sequences are found not only in mammalian mRNAs, but also in plant mRNA (van de Walle, 1973; Manahan et al., 1973; Higgins et al., 1973), slime mould mRNA (Firtel et al., 1972) and yeast mRNA (McLaughlin et al., 1973; Reed & Wintersberger, 1973), it seems likely that the presence of poly(A) is necessary to

fulfil some essential role in all types of eukaryotic cells. One possibility stems from the observation that most of the poly(A) attached to HnRNA in the nucleoplasm appears to reach the cytoplasm as part of mRNA molecules (Jelinek et al., 1973). Moreover, if poly(A) synthesis is inhibited by use of the drug cordycepin (3'-deoxyadenosine), then the appearance of mRNA in polysomes is greatly reduced (Penman et al., 1970; Darnell et al., 1971a). It seems likely therefore that the poly(A) segment has some role in promoting transport of mRNA into the cytoplasm. Because poly(A) is also found in the mRNA of viruses which replicate in the cytoplasm, it is possible that poly(A) segments may have a further role in the process of translation of the mRNA (Kates, 1970; Yogo & Wimmer, 1972; Eaton et al., 1972). In this respect, it is interesting that the size of the poly(A) sequence becomes shorter with age (Mendecki et al., 1972; Sheiness & Darnell, 1973). Furthermore, specific proteins may be bound to the poly(A) segment of polysomal mRNA (Kwan & Brawerman, 1972; Blobel, 1973). Additional studies on the cytoplasmic functions, if any, of poly(A) are clearly required.

As regards the short poly(A) (oligo A) and U-rich regions of HnRNA there is no firm evidence for any particular function. They may comprise transcription products of regulatory sequences in DNA. On the other hand, they may themselves be control sites which are involved in the modification of HnRNA molecules by, for example, nucleases or polymerases. Little is also known of the function of the G-rich regions in polyribosomal RNA. One possibility is that these regions might be binding sites for particular proteins involved in the translation process.

In summary, it is probable that at least some of the homopolymer-rich and the pure homopolymer sequences observed in this study have a definite role to play in nucleic acid metabolism, either in binding regulatory elements or in being transcribed into RNA species containing homopolymer-rich sequences. It should be emphasised, however, that since there are such a large number of sequences present of different sizes and different degrees of dA-richness and dG-richness, sequences with different characteristics may have different functions. Nonetheless, at present, it is not possible to assign with certainty a particular function to any of these various types of homopolymer-rich region in DNA.

G. Future work

The work described in this report might be continued in a number of ways.

(1) By isolation of hybrids of mRNA with fragments of DNA, it might be possible to determine the position of the homopolymer-rich sequences in relation to structural genes.

(2) Purification of DNA·poly(U) hybrids to produce a preparation enriched for dA-rich sequences would permit studies on the binding of various proteins e.g. histones or acidic proteins.

(3) Studies using the endonuclease from Neurospora crassa would permit more accurate analysis of the properties of the pure deoxyhomopolymeric sequences.

(4) Preparations of [^{3}H] poly(U) and [^{3}H] poly(C) of higher specific activity could be used to facilitate study of in situ hybrids.

(5) The content of dA-rich and dG-rich regions in various species might be further investigated. In particular,

it would be of interest to compare DNA from normal cells with that from neoplastic or embryonic tissues.

REFERENCES

Adesnick, M., Salditt, M., Thomas, W. & Darnell, J.E. (1972) J. Mol. Biol. 71, 21-30

Adler, A.J., Schaffhausen, B., Langan, T.A. & Fasman, G.D. (1971) Biochemistry 10, 909-913

Allfrey, V.G. & Mirsky, A.E. (1963) Cold Spring Harbor Symp. Quant. Biol. 28, 247-262

Allfrey, V.G., Littau, V.C. & Mirsky, A.E. (1963) Proc. Nat. Acad. Sci. U.S.A. 49, 414-421

Allfrey, V.G., Faulkner, R. & Mirsky, A.E. (1964) Proc. Nat. Acad. Sci. U.S.A. 51, 786-793

Aloni, Y. (1973) Nature (London) New Biol. 243, 2-6

Amalric, F., Bernard, S. & Simard, R. (1973) Nature (London) New Biol. 243, 38-41

Arion, V.I. & Georgiev, G.P. (1967) Proc. Nat. Acad. Sci. U.S.S.R. 172, 716

Arrighi, F.E. & Hsu, T.C. (1971) Cytogenetics 10, 81-86

Arrighi, F.E., Hsu, T.C., Saunders, P. & Saunders, G.F. (1970) Chromosoma 32, 224-236

Barr, G.C. & Butler, J.A.V. (1963) Nature (London) 199, 1170-1172

Bautz, E.K.F. & Hall, B.D. (1962) Proc. Nat. Acad. Sci. U.S.A. 48, 400-408

Becker, Y. (1972) J. Theoret. Biol. 34, 47-59

Bekhor, I., Bonner, J. & Dahmus, G.K. (1969) Proc. Nat. Acad. Sci. U.S.A. 62, 271-277

