Renaturation Rate Studies of a Single Family of Interspersed Repeated Sequences in Human Deoxyribonucleic Acid†

Frank P. Rinehart, Thomas G. Ritch, Prescott L. Deininger, and Carl W. Schmid*

ABSTRACT: We have investigated the renaturation kinetics of a single family of cloned interspersed repeated sequences isolated from human DNA. Cross-renaturation studies of individual cloned sequences reveal heterogeneity in both the renaturation rate and the thermal stability of heteroduplexes formed from members of this family of sequences. However, cloned members of this family all renature with approximately the same number of copies in the human genome, demonstrating that they are a single family of sequences by the criterion of DNA renaturation kinetics. When a single cloned member of the family is renatured with total human DNA as a function of temperature, the thermal stability of the renatured heteroduplexes is found to be independent of the renaturation temperature over the range 25 °C below to 4 °C above the melting temperature. Further, the number of genomic copies with which this cloned family member reacts is also independent of the renaturation temperature from 25 °C below up to the melting temperatures. These observations demonstrate a remarkable degree of homogeneity in the evolutionary sequence divergence of members of this family. These results also demonstrate that renaturation kinetics can accurately measure the number of genomic copies of interspersed repeated DNA sequences.

The presence of repeated sequences in eukaryotic DNAs was first recognized by their relatively rapid rate of duplex renaturation (Britten & Kohne, 1968). To a first approximation, the genomic copy number of these repeated sequences is expected to be proportional to their second-order renaturation rate constant (Britten & Kohne, 1968; Wetmur & Davidson, 1968). Unfortunately, interpretations of this measurement are complicated by the evolutionary divergence of repeated DNA sequences (Britten & Kohne, 1968). Repeated sequences are not necessarily exact copies of the same sequence, but are usually similar base sequences which have diverged from a common ancestor. For this reason it is useful to regard a group of repeated sequences as a family of related but distinct members. It has always been recognized that renaturation kinetics might provide a different definition of a family of sequences than some other criterion, such as base sequence (Britten & Kohne, 1968).

Intuitively, the sequence divergence of members of a family would slow down the rate with which they cross-renature. This retardation of the renaturation rate would lead to an underestimate of the family's copy number and a corresponding overestimate of its sequence complexity. Several laboratories have calibrated the effect of base pair mismatching on the renaturation rate of DNA (Hutton & Wetmur, 1973; McCarthy & Farquhar, 1972; Bonner et al., 1973; Marsh & McCarthy, 1974). Most of these calibrations were performed on chemically modified DNAs to simulate mispairing. The results of these studies are in fair agreement with each other. As a rough rule, a 10% mispairing in the renatured heteroduplex corresponds to a 2-fold reduction in its renaturation rate (Bonner et al., 1973).

Although the results of these calibrations agree, they do not completely explain anomalies in the renaturation rates of several clustered repetitive DNAs. For example, Southern (1970) found that guinea pig satellite DNA has a renaturation rate corresponding to a complexity of 10^5 nucleotides (nt), whereas the pyrimidine tracts of this satellite correspond to a ten-nucleotide ancestral sequence. The number of copies of African green monkey α satellite estimated from its renaturation rate is also about 2 orders of magnitude too low (Singer & Donehower, 1979). It is conceivable that chemically modified DNAs may be a poor model system to simulate mispairing in evolutionarily modified DNAs. As one illustration of this point, two related Drosophila satellites, which consist of a tandem array of a seven-nucleotide subunit, share five out of seven bases (Gall et al., 1973). Although these satellites share 72% homology, they do not cross-renature under normal renaturation conditions, and by the criterion of renaturation they are segregated into separate families of sequences (Blumenfeld, 1973). Special sequence effects of this type could result in an underestimation of the copy number of repeated DNAs and an overestimation of their complexity, even when their renaturation rates are corrected according to the calibrations deduced for chemically modified DNAs.

We have recently found that most of the 300 nucleotide interspersed repeats in human DNA belong to a single family of sequences, the Alu family (Houck et al., 1979). Although DNA renaturation kinetics provided the first clue for the existence of this family, it is not certain that all members of the family can cross-renature. It is also possible that the renaturation kinetics of the interspersed repeats might be subject to the same anomalies and special effects described above for satellite DNAs. For example, the presence of the Alu family in both hamster and mouse is strong evidence for its presence in the rodent genome (Jelinek et al., 1980; Krayev et al., 1980). However, careful studies of the renaturation rate of rat DNA do not reveal the presence of a single ubiquitous Alu family of repeated sequences (Pearson et al., 1978).

The issue to be tested is whether a group of interspersed repeats, which are identified as a single family by their base sequences, are also identified as a single family by DNA renaturation. One way to approach this issue is to determine whether classes of sequences are homogeneously or heterogeneously diverged from their common ancestor (Bendich & Anderson, 1977). In the case of homogeneous divergence, each

† From the College of the Virgin Islands, Division of Science and Mathematics, St. Thomas, US, VI 00801 (F.P.R.), Department of Chemistry, University of California at Davis, Davis, California 95616 (T.G.R. and C.W.S.), and Department of Pediatrics, University of California at San Diego, La Jolla, California 92093 (P.L.D.). *Received November 14, 1980.* This work was supported by U.S. Public Health Grant GM 21346. C.W.S. was an American Cancer Society Scholar in Cancer Research, and P.L.D. was supported by U.S. Public Health Service Training Grant GM07199.

0006-2960/81/0420-3003$01.25/0

Blur 1/Blur 8

7x20x3xx3xx1x3xxxx12xxx6xx4xx1x2x2x1i1x7dx4 (101 bases overlap, 76% homology).

Blur 2/Blur 8

7x1xx5x6xx3x3xx3x2x3x10xx8x2x15x11x16x4xx5xxx2xx4x2x10x11x3x6x3x41x5xx22xxx 1xx3x2d712xx6x4x16 (265 bases overlap, 81.5% homology).

Blur 11/Blur 8

14x7i8xx7ixx10x10xx27x2x1x6x1x15xx9x2x1xx4x13xx10x3x5xx3x2x1i1x4xx15x5x2x 1xx3x3x3x2ix9x21 (267 bases overlap, 83.5% homology).

Blur 19/Blur 8

2x1x14x3xx1xxx1x6x1x8x1xx1x1i2x6ddddddddddx4x2x2x2x2x4x1xiiiiii1xx1xx 1xx3xx1xxx4x1x2x5xxxxx3x6x3x6x3x4i1x12x14x2x1xx14ix1xx6xx2 (234 bases overlap, 73% homology).

FIGURE 1: Map of heteroduplex structures formed by renatured Blur clones as predicted from their base sequences. The numbers indicate lengths of homologous runs and x indicates a base-pair mismatch. d indicates where a clone has a deletion compared with Blur 8 and i indicates an inserted base relative to the sequence of Blur 8. The total number of overlapping bases between each clone and Blur 8 and the percent sequence homology are indicated at the end of each heteroduplex map. The lengths of the clones determined by their sequences are Blur 1, 156 bases; Blur 2, 301 bases; Blur 8, 266 bases; Blur 11, 281 bases; and Blur 19, 241 bases.

sequence renatures with the same rate constant to form duplexes with the same extent of mispairing, so that all members are identified as belonging to a single family. Increasing the renaturation temperature affects the renaturation rate of all members by the same factor but does not change the size of the homogeneous family until the melting temperature of the resulting duplex is exceeded. In contrast, individual members of a heterogeneously divergent family cross-renature with distinct rate constants. As the renaturation temperature is increased, highly divergent family members of a heterogeneous family no longer cross-renature to form highly mispaired low thermal stability duplex structures. In effect the copy number of a heterogeneous family decreases at higher renaturation temperatures as it segregates into non-cross-renaturable subgroups. The two Drosophila satellites described above are extreme examples of a heterogeneously diverged family; Bendich & Anderson (1977) and Bouchard & Swift (1977) have found that repetitive DNA from several plants is homogeneously diverged.

Materials and Methods

DNA Samples. Individual members of the human Alu family of renatured interspersed repeats have been cloned into the *Bam*HI site of *E. coli* plasmid pBR 322 as previously described (Rubin et al., 1980). The clones (Blur 1, 2, 8, 11, and 19) were purified as closed circular molecules by two rounds of ethidium bromide/CsCl density gradient sedimentation equilibrium (Clewell, 1972).

The clone Blur 8 was arbitrarily selected to drive the renaturation of the radiolabeled cloned Alu family members as described below. This driver DNA, which includes both the plasmid and human insert sequences, was sheared by several passages through a French press to a single-strand length of 350 nucleotides, as estimated by alkaline gel electrophoresis. The heteroduplex structures formed between Blur 8 and the other clones as predicted by their base sequences are depicted in Figure 1.

Total human nuclear DNA was isolated from placenta as previously described (Schmid & Deininger, 1975). This sample was sheared in a French press to a single-strand length of 450 nucleotides.

The human inserts in the Blur clones were ^{32}P end labeled as follows: Plasmid clones were restricted with *Bam*HI endonuclease to excise the human inserts. The resulting linear molecules were end labeled with a mixture of the four deoxynucleotide [α-^{32}P]triphosphates (Amersham) using the Klenow fragment of *E. coli* DNA polymerase I (Friedman et al., 1979). The resulting solution was fractionated on a Sephadex G-200 column to separate the excluded DNA from the included triphosphates. Human insert DNA was separated from linear plasmid DNA by neutral agarose gel electrophoresis.

Renaturation Studies and Melting Studies. Prior to renaturation, ^{32}P-labeled inserts were mixed with an excess of sheared human DNA, or the sheared Blur 8 driver DNA, or for purposes of a self-renaturation control no added DNA. For minimization of variability in renaturation studies, single stock solutions of Blur 8 driver DNA and of sheared human DNA were employed thoughout the project. Identical recipes and time points were used for all the comparison studies of different clones and for the temperature studies of individual clones. In most cases the excess of driver Alu family complements to tracer DNA was at least a 1000:1 molar ratio. In one case, Blur 19, it was necessary to use a lower molar ratio (57:1) to achieve the lower C_0t values. Even in the worst case of Blur 19 a molar ratio of 1500:1 was used for later C_0t points and, as demonstrated by the control, self-renaturation of the tracer was not a problem over the time interval studied. These DNA mixtures in 0.025 M sodium phosphate buffer were denatured by immersion in boiling water for 2 min and quenched on ice. The samples were subsequently adjusted to 0.12 M sodium phosphate and brought to the renaturation temperature.

The extent of renaturation was assayed by hydroxylapatite chromatography. In most cases the column assay temperature was identical with the renaturation temperature. However, in the experiment of Figure 4 for which the renaturation rate was studied at and above the DNA melting temperature, a column assay temperature of 60 °C was employed. Cherenkov counting was employed to monitor the radiolabel.

DNA melting temperatures were determined by thermal elution of renatured DNA from hydroxylapatite. In most cases the DNA was eluted in 0.12 M phosphate buffer; however, in the experiment of Figure 5, 0.1 M phosphate buffer was employed. As described by Martinson (1973), double-strand DNA can desorb without denaturation from hydroxylapatite at higher temperatures. The use of the lower buffer concentration is expected to provide a more accurate and slightly higher measure of the duplex melting temperature (Fox et al., 1980). Bonner et al. (1973) employ the thermal elution of duplex DNA from hydroxylapatite in deriving their empirical calibrations of the effect of mismatching on the DNA renaturation rate. Since we use the same assay in this work, their empirical calibrations should accurately apply to these results.

Theory and Analysis. The integrated rate equation for the second-order renaturation of single strand DNA is

$$\frac{C}{C_0} = \frac{1}{(1 + k_2 C_0 t)} \quad (1)$$

where C is the concentration of single-strand DNA in molar units, C_0 is its initial concentration, k_2 is a second-order rate constant, and t is the time in seconds (Britten & Kohne, 1968; Wetmur & Davidson, 1968).

In addition we are interested in the case where a radiolabeled tracer DNA having a concentration T reacts with a driver DNA having a much higher concentration, C. For this case

$$\frac{dT}{dt} = -k_T CT \tag{2}$$

where k_T is the second-order rate constant for formation of a heteroduplex between the tracer and driver DNAs. We assume that the tracer concentration is so low that its second-order self-renaturation is negligible. The integrated rate obtained after substituting eq 1 into eq 2 is

$$\frac{T}{T_0} = \left[\frac{1}{(1 + k_2 C_0 t)}\right]^{k_T/k_2} \tag{3a}$$

or

$$\ln \frac{T}{T_0} = \frac{-k_T}{k_2} \ln (1 + k_2 C_0 t) \tag{3b}$$

where T_0 is the initial concentration of the tracer DNA (Bonner et al., 1973; Hinnebush et al., 1978). One experimental aim in this work is to estimate the ratio of rate constants for heteroduplex and homoduplex formation. This ratio is estimated as the least-squares slope of the straight line plot suggested by eq 3b. In evaluating this slope we also allowed for the presence of a nonrenaturable contaminant in preparations of the radiolabeled tracers. In most cases, this was a small amount of the sample which could be accurately estimated by exhaustively renaturing the tracer with total human DNA. The maximum value for this contaminant, 19%, occurred in the case of Blur 11. For other cases it was ~5%. Except for this one independent estimate, which is a constant for each preparation, there are no adjustable parameters in the determination of the least-squares slope of the line corresponding to eq 3b. In a few cases there is systematic error between early time renaturation points and the result of this least-squares procedure which weights later time points more heavily. Otherwise, the results of the simple theory of eq 3a and the least-squares procedure to estimate k_T/k_2 agree with the experimental data (see Results). Although more accurate representations of the data would be possible with more detailed equations, we elected this method as being ideal for detecting relative trends in the data; it is simple and unbiased in application. Equation 3a is based on the assumption that the driver DNA is depleted with second-order kinetics. This is not exactly true (Smith et al., 1975). Single-stranded tails on the ends of these renatured driver molecules may be available for renaturation with tracer molecules. For well-matched complementary sequences, the rate constant for the renaturation of a single-strand tracer with these tails is typically retarded by more than an order of magnitude as compared to its renaturation with a free single strand (Smith et al., 1975). This retardation results from shorter sequence complements and possibly steric effects on the single-stranded tails of duplex molecules. Tails will only renature with homologous tracer sequences. When the DNA fragment length is greater than the repeat length, the tails are expected to contain homologous DNA less often. The analogous retardation in rate constant has not been calibrated for the case of imperfect complements but is presumably even more severe. We have found (vide infra) that the renaturation data for imperfect complements qualitatively resembles the shape of curves drawn for eq 3a and does not resemble theoretical curves predicted for the contribution of single-stranded tails to the renaturation of tracer (Smith et al., 1975). This difference suggests that renaturation of single-stranded tails does not make an important contribution to the formation of heteroduplexes and suggests that the rate of heteroduplex formation with single-strand tails is severely retarded. Accordingly, we will ignore the possible contribution of these single-strand tails toward the kinetics of heteroduplex formation. This assumption may lead to an overestimate in the rate constant for heteroduplex formation.

In a few cases we also measured the renaturation rate at or above the duplex melting temperature where the reverse reaction is significant and the overall process goes toward equilibrium. For this case we find

$$k_2 C_0 t = \frac{1}{F} \ln \left[\left(\frac{\frac{C}{C_0} + \frac{R}{2(1 + 1/R)} + \frac{1}{2}F}{\frac{C}{C_0} + \frac{R}{2(1 + 1/R)} - \frac{1}{2}F} \right) \Bigg/ \left(\frac{1 + \frac{R}{2(1 + 1/R)} + \frac{1}{2}F}{1 + \frac{R}{2(1 + 1/R)} - \frac{1}{2}F} \right) \right] \tag{4}$$

where

$$F = \left(\frac{R^2}{(1 + 1/R)^2} + \frac{4R}{(1 + 1/R)} \right)^{1/2}$$

and R is the mass ratio of single-strand and double-strand DNA at equilibrium (Benson, 1960).

Results

The renaturation rate of a radiolabeled single member of the Alu family, Blur 8 insert, driven with an excess of sheared Blur 8 (plasmid and insert), was studied as a function of temperature (Figure 2A). As expected for a single sequence driving itself, the renaturation of this clone is accurately described as a single second-order transition. The clone's absolute renaturation rate is consistent with its sequence complexity (Britten & Kohne, 1968; Wetmur & Davidson, 1968). The temperature dependence of its renaturation rate also agrees with similar studies on other systems (Wetmur & Davidson, 1968; Bonner et al., 1973).

In contrast to Blur 8's second-order kinetics of self-renaturation, the cross-renaturation of another labeled Alu family member, Blur 2 insert, driven with Blur 8 shows marked deviations from second-order kinetics (Figure 2B). In these experiments the self-renaturation of the driver, Blur 8, is a second-order reaction which competes with the cross-renaturation of the driver and the tracer, Blur 2. In effect the driver concentration is rapidly depleted so that the renaturation kinetics of the tracer lags behind. This deviation from simple second-order kinetics becomes even more pronounced at higher temperatures approaching the melting temperature of the Blur 2–Blur 8 heteroduplex (~72 °C; see below). An alternative explanation of result of this temperature study is that only a decreasing fraction of the probe is capable of heteroduplex renaturation at higher temperatures. This explanation is unlikely since the probe is a homogeneous DNA sample so that all probe molecules should be equivalent in their renaturation properties. It should also be noted that as judged by gel electrophoresis the probe molecules remain intact during renaturation at even the highest temperatures.

The kinetics of heteroduplex formation in the presence of a competing second-order reaction can be described by eq 3a which is derived for this situation (Figure 2B) (Bonner et al., 1973; Hinnebush et al., 1978). The renaturation rate constant for heteroduplex formation between the driver, Blur 8, and several labeled human inserts has been analyzed by use of this equation as illustrated for the case of Blur 2 (Figure 2B). This

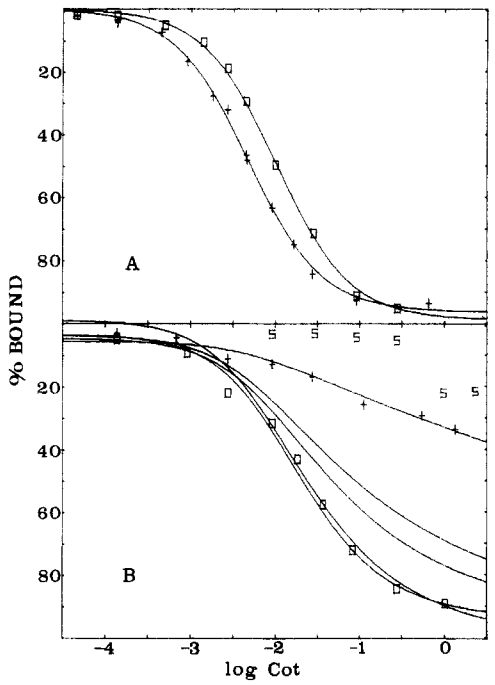

FIGURE 2: (A) Renaturation of ^{32}P Blur 8 insert DNA with Blur driver DNA in 0.12 M phosphate buffer. Data points are shown for studies at 50 (□) and 70 °C (+). The data for other temperature studies are omitted for clarity. The least-squares second-order rate constants for these studies are 95 M^{-1} s^{-1}, 50 °C; 151 M^{-1} s^{-1}, 60 °C; 210 M^{-1} s^{-1}, 65 °C; 206 M^{-1} s^{-1}, 70 °C; 140 M^{-1} s^{-1}, 75 °C; and 115 M^{-1} s^{-1}, 80 °C. For comparison the melting temperature of the renatured Blur 8 homoduplex is 88 °C. (B) Renaturation of ^{32}P Blur 2 insert DNA driven with Blur 8 DNA in 0.12 M phosphate buffer. Data points are shown for studies at 50 (□) and 70 °C (+). The data points at other temperatures are omitted for clarity. Curves in order of decreasing rate are least-square solutions of 3b for 50, 55, 60, 65, and 70 °C studies. Values for the ratios of rate constant of Blur 8 driver ^{32}P Blur 2 DNA heteroduplex to Blur 8 driver ^{32}P Blur 8 homoduplex in 3a are, respectively, 0.644, 0.452, 0.325, 0.239, 0.069. Points denoted by S show the amount of self-renaturation of ^{32}P Blur 2 DNA during time periods equivalent to the value of C_0t at which they are plotted. For comparison, the melting temperature of the renatured Blur 2 Blur 8 heteroduplex is approximately 72 °C.

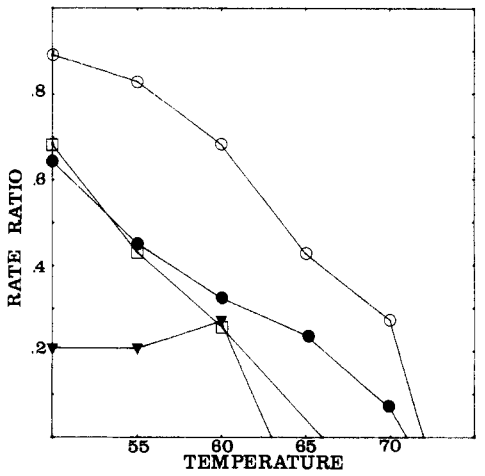

FIGURE 3: The points represent the ratio of the rate constant for renaturation of the indicated ^{32}P-labeled Alu family members driven by Blur 8 driver to the rate constant for renaturation of ^{32}P-labeled Blur 8 DNA driven by Blur 8 driver. (○) Blur 11, (●) Blur 2, (□) Blur 1, (▼) Blur 19. The lines intersect the abcissa at the melting temperature of the ^{32}P-labeled Alu family member Blur 8 driver heteroduplex from hydroxylapatite. For comparison the homologous duplexes would have a T_m of approximately 88 °C.

quantity expressed as the ratio of the renaturation rates of heteroduplex to homoduplex formation has been studied as a function of temperature (eq 3b, Figure 3). The melting temperature of each heteroduplex has also been determined and is indicated as an intercept on the abscissa in Figure 3.

There is the possibility of considerable experimental uncertainty in these estimates of the rate constant ratio. However, we believe the data are reliable enough to conclude that there are differences in both the renaturation rates and the thermal stabilities of the different Blur heteroduplexes (Figure 3). All rate constant ratios are less than unity. Further, there is a general decrease in the rate ratio for a particular heteroduplex at higher temperatures. There is also a rough correspondence between the heteroduplex melting temperatures, the extent of base pair mismatching, and their renaturation rates (Figures 1 and 3). All of these observations are qualitatively consistent with the known effects of sequence mismatching on the renaturation rate of DNA (Bonner et al., 1973). There are, however, some quantitative differences between these rate constant ratios and calibrations of the effects of sequence mismatching on the renaturation rates. When these ratios are corrected for both length effects and base pair mismatching, the corrected ratios are significantly greater than unity with an average value of approximately two. The details of these corrections are described below (Bonner et al., 1973; Hinnebush et al., 1978). The failure of these corrections to adjust the observed ratios to exactly unity implies there is some systematic error in either our data or the calibration of the effects of mismatching or that these calibrations do not exactly apply to the system studied here. As described under Theory and Analysis we ignore the contribution of tails to the formation of heteroduplex structures. As a consequence our values for these ratios should be regarded as potential overestimates. As another potential source of error the calibrations for mismatching are based on studies of randomly sheared DNA fragments which are modified at random sites, whereas in this study rate constants for formation of specific heteroduplexes are measured. This possibility has been previously discussed (Bonner et al., 1973). We do not know which of these three possibilities correctly explains the 2-fold discrepancy in the corrected rate ratio. This discrepancy does not affect any of the conclusion which we draw from this study. Marsh & McCarthy (1974) have also calibrated the temperature dependence of the effect of duplex mismatching on the renaturation rate. Comparing the results of their calibration to the data of Figure 3, we find that in one case (Blur 11) the agreement is excellent but that in the other three cases their calibration tends to underestimate the renaturation rate depression by about 2-fold. Marsh and McCarthy used UV hypochromism as a measure of both the renaturation rate and duplex thermal stability, whereas Bonner et al. (1973) based their calibrations in part on studies of radiolabeled DNA samples and the use of hydroxylapatite binding assays. For this reason it seems more appropriate to use the calibrations proposed by Bonner et al. in interpreting the present results.

In principle, the number and arrangement of base pair mismatches in the Blur heteroduplexes (Figure 1) determine the number and size of available nucleation sites for heteroduplex formation and could be used to predict the relative renaturation rates as a function of temperature for various clones (Wetmur & Davidson, 1968; Hutton & Wetmur, 1973). In practice we have been unable to use this sequence information to predict these renaturation rates accurately. It is interesting that the thermal stabilities of the Blur clone heteroduplexes are in fair agreement with their base-pairing fidelity (Figures 1 and 3). Using the data of Figure 1 and taking 88 °C as the melting temperature of homoduplexes, we predict melting temperatures of 64, 69.5, 71.5, and 61 °C for the

Table I: Copy Number of Individual Members of Alu Family Measured by Renaturation with Sheared Nuclear Cell DNA

family member	rate constant[a] (M^{-1} s^{-1})	apparent copy no.[b]	hydroxylapatite melting temp of DNA renatured at 60 °C (°C)	optimum temp correction[c]	base-pair mismatch correction[d]	length correction[e]	corrected copy no.[f]
Blur 1	23.9	0.49×10^5	65	1.0	4.89	2.31	0.55×10^6
Blur 2	75.8	1.57×10^5	70	1.0	3.56	1.52	0.85×10^6
Blur 8	55.7	1.15×10^5	70	1.0	3.39	1.68	0.62×10^6
Blur 11	119	2.46×10^5	73	0.9	2.83	1.61	1.01×10^6
Blur 19	34.2	0.71×10^5	67	1.0	4.32	1.72	0.53×10^6

[a] Measured in 0.12 M phosphate at 60 °C. [b] The rate constant divided by the rate constant for 600 nt long single-copy DNA fragments in 0.12 M phosphate at 60 °C, $k = 4.85 \times 10^{-4}$ (Schmid & Deininger, 1975; Holland et al., 1980). [c] Estimated from Figure 4 of Bonner et al. (1973). T_m of native (i.e., undenatured) Alu family members is 88 °C. [d] Based on an estimate of a factor of two decrease in the rate constant for each 10 °C difference in the thermal stability of native and renatured DNA (Bonner et al., 1973). [e] Calculated from eq 7 of Hinnebusch et al. (1978). Correcting the rate constant for a tracer of length L (Figure 1) driven by sheared nuclear DNA of length 450 nucleotides to a rate constant for a tracer of 600 nucleotides in length driven by sheared nuclear DNA of the same length; 600 nucleotides was the length of sheared nuclear DNA used to measure the single-copy rate constant (Schmid & Deininger, 1975). [f] Apparent copy number times product of three correction factors (columns 3, 5, 6, and 7).

heteroduplexes format between Blur 8 and Blurs 1, 2, 11, and 19, respectively, using the relationship of 1 °C depression in melting temperature for each 1% base-pair mismatching (Bonner et al., 1973). These predications compared favorably with the values indicated in Figure 3.

The differences between the renaturation rates and thermal stabilities of the heteroduplexes are a consequence of sequence heterogeneity among Alu family members. As an extreme result of such heterogeneity one might imagine that individual members of the family may only renature with a select subgroup of the entire family. In this event, renaturation rate kinetics would not provide a meaningful measure of the copy number of the family as illustrated by the example of the Drosophila satellite DNAs described in the introduction.

To examine this issue we determined the renaturation rate of each of the radiolabeled Blur clone inserts when driven by total human DNA at 60 °C. In each case the kinetics are accurately described by a single second-order reaction. The rate constants for these reactions are listed in Table I along with the corresponding heteroduplex melting temperatures. The rate constants for five cloned inserts differ by 5-fold which should correspond to a similar range in apparent copy number (Table I). To estimate the true copy number there are at least three corrections which must be applied to this apparent copy number: First, the optimum temperature for renaturation of a heteroduplex occurs at about 25 °C below the mean of the heteroduplex and homoduplex melting temperatures; for a homoduplex the optimum occurs at about 25 °C below its melting temperature (Bonner et al., 1973). In this particular experiment the renaturation temperature (60 °C) happens to be 5 °C above the optimum for heteroduplex formation and 5 °C below the optimum for homoduplex formation so that the correction is negligible (Table I). Second, the repetitive heteroduplex structures are mismatched so that their renaturation rate is slowed relative to the well-paired single copy sequences. This is estimated to be a 2-fold reduction in the renaturation rate for each 10% of base-pair mismatching in a heteroduplex (Bonner et al., 1973). Third, there is a difference in the lengths of the Blur clone tracers and the length of the fragments used to measure the single copy renaturation rate (Schmid & Deininger, 1975). The effect of DNA fragment length on the renaturation rate has been exactly formulated (Hinnebush et al., 1978). Upon correcting the apparent copy number for all three effects, we find only a 2-fold range in the corrected copy number, $(0.5–1.0) \times 10^6$ (Table I). There is considerable uncertainty in our experimental results as well as the corrections we applied to these results. Consequently, a factor of 2 as the extreme range in the corrected copy numbers may well be within experimental uncertainty.

We conclude that each of the labeled cloned inserts renatures with the same large number of genomic copies of Alu family members. The number is in fact large enough to account for essentially all of the 300 nt interspersed repeats in human DNA (see Discussion). This implies that each member of the Alu family can cross-renature with essentially all other members. Thus, by the criterion of renaturation kinetics we conclude that the Alu family is a single family of sequences and its exact copy number can be estimated from its renaturation rate (Discussion).

There is a possible contradiction between the results of Figure 3 and those of Table I. The results of Figure 3 imply that at least at some renaturation temperature some individual Alu family members can no longer cross-renature so that the overall family would consist of nonrenaturable subgroups. In contrast, the results of Table I imply that at 60 °C each cloned family member reacts with essentially the entire family. We have distinguished between these alternatives by examining the renaturation rate of Blur 8 with total human DNA as a function of temperature.

The reaction between labeled Blur 8 and sheared human DNA follows simple second-order kinetics when renatured at 50, 60 (data not shown), and 70 °C (Figure 4). Remarkably, the resultant duplexes formed at these three renaturation temperatures have indistinguishable melting profiles (Figure 5). Thus, the Alu family appears to be a homogeneously divergent family of sequences. Heterogeneity in this family would be revealed by both a decrease in the copy number of the family at higher renaturation temperatures and an increase in the thermal stability of those duplexes in the thermal stability of those duplexes which can renature at higher temperatures (Bendich & Anderson, 1977). As discussed below, the copy number of the family is also constant for renaturation temperatures of 50–75 °C. In these experiments the melting temperature, 75 °C, was measured in 0.10 M phosphate buffer (Table II), whereas in previous experiments the melting temperature, 70 °C, was measured in 0.12 M phosphate buffer (Table I). This difference does not affect any of the conclusions we reach in this work.

As an even more stringent test of the homogeneity of the Alu family, we studied its renaturation at both the melting temperature of 75 °C and above the melting temperature at

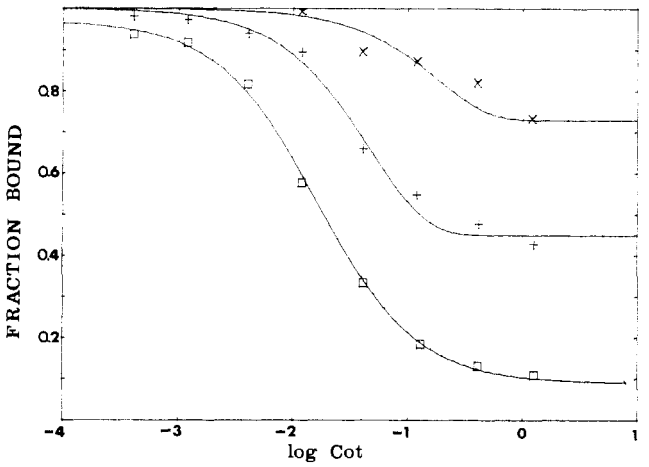

FIGURE 4: Renaturation of ^{32}P Blur 8 insert DNA with sheared nuclear driver DNA in 0.12 M phosphate buffer. The data points shown are for studies at 70 (□), 75 (+), and 78.8 °C (×). Rate constants for these and other temperature studies are summarized in Table II. The curve drawn for the 70 °C data is a least-squares second-order fit (eq 1). The curves for the 75 and 78.8 °C data are for a second-order approach to equilibrium (eq 4) with rate constants of 12.5 and 1.5 M^{-1} s^{-1} and equilibrium single-strand to double-strand mass ratios of 0.818 and 2.70, respectively.

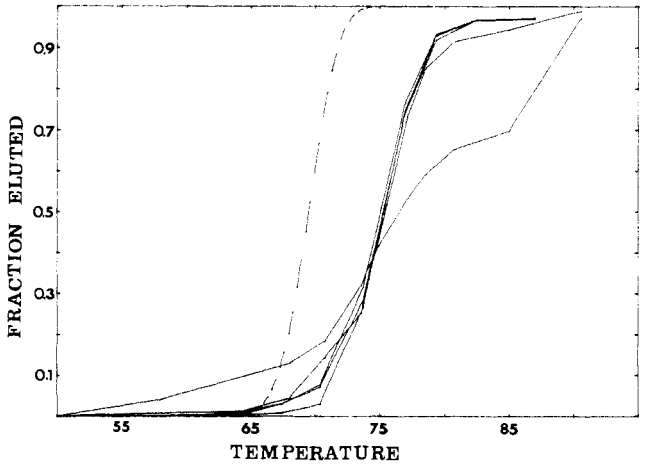

FIGURE 5: Thermal stabilities of ^{32}P Blur 8 insert DNA sheared nuclear DNA heteroduplexes renatured at various temperatures to C_0t = 0.4. Thermal elution was performed in 0.1 M phosphate buffer for reasons described in the text; consequently these melting temperatures cannot be directly compared to the values cited in Table I. The curves for the renaturation products formed at 50, 60, 70, and 75 °C are indistinguishable. A portion of the heteroduplex renatured at 78.8 °C melts at a higher temperature as described in the text. The dashed curve is for a Poisson distribution of mutations in Alu family members which is described in the text.

78.8 °C (Figure 5). The renaturation at these high temperatures no longer follows simple second-order kinetics but presumably is instead a reversible reaction which achieves an equilibrium between single and double strands (Figure 4). The kinetics at these temperatures can be satisfactorily represented by eq 4, which is derived for this reversible case. The data for the 78.8 °C renaturation are not very accurate because of the small amount of renaturation and the obvious need for very precise temperature control at this point in the thermal transition of the duplex (Figures 4 and 5). We regard the data at 75 °C as being more reliable. It is noteworthy that the renaturation at 75 °C levels off to an equilibrium duplex concentration, 55%, which is consistent with its thermal stability (Figures 4 and 5). The thermal stability of the 75 °C renaturation product is indistinguishable from the duplexes renatured at lower temperatures, 50–70 °C (Figure 5). Even

Table II: Renaturation Rate of ^{32}P-Labeled Clone 8 Insert with Sheared Nuclear DNA

renaturation temp (°C)	rate constant (M^{-1} s^{-1})	optimum temp correction[a] (°C)	corrected copy no. × 10^{-3}[b]
50	38.3	1.1	360
60	55.7	1	476
70	60	1.4	718
75	12.5	~2.5	~267

[a] Estimated from Figure 4 of Bonner et al. (1973) taking 88 °C as the melting temperature of the native, i.e., unrenatured, Alu family. For the 75 °C correction it was necessary to extrapolate the results of Bonner et al. so that the correction is approximate. [b] Calculated as the rate constant (column 2) times the optimum temperature correction (column 3) multiplied by the factor 1.68 × 2.46/(4.83 × 10^{-4}). The value of 1.68 is the length correction described in Table I. The value of 2.46 is a correction for base-pair mismatching estimated from the rule of a 2-fold rate reduction for 10% mismatching (Bonner et al., 1973). The value of 4.83 × 10^{-4} is the renaturation rate constant of approximately 600 nt single-copy DNA fragments measured in 0.12 M phosphate at 60 °C (Schmid & Deininger, 1975; Holland et al., 1980).

when the DNA is renatured at 3.8 °C above its melting temperature, the duplexes which form have the same thermal stability component (Figure 5). In addition, the 78.8 °C renaturation product includes a minor high-temperature component which we attribute to the self-renaturation of the labeled Blur 8 tracer (Figure 5). Self-renaturation of the tracer is expected to be more evident in the 78.8 °C renaturation product, first because there is less total renaturation at 78.8 °C so that the relative amount of self-renaturation is enhanced and second because the tracer concentration is not effectively depleted by renaturation with human DNA so that it will self-renature more extensively. Except for this minor experimental problem, we conclude that the thermal stability of the renatured Alu family duplex is independent of renaturation temperature from 25 °C below to 4 °C above its melting temperature. This is especially strong evidence that the Alu family is homogeneous in its sequence divergence.

The dependence of the copy number of the Alu family on its renaturation temperature is also a measure of its homogeneity. Although the apparent copy number of the family dramatically decreases at higher renaturation temperatures (Table II), most of this decrease can be attributed to the general effect of temperature on the renaturation rate of DNA (Bonner et al., 1973). By use of the calibration of Bonner et al. (1973), the apparent copy number can be corrected to the value which would be observed at the optimum renaturation temperature. Upon applying this correction, we find that the copy number is roughly constant over a 25 °C range in renaturation temperature (Table II). Heterogeneity in the sequence divergence of the family members would result in a decreasing family size at higher renaturation temperatures (Bendich & Anderson, 1977). This is further evidence that the Alu family is essentially homogeneous in its sequence divergence.

The finding that the Alu family is homogeneous in its sequence divergence does not necessarily mean that each member of the family has diverged in sequence by exactly the same amount. Instead, this finding implies that the family is not segregated into discernable subgroups which cross-renature at different rates. Base sequence studies of ten cloned Alu family members show that they all share a recognizable consensus sequence but that individual members have diverged by an average of approximately 13% from this common sequence (Rubin et al., 1980; Dininger et al., 1980). Blur 8,

as an example, shows 10.9% divergence form the consensus sequence. The positions of the mutations within individual member sequences occur at random positions and do not identify discernable subgroups. The mutations in the Alu family might be represented by a Poisson distribution of mutations relative to their average consensus sequence. We have estimated the number of base-pair mismatches in heteroduplexes constructed between the entire Alu family, which is taken as a Poisson distribution of mutations having an average of 10% divergence, and a single clone of the Alu family, which has arbitrarily been assigned a divergence of 8%. For simplicity we have assumed that the number of heteroduplex base-pair mismatches is the sum of the number of mutations in both strands. [This assumption leads to a slight overestimate in the number of mismatches in those heteroduplexes which have the highest degree of mismatching. The thermal stability of these heteroduplexes can be estimated by assuming a 1 °C depression for each 1% mismatching in the duplex DNA (Bonner et al., 1973).] The difference in melting temperatures between the observed and predicted curves (Figure 5) is not important for the present purposes but may be attributed to some of the uncertainty in the rule of 1 °C depression for every 1% base-pair mismatching (Bonner et al., 1973). The width of the melting range for a Poisson distribution of mutations is even narrower than the observed melting range of the Blur 8 Alu family heteroduplexes (Figure 5). Because the widths of the theoretical and observed transitions are approximately equal, we conclude that a "homogeneously" diverged family of sequences could be composed of member sequences which have randomly diverged to different extents from a common sequence. Our only requirement is that a homogeneous family cannot be composed of special sequence variation subgroups which discernably differ in their rates of cross-renaturation and in heteroduplex thermal stability (Bendich & Anderson, 1977). In effect, the thermal stability and renaturation rate studies do not distinguish between a homogeneously divergent and a randomly divergent family of sequences.

Discussion

There are two principal results from this study: the Alu family is found to be essentially homogeneously divergent in base sequence, whereas heteroduplexes formed between individual members of the family are heterogeneous in both thermal stability and renaturation rate. It is possible to reconcile these findings. One case reflects average properties, whereas the other case reflects individual distinctions. Base sequence studies show that individual cloned 300 nt Alu family members resemble a common consensus sequence (Rubin et al., 1980; Deininger et al., 1980). These same studies also show that mutations are randomly distributed throughout the sequence with an average mutational frequency of 10%. Presumably, there is a distribution of mutations among the members of the Alu family. Heteroduplexes constructed from individual cloned members of such a family would have different base-pair fidelities and would also be heterogeneous in the positions of these mutations (Figure 1). Consequently, we expect these heteroduplexes to differ somewhat in both their thermal stabilities and renaturation rates, as is the case observed here (Figure 3). Since individual clones resemble the consensus sequence to different degrees, we also expect heteroduplexes formed between individual clones and the entire Alu family to differ somewhat in both their renaturation rates and thermal stabilities. This is also the observed case (Table I).

In contrast to these two observations, the majority of heteroduplexes formed between an individual clone and the ~600 000 members (see below) of the Alu family are rather similar in their renaturation rate and thermal stability so that the family appears to be remarkably homogeneous. Although the effect of mismatched base pairs on thermal stability is large (a 1 °C depression per 1% mismatched), the predicted melting range of heteroduplexes with a Poisson distribution of mutations is rather narrow. We would be unable to detect any heterogeneity in the melting temperatures of such heteroduplexes. Renaturation rate kinetics, which undergoes only a 2-fold change for 10% mismatching, is even more insensitive to the effects of mismatching than the thermal stability. It is very unlikely that this heterogeneity would be apparent in the temperature studies of the copy number (Table II). In other words, when we compare any one cloned sequence to the entire Alu family, most of the other family members have diverged from it by exactly the same extent when assayed by the crude techniques of melting temperature and renaturation rate.

The Blur clones studied here were isolated as 300 nt heteroduplexes from S_1 digests of renatured repeated human DNA (Rubin et al., 1980). It is conceivable that this procedure selects for a subset of the Alu family which forms especially well-paired 300 nt S_1 resistent heteroduplexes. Further, this heteroduplex is corrected to a homoduplex by the *E. coli* repair system so that the final sequence is probably an average of the two heteroduplex sequences. Both the isolation and repair of the Blur clones could select for sequences which more closely resemble each other than random members of the Alu family. According to this argument the Blur clones would provide an underestimate of the observed heterogeneity in renaturation rate and thermal stability of Alu family heteroduplexes. In base-pairing studies with total human DNA, however, the entire complement of Alu family members is available for renaturation with the Blur clone. Consequently, the finding that the entire family is homogeneously divergent is not likely to be affected by this possible bias in selecting the Blur clones.

The finding that heteroduplexes of individual cloned sequences differ in their thermal stability and renaturation rate is of technical interest. It is useful to employ individual cloned sequences as hybridization probes to detect other members of a repetitive sequence family. It is possible, particularly at higher temperatures, that certain cloned sequences would not cross-hybridize and would not be recognized as belonging to the same family of sequences.

The finding that the entire family of sequences is homogeneously divergent is of some biological significance. One could imagine sequence subgroups of the family would be fine-tuned to perform specialized functions. If such subgroups were identified by gross modifications of certain regions in the sequence, the Alu family would be heterogeneously divergent. Instead we find here that the Alu family is, practically speaking, homogeneous in its divergence. Presumably, these putative subgroups could still be distinguished by minor or highly localized specific sequence modifications. Members of a homogeneous family have all experienced the same evolutionary history, such as random mutational divergence from a common ancestral sequence (Bendich & Anderson, 1977).

The effect of base-pair mismatching is to slow the renaturation rate of DNA (Britten & Kohne, 1968; Bonner et al., 1973; Marsh & McCarthy, 1974). For a homogeneously divergent family of repeated sequences, the renaturation rate of each member would be slowed by about the same amount (Bendich & Anderson, 1977). For a homogeneously divergent family, we can therefore in principle estimate its true copy

number by correcting its observed renaturation rate for the effects of base pair mismatching.

We have estimated a broad range in values, 0.27 to 1.0 million copies, for the repetition frequency of the Alu family in the human genome (Tables I and II). In addition to this broad range of values, there is the additional possibility of some systematic error in either our data or one of the several corrections we employed to adjust the apparent copy number. The cause of our concern is that, when we apply the same correction factors to the renaturation rates of the heteroduplexes formed between Blur 8 and the other Blur clones, we find that the corrected rate is up to 2-fold too fast. Accordingly, our range of 0.27 to 1.0 million copies for the Alu family might also be up to a factor of 2 too large. Keeping the possibility of a systematic error in mind, we find an average value of 0.6 ± 0.2 (rms) million copies of this sequence (Tables I and II).

600 000 Alu family sequences could account for all of the interspersion of 300 nt repeats in human DNA. This repetition frequency corresponds to 7.2% ($0.6 \times 10^6 \times 300$ bp/2.5×10^9 bp) of the human genome which is taken to have a complexity of 2.5×10^9 bp (Lewin, 1974). This value of 7.2% agrees with our previous estimate that the 300 nt long interspersed repeats in the "50 000"-fold repetitive sequences class, which includes the Alu family, would make up at most only 7% of the genome (Houck et al., 1979). If these 0.6×10^6 copies were randomly distributed throughout the genome they would be spaced at an average distance of 4200 bp [2.5×10^9 bp/(0.6×10^6)]. For comparison, there are at least 7 Alu family members located within a 65 000-bp region of the human β-globin gene cluster corresponding to an average spacing of about 9000 bp (Fritsch et al., 1980). However, the interspersion of 300 nt repeats is not uniform throughout the human genome. About 60% of the human genome is occupied by close interspersion of 2000 nt single-copy sequences with 300 nt repeated sequences; repeated sequences are more distantly interspersed in the remainder of the genome (Schmid & Deininger, 1975; Deininger & Schmid, 1976). If we restrict the 0.6 million Alu family sequences to 60% of the genome, they would then have an average spacing of 2500 nt [$0.6 \times 2.5 \times 10^9$ (0.7×10^6)]. We conclude from these comparisons that the copy number of the Alu family is sufficient to account for all of the interspersion of 300 nt repeated sequences in human DNA. This does not preclude the existence of other 300 nt interspersed repeated sequences in human DNA but merely indicates that they are much less abundant than the Alu family members.

As discussed in the introduction, it is reasonably certain that Alu family sequences are found in the rodent genome. However, the renaturation rate of rat DNA, selected as a particularly well-studied rodent example, does not suggest the presence of a highly abundant Alu family of sequences (Pearson et al., 1978). Since in this work we find that carefully interpreted renaturation rates can be an accurate if not precise measure of the copy number of the human Alu family, we must conclude that Alu family sequences are either much less abundant in rat than in human DNA or that the renaturation rate of the rat Alu family is subject to some anomaly which does not influence the renaturation rate of human DNA.

Acknowledgments

We wish to thank Mike Fox for examining the length of DNA fragments after renaturation and Dr. Catherine Houck for stimulating critical discussions.

References

Bendich, A. J., & Anderson, R. S. (1977) *Biochemistry 16*, 4655.

Benson, S. W. (1960) *The Foundations of Chemical Kinetics*, p 30, McGraw-Hill, New York.

Blumenfeld, M. (1973) *Cold Spring Harbor Symp. Quant. Biol. 27*, 423.

Bonner, T. I., Brenner, D. J., Neufeld, B. R., & Britten, R. J. (1973) *J. Mol. Biol. 81*, 123.

Bouchard, R. A., & Swift, H. (1977) *Chromosoma 61*, 317.

Britten, R. J., & Kohne, D. (1968) *Science (Washington, D.C.) 161*, 529.

Clewell, D. B. (1972) *J. Bacteriol. 110*, 667.

Deininger, P. L., & Schmid, C. W. (1976) *J. Mol. Biol. 106*, 773.

Deininger, P., Jolly, D., Friedman, T., Rubin, C. M., Houck, C. M., & Schmid, C. W. (1980) in *Mechanistic Studies of DNA Replication and Recombination* (Alberts, B., & Fox, C. F., Eds.) (in press).

Fox, G. M., Umeda, J., Lee, R. K.-Y., & Schmid, C. W. (1980) *Biochim. Biophys. Acta* (in press).

Friedman, T., Esty, A., LaPorte, P., & Deininger, P. (1979) *Cell (Cambridge, Mass.) 17*, 715.

Fritsh, E. F., Lawn, R. M., & Maniatis, T. (1980) *Cell (Cambridge, Mass.) 19*, 959–972.

Gall, J. D., Cohen, E. H., & Atherton, D. D. (1973) *Cold Spring Harbor Symp. Quant. Biol. 28*, 417.

Hinnebush, A. G., Clark, V. E., & Klotz, L. C. (1978) *Biochemistry 17*, 1521.

Holland, C. A., Mayrand, S., & Pederson, T. (1980) *J. Mol. Biol. 138*, 755.

Houck, C. M., Rinehart, F. P., & Schmid, C. W. (1979) *J. Mol. Biol. 132*, 289.

Hutton, J. R., & Wetmur, J. G. (1973) *Biochemistry 12*, 558.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L., & Schmid, C. W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1398.

Krayev, A. S., Kramerov, D. A., Skryabin, K. G., Ryskov, A. P., Bayer, A. A., & Georgiev, G. P. (1980) *Nucