- Bolle, A., Epstein, R. H., Salser, W. & Geiduschek, E. P. (1968). *J. molec. Biol.* **31**, 325.
- Bremer, H., Konrad, M. & Bruner, R. (1966). *J. molec. Biol.* **16**, 104.
- Chamberlin, M. & Berg, P. (1962). *Proc. nat. Acad. Sci., Wash.*, **48**, 81.
- Cohen, S. N. & Hurwitz, J. (1967). *Proc. nat. Acad. Sci., Wash.*, **57**, 1759.
- Colvill, A. J. E., Kanner, L., Tocchini-Valentini, G. P., Sarnat, M. T. & Geiduschek, E. P. (1965). *Proc. nat. Acad. Sci., Wash.*, **53**, 1140.
- Fox, C. F. & Weiss, S. B. (1964). *J. biol. Chem.* **239**, 175.
- Fraser, S. & Jerrel, J. (1953). *J. biol. Chem.* **205**, 291.
- Furth, J. J., Hurwitz, J. & Goldman, M. (1961). *Biochem. biophys. Res. Commun.* **4**, 362.
- Geiduschek, E. P., Moohr, J. W. & Weiss, S. B. (1962). *Proc. nat. Acad. Sci., Wash.*, **48**, 1078.
- Geiduschek, E. P., Nakamoto, T. & Weiss, S. B. (1961). *Proc. nat. Acad. Sci., Wash.*, **47**, 1405.
- Geiduschek, E. P., Snyder, L., Colvill, A. J. E. & Sarnat, M. (1966). *J. molec. Biol.* **19**, 541.
- Gillespie, D. & Spiegelman, S. (1965). *J. molec. Biol.* **12**, 829.
- Green, M. H. (1964). *Proc. nat. Acad. Sci., Wash.*, **52**, 1388.
- Hall, B. D., Nygaard, A. P. & Green, M. H. (1964). *J. molec. Biol.* **9**, 143.
- Hurwitz, J., Furth, J. J., Anders, M. & Evans, A. (1962). *J. biol. Chem.* **237**, 2752.
- Jones, O. W. & Berg, P. (1966). *J. molec. Biol.* **22**, 199.
- Khesin, R. B., Gorlenko, Z. M., Shemyakin, M. F., Bass, I. A. & Prozorov, A. A. (1963). *Biokhimiya*, **28**, 1070.
- Luria, S. E. (1965). *Biochem. biophys. Res. Commun.* **18**, 735.
- Maitra, U., Nakata, Y. & Hurwitz, J. (1967). *J. biol. Chem.* **242**, 4908.
- Mandell, J. D. & Hershey, A. D. (1960). *Analyt. Biochem.* **1**, 66.
- Nakamoto, T., Fox, C. F. & Weiss, S. B. (1964). *J. biol. Chem.* **239**, 167.
- Richardson, J. P. (1966). *J. molec. Biol.* **21**, 83.
- Robinson, W. S., Hsu, W. T., Fox, C. F. & Weiss, S. B. (1964). *J. biol. Chem.* **239**, 2944.
- Steck, T. L., Caicuts, M. J. & Wilson, R. G. (1968). *J. biol. Chem.* **243**, 2769.
- Weiss, S. B. & Nakamoto, T. (1961). *Proc. nat. Acad. Sci., Wash.*, **47**, 1400.

APPENDIX

Methods for the Analysis of Deoxyribonucleic Acid-Ribonucleic Acid Hybridization Data

BY JOHN O. BISHOP, FORBES W. ROBERTSON, JAMES A. BURNS AND
MARIALUISA MELLI

Department of Genetics, University of Edinburgh

Our immediate aim is a qualitative characterization of cRNA*. To do this, we have asked two types of question. (1) What proportion of all the sequences of the DNA are represented by complementary RNA copies? (2) What proportion of the different RNA sequences in one RNA preparation are also present in a second? These depend on two types of experiment. (1) Saturation of the DNA by excess of RNA. If every DNA sequence that has been transcribed hybridizes fully with complementary RNA, the proportion of the DNA transcribed is obtained by measuring the amount of RNA hybridized to unit amount of DNA. In practice, the precondition is not fulfilled, and means have to be found of overcoming this. (2) Hybridization competition (Nygaard & Hall, 1964). A labelled RNA preparation is diluted by an unlabelled preparation. The common sequences are diluted out completely if a sufficiently high concentration

of the unlabelled sequences can be added. The labelled RNA hybrid at this point represents those sequences present in the labelled RNA and absent from the unlabelled RNA.

When the RNA contains a single sequence (e.g. 16S rRNA) or a uniform distribution of sequences (e.g. partially degraded rRNA) interpretation of the data is relatively straightforward. When the RNA is a complex mixture of many different sequences at different relative concentrations, several serious problems of interpretation arise. (1) If the number of sequences is very large, the attainable concentration of each may be restricted by the limited solubility or availability of RNA (Bishop, 1969b; Melli & Bishop, 1969). (2) The rarer species in a complex mixture may be overlooked because their relative concentration is very low. (3) These difficulties are increased if a large proportion of the RNA is inert, or hybridizes with a very small part of the DNA (e.g. rRNA in an mRNA preparation). (4) The presence of inert RNA or rRNA prevents measurement of the absolute concentration of the RNA sequences that are responsible for most of the

* Abbreviations: cRNA, RNA synthesized *in vitro* by using purified RNA polymerase and a DNA primer; mRNA, messenger RNA; rRNA, ribosomal RNA; tRNA, transfer RNA.

hybridization. In hybridization-competition experiments, the shape of the dilution curve is determined partly by the sequence homology of the competing mixtures, partly by the concentrations of the hybridizing sequences. Estimation of the first is complicated by the impossibility of measuring the second.