Bell, D., Wilkie, N. & Subak-Sharpe, J.H. (1971) J. Gen. Virol. 13, 463-475

Bellet, A.J.D., Younghusband, H.B. & Primrose, S.B. (1973) Virology 50, 35-44

Bernardi, G. (1965) Nature (London) 206, 779-783

Bernhardt, D. & Darnell, J.E. (1969) J. Mol. Biol. 42, 43-56

Bick, M.D., Huang, H.L. & Thomas, C.A. (1973) J. Mol. Biol. 77, 75-84

Birnboim, H.C., Mitchel, R.E.J. & Strauss, N.A. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 2189-2192

Birnstiel, M.L., Spiers, J., Purdom, I., Jones, K. & Loening, U.E. (1968) Nature (London) 219, 454-463

Birnstiel, M.L., Sells, B.H. & Purdom, I.F. (1972) J. Mol. Biol. 63, 21-39

Bishop, J.O. & Rosbach, M. (1973) Nature (London) New Biol. 241, 204-207

Bishop, J.O., Pemberton, R. & Baglioni, C. (1972) Nature (London) New Biol. 235, 231-234

Blobel, G. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 924-928

Boedtker, H. (1968) J. Mol. Biol. 35, 61-70

Bolton, E.T. & McCarthy, B.J. (1962) Proc. Nat. Acad. Sci. U.S.A. 48, 1390-1397

Bonner, J., Dahmus, M.E., Famborough, D., Huang, R.C., Marushige, K. & Tuan, D.Y.H. (1968) Science 159, 47-56

Borst, P. & Grivel, L. (1971) FEBS Lett. 13, 73-88

Borst, P. & Ruttenberg, G.J.C.M. (1972) Biochim. Biophys. Acta 259, 313-320

Bostock, C.J. & Prescott, D.M. (1971) Exp. Cell Res. 64, 267-274

Bram, S. (1971) Nature (London) New Biol. 232, 174-176

Bram, S. (1972) Biochimie 54, 1003-1011

Bram, S. & Tougard, P. (1972) Nature (London) New Biol. 239, 128-131

Britten, R.J. & Davidson, E.H. (1969) Science 165, 349-358

Britten, R.J. & Kohne, D.E. (1968) Science 161, 529-540

Britten, R.J. & Smith, J. (1970) Carnegie Inst. Wash. Yearbook 68, 378-386

Britten, R.J., Pavich, M. & Smith, J. (1970) Carnegie Inst. Wash. Yearbook 68, 400-402

Brown, D.D. & Dawid, I.B. (1968) Science 160, 272-280

Brown, D.D. & Weber, C.S. (1968) J. Mol. Biol. 34, 661-680

Brown, D.D., Wensink, P.C. & Jordan, E. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 3175-3179

Brown, D.D., Wensink, P.C. & Jordan, E. (1972) J. Mol. Biol. 63, 57-73

Brownlee, G.G., Sanger, F. & Barrell, B.G. (1968) J. Mol. Biol. 34, 379-412

Brownlee, G.G., Cartwright, E.M., Cowan, N.J., Jarvis, J.M. & Milstein, C. (1973) Nature (London) New Biol. 244, 236-240

Burdon, R.H. (1963) Biochem. Biophys. Res. Commun. 11, 472-476

Burdon, R.H. (1967) Biochem. J. 104, 186-197

Burdon, R.H. (1971a) Biochim. Biophys. Acta 232, 359-370

Burdon, R.H. (1971b) Prog. Nucleic Acid Res. Mol. Biol. 11, 33-79

Burdon, R.H. & Clason, A.E. (1969) J. Mol. Biol. 39, 113-124

Burdon, R.H. & Douglas, J.T. (1974) Nucleic Acid Res. 1, 97-105

Burdon, R.H. & Garven, E.V. (1971) Biochim. Biophys. Acta 232, 371-378

Burdon, R.H. & Shenkin, A. (1972) FEBS Lett. 24, 11-14

Burdon, R.H., Martin, B.T. & Lal, B.M. (1967) J. Mol. Biol. 28, 357-371

Burr, H. & Lingrel, J.B. (1971) Nature (London) New Biol. 233, 41-43

Burton, K. & Petersen, G.B. (1960) Biochem. J. 75, 17-27

Busby, D.W.G., House, W. & McDonald, J.R. (1964) in Virological Techniques, 218 pp. Churchill: London

Callan, H.G. (1967) J. Cell Sci. 2, 1-7

Callan, H.G. & Lloyd, L. (1960) Phil. Trans. Roy. Soc. (London) 243, 135-219

Caspersson, T., Farber, S., Foley, G.E., Kubynowski, J., Modest, E.F., Simonsson, E., Wagh, U. & Zech, L. (1968) Exp. Cell Res. 49, 219-225

