PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

DNA-RNA Hybridization

J. O. Bishop, J. S. Beckman, M. S. Campo, N. D. Hastie, M. Izquierdo and S. Perlman

Phil. Trans. R. Soc. Lond. B 1975 **272**, 147-157
doi: 10.1098/rstb.1975.0077

Phil. Trans. R. Soc. Lond. B. **272**, 147–157 (1975) [147]
Printed in Great Britain

DNA–RNA hybridization

By J. O. Bishop, J. S. Beckmann, M. S. Campo, N. D. Hastie, M. Izquierdo and S. Perlman
Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN

Interest in nucleic acid hybridization stems mainly from its great power as a tool in biological research. It is used in several quite distinct ways. Because of the high degree of specificity that they show, hybridization techniques can be used to measure the amount of one specific sequence within a very heterogeneous mixture of sequences. Measurements of $1/10^6$–10^7 have been recorded. In extension of this, various properties of a specific sequence can often be studied. Secondly, because the kinetics of nucleic acid hybridization are quite well understood, it can be used to characterize both a pure sequence and a very complex mixture of sequences, like the genome of a vertebrate. Thirdly, again because of its specificity, it can be used to measure homologies between different populations of nucleic acids. Lastly, in conjunction with other techniques, it can be used as a basis for the fractionation of nucleic acid populations and the purification of specific sequences. Specific examples of these applications are given, with special reference to the organization of the genome in higher eukaryotes.

Introduction

As a tool for the study of the nucleic acids of multicellular organisms, DNA–RNA hybridization has matured only recently. True, the technique found early application in studies of ribosomal DNA (see Birnstiel, Chipchase & Spiers 1971). Only recently, however, have techniques emerged which allow meaningful measurements to be made on non-repetitive DNA sequences and their RNA products and on the relations between non-repetitive and moderately repetitive DNA sequences. The turning point was the work of Wetmur & Davidson (1968) and more particularly of Britten & Kohne (1968) which showed unequivocally that eukaryotic DNA contains both repetitive and non-repetitive DNA sequences, and also that the non-repetitive sequences are represented about once per haploid eukaryote genome.

The importance of these discoveries was threefold. They showed that repetitive sequences form a prominent, and therefore possibly important part of the genome of the vast majority of eukaryotes. They showed also that eukaryotes contain a significant proportion of non-repetitive DNA sequences. Lastly, they showed that there is an inverse relation between the sequence complexity and the reaction rate of nucleic acids. At that time this was not widely recognized, with the result that erroneous conclusions had been drawn from a large number of quite inadequate DNA–RNA hybridization experiments. This situation quickly changed. However, only with the advent of new methods did significant progress come about. The most important of these new methods would seem to be (*a*) the hybridization of trace amounts of labelled DNA with an excess of unlabelled RNA (Kohne 1968; Davidson & Hough 1971; Bishop, Morton, Rosbash & Richardson 1974) and (*b*) the hybridization of trace amounts of labelled RNA with an excess of unlabelled DNA (Melli *et al.* 1971; Gelderman, Rake & Britten 1971; Greenberg & Perry 1971; Bishop 1972).

The complexity rule

The overriding advantage of these reactions, which are carried out entirely in solution, is that they follow simple kinetic rules. In particular they obey the *complexity rule*, referred to above, which states that, other things being equal, the rate of a hybridization reaction is inversely proportional to the complexity of the majority reagent. As a result, important conclusions can be drawn from their kinetics. An example of this is given in figure 1, which shows two pseudo-first-order reactions driven by an excess, respectively, of rabbit haemoglobin messenger RNA (Hb mRNA) and RNA from poliomyelitis virus (polio RNA). In each case the labelled reagent was a partial DNA copy (cDNA) of the same RNA, synthesized by means of avian myeloblastosis virus reverse transcriptase. The reactions approximate first-order form, so that rate-constants (k_2) may be calculated with a measure of confidence. These are shown in table 1.

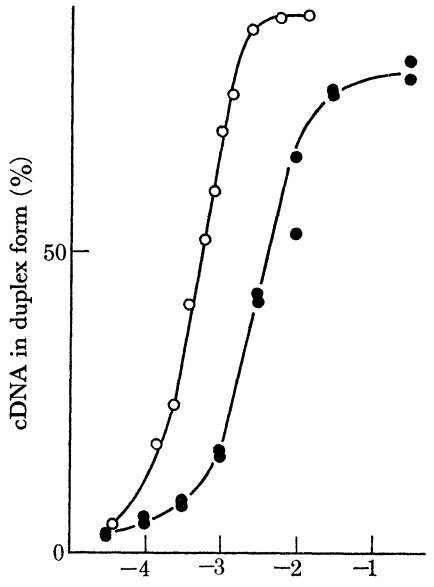

FIGURE 1. Kinetic standard for RNA-driven hybridization experiments. RNA purification, synthesis of cDNA and conditions for the assay of RNA–cDNA duplex using nuclease S1 are described by Bishop & Freeman (1974). The reactions were carried out in 1 mM EDTA–0.24 M phosphate buffer containing equimolar amounts of Na_2HPO_4 and NaH_2PO_4 (PEB) at 70 °C. ○, reaction between rabbit Hb mRNA and complementary cDNA; ●, reaction between polio mRNA and complementary cDNA.