In this Appendix we describe methods for analysing DNA-RNA hybridization results when the RNA is a complex mixture of sequences. These take account of the difficulties described above, and in some cases provide means of avoiding them. The methods also take account of the fact that the rate of DNA-RNA hybridization, like DNA renaturation, is limited by the collision frequency of complementary nucleotide sequences (Bishop, 1969b). Methods previously suggested for the interpretation of saturation (McCarthy & Bolton, 1964) and competition (Bolle, Epstein, Salser & Geiduschek, 1968) data do not do so, and we believe them to be in error for this reason.

Symbols

The symbols used are defined as follows:

- t , time (hr.);
- h , concentration of labelled RNA ($\mu\text{g./ml.}$);
- g , concentration of unlabelled RNA ($\mu\text{g./ml.}$);
- r , hybridization ($\mu\text{g. of RNA hybridized}/\mu\text{g. of DNA recovered}$);
- $r_{\text{max.}}$, maximum observed hybridization;
- r_s , maximum obtainable hybridization, estimated by extrapolating to infinite RNA concentration (r_s = the saturation value);
- a , specific radioactivity of hybridizing RNA (c.p.m./ $\mu\text{g. of RNA}$);
- c_0 and c , hybridization, expressed as c.p.m./ $\mu\text{g. of DNA recovered}$; c_0 is the hybridization due to labelled RNA alone, c is the hybridization observed in the presence of unlabelled competing RNA: $c_0 = ra$;
- C , competition of labelled RNA by unlabelled RNA, as a proportion: $C = (c_0 - c)/c_0$;
- p , the RNA sequences of the labelled RNA preparations that are competed against by sequences of the unlabelled RNA in a hybridization-competition experiment, expressed as a proportion of all sequences in the labelled preparation;
- q , proportion of RNA sequences that are not competed against: $q = 1 - p$;
- p_c , proportion of the actual hybridization due to labelled RNA alone (c) that, in a hybridization-competition experiment, is found to be competed against by the unlabelled RNA by extrapolating to infinite concentration of unlabelled RNA;
- q_c , $1 - p_c$;

F , the concentration factor, equal or approximately equal to the ratio (proportion of the total mass of the labelled RNA present in sequences the hybridization of which is competed against by unlabelled RNA)/(proportion of the total mass of the unlabelled RNA present in sequences that compete against labelled RNA sequences);

K , a constant, the slope of the plot of $1/r$ against $1/h$;

K_0 and K_1 , constants;

$r_t, r_{st}, c_t, c_{ot}, K_t, K_{ot}$, have the meanings ascribed to them above except that they refer to one RNA sequence in a mixture of sequences;

bold type, where data and parameters simulating experimental data and parameters are produced by means of model systems, the same symbols are used, but in **bold face type**

Measurement of the proportion of the DNA sequence to which sequences in an RNA preparation are complementary

Immobilization of the denatured DNA on membrane filters by the method of Gillespie & Spiegelman (1965) prevents DNA renaturation, which would otherwise decrease DNA-RNA hybridization. We anneal immobilized ^{32}P -labelled DNA with different concentrations of uniformly ^{14}C -labelled cRNA. The amount of RNA hybridized per unit of DNA is calculated from the specific radioactivities of the RNA and DNA. This is equivalent to the proportion of the DNA hybridized. By extrapolating to infinite RNA concentration (see below) we find the saturation value (r_s), which is the apparent hybridization at infinite RNA concentration. For various reasons this may underestimate the hybridization that would be observed if a truly saturating RNA concentration were available, here called the hybridization value. Either the saturation value or both values may underestimate the complementarity, which is the proportion of the DNA sequence that is represented by complementary RNA copies. Here we (1) describe a method for estimating the saturation value, (2) examine directly the relationship between the saturation value and the complementarity in the case of cRNA, (3) examine the effect on this relationship of the self-complementary sequences present in cRNA and (4) examine by means of a model system the relationship between the saturation value and the hybridization value in cases where the different sequences of the RNA form a frequency distribution.

Estimation of the saturation value. The saturation value has usually been estimated visually from the saturation curve. We consider that this method is open to objection. There is good evidence (Nygaard & Hall, 1964; Bishop, 1969b) that DNA-RNA

hybridization is rate-limited by the collision frequency of complementary sequences. As the number of complementary DNA sites is decreased by hybridization, the reaction will become progressively slower, so that the hybridization of the DNA will be complete only at infinite time or at infinite RNA concentration. When the RNA contains very few sequences, the concentration of each can be very high, and the saturation point estimated by inspection may accurately represent the saturation value. However, when the RNA contains many different sequences, and particularly if it is greatly diluted by rRNA, a close approach to the saturation value may be impossible. A second objection is that visual inspection of the saturation curve is very subjective. The highest RNA concentration used often appears to be close to saturation simply because the shape of the curve makes visual extrapolation difficult (see Fig. 3, curve 1). For the same reason, a change of scale of the abscissa (RNA concentration) can change the visual interpretation of the curve.

We have sought an alternative method of estimating the saturation value. In principle, if the concentration-dependence of the reaction can be described by a straight line, a saturation value can be found by extrapolating to infinite RNA concentration. In practice, we have found that a linear relationship is obtained when the reciprocal of hybridization is plotted against the reciprocal RNA concentration [Fig. 2 and Table 2 of the main paper (Bishop & Robertson, 1969) and Fig. 1 and Tables 1 and 2].

The reciprocal plot can be described by $1/r = (K/h) + K_1$. When h is infinite $r = r_s$ and $(k/h) = 0$. Making the assumption that linear extrapolation is valid we put $K_1 = 1/r_s$ and:

$$1/r = K/h + 1/r_s \quad (1)$$

This equation may be used to estimate r_s by extrapolation. However, a more useful form is:

$$h/r = h/r_s + K \quad (2)$$

r_s is obtained as the reciprocal of the slope of h/r against h , together with a standard error, if desired.

The specific radioactivity of the hybridizing

sequences in an mRNA preparation is not usually known. In this case we write:

$$h/c = h/c_s + K^1 \quad (3)$$

Examples of the use of eqn. (2) are shown in Fig. 2 of the main paper (Bishop & Robertson, 1969) and Fig. 1. To determine whether its use was restricted to a particular annealing time, bacteriophage T4 DNA was annealed with different concentrations of *Micrococcus* cRNA for 1, 2 and 12 hr. A straight-line reciprocal plot was found at each time by using either eqn. (1) or eqn. (2). Saturation values were estimated by using eqn. (2) and found not to differ significantly (Table 1).

Relationship between saturation value and complementarity. Some of the DNA may fail to hybridize with complementary RNA, even at infinite RNA concentration. Therefore, the saturation value is not necessarily an estimate of the proportion of the DNA that is complementary to sequences present in the cRNA. We call the ratio of complementarity to the saturation value the hybridization factor. An RNA specimen having a known homology with the DNA is a necessary control for these experiments. The most likely source of such RNA is the DNA-RNA hybrid that is synthesized when RNA polymerase is incubated with a single-stranded DNA template (Sinsheimer & Lawrence, 1964; Chamberlin & Berg, 1964; Hayashi, Hayashi & Spiegelman, 1965). With a template of single-stranded ϕ X174 DNA, *Escherichia coli* RNA polymerase synthesized a DNA-RNA hybrid containing approximately equal amounts of RNA and DNA. We used denatured T4 DNA as template for both *Micrococcus* RNA polymerase and *E. coli* RNA polymerase. In both cases the maximum RNA/DNA ratio that could be attained in the enzymically synthesized hybrid was about 0.5:1 (Bishop, 1969a). The 0.5:1 hybrid is a less satisfactory control than a 1:1 hybrid would be. Instead of an absolute estimate, it gives a maximum estimate of the success of hybridization. In a 0.5:1 hybrid, 50% of the DNA is hybridized with RNA. If sequences are selected for transcription, at least 50% of the DNA sequences have been transcribed. On the other hand, if sequences are not selected, it may be that

Table 1. *Effect of annealing time on the saturation value*

Bacteriophage T4 DNA was annealed with five concentrations (5.3, 6.6, 9.1, 13.3 and 26.6 μ g./ml.) of *Micrococcus* cRNA for 1, 2 or 12 hr.

Annealing time (hr.)	Slope of reciprocal plot ($1/r_s$)	Saturation value (r_s) (μ g. of RNA/ μ g. of DNA)	$r_{\max.}$	$r_{\max.}/r_s$
1	2.38 ± 0.21	0.42	0.22	0.52
2	2.50 ± 0.34	0.40	0.27	0.67
12	2.47 ± 0.03	0.41	0.32	0.78

half the copies of every DNA sequence have been transcribed. In this situation, if the isolated RNA hybridized with say, 25% of the DNA, it would be possible to conclude that the minimum value of the hybridization factor was 2. It might, however, be as great as 4.

cRNA was synthesized with *Micrococcus* RNA polymerase with denatured ^{32}P -labelled bacteriophage T4 DNA as template and $[^{14}\text{C}]\text{ATP}$ as substrate. A sample was centrifuged to equilibrium in a preparative caesium sulphate density gradient (initial concentration 43.4% caesium sulphate) to separate the DNA-RNA hybrid from free RNA (Bishop, 1969a). The RNA/DNA ratio of the hybrid was estimated from the radioactivity in the hybrid band, by using the known specific radioactivities of the ^{32}P -labelled DNA and the $[^{14}\text{C}]\text{ATP}$ substrate, and assuming the RNA to have the same average base composition as bacteriophage T4

DNA. After treatment with ribonuclease the RNA/DNA ratio of the hybrid was 0.54. Without the removal of ribonuclease-sensitive RNA the ratio was about 25% greater. Assuming that ribonuclease-sensitive RNA is not base-paired to DNA, we take 54% as a minimum estimate of the percentage of DNA sequences transcribed. After heat denaturation (90° for 10 min.) RNA was prepared from the reaction mixture in the usual way. A reciprocal plot of the saturation of bacteriophage T4 DNA by this RNA is shown in Fig. 1. The saturation value estimated from the reciprocal of the slope was 0.28. In this case, then, the saturation value is 52% of the minimum complementarity between DNA and RNA.

The results of three separate hybridization experiments are shown in Table 2. In the two additional cases the RNA/DNA ratio of the enzymic hybrid was not measured. It can, however, be taken to be close to 0.5 since the product/primer ratio after RNA synthesis was in each case greater than 1 (Bishop, 1969a).

From these results we find the minimum hybridization factor to be about 2. If we take it that the factor is the same for different types of cRNA preparations we can calculate the maximum factor to be about 2.5 from the fact that cRNA made on the native bacteriophage T4 template with *Micrococcus* RNA polymerase hybridizes with 38% of bacteriophage T4 DNA (Table 2). If this cRNA were complementary to all of the bacteriophage T4 DNA the factor would be 2.6. In round figures, then, we estimate the hybridization factor to be between 2 and 2.5 for cRNA.

Whether only 40–50% of hybridizable DNA sequences actually do hybridize under our conditions is open to question. Because it decreases non-specific binding of RNA and makes results more reproducible, we have always treated the membranes with ribonuclease, although this is known to decrease apparent hybridization by up to one-half (Nygaard & Hall, 1964). We have made the assumption that ribonuclease attacks only RNA strands that are not base-paired with DNA. This

Fig. 1. Hybridization between bacteriophage T4 DNA and cRNA isolated from enzymically synthesized DNA-RNA hybrid. The ordinate, h/r , shows the ratio of RNA concentration to hybridization. The abscissa, h , shows the RNA concentration ($\mu\text{g./ml.}$). The slope of the line is 3.53 ± 0.13 .

Table 2. Hybridization with cRNA synthesized on a denatured bacteriophage T4 DNA template

In each case bacteriophage T4 DNA was annealed with five concentrations of cRNA. The concentration ranges were 1.1–8.8 $\mu\text{g./ml.}$, 0.75–12 $\mu\text{g./ml.}$ and 1.2–14.5 $\mu\text{g./ml.}$ respectively in Expts. 1, 2 and 3.

Expt. no.	RNA synthesis product/primer ratio	Hybridization			
		Slope of reciprocal plot ($1/r_s$)	Saturation value (r_s)	$r_{\text{max.}}/r_s$	Complementarity
1	1.35	3.53 ± 0.13	0.28	0.78	0.54
2	1.8	4.35 ± 0.51	0.23	0.78	—
3	1.45	3.31 ± 0.32	0.30	0.83	—

view is supported by the fact that ribonuclease attacks only about 20% of the RNA in the enzymically synthesized RNA-DNA hybrid from bacteriophage T4 (Bishop, 1969b), suggesting 20% as an upper limit to the amount of correctly base-paired RNA that may be attacked by ribonuclease.

We know of only one other situation that gives an estimate of the hybridization factor. Landy, Abelson, Goodman & Smith (1967) and Andoh & Ozeki (1967) hybridized tyrosine tRNA with the DNA of the transducing bacteriophage $\phi 80\text{dsu}_{\text{III}}^+$. In both cases it was visually estimated that 1 tyrosine tRNA molecule hybridized with each bacteriophage DNA molecule. Using eqn. (2) we derived an estimate from the combined data of Landy *et al.* (1967) and Andoh & Ozeki (1967) of 1.15 tRNA molecules/molecule of bacteriophage DNA. It has recently been found that bacteriophage $\phi 80\text{dsu}_{\text{III}}^+$ contains not one but two genes for tyrosine tRNA (R. L. Russell, J. N. Abelson, A. Landy, M. L. Gelfer, S. Brenner & J. D. Smith, personal communication). This leads to an estimate of 1.74 for the hybridization factor. cRNA contains at least 40% of self-complementary sequences (Table 1). Calculations based on the rates of hybridization and self-annealing of cRNA show that the self-complementarity will tend to increase the value of the hybridization factor. We therefore favour the lower factor for non-self-complementary RNA.

Effect of the RNA-sequence frequency distribution. When the RNA contains a mixture of different sequences, each hybridizes with the DNA independently of the others (Bishop, 1969b). If different sequences are present at different relative concentrations, each will show a different dependence on the overall RNA concentration. The observed saturation curve will be the sum of these different reactions. We want to know to what extent the frequency distribution will bias the estimate of the hybridization value obtained from the saturation value.

A complete answer to this question requires a knowledge of the frequency distribution of the RNA sequences, and this is not known. We have approached the problem by constructing a model system, and examining the behaviour of several notional frequency distributions. The model is based on a DNA molecule with 100 different sites, about the same as the number of cistrons in bacteriophage T4 DNA. The hybridization of each RNA sequence is taken to follow a relationship like that given in eqn. (1):

$$1/r_i = K_i/h + 1/r_{st} \quad (4)$$

The grounds for this are empirical. Linear plots of this sort have been published by Lavallé & De Hauwer (1968) for the hybridization of *E. coli* RNA with the DNA of bacteriophage $\phi 80\text{pt}_1$. Most of

the RNA that hybridizes with this DNA is tryptophan-operon mRNA (Imamoto & Yanofsky, 1967). The data of Landy *et al.* (1967) and Andoh & Ozeki (1967) on the hybridization of tyrosine tRNA also fit a straight line. Lastly, we have analysed much of the available data on the hybridization of bacterial rRNA. With very few exceptions, reciprocal plots produced straight lines.

The constant, K_i , in eqn. (4) takes a different value for each RNA sequence. The value was set to be inversely proportional to the relative concentration of the sequence. At infinite RNA concentration all DNA sites were assumed to be saturated, and all DNA sites were the same 'size'. Hence, r_{st} took the same value (0.01) for each RNA sequence. In this way, 100 simultaneous equations were generated for each RNA-sequence frequency distribution. A computer programme was used to obtain and summate the 100 values of r_i ($r = \sum r_i$) for chosen ranges of h (RNA input), and to analyse the numbers so produced by means of eqn. (2) to obtain estimates of r_s . Three ranges of h were examined, covering three successive orders of magnitude. Each range was divided into ten equal increments of h .

We first compared an RNA in which the 100 sequences were equal in frequency (uniform distribution) with an RNA in which the frequencies were continuously distributed. This distribution was designed so that the frequency distribution of $\log(\text{concentration})$ was Gaussian. A sevenfold range of concentrations fell within 1 s.d., and a 50-fold range within 2 s.d. (Fig. 2). The results are shown in Fig. 3 and Table 3. As expected, when the frequencies of the different sequences were equal, eqn. (2) always extrapolated to the hybridization value,

Fig. 2. Continuous RNA distribution model. The abscissa, C/C' , shows the concentration of sequences relative to the mode. The ordinate shows the relative number of sequences.

regardless of range. With the continuous distribution, the extrapolated saturation value varied in the different ranges. Nevertheless, the deviation of the reciprocal plot from linearity was never great enough to be detected by hybridization techniques. The standard error introduced by an average 5% 'experimental' error randomly applied to the points obscured the standard error caused by the deviation from linearity (compare Table 3, 'no error' and '5% error'). This is an important result, because it shows that the linearity of the plot is not an indication of whether or not the RNA-sequence frequency distribution is uniform. The saturation value approached the hybridization value (1 in this case) more closely in the higher concentration ranges. Unfortunately, in the experimental situation there is no direct way of assessing the range in which the experiment has been executed (see Fig. 3). However, a rough guide is provided by the relationship

Fig. 3. Simulation of DNA hybridization by using uniform and continuous model RNA frequency distributions. The ordinate, r , shows the hybridization. The abscissa, h (arbitrary units), shows the RNA concentration. Three orders of magnitude of h are shown. For curves 2 abscissa values are 10 times, and for curves 3 100 times those shown. $\times$, Uniform distribution; o , continuous distribution.

between the saturation value and the highest experimental point on the curve ($r_{\max.}$, Table 3). As the concentration range increases, the ratio $r_{\max.}/r_s$ increases. With the wide frequency distribution we have examined here, a reasonable estimate of the hybridization value is obtained when this ratio approaches 0.9.

The model was also used to examine the effects of discontinuous distributions. Two distributions were studied. In one (D1) the frequencies of half the RNA sequences (50) were decreased tenfold, in the other (D2) 100-fold. The results are shown in Table 4. Once again deviations from linearity were obscured by the effects of a 5% experimental error. The D1 distribution gave results very similar to the continuous distribution. The D2 distribution gave a saturation value within 10% of the hybridization value only in the highest range. In this case the ratio $r_{\max.}/r_s$ was not a good guide to the accuracy of the estimate. Even with a ratio of 0.9, the estimate could be much closer to the hybridization value due to the concentrated half of the RNA sequences (0.5) than to the hybridization value of 1.0 representing all the sequences.

It is important to note that we have here examined rather extreme distributions in the context of bacteriophage T4-specific RNA. Particularly in the case of cRNA we do not suspect the existence of such wide concentration differences. With the exception of the discontinuous D2 distribution it is clear that a reasonably good estimate of the hybridization value is obtained when the ratio $r_{\max.}/r_s$ is in the region of 0.9. In effect, on the basis of this criterion, RNA sequences that are at a concentration 100-fold lower than the mean may be seriously underestimated.

Measurement of the sequence homology between different RNA preparations by hybridization competition

Straightforward interpretation of competition experiments demands the fulfilment of two conditions: (a) the use of a concentration of labelled RNA

Table 3. *Effect of a continuous distribution of RNA sequence frequencies on the saturation value*

Numbers were obtained by using a model of DNA-RNA hybridization, which is described in the text, and analysed by means of eqn. (2).

RNA concentration range	Uniform distribution		Continuous distribution			
	$r_{\max.}$	r_s	$r_{\max.}$	r_s (No error)	r_s (5% error)	$r_{\max.}/r_s$ (No error)
$\times 1$	0.50	1.00	0.41	0.73 ± 0.01	0.71 ± 0.05	0.56
$\times 10$	0.90	1.00	0.83	0.94 ± 0.006	0.94 ± 0.03	0.88
$\times 100$	0.99	1.00	0.98	1.00 ± 0.001	0.96 ± 0.05	0.98

Table 4. *Effect of discontinuous distributions on the saturation value*

Numbers were obtained by using a model of DNA-RNA hybridization as described in the text, and analysed by means of eqn. (2). Distribution D1 contains equal numbers of RNA sequences at relative concentrations 1 and 0.1. Distribution D2 contains equal numbers at concentrations 1 and 0.01.

RNA concentration range	Distribution D1				Distribution D2			
	$r_{\max}$	r_s (No error)	r_s (5% error)	$r_{\max}/r_s$	$r_{\max}$	r_s (No error)	r_s (5% error)	$r_{\max}/r_s$
$\times 1$	0.40	0.67 ± 0.005	0.73 ± 0.04	0.60	0.34	0.52 ± 0.001	0.51 ± 0.03	0.65
$\times 10$	0.80	0.90 ± 0.01	0.92 ± 0.04	0.88	0.56	0.60 ± 0.01	0.59 ± 0.02	0.93
$\times 100$	.97	1.00 ± 0.001	0.96 ± 0.05	0.97	0.82	0.89 ± 0.01	0.91 ± 0.04	0.92

sufficient to saturate complementary DNA sites; (b) the use of a concentration of unlabelled RNA sufficient to dilute out identical labelled RNA sequences completely. As we pointed out above, it is impossible to satisfy these conditions completely. Approximations may be achieved, but their subjective evaluation is likely to be mistaken.

We have developed methods of analysing competition data that do not depend on the assumption of saturation. However, the relationships are developed for the case of uniformly distributed (or homogeneous) RNA preparations, and their usefulness is less clear where the frequency distributions are non-uniform. The effects of such frequency distributions are discussed in the following section.

Of the two conditions for straightforward interpretation of competition data, the more likely to be achieved is the near-saturation of DNA sites by labelled RNA. If this condition is fulfilled, the approach to the competition plateau can be used to estimate sequence homology. The shape of the approach curve depends on the concentration factor (F) and the homology (p). Competition of the common RNA fraction should follow a simple dilution relationship from which we derive the expression:

$$g/C = g/p + Fh/p \quad (5)$$

By plotting g/C against g , the most probable value of $1/p$ can be estimated by regression analysis. If the two RNA preparations hybridize with the same part of the DNA, $p = 1$, and eqn. (5) reduces to $g/C = g + Fh$. This is identical with the relationships used by Landy *et al.* (1967), Mangiarotti, Apirion, Schlessinger & Silengo (1968), Lavallé & De Hauwer (1968) and Stubbs & Hall (1968) to estimate the relative concentrations of different uniformly distributed RNA preparations.

In many cases even an apparent saturation of the DNA by labelled RNA is difficult or impossible to achieve. Under these circumstances the linear relationship found to obtain between $1/r$ and $1/h$ permits the derivation of an extension of eqn. (5). From eqn. (1), setting a as the specific radioactivity of the labelled RNA, we can write the radioactivity

in the DNA-RNA hybrid, $c_0 = ra$ and $c_0 = (p_c + q_c)(r_s ha)/(h + Kr_s)$.

Keeping the amount of labelled RNA constant, and varying the amount of unlabelled competitor (g), the concentration and specific radioactivity of the uncompetited RNA will be unchanged. The effective concentration of the competed RNA becomes $h + (g/F)$ and the specific radioactivity, $ah/(h + g/F)$. Then:

$$c = \frac{p_c r_s ah}{h + g/F + Kr_s} + \frac{q_c r_s ah}{h + Kr_s}$$

The proportional competition $C = (c_0 - c)/c_0$ and:

$$g/C = g/p_c + F(h + Kr_s)/p_c \quad (6)$$

When g/C is plotted against g , the slope of the line gives $1/p_c$.

Effects of RNA-sequence frequency distributions

Eqns. (5) and (6) are developed on the assumption of uniform frequency distributions. If the distributions are not uniform the estimate of homology will be disturbed in two ways. (a) The value p_c represents the partition of c_0 into sequences that are present in the unlabelled RNA and sequences that are not. If the labelled RNA has a uniform distribution, c_0 represents an equal partial hybridization of all sequences, and the partition of c_0 will equal the partition of r_s , i.e. $p_c = p$. If the labelled RNA sequences do not have a uniform distribution, c_0 will contain a disproportionate proportion of the more concentrated sequences and p_c will not equal p (unless, by chance, the frequency distributions of the competed and uncompetited sequences are similar). (b) If either of the two competing RNA preparations does not have a uniform distribution, the concentration factor, F , becomes a complex function dependent on the precise nature of the distributions. The interpretation of the shape of the competition curve based on a simple definition of F (eqns. 5 and 6) may then give an inaccurate estimate of p_c .

It is not possible to measure the magnitude of

these errors without knowledge of the frequency distributions. We therefore used the computer model to study the effects of different frequency distributions, in an attempt to define the circumstances under which accurate estimates are obtained. The same four distributions were tested against each other in all possible combinations. Each competition curve was generated by adding 11 hypothetical concentrations of unlabelled RNA, covering a range from 0.5 times to 10.5 times the concentrations of labelled RNA in ten equal increments. The computer was programmed to calculate for each RNA sequence (from eqn. 3), $c_{oi} = 0.01 h / (0.01 K_{oi} + h)$, from which $c_o = \sum_{i=1-100} c_{oi}$ was calculated and for each input amount of unlabelled RNA $c_i = 0.01 h / (0.01 K_i + h + g)$ and $c = \sum_{i=1-100} c_i$. $C = (c_o - c) / c_o$ and values of p_c were calculated from g and C by using eqn. (6). K_{oi} and K_i reflected the frequencies of sequence i in the two competing distributions. K_{oi} was equal to 100 divided by the frequency of i (as a proportion) in the labelled preparation. K_i was equal to 100 divided by its frequency in mixture ($h + g$).

Each hypothetical labelled RNA distribution was examined at two input concentrations. These concentrations were characterized (Table 5) by means of the ratio $r_{\max.}/r_s$ derived from the saturation curves. These ratios are the only means of comparing actual experimental data with the values given in Table 5. When the RNA is not uniformly labelled we use the ratio c_o/c_s , which is analogous to $r_{\max.}/r_s$, to characterize the concentration of labelled RNA. Values of $r_{\max.}$ are also shown in Table 5, but of course these are not known in absolute terms for real experiments.

Since all available RNA sequences are present in each distribution, the actual homology is 1 in every case, and we are interested in seeing how closely p_c approaches 1. The results obtained are shown in Table 5. Once again the standard error introduced by an average 5% 'experimental' error completely obscured the small standard error due to non-linearity. The linearity of the plot is not, then, a guide to the accuracy of the estimate. Only the 'without error' values are shown in Table 5.

In every case of competition by the uniformly distributed RNA, p_c was within 3% of the homology (1). In every case of competition by an identical competitor, the result was within 5% (values underlined). In the competition of uniformly distributed RNA by continuously distributed RNA, the error was 15% with the lower labelled-RNA input, 7% with the higher. As in the saturation experiments, greater errors are introduced if discontinuous distributions are present. Competition of 'labelled' discontinuous distributions produced errors up to 20%. Competition by 'unlabelled' discontinuous distributions produced greater errors. With distribution D2 the result is an approximation to the concentrated half of the distribution. With wider discontinuous distributions, this approximation would improve. In much the same way, when it is the discontinuous distribution that is being competed against, the less-concentrated sequences do not contribute to c_o , and the competition will be close to 1 even if they are absent from the competitor. Thus any part of an RNA-sequence distribution at a concentration 100-fold or more lower than the mean is very likely to be overlooked. For most purposes this is a satisfactory situation. The greatest source of potential error is clearly a discontinu-

Table 5. *Effect of continuous and discontinuous distributions on the estimation of sequence homology by competition experiments*

The right-hand side of the table shows values of p_c obtained by using the model of DNA-RNA hybridization described in the text. The upper and lower parts show the effects of competing against two different input concentrations of labelled RNA, which are characterized in terms of saturation concentrations by $r_{\max.}$ and $r_{\max.}/r_s$. These represent the highest concentrations of range $\times 1$ and $\times 10$ (Tables 3 and 4). The range of 'unlabelled' RNA concentrations was in each case from 0.5–10.5 times the 'labelled' RNA concentration. Values of $p_c \pm$ s.e.m. were derived by using eqn. (6). The symbols U, C, D1 and D2 refer respectively to uniform, continuous and the two discontinuous RNA sequence frequency distributions.

Labelled RNA			Unlabelled RNA			
Type	$r_{\max.}$	$r_{\max.}/r_s$	U	C	D1	D2
U	0.50	0.50	1.00	0.85 $\pm$ 0.01	0.78 $\pm$ 0.01	0.54 $\pm$ 0.003
C	0.41	0.56	0.98 $\pm$ 0.002	0.96 $\pm$ 0.004	0.77 $\pm$ 0.01	0.54 $\pm$ 0.003
D1	0.40	0.60	0.99 $\pm$ 0.001	0.82 $\pm$ 0.01	0.95 $\pm$ 0.003	0.87 $\pm$ 0.001
D2	0.34	0.65	1.00 $\pm$ 0.000	0.82 $\pm$ 0.01	0.94 $\pm$ 0.000	0.98 $\pm$ 0.000
U	0.90	0.90	1.00	0.93 $\pm$ 0.01	0.88 $\pm$ 0.01	0.59 $\pm$ 0.006
C	0.83	0.88	0.97 $\pm$ 0.004	1.00 $\pm$ 0.001	0.87 $\pm$ 0.01	0.60 $\pm$ 0.008
D1	0.80	0.88	0.97 $\pm$ 0.006	0.89 $\pm$ 0.01	0.99 $\pm$ 0.001	0.81 $\pm$ 0.008
D2	0.56	0.93	0.98 $\pm$ 0.006	0.87 $\pm$ 0.001	1.00 $\pm$ 0.000	0.97 $\pm$ 0.002

ous distribution between types D1 and D2. At the moment we have no method of detecting such distributions.

The most encouraging result to emerge from this study is the series of rather accurate values obtained when an RNA-sequence distribution is competed against by an identical distribution. Significant deviations from $p_c = 1$, therefore, will strongly suggest non-identity. Values of p_c between 1 and 0.8 can be produced readily by both qualitative or quantitative differences, and as p_c moves further away from 1 the probability increases that the difference is qualitative.

Alternative methods

Other interpretations of saturation (McCarthy & Bolton, 1964) and competition (Bolle *et al.* 1968; Lavallé & De Hauwer, 1968) experiments have been proposed. It is not our purpose to present detailed criticisms at this time; but we feel it is relevant to put forward our main reasons for seeking alternative methods.

Bolton & McCarthy (1962) described a hybridization technique with the use of agar to trap the denatured DNA. When the DNA-agar is annealed with different amounts of RNA, the plot of hybridization against RNA concentration follows a curve similar to the curves observed with the nitrocellulose-membrane method. McCarthy & Bolton (1964) suggested that the deviation from linear proportionality was mainly due to variation in the relative frequencies of different RNA sequences. It was thought that at intermediate RNA concentrations those complementary DNA and RNA sequences that happened to be equal in absolute concentration would react to completion. The unhybridized DNA sites would all correspond to RNA sequences that were less concentrated, and the unhybridized RNA to DNA sites that were saturated. Thus this interpretation, which has been taken up by Bolle *et al.* (1968), Kennell (1968) and Kennell & Kotoulas (1968), assumes that, when a DNA sequence and its complementary RNA copy are initially equal in concentration, they proceed to react to completion. We disagree with this assumption because the rate of DNA-RNA hybridization is limited by the concentrations of the reactants (Nygaard & Hall, 1964; Bishop, 1969b). Consequently, slowing of the reaction due to withdrawal of the reactants into the hybrid is maximal when their amounts are equal. Secondly, if the interpretation were correct, it would follow that the approach to the saturation plateau should be a straight line when a single RNA sequence (or an equimolar mixture of sequences) is annealed with DNA. We pointed out above that a number of homogeneous and nearly homogeneous RNA species of this type have recently been studied.

In each case a curvilinear approach to the saturation plateau was observed. Of particular importance is the very clear case of tyrosine tRNA studied by Landy *et al.* (1967) and by Andoh & Ozeki (1967).

The methods described by Lavallé & De Hauwer (1968) are based on a mathematical relationship identical in effect with our eqn. (1). However, they differ from us in assuming that the equation is generated by an equilibrium between free RNA and DNA, and DNA-RNA hybrid. As shown in Table 1, eqn. (1) fits data obtained at a number of reaction times, and therefore does not exclusively describe equilibrium conditions. However, although we do not favour an equilibrium hypothesis, data that would unequivocally disprove it are not available.

Conclusions

(1) The methods we have described here (essentially, eqns. 1, 2, 3 and 6) seem to be very suitable for the analysis of hybridization data when there is good reason to believe that the RNA that is hybridizing is either a single homogeneous species or a mixture of species at equal or very similar concentrations. Under these circumstances they make unnecessary the attainment of apparently saturating RNA concentrations.

(2) When a mixture of RNA sequences is being studied, and the different sequences are not known to be at equal or very similar concentrations, conclusions have to be qualified by reference to possible distributions. It is clear that RNA sequences present in concentrations lower than the mean concentration of the hybridizing sequences by more than 100-fold are unlikely to be detected. Difficulties of interpretation arise with discontinuous distributions where the lower concentration is 10–100-fold less than the mean. In Tables 3, 4 and 5 we provide a rough framework within which experimental results may be assessed. Here we present an assessment of the hybridization experiments described in this paper.

(a) The interpretation of the experiments with cRNA synthesized on denatured DNA is particularly important because they provide an estimate of the hybridization factor. In the three experiments shown in Table 2, the ratio $r_{\max}/r_s$ was 0.78, 0.83 and 0.78 respectively. From Tables 3 and 4 we conclude that either r_s is within 10% of the hybridization value, or the cRNA contains a discontinuous distribution similar to type D2. If this is so, about 80% of r_s is due to the more concentrated part of the distribution. For Expt. 1 (Table 2) this is 0.22. Making the most unfavourable assumption, that this represents copies of all of 22% of the DNA sequences, the widest discontinuous distribution is given if the remaining complementarity between RNA and DNA (0.54–0.22=

0.32) represents fewer copies of the remaining DNA sequences ($1 - 0.22 = 0.78$). The relative frequency of these would be $0.32/0.78 = 0.41$. Thus, in this case, the distribution cannot be sufficiently discontinuous to introduce an error greater than 10% in r_s .

(b) Comparison between *Micrococcus* cRNA and *E. coli* cRNA. In the saturation experiments [Fig. 2 and Table 2 of the main paper (Bishop & Robertson, 1969)], the ratio $r_{\max}/r_s$ reaches values of 0.84 and 0.85 respectively. From Tables 3 and 4, this suggests that the estimates of r_s are within 10%, unless the distributions are discontinuous and tending towards type D2 rather than D1. In the competition of labelled *Micrococcus* cRNA by unlabelled *E. coli* cRNA [Fig. 3 of the main paper (Bishop & Robertson, 1969)] the concentration of labelled *Micrococcus* cRNA used gave $r_{\max}/r_s = 0.51$, and $p_c \sim 0.73$. From Table 5, either the observed difference is qualitatively real, or part of *Micrococcus* cRNA that is not discontinuously distributed is being competed against by *E. coli* cRNA that is discontinuously distributed and tending towards D2.

(c) Competition between labelled 16–20 min. mRNA or labelled *Micrococcus* cRNA and unlabelled 4 min. mRNA [Figs. 5 and 6 of the main paper (Bishop & Robertson, 1969)]. $c_0/c_s = 0.44$ and 0.50 respectively; $p_c \sim 0.57$ and 0.56 respectively. From Table 5 either the differences are qualitatively real, or about half the competing RNA is distributed like type D2, i.e. about half the sequences are 100-fold lower in concentration than the remainder.

(d) Competition between labelled 16–20 min. mRNA and unlabelled *Micrococcus* cRNA. $c_0/c_s = 0.44$ and $p_c \sim 0.79$. The difference is evidently real, but could be produced by a distribution like type D1 in the *Micrococcus* cRNA. Thus we cannot conclude that *Micrococcus* cRNA fails to transcribe any late mRNA sequences.

(3) Use in more complex systems. Although hybridization-saturation and -competition experiments with non-uniform RNA sequence distributions raise difficulties of interpretation, it seems likely that they will continue to be used for the characterization of RNA preparations. Our experiments and models show principally that, with the use of appropriate methods, observed differences

between RNA preparations are always real. There is a difficulty, however, in determining whether the differences are qualitative or quantitative (distributional) differences. Similarly, widely different distributions may go undetected.

Difficulties are likely to be increased in working with systems more complex than bacteriophage T4. In saturation experiments, with cRNA and rat DNA, for example, most of the complementary DNA sequences do not hybridize because the number of RNA sequences is so great that the concentration of each is too low (Melli & Bishop, 1969). As the genetic and physiological complexity of the system increases, the existence of wider RNA-sequence distributions is to be expected, with correspondingly greater effects on saturation and competition experiments.

J. O. B., J. A. B. and M. M. are members of the Medical Research Council Epigenetics Research Group, and F. W. R. is a member of the Agricultural Research Council Unit of Animal Genetics.

REFERENCES

- Andoh, T. & Ozeki, H. (1967). *Proc. nat. Acad. Sci., Wash.*, **59**, 782.
- Bishop, J. O. (1969a). *Biochim. biophys. Acta*, **174**, 636.
- Bishop, J. O. (1969b). *Biochem. J.* **113**, 805.
- Bishop, J. O. & Robertson, F. W. (1969). *Biochem. J.*, **115**, 353.
- Bolle, A., Epstein, R. H., Salser, W. & Geiduschek, E. P. (1968). *J. molec. Biol.* **31**, 325.
- Bolton, E. & McCarthy, B. J. (1962). *Proc. nat. Acad. Sci., Wash.*, **48**, 400.
- Chamberlin, M. & Berg, P. (1964). *J. molec. Biol.* **8**, 297.
- Gillespie, D. & Spiegelman, S. (1965). *J. molec. Biol.* **12**, 829.
- Hayashi, M. N., Hayashi, M. & Spiegelman, S. (1965). *Biophys. J.* **5**, 231.
- Imamoto, F. & Yanofsky, C. (1967). *J. molec. Biol.* **28**, 1.
- Kennell, D. (1968). *J. molec. Biol.* **34**, 85.
- Kennell, D. & Kotoulas, A. (1968). *J. molec. Biol.* **34**, 71.
- Landy, A., Abelson, J., Goodman, H. M. & Smith, J. D. (1967). *J. molec. Biol.* **29**, 457.
- Lavallé, R. & De Hauwer, G. (1968). *J. molec. Biol.* **37**, 269.
- McCarthy, B. J. & Bolton, E. (1964). *J. molec. Biol.* **8**, 184.
- Mangiarotti, G., Apirion, D., Schlessinger, D. & Silengo, L. (1968). *Biochemistry*, **7**, 456.
- Melli, M. & Bishop, J. O. (1969). *J. molec. Biol.* **40**, 117.
- Nygaard, A. P. & Hall, B. D. (1964). *J. molec. Biol.* **9**, 125.
- Sinsheimer, R. L. & Lawrence, M. (1964). *J. molec. Biol.* **8**, 289.
- Stubbs, J. D. & Hall, B. D. (1968). *J. molec. Biol.* **37**, 289.