Chamberlin, M.J. (1965) Fed. Proc. 24, 1446-1457

Chamberlin, M., McGrath, J. & Waskell, L. (1970) Nature (London) 228, 227-231

de la Chapelle, A., Schroder, J., Selander, R-K. & Stenstrand, K. (1973) Chromosoma 42, 365-382

Clark, R.J. & Felsenfeld, G. (1972) Nature (London) New Biol. 240, 226-229

Clarke, P.R. & Byvoet, P.B. (1969) Fed. Proc. 28, 843

Clarkson, S.G., Birnstiel, M.L. & Purdom, I.F. (1973) J. Mol. Biol. 79, 411-430

Combard, H. & Vendrely, R. (1970) Biochem. J. 118, 875-881

Comings, D.E. (1972) Advan. Hum. Genet. 3, 237-431

Comings, D.E. & Berger, R.O. (1969) Biochem. Genet. 2, 319-333

Comings, D.E., Avelino, E., Okada, T.A. & Wyandt, H.E. (1973) Exp. Cell Res. 77, 469-493

Corneo, G., Ginelli, E. & Polli, E. (1970) J. Mol. Biol. 48, 319-327

Corneo, G., Zardi, L. & Polli, E. (1972) Biochim. Biophys. Acta 269, 201-204

Cox, E.C. (1972) Nature (London) New Biol. 239, 133-134

Crick, F.H.C. (1958) Symp. Soc. Exp. Biol. 12, 138-163

Crick, F.H.C. (1971) Nature (London) 234, 25-27

Dahmus, M.E. & Bonner, J. (1970) Fed. Proc. 29, 1255-1264

Darnell, J.E. (1968) Bact. Rev. 32, 262-290

Darnell, J.E., Philipson, L., Wall, R. & Adesnick, M. (1971a) Science 174, 507-510

Darnell, J.E., Wald, R. & Tushinski, R.T. (1971b) Proc. Nat. Acad. Sci. U.S.A. 68, 1321-1325

Darnell, J.E., Jelinek, W.R. & Molloy, G.R. (1973) Science 181, 1215-1221

Davidson, E.H., Hough, B.R., Amenson, C.S. & Britten, R.J. (1973) J. Mol. Biol. 77, 1-24

Davidson, J.N. (1972) The biochemistry of the nucleic acids, 396 pp. 7th Ed. Chapman & Hall: London

Davies, D.R. (1967) Ann. Rev. Biochem. 36, 321-364

Davies, H.G. & Small, J.V. (1968) Nature (London) 217, 1122-1125

Dawid, I.B., Brown, D.D. & Reeder, R.H. (1970) J. Mol. Biol. 51, 341-360

De Lange, R.J., Famborough, D.M., Smith, E.L. & Bonner, J. (1969) J. Biol. Chem. 244, 319-334

De Lange, R.J., Hooper, J.A. & Smith, E.L. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 882-884

Dev, V.G., Warburton, D., Miller, O.J., Miller, D.A., Erlanger, B.F. & Beiser, M. (1972) Exp. Cell Res. 74, 288-293

Diggleman, H., Faust, C.H. & Mach, B. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 693-696

Drets, M.E. & Shaw, M.W. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 2073-2077

Dupraw, E.J. (1970) DNA and chromosomes 340 pp. Holt, Rinehart & Winston: New York

Dupraw, E.J. & Bahr, G.F. (1969) Acta Cytol. 13, 188-205

Dutrillaux, B. & Lejeune, J. (1971) C.R. Acad. Sci. (Paris) 272, 2638-2640

Earle, W.R. (1943) J. Nat. Cancer Inst. 4, 165-212

Eaton, B.T., Donaghue, T.P. & Faulkner, P. (1972) Nature (London) New Biol. 238, 109-111

Edmonds, M. & Abrams, R. (1960) J. Biol. Chem. 235, 1142-1149

Edmonds, M. & Caramela, M.G. (1969) J. Biol. Chem. 244, 1314-1324

Edmonds, M., Vaughan, M.H. & Nakazato, H. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 1336-1340

Elgin, S.C.R., Froehner, S.C., Smart, J.E. & Bonner, J. (1971) Adv. Cell. Mol. Biol. 1, 1-57

Ellison, J.R. & Barr, H.J. (1972) Chromosoma 36, 375-390

Elton, R.A. (1973) Nature (London) New Biol. 243, 287-288

Evenson, D.P., Mego, W.A. & Taylor, J.H. (1972) Chromosoma 39, 225-235

Fabricius, H. (1621) in The Embryological Treatises of Fabricius. Translation by H.B. Adelman. Cornell University Press, 1942

Firtel, R.A. & Lodish, H.F. (1973) J. Mol. Biol. 79, 295-314

Firtel, R.A., Jacobson, A. & Lodish, H.F. (1972) Nature (London) New Biol. 239, 225-228

Flamm, W.G., McCallum, M. & Walker, P.M.B. (1967) Proc. Nat. Acad. Sci. U.S.A. 57, 1729-1734

Flamm, W.G., Walker, P.M.B. & McCallum, M. (1969) J. Mol. Biol. 40, 423-443

Flamm, W.G., Bernheim, N.J. & Brubaker, P.E. (1971) Exp. Cell Res. 64, 97-104

Frenster, J.H., Allfrey, V.G. & Mirsky, A.E. (1963) Proc. Nat. Acad. Sci. U.S.A. 50, 1026-1032

Gall, J.G. (1969) Genetics 61, 121-132

Gall, J.G. & Pardue, M.L. (1971) Methods in Enzymology Vol. XXI, 470-480. Ed. L. Grossman & K. Moldave. Acad. Press: New York

Gall, J.G., Cohen, E.H. & Polan, M.L. (1971) Chromosoma 33, 319-344

Georgiev, G.P. (1969) J. Theoret. Biol. 25, 473-490

Georgiev, G.P., Ananieva, L.N. & Kozlov, J.V. (1966) J. Mol. Biol. 22, 365-371

Georgiev, G.P., Ryskov, A.P., Coutelle, C., Mantieva, V.L. & Avakyan, E.R. (1972) Biochim. Biophys. Acta 259 259-283

Gilbert, W. & Maxam, A. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3581-3584

Gillespie, D. (1968) Methods in Enzymology Vol. XII B, 641-668. Eds. L. Grossman & K. Moldave. Acad. Press: New York

Gillespie, D. & Spiegelman, S. (1965) J. Mol. Biol. 12, 829-842

Gilmour, R.S. & Paul, J. (1969) J. Mol. Biol. 40, 137-139

Gilmour, R.S. & Paul, J. (1970) FEBS Lett. 9, 242-244

Gilmour, R.S. & Paul, J. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 3440-3442

Greenberg, J.R. & Perry, R.P. (1972) J. Mol. Biol. 72, 91-98

Guha, A. & Szybalski, W. (1968) Virology 34, 608-616

Gurdon, J.B. & Uehlinger, V. (1966) Nature (London) 210, 1240-1241

Hall, B.D. & Spiegelman, S. (1961) Proc. Nat. Acad. Sci. U.S.A. 47, 137-146

Harrison, P.R., Hell, A., Birnie, G.D. & Paul, J. (1972) Nature (London) 239, 219-221

Hay, E.D. & Revel, J.P. (1963) J. Cell Biol. 16, 29-51

Heitz, E. (1933) Z. Zellforsch. 20, 237-285

Hennig, W., Hennig, I. & Stein, H. (1970) Chromosoma 32, 31-63

Heyden, B., Nusslein, C. & Schaller, H. (1972) Nature (London) New Biol. 240, 9-12

Higgins, T.V.J., Mercer, J.F.B. & Goodwin, P.B. (1973) Nature (London) New Biol. 246, 68-70

Hirsch, M. & Penman, S. (1973) J. Mol. Biol. 80, 379-391

Hirschmann, S.Z. & Felsenfeld, G. (1966) J. Mol. Biol. 16, 347-358

Holmes, D.S. & Bonner, J. (1973) Biochemistry 12, 2330-2338

Holmes, D.S., Mayfield, J., Sander, G. & Bonner, J. (1972) Science 177, 72-74

Holmquist, R., Jukes, T.H. & Pangburn, S. (1973) J. Mol. Biol. 78, 91-116

Holoubek, V. (1966) J. Cell Biol. 31, 49A

Horgen, P.A. & Griffin, D.H. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 338-341

House, W. & Wildy, P. (1968) Lab. Pract. 14, 594-601

Houssais, J.F. & Attardi, G. (1966) Proc. Nat. Acad. Sci. U.S.A. 56, 616-623

Hradecna, Z. & Szybalski, W. (1967) Virology 32, 633-643

Huang, R.C. & Bonner, J. (1962) Proc. Nat. Acad. Sci. U.S.A. 48, 1216-1222

Huang, R.C. & Bonner, J. (1965) Proc. Nat. Acad. Sci. U.S.A. 54, 960-967

Huang, R.C. & Huang, P.C. (1969) J. Mol. Biol. 39, 365-378

Huang, R.C., Bonner, J. & Murray, K. (1964) J. Mol. Biol. 8, 54-64

Huang, R.C.C., Maheshwari, N. & Bonner, J. (1960) Biochem. Biophys. Res. Commun. 3, 689-694

Huberman, J.A. (1973) Ann. Rev. Biochem. 42, 355-378

Huberman, J.A. & Attardi, G. (1966) J. Cell Biol. 31, 95-105

Huberman, J.A. & Riggs, A.D. (1966) Proc. Nat. Acad. Sci. U.S.A. 55, 599-606

Hunt, J.A. (1973a) Biochem. J. 131, 315-325

Hunt, J.A. (1973b) Biochem. J. 131, 327-333

Ifft, J.B., Voet, D.H. & Vinograd, J. (1961) J. Phys. Chem. 65, 1138-1145

Iwai, K., Ishikawa, K. & Hayashi, H. (1970) Nature (London) 226, 1056-1058

Jacob, F. & Monod, J. (1961) J. Mol. Biol. 3, 318-356

Jeanteur, P., Amaldi, F. & Attardi, G. (1968) J. Mol. Biol. 33, 757-775

Jelinek, W., Adesnick, M., Salditt, M., Sheiness, D., Wall, R., Molloy, G., Philipson, L. & Darnell, J.E. (1973) J. Mol. Biol. 75, 515-532

Jones, K.W. & Robertson, F.W. (1970) Chromosoma 31, 331-345

Jones, K.W., Bishop, J.O. & Brito-da-Cunha, A. (1973) Chromosoma 43, 375-390

Kates, J. (1970) Cold Spring Harbor Symp. Quant. Biol. 35, 743-752

Kates, J. & Beeson, J. (1970) J. Mol. Biol. 50, 19-33

Kavenoff, R. & Zimm, B.H. (1973) Chromosoma 41, 1-27

Kedinger, C., Gissinger, F., Gniazdowski, M., Mandel, J. & Chambon, P. (1972) Europ. J. Biochem. 28, 269-276

Kennell, D.E. (1970) Prog. Nucleic Acid Res. Mol. Biol. 11, 259-301

King, J.L. & Jukes, T.H. (1969) Science 164, 788-798

Kit, S. (1961) J. Mol. Biol. 3, 711-716

Kizer, P.E. & Saunders, G.F. (1972) Biochemistry 11, 1562-1568

Kleiman, L. & Huang, R.C. (1971) J. Mol. Biol. 55, 503-521

Kleinsmith, L.J. & Allfrey, V.G. (1969) Biochim. Biophys. Acta 175, 123-135

Kohne, D.E. (1968) Biophys. J. 8, 1104-1118

Koslov, Y.V. & Georgiev, G.P. (1970) Nature (London) 228, 245-247

Kretsinger, R.H. & Garmany, G.P. (1970) Biochim. Biophys. Acta 224, 651-653

Kubinski, H. & Rose, J.A. (1967) Proc. Nat. Acad. Sci. U.S.A. 57, 1720-1725

Kubinski, H., Opara-Kubinska, Z. & Szybalski, W. (1966) J. Mol. Biol. 20, 313-329

Kunitz, M. (1940) J. Gen. Physiol. 24, 15-32

Kurnit, D.M., Schildkraut, C.L. & Maio, J.J. (1972) Biochim. Biophys. Acta 259, 297-312

Kwan, S. & Brawerman, G. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 3247-3250

Laird, C.D. (1971) Chromosoma 32, 378-406

Lane, C.D., Gregory, C.M., Iyazumi, T. & Scherrer, K. (1973) Nature (London) New Biol. 243, 78

Langan, T.A. (1969) J. Biol. Chem. 244, 5763-5765

Laskey, R.A. & Gurdon, J.B. (1970) Nature (London) 228, 1332-1334

Lee, C.S. & Thomas, C.A. (1973) J. Mol. Biol. 77, 25-42

Lee, S.Y., Mendecki, J. & Brawerman, G. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 1331-1335

Le Talaer, J.Y. & Jeanteur, P. (1971) FEBS Lett. 12, 253-256

Levy, S. & Simpson, R.T. (1973) Nature (London) New Biol. 241, 139-141

Lim, L. & Canellakis, E.S. (1970) Nature (London) 227, 710-712

Lim, L., Canellakis, Z.N. & Canellakis, E.S. (1970) Biochim. Biophys. Acta 209, 128-138

Lin, S.Y. & Riggs, A.D. (1970) Nature (London) 228, 1185-1186

Lindberg, U. & Darnell, J.E. (1970) Proc. Nat. Acad. Sci. U.S.A. 65, 1089-1096

Littau, V.C., Allfrey, V.G., Frenster, J.H. & Mirsky, A.E. (1964) Proc. Nat. Acad. Sci. U.S.A. 52, 93-100

Lodish, H.F. (1973) Personal Communication

Loeb, A.L. (1969) J. Biol. Chem. 244, 1672-1681

Loening, U.E. (1967) Biochem. J. 102, 251-257

Loening, U.E. (1969) Biochem. J. 113, 131-138

McCarthy, B.J. & Church, R.B. (1970) Ann. Rev. Biochem. 39, 131-150

McCarthy, B.J. & Hoyer, B.H. (1964) Proc. Nat. Acad. Sci. U.S.A. 52, 915-922

McLaughlin, C.S., Warner, J.R., Edmonds, M., Nakazato, H. & Vaughan, M.H. (1973) J. Biol. Chem. 248, 1466-1471

McPherson, I.A. & Stoker, M.G.P. (1962) Virology 16, 147-151

Maden, B.E.H. (1971) Prog. Biophys. Mol. Biol. 22, 127-177

Maio, J.J. (1971) J. Mol. Biol. 56, 579-595

Maitra, U., Cohen, S.N. & Hurwitz, J. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 113-122

Manahan, C.O., App, A.A. & Still, C.S. (1973) Biochem. Biophys. Res. Commun. 53, 588-595

Mandell, J.D. & Hershey, A.D. (1960) Analyt. Biochem. 1, 66-77

Mansbridge, J.N., Crossley, J.A., Lanyon, W.G. & Williamson, R. (1973) Biochem. Soc. Trans. 1, 588-589

Markham, R. & Smith, J.D. (1952) Biochem. J. 52, 552-565

Marmur, J. (1961) J. Mol. Biol. 3, 208-218

Marshall, R. (1972) Chromosoma 37, 395-404

Marshall, S. & Gillespie, D. (1972) Nature (London) New Biol. 240, 43-45

Mattoccia, E. & Comings, D.E. (1971) Nature (London) New Biol. 229, 175-176

Mazen, A. & Champagne, M. (1968) Bull. Soc. Chim. Biol. 50, 1601-1603

Melli, M. & Pemberton, R.E. (1972) Nature (London) New Biol. 236, 172-174

Mendecki, J., Lee, S.Y. & Brawerman, G. (1972) Biochemistry 11, 792-798

Michelson, A.M., Massoulie, J. & Guschlbauer, W. (1967) Prog. Nucleic Acid Res. Mol. Biol. 6, 83-141

Milman, G., Langridge, R. & Chamberlin, M.J. (1967) Proc. Nat. Acad. Sci. U.S.A. 57, 1804-1810

Mirsky, A.E. & Ris, H. (1951) J. Gen. Physiol. 34, 475-492

Miyazawa, Y. & Thomas, C.A. (1965) J. Mol. Biol. 11, 223-237

Molloy, G.R. & Darnell, J.E. (1973) Biochemistry 12, 2324-2329

Molloy, G.R., Sporn, M., Kelley, D.E. & Perry, R.P. (1972a) Biochemistry 11, 3256-3260

Molloy, G.R., Thomas, W.L. & Darnell, J.E. (1972b) Proc. Nat. Acad. Sci. U.S.A. 69, 3684-3688

Mowshowitz, D.B. (1970) J. Mol. Biol. 50, 143-157

Muller, H.J. (1967) in Heritage from Mendel p. 419 Ed. R.A. Brink. Univ. of Wisconsin Press

Mushynski, W.E. & Spencer, J.H. (1970a) J. Mol. Biol. 52, 91-106

Mushynski, W.E. & Spencer, J.H. (1970b) J. Mol. Biol. 52, 107-120

Nakazato, H., Kopp, D.W. & Edmonds, M. (1973) J. Biol. Chem. 248, 1472-1476

Nygaard, A.P. & Hall, B.D. (1964) J. Mol. Biol. 9, 125-142

O'Brien, S.J. (1973) Nature (London) New Biol. 242, 52-54

Ogawa, Y., Quagliarotti, G., Jordan, J., Taylor, C.W., Starbuck, W.C. & Busch, H. (1969) J. Biol. Chem. 244, 4387-4392

Ohba, Y. (1966) Biochim. Biophys. Acta 123, 84-90

Ohno, S. (1971) Nature (London) 234, 134-137

Ohta, T. & Kimura, M. (1971) Nature (London) 233, 118-119

Opara-Kubinska, Z., Kubinski, H. & Szybalski, W. (1964) Proc. Nat. Acad. Sci. U.S.A. 52, 923-930

Pardon, J.F. & Wilkins, M.H.F. (1972) J. Mol. Biol. 68, 115-124

Pardue, M.L. & Gall, J.G. (1970) Science 168, 1356-1358

Patel, G., Howk, R. & Wang, T.Y. (1967) Nature (London) 215, 1488-1489

Paul, J. (1971) Symp. Soc. Exp. Biol. 25, 117-126

Paul, J. (1972) Nature (London) 238, 444-446

Paul, J. & Gilmour, R.S. (1968) J. Mol. Biol. 34, 305-316

Pemberton, R.E. & Baglioni, C. (1972) J. Mol. Biol. 65, 531-535

Pene, J.J., Knight, E. & Darnell, J.E. (1968) J. Mol. Biol. 33, 609-623

Penman, S., Rosbach, M. & Penman, M. (1970) Proc. Nat. Acad. Sci. U.S.A. 67, 1878-1885

Perkowska, E., Macgregor, H.C. & Birnstiel, M.L. (1968) Nature (London) 217, 649-650

Petes, T.D. & Fangman, W.L. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 1188-1191

Philipson, L., Wall, R., Glickman, G. & Darnell, J.E. (1971) Proc. Nat. Acad. Sci. U.S.A. 68, 2806-2809

Pilet, J. & Brahms, J. (1972) Nature (London) New Biol. 236, 99-100

Platz, D., Kish, V.M. & Kleinsmith, L.J. (1970) FEBS Lett. 12, 38-40

Pyeritz, R.E. & Thomas, C.A. (1973) J. Mol. Biol. 77, 57-73

Quetier, F. & Vedel, F. (1973) Biochem. Biophys. Res. Commun. 54, 1326-1334

Rall, S.L. & Cole, R.D. (1971) J. Biol. Chem. 246, 3589-3591

Reed, J. & Wintersberger, E. (1973) FEBS Lett. 32, 213-217

Reeder, R.H. & Brown, D.D. (1970) J. Mol. Biol. 51, 361-377

Richards, B.M. & Pardon, J.F. (1970) Exp. Cell Res. 62, 184-196

Richardson, C.C. (1966) in Procedures in Nucleic Acid Research Vol. 1, p. 212-223 Ed. G.L. Cantoni & D.R. Davies. Harper & Row: New York

Riley, M., Maling, B. & Chamberlin, M.J. (1966) J. Mol. Biol. 20, 359-389

Roberts, J.W. (1969) Nature(London) 224, 1168-1174

Roeder, R.G. & Rutter, W.J. (1970) Proc. Nat. Acad. Sci. U.S.A. 65, 675-682

Rubin, G.M. (1973) J. Biol. Chem. 248, 3860-3875

Rudner, R., Le Doux, M. & Mazelis, A. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 2745-2749

Rush, M.G., Kleinschmidt, A.K., Hellman, W. & Warner, R.C. (1967) Proc. Nat. Acad. Sci. U.S.A. 58, 1676-1683

Rush, M.G., Eason, R. & Vinograd, J. (1971) Biochim. Biophys. Acta 228, 585-594

Rust, P. (1970) Biochem. Biophys. Res. Commun. 39, 455-460

Scherrer, K. & Marcaud, L. (1965) Bull. Soc. Chim. Biol. 47, 1697-1713

Schildkraut, C. & Lifson, S. (1965) Biopolymers 3, 195-208

Schildkraut, C., Marmur, J. & Doty, P. (1962) J. Mol. Biol. 4, 430-443

Schmidt, G. & Thannhauser, S.J. (1945) J. Biol. Chem. 161, 83-89

Schnedl, W. (1971) Nature (London) New Biol. 233, 93-94

Schreck, R.R., Warburton, D., Miller, O.J., Beiser, S.M. & Erlanger, B.F. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 804-807

Sekeris, C.E., Sekeri, K.E. & Gallwitz, D. (1967) Z. Physiol. Chem. 348, 1660-1666

Selander, R-K. & de la Chapelle, A. (1973) Nature (London) New Biol. 245, 240-244

Shannon, M.P., Kaufman, T.C. & Judd, B.H. (1970) Genetics 64, s58

Shapiro, H.S. (1967) Methods in Enzymology Vol. XIIA, 205-212 Eds. L. Grossman & K. Moldave. Acad. Press: New York

Sheiness, D. & Darnell, J.E. (1973) Nature (London) New Biol. 241, 265-268

Sheldrick, P. & Szybalski, W. (1967) J. Mol. Biol. 29, 217-228

Shishido, K. & Ikeda, Y. (1972) FEBS Lett. 25, 353-356

Silverstein, S., Bachenheimer, S.L., Frenkel, W. & Roizman, B. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 2101-2104

Smart, J.E. & Bonner, J. (1971) J. Mol. Biol. 58, 675-684

Smillie, E.J. & Burdon, R.H. (1970) Biochim. Biophys. Acta 213, 248-250

Smythies, J.R., Benington, F. & Morin, R.D. (1972) J. Theoret. Biol. 37, 151-158

Sneider, T.W. (1971) J. Biol. Chem. 246, 4774-4783

Soeiro, R., Birnboim, H.C. & Darnell, J.E. (1966) J. Mol. Biol. 19, 362-372

Solari, A.J. (1971) Exp. Cell Res. 67, 161-170

Southern, E.M. (1970) Nature (London) 227, 794-798

Southern, E.M. (1971) Nature (London) New Biol. 232, 82-85

Spelsberg, T.C. & Hnilica, L.S. (1971) Biochim. Biophys. Acta 228, 202-211

Spencer, J.H. & Chargaff, E. (1963) Biochim. Biophys. Acta 68, 18-27

Stedman, E. & Stedman, E. (1950) Nature (London) 166, 780-781

Stevens, R.H. & Williamson, A.R. (1972) Nature (London) 239, 143-146

Stevens, R.H. & Williamson, A.R. (1973) Nature (London) New Biol. 245, 101-103

Studier, F.W. (1965) J. Mol. Biol. 11, 373-390

Sturtevant, A.H. (1951) in Genetics in the 20th Century, p. 101 Ed. L.C. Dunn. McMillan: New York

Subak-Sharpe, J.H. (1967) Brit. Med. Bull. 23, 161-168

Sugano, N., Olson, M.O.J., Yeoman, L.C., Johnson, B.R., Taylor, C.W., Starbuck, W.C. & Busch, H. (1972) J. Biol. Chem. 247, 3589-3591

Summers, W.C. & Szybalski, W. (1968) Virology 34, 9-16

Sumner, A.T., Evans, H.J. & Buckland, R.A. (1971) Nature (London) New Biol. 232, 31-32

Sutton, W.D. (1972) Nature (London) New Biol. 237, 70-71

Sutton, W.D. & McCallum, M. (1971) Nature (London) New Biol. 232, 83-85

Suzuki, Y., Gage, L.P. & Brown, D.D. (1972) J. Mol. Biol. 70, 637-649

Szybalski, W.C., Kubinski, H. & Sheldrick, P. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 123-127

Tapiero, H., Caneva, R. & Schildkraut, C.L. (1972) Biochim. Biophys. Acta 272, 350-360

Teng, C.S., Teng, C.T. & Allfrey, V.G. (1971) J. Biol. Chem. 246, 3597-3609

Thomas, C.A. (1971) Ann. Rev. Gen. 5, 237-256

Thomas, C.A. & Dancis, B.M. (1973) J. Mol. Biol. 77, 43-55

Thomas, C.A., Zimm, B.H. & Danois, B.M. (1973) J. Mol. Biol. 77, 85-99

Tischler, P.V. & Epstein, C.J. (1968) Anal. Biochem. 22, 89-98

Tobia, A.M., Schildkraut, C.L. & Maio, J.J. (1970). J. Mol. Biol. 54, 499-515

Tobler, H. (1972) in Nucleic Acid Hybridisation in the Study of Cell Differentiation, p. 1-9. Ed. H. Ursprung. Springer-Verlag: Berlin

Travers, A. (1971) Nature (London) New Biol. 229, 69-74

Travers, A. & Burgess, R.R. (1969) Nature (London) 222, 537-540

Tunis-Schneider, M.J.B. & Maestre, M.F. (1970) J. Mol. Biol. 52, 521-541

Uchida, T. & Egami, F. (1967) in Methods in Enzymology Vol. XIIA, p. 239-247. Eds. L. Grossman & K. Moldave Acad. Press: New York

Ullman, J.S. & McCarthy, B.J. (1973) Biochim. Biophys. Acta 294, 416-424

Volkin, E. & Cohn, W.E. (1953) J. Biol. Chem. 205, 767-782

Wainwright, S.D. (1972) Control Mechanisms in Protein Synthesis, p. 344-392 Columbia Univ. Press

Walker, P.M.B. (1969) Prog. Nucleic Acid Res. Mol. Biol. 9, 301-326

Walker, P.M.B. & McLaren, A. (1965) Nature (London) 208, 1175-1179

Wall, R. & Darnell, J.E. (1971) Nature(London) New Biol. 232, 73-76

van de Walle, C. (1973) FEBS Lett. 34, 31-34

Waring, M. & Britten, R.J. (1966) Science 154, 791-794

Warner, J., Soeiro, R., Birnboim, H.C. & Darnell, J.E. (1966) J. Mol. Biol. 19, 349-361

Watson, J.D. & Crick, F.H.C. (1953) Nature (London) 171, 737-738

Weinberg, R.A. (1973) Ann. Rev. Biochem. 42, 329-354

Weinberg, R.A. & Penman, S. (1970) J. Mol. Biol. 47, 169-178

Weinberg, R.A., Loening, U.E., Willems, M. & Penman, S. (1967) Proc. Nat. Acad. Sci. U.S.A. 58, 1088-1095

Weinberg, R.A., Ben-Ishai, Z. & Newbold, J.E. (1972) Nature (London) New Biol. 238, 111-113

Weisblum, B. & de Haseth, P.L. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 629-632

Weiss, S.B. (1960) Proc. Nat. Acad. Sci. U.S.A. 46, 1020-1030

Wilkie, N., Ustacelebi, S. & Williams, J.F. (1973) Virology 51, 499-503

Williamson, R. (1972) J. Med. Genet. 9, 348-355

Williamson, R., Drewienkiewicz, C.E. & Paul, J. (1973) Nature (London) New Biol. 241, 66-68

Yanofsky, S.A. & Spiegelman, S. (1962) Proc. Nat. Acad. Sci. U.S.A. 48, 1069-1078

Yasmineh, W.G. & Yunis, J.J. (1971) Exp. Cell Res. 64, 41-48

Yogo, Y. & Wimmer, E. (1972) Proc. Nat. Acad. Sci. U.S.A. 69, 1877-1882

Zubay, G. & Doty, P. (1959) J. Mol. Biol. 1, 1-20