TABLE 1. KINETIC STANDARDS FOR RNA-DRIVEN HYBRIDIZATION REACTIONS

RNA	complexity (N) nucleotides	length of tracer (L) nucleotides	$R_0t_{\frac{1}{2}}$ mol s l^{-1}	k_2 l mol^{-1} s^{-1}	k_n l mol^{-1} s^{-1}
rabbit Hb mRNA	1320	470	5.8×10^{-4}	1190	7.2×10^4
polio RNA	8120	375	2.8×10^{-3}	246	10.3×10^4
EMC RNA	8120	375	4.1×10^{-3}	168	7.0×10^4
				average	8.2×10^4

The lengths of the cDNA tracers were measured by sedimentation in alkaline sucrose gradients (Bishop & Freeman 1974). Values of $R_0t_{\frac{1}{2}}$ (the R_0t value at which the transition mid-point occurs) are taken from figure 1, and from similar experiments with encephalomyelitis virus (EMC) RNA. k_2 is calculated from the relation: $k_2 = 0.69/R_0t_{\frac{1}{2}}$ (Bishop 1972) and $k_n = k_2N/L^{\frac{1}{2}}$ (Hutton & Wetmur 1973).

The nucleation rate-constant, k_n, is a measure of the rate of duplex formation per nucleotide, and also takes account of variations in k_2 due to differences in the lengths (L) of the smaller molecular species, in this case the cDNA. k_n is approximately constant for the three RNA preparations, with a mean value of 8.2×10^4. This is equivalent to 9.2×10^4 in 0.4 M Na^+, which compares with Hutton & Wetmur's (1973) value of 13×10^4 for ϕX174 DNA and complementary RNA, and an average value of $16–17 \times 10^4$ for the renaturation of DNA from a number of different sources (Hutton & Wetmur 1973).

Specificity in hybridization reactions

It is well known that the specificity of nucleic acid interactions depends upon the reaction conditions. At temperatures near the rate optimum in any given case, on the order of 10% mismatching of base-pairs is tolerated (see, for example, Bonner, Bremner, Neufeld & Britten 1973). The rate of renaturation of sequences mismatched to this extent is reduced by a factor of only two relative to the renaturation of perfectly matched sequences. Thus, sequences differing in less than 10% of their bases will certainly form duplexes under the usual reaction conditions. The extent to which mismatching is tolerated can be reduced by reducing the concentration of monovalent cation or increasing the temperature. The penalty for this, however, is a fall in the reaction rate which in many instances, particularly in eukaryote systems, would be quite unacceptable.

Fortunately, the problem often does not arise. In many applications there will be no sequences present which cross-react with the sequences under study. The examples which follow suggest that this is true of the DNA sequences which encode the structure of adult haemoglobin both in the duck and in the rabbit.

Estimating and detecting DNA sequences which form a minute part of the genome

Until recently, estimates of the number of DNA sequences which encode the structure of haemoglobin were rather extravagant. In 1972 Bishop, Pemberton & Baglioni and Bishop & Pemberton showed that the repetition frequency of haemoglobin genes in the duck is less than 5, and this was rapidly confirmed, with a reduction in frequency, both in the duck (Packman, Aviv, Ross & Leder 1972; Bishop & Rosbash 1973) and in the mouse (Harrison, Hell, Birnie & Paul 1972). A further refinement in technique (Bishop & Freeman 1974) allows us to assert that in the duck and the rabbit one structural gene encodes each adult haemoglobin polypeptide chain. The critical experiments are shown in figure 2. These experiments utilize cDNA probes prepared against highly purified Hb mRNA. Different amounts of probe were annealed with a constant amount of homologous DNA to very high C_0t values, 1000 in the case of the duck, 4000 in that of the rabbit. C_0t means the product of initial DNA concentration (C_0) and time (t) (mol s l^{-1}). Under our annealing conditions the C_0t at which non-repeated DNA sequences are expected to be half renatured ($C_0t_{\frac{1}{2}}$) is about 400 for the duck and about 900 for the rabbit. The $C_0t_{0.9}$, at which 90% of non-repeated sequences are renatured, is an order of magnitude higher, 4000 and 9000 for duck and rabbit respectively. However, as the ratio of the cDNA probe to its DNA complement rises, the rate of the reaction increases and it terminates earlier. The cDNA probe is single-stranded, but the cellular DNA contains both complementary strands. Thus, as the ratio of cDNA to cellular DNA is increased, the ratio of one complement to the other is increased; the proportion of the major complement which can find a minor complement falls, but at the same time the specific radioactivity of the major

complement rises, since the cDNA is labelled and the cellular DNA is not. The outcome of these considerations is the relation $P = xD_0^c/(1-x)$ where P represents the amount of minor complement (complementary to the cDNA sequence), x the proportion of cDNA in duplex form, and D_0^c the amount of cDNA added to the reaction (Bishop & Freeman 1974). This relation applies strictly only if the reaction has effectively gone to completion. Since we have chosen conditions (C_0t) in which the renaturation of the DNA is incomplete, the reaction will go to completion only

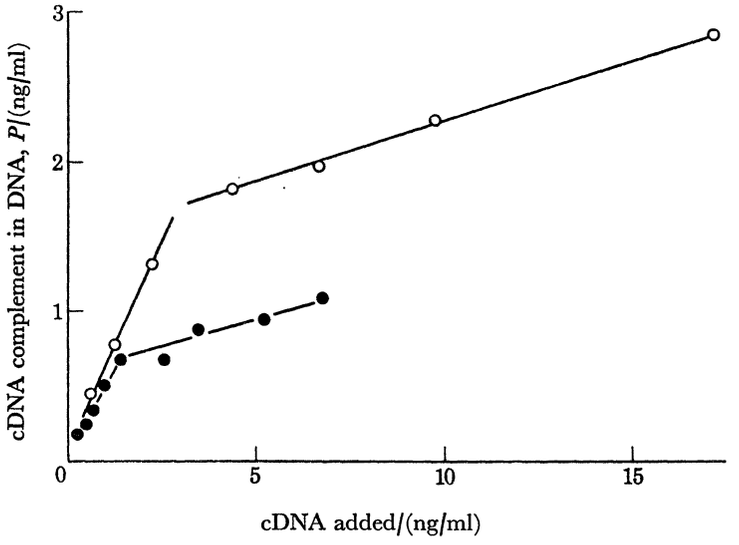

FIGURE 2. Titration of complementary cellular DNA sequences with a moderate excess of cDNA prepared against duck and rabbit Hb mRNA.

Reactions were carried out and assayed under the conditions described in figure 1. Duck and rabbit cellular DNA were at concentrations of 6.05 mg/ml and 5 mg/ml, and the samples were annealed, respectively, for 24 and 72 h. The mean sizes of the cDNA molecules are given in table 2.

TABLE 2. NUMBER OF STRUCTURAL GENES ENCODING EACH POLYPEPTIDE CHAIN OF ADULT DUCK AND RABBIT HAEMOGLOBIN

(C denotes the analytical complexity of the DNA in question; D_0, the DNA concentration; l, the estimate of the amount of DNA per genome which is complementary to the cDNA is given by $l = CxD_0^c/1-x)D$, where x is taken at the transition points seen in figure 2.)

	duck		
	expt. 1	expt. 2	rabbit
l ($\times 10^{-5}$)	3.2	2.6	2.9
M ($\times 10^{-5}$)	1.7	1	1.5
$1/M$	1.9	2.6	1.9

when the majority strand is in some excess. At this point, if the probe were completely pure (i.e. complementary only to Hb mRNA) we would expect the function $xD_0^c/(1-x)$ to maintain a constant value, irrespective of increase in D_0^c. As figure 2 shows, this expectation is not quite fulfilled. What happens is that $kD_0^c/(1-x)$ rapidly reaches a transition point beyond which increases in D_0^c produce much smaller increases in $xD_0^c/(1-x)$. We attribute this second phase to sequences contaminating the cDNA preparation, due to minor RNA species contaminating even these highly purified Hb mRNA preparations.

With this provision, the experimental results point to the existence of a single DNA sequence specifying each haemoglobin polypeptide chain in both the duck and the rabbit. Adult duck

haemoglobin has three distinct polypeptide chains, two α-type and one β-type and analysis of the sort shown in figure 1 (unpublished) confirms this. Adult rabbit haemoglobin has two distinct polypeptide chains. The analysis of figure 1, which is based on this assumption, rather confirms it. Table 2 shows, then, that in both the rabbit and the duck, a single structural gene encodes the primary sequence of each polypeptide chain of adult haemoglobin.

Three cDNA sequences, each of molecular mass 1.7×10^5, make 5.1×10^5, or 1 part in 1.6×10^6 of duck DNA. Two cDNA sequences, each of mol. mass 1.5×10^5 make up 1 part in 6×10^6 of rabbit DNA.

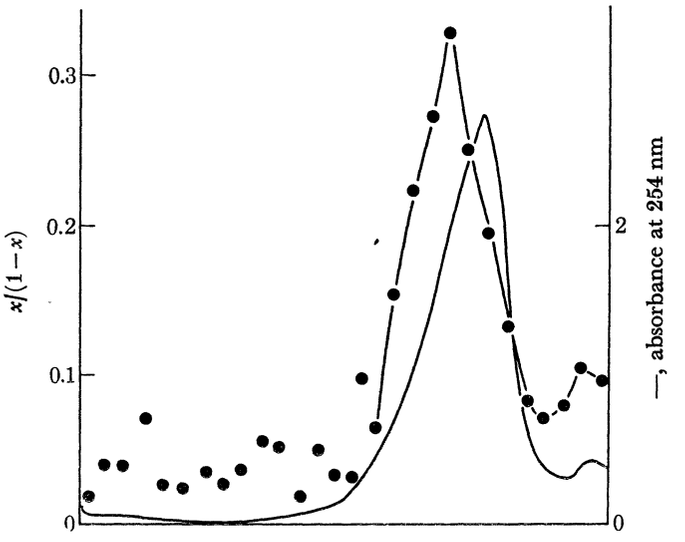

FIGURE 3. Size of the Eco R1 fragments of duck DNA which contain the haemoglobin structural genes.

Duck nuclear DNA, prepared by methods designed to avoid shearing and nuclease attack, was completely digested using restriction nuclease Eco R1. Digestion was controlled by monitoring the digestion of bacteriophage λ DNA. The fragments were denatured in 0.2 M NaOH, loaded on a 15–40 % sucrose gradient containing 0.2 M NaOH and 0.5 % SLS, and centrifuged for 20.5 h at 25000 rev/min in the Spinco no. 41 rotor at 20 °C. Individual fractions were exhaustively dialysed against 0.3 M NaCl–10 mM Na-acetate, pH 5, and then precipitated with 3 volumes of ethanol. The DNA was collected by centrifugation, dried and dissolved in 10 mM NaCl–10 mM tris, pH 7.5. Reaction and assay conditions were as for figure 1. The modal point of the absorbance peak was estimated to be 17.2*S* by analytical centrifugation under alkaline conditions (Studier 1965). Sedimentation was from right to left. The modal point of the fragments which react with Hb-specific cDNA is estimated to sediment at 20.9*S*.

An obvious application of this method is the characterization of DNA fragments which carry the haemoglobin structural genes. In an earlier publication (Bishop & Freeman 1974) it was shown that the duck haemoglobin genes are adjacent to repetitive DNA sequences. More recently, we have attempted to characterize the duck DNA fragments produced by the restriction enzyme Eco R1 (Smith & Nathans 1973) which contain the haemoglobin structural genes. These fragments appear, by sedimentation analysis in alkaline sucrose gradients, to have a single-strand molecular mass of about 3×10^6 (figure 3). This has been confirmed by agarose gel electrophoresis of native restricted DNA fragments (unpublished results). In each case, the modal size of the overall distribution of DNA fragments was not very much less.

The buoyant density of Eco R1 fragments of duck DNA is shown in figure 4. The mean density of the fragments carrying the haemoglobin structural gene suggests a (G+C)-content about 2 % greater than the modal value for duck DNA (about 43 %). These experiments show

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

that purification of the appropriate Eco R1 fragments on the basis of size and (G + C)-content is most unlikely.

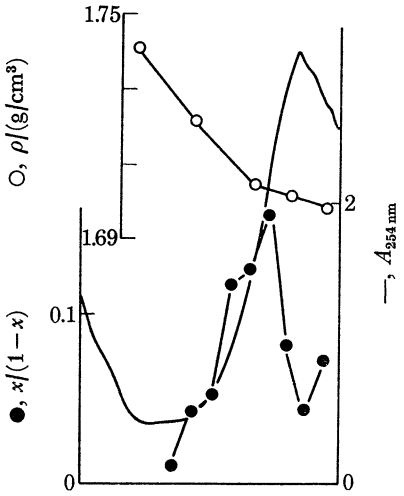

FIGURE 4. Buoyant density of the Eco R1-fragments of duck DNA which contain the haemoglobin structural genes.

Duck DNA, prepared and digested as described in the legend to figure 3, was adjusted to 56.71 % CsCl (ρ_0 = 1.71) and 5 ml portions, overlaid with paraffin, were centrifuged for 60 h at 40000 rev/min at 25 °C in the Spinco no. 50 angle rotor. Fractions were collected through a u.v. (254 nm) monitor, exhaustively dialysed against 0.3 M NaCl–10 mM Na-acetate, pH 5, and precipitated with 3 volumes of ethanol at −20 °C. Reactions with Hb-specific cDNA were carried out as for figure 2. The densities of gradient fractions were measured by weighing aliquots in a 50 μl Lang–Levy constriction pipette.

The estimation of sequence complexity in mRNA populations

Sequence complexity measurements of DNA, by means of DNA renaturation, are quite common (e.g. Britten & Kohne 1968; Laird *et al.* 1974). Similar measurements of RNA populations are somewhat less so (Davidson & Hough 1971; Brown & Church 1972; Turner & Laird 1973; Goldberg, Galau, Britten & Davidson 1973). Recently, we observed that the poly(A)-containing mRNA population of HeLa cells falls into three discrete abundance classes (Bishop *et al.* 1974). These experiments, which were carried out by hybridizing an excess of poly(A)-containing mRNA with cDNA prepared against the same material, provide the basis for a number of studies now in progress, aimed at elucidating the differences between the mRNA populations of different cell-types and of different developmental stages.

One such study is based on *Drosophila melanogaster*. The polytene chromosomes of *D. melanogaster* salivary glands contain 4000–5000 bands which can mostly be visualized in the light microscope. It has been suggested that each of these contains a single structural gene (see Judd & Young 1974; Lefevre 1974) but there is some dissent (see O'Brien 1973). We are approaching this question by measuring the complexity of mRNA populations derived from *D. melanogaster* embryos, larvae, pupae and imagos and to this we are adding cross-hybridization experiments between mRNA populations from the different stages. Our results, so far, suggest that the total number of mRNA sequences expressed within the life-cycle of the fly will be very similar to the number of bands seen in the polytene chromosomes of the larval salivary glands. An interesting example of this work is shown in figure 5. The basic material of these studies is Schneider's L3 cell line. The reaction between mRNA of these cells and homologous cDNA has a three-phase

transition, showing that the mRNA population contains three distinguishable abundance classes (see table 3). Surprisingly, total embryo mRNA seems to contain the same classes, in roughly the same proportion. This was confirmed by annealing an excess of L3 cell mRNA with cDNA prepared against embryo mRNA. As far as the sequences transcribed by AMV

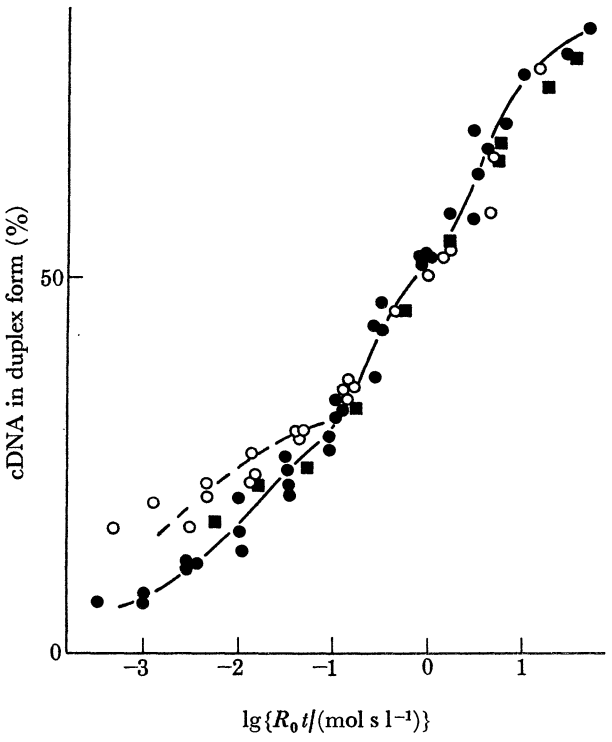

FIGURE 5. Reactions between mRNA and mRNA-specific cDNA from total *D. melanogaster* embryos and Schneider's L3 cell line.

Cytoplasm was prepared by the method of Spradling, Penman, Campo & Bishop (1974) with the addition of 2 units/ml of rat liver RNAase inhibitor (obtained from G. D. Searle Ltd). mRNA of admixed HeLa cells showed the usual profile in sucrose gradients. Reaction and assay conditions were as for figure 1. ●, L3-cell line mRNA in excess over homologous cDNA; ○, embryo mRNA and homologous cDNA; ■, cell-line mRNA in excess over embryo cDNA.

TABLE 3. NUMBER OF mRNA SEQUENCES IN SCHNEIDER'S L3 *D. MELANOGASTER* CELLS

(The cells were grown in Shield's medium and harvested in log-phase growth (2×10^6 cells/ml). The $R_0 t_{\frac{1}{2}}$ values shown in figure 5 are based on the assumption that poly(A) forms 10% of *Drosophila* mRNA. Those in the table are adjusted to the more likely assumption that the value is 5%.)

component	$R_0 t_{\frac{1}{2}}$ / mol s l^{-1}	proportion of total (P)	$R_0 t_{\frac{1}{2}} P$	no. of 4×10^5 mol. mass sequences
1	9×10^{-3}	0.25	2.3×10^{-3}	4
2	0.6	0.25	0.15	180
3	9	0.35	3.1	3700

reverse transcriptase are concerned, total embryo mRNA appears to be essentially identical in composition to the mRNA of L3 cells. The total number of different mRNA sequences in each case is about 2500. So far, after examining embryos, larvae and pupae, we have seen no indication that the total number of poly(A)-containing mRNA sequences in *Drosophila* is greater than the number of chromomeres.

A second example of this type of analysis is shown in figure 6. Immature duck red blood cells are highly specialized. Their nuclei are in the process of progressive inactivation, and within a matter of days they are withdrawn from circulation and destroyed.

Analysis of the cytoplasmic poly(A)-containing mRNA from these cells reveals a major and a minor component. The $R_0 t_{\frac{1}{2}}$ of the major component is consistent (within a factor of two) with the interpretation that it is entirely due to the three species of mRNA which specify the polypeptide chains of adult duck haemoglobin. The minor component suggests that these cells contain about 100 other mRNA sequences (assuming their average molecular mass to be 6×10^5). A similar experiment with nuclear poly(A)-containing RNA (figure 6) suggests that the nuclear population is only 2 to 3 times more complex than the cytoplasmic.

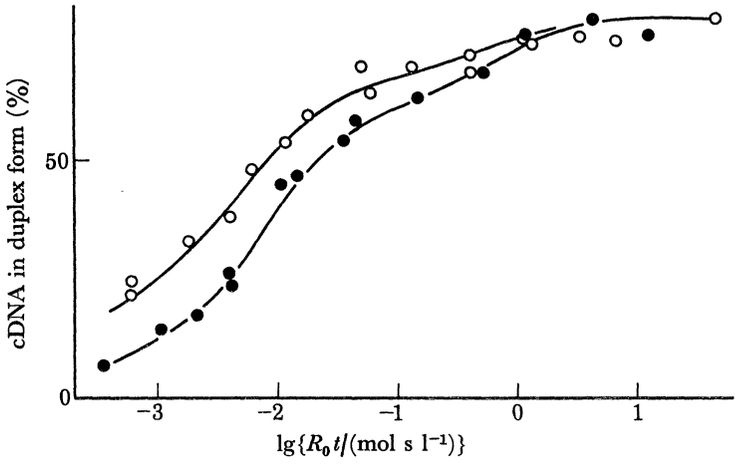

FIGURE 6. Reactions between poly(A)-containing RNA from the nuclei and cytoplasm of immature duck red blood cells and cDNA prepared against the cytoplasmic RNA. ○, cytoplasmic RNA; ●, nuclear RNA.

An indication of the depth with which such analyses can be carried out is given in figure 7. Analysis of the mRNA population from a rat cell line shows the existence of two abundance classes, containing about 150 and 6000 sequences of average length. It has been observed that some mRNA sequences are complementary to moderately repetitive sequences in rat DNA while others are complementary to non-repetitive sequences (Campo & Bishop 1974). It was therefore of interest to ask whether those sequences which are transcribed from repetitive DNA are also the most abundant. This was done by separating the cDNA population into two classes, one complementary to repetitive and the other to non-repetitive DNA sequences, and reacting these separately with the mRNA. The results clearly show that all of the sequences which are complementary to repetitive DNA sequences fall into the abundant class of mRNA. The sequences which are complementary to non-repetitive DNA sequences are found mainly in the non-abundant class, but a significant proportion are abundant.

Homologies between mRNA populations

Although we know that different tissues of an organism contain different populations of protein molecules, the extent of the differences is difficult to quantitate. Furthermore, it is not yet clear whether differences in the protein populations reflect differences in the mRNA populations. As we have seen, molecular hybridization methods can provide us with an overall

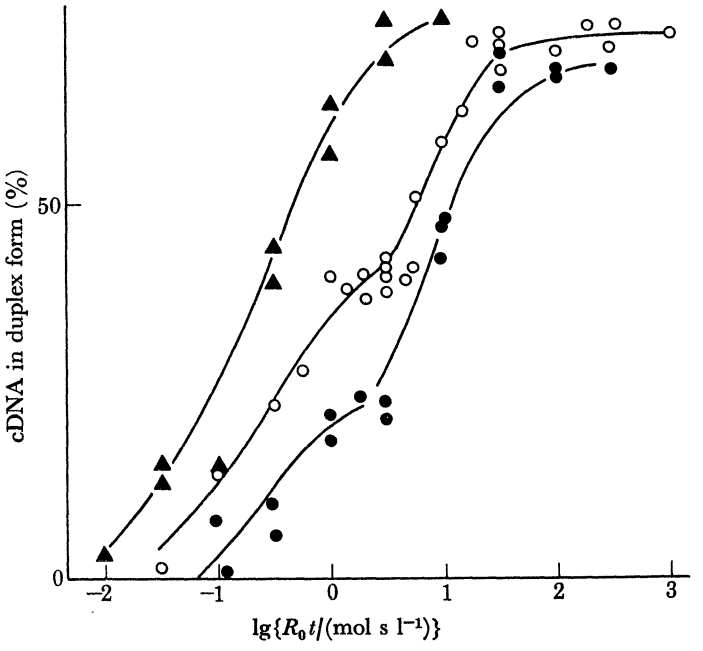

FIGURE 7. Analysis of the poly(A)-containing mRNA population of a rat cell line.

Methods are described in Campo & Bishop (1974). In each case the reaction was driven with an excess of poly(A)-containing mRNA. ○, the reaction of unfractionated cDNA; ●, reaction of cDNA complementary to non-repetitive DNA sequences (that which had not renatured with an excess of cellular DNA after annealing to a C_0t of 200); ▲, reaction of cDNA complementary to repetitive DNA sequences (that which renatured with an excess of cellular DNA after annealing to a C_0t of 20).

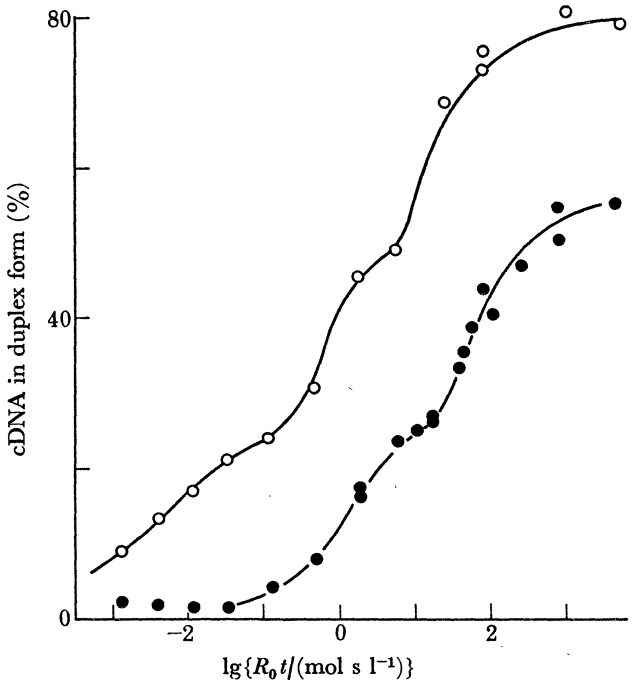

FIGURE 8. Reaction between poly(A)-containing mRNA from liver and cDNA prepared against liver and brain mRNA. ○, cDNA prepared against liver mRNA; ●, cDNA prepared against brain mRNA.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

description of an mRNA population. This can also be used to measure the similarities and differences between populations. Figure 8 shows an example of this.

Liver mRNA shows three abundance components. The first of these corresponds to a small number (about 10) of mRNA sequences of average size. If these sequences are present in brain tissue, they must be in much smaller proportion to the total, because no corresponding component is observed in the cross-reaction between liver mRNA and cDNA prepared against brain mRNA. The second and third transitions of liver mRNA, although not yet fully defined, suggest that they correspond to about 300 and about 15000 different sequences, respectively. Strikingly, the reaction of liver mRNA with cDNA prepared against brain mRNA suggests that about 70% of the mRNA population of brain is made up of sequences which are also present in liver.

Digestion of duck DNA by the restriction enzyme Eco R1 was kindly carried out by Dr K. W. Murray of the Department of Molecular Biology, University of Edinburgh. The experiments reported here were supported by grants from the Cancer Research Campaign and the Science Research Council. S. P. was supported by the Helen Hay Whitney Foundation, J. B. by the European Molecular Biology Organization and M. S. C., partly by Consiglio Nationale de la Ricerche, Roma. Avian myoblastosis virus was kindly supplied by Dr J. W. Beard, Duke University, Durham, N.C.

References (Bishop *et al.*)

Birnstiel, M. L., Chipchase, M. & Spiers, J. 1971 *Prog. Nucleic Acid Res. molec. Biol.* **11**, 351.
Bishop, J. O. 1972 *Acta endocrinol.* (suppl.) **161**, 247.
Bishop, J. O. & Freeman, K. B. 1974 *Cold Spring Harbor Symp. quant. Biol.* **38**, 707.
Bishop, J. O. & Pemberton, R. E. 1972 *F.E.B.S. Symp.* **24**, 25.
Bishop, J. O. & Rosbash, M. 1973 *Nature New Biol.* **241**, 204.
Bishop, J. O., Morton, J. G., Rosbash, M. & Richardson, M. 1974 *Nature, Lond.* **250**, 199.
Bishop, J. O., Pemberton, R. E. & Baglioni, C. 1972 *Nature New Biol.* **235**, 231.
Bonner, T. I., Bremner, D. J., Neufeld, B. R. & Britten, R. J. 1973 *J. molec. Biol.* **81**, 123.
Britten, R. J. & Kohne, D. E. 1968 *Science, N.Y.* **161**, 529.
Brown, I. R. & Church, R. B. 1972 *Devl Biol.* **29**, 73.
Campo, M. S. & Bishop, J. O. 1974 *J. molec. Biol.* **90**, 649.
Davidson, E. H. & Hough, B. R. 1971 *J. molec. Biol.* **56**, 491.
Gelderman, A. H., Rake, A. V. & Britten, R. J. 1971 *Proc. natn. Acad. Sci. U.S.A.* **68**, 72.
Goldberg, R. B., Galau, G. A., Britten, R. J. & Davidson, E. H. 1973 *Proc. natn. Acad. Sci. U.S.A.* **70**, 3516.
Greenberg, J. R. & Perry, R. P. 1971 *J. Cell Biol.* **50**, 774.
Harrison, P. R., Hell, A., Birnie, G. D. & Paul, J. 1972 *Nature, Lond.* **239**, 219.
Hutton, J. R. & Wetmur, J. G. 1973 *J. molec. Biol.* **77**, 495.
Judd, B. H. & Young, W. M. 1974 *Cold Spring Harbor Symp. quant. Biol.* **38**, 573.
Kohne, D. E. 1968 *Biophys. J.* **8**, 1104.
Laird, C. D., Chooi, W. Y., Cohen, E. H., Dickson, E., Hutchinson, N. & Turner, S. H. 1974 *Cold Spring Harbor Symp. quant. Biol.* **38**, 311.
Lefevre, G. 1974 *Cold Spring Harbor Symp. quant. Biol.* **38**, 591.
Melli, M., Whitfield, C., Rao, K. V., Richardson, M. & Bishop, J. O. 1971 *Nature New Biol.* **231**, 8.
O'Brien, S. J. 1973 *Nature, Lond.* **242**, 52.
Packman, S., Aviv, H., Ross, J. & Leder, P. 1972 *Biochem. biophys. Res. Commun.* **49**, 813.
Smith, H. O. & Nathans, D. 1973 *J. molec. Biol.* **81**, 419.
Spradling, A., Penman, S., Campo, M. S. & Bishop, J. O. 1974 *Cell* **3**, 23.
Studier, F. W. 1965 *J. molec. Biol.* **11**, 373.
Turner, S. H. & Laird, C. D. 1973 *Biochem. Genet.* **10**, 263.
Wetmur, J. G. & Davidson, N. 1968 *J. molec. Biol.* **31**, 349.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Discussion

A. D. B. Malcolm (*Department of Biochemistry, Glasgow University*)

Dr Bishop, you mentioned that DNA hybridization would tolerate up to 10 % mismatching of base pairs. Presumably this figure does not include the G–U 'wobble' pair since this is thermodynamically stable. Presumably too short a deletion could be allowed for by a looping out in the other DNA strand.

After allowing for all these factors what do you estimate is the possible error in your calculations of gene frequencies?

J. O. Bishop

It is not really clear that the question of specificity arises. The populations we are dealing with are of quite low complexity relative to eukaryotic DNA. The *Drosophila* and mouse mRNA populations are equivalent to about one and five times the *E. coli* genome, respectively. In DNA-driven experiments at least 80 % of each cDNA preparation hybridizes with 'single-copy' DNA sequences. If these measurements could be made with sufficient precision, they would preclude the possibility of mismatching within that major fraction. However, I doubt whether the data can exclude a difference of up to twofold in rate. It is worth mentioning that if mismatching is occurring its effect would be to reduce our estimates of gene frequencies relative to the real values.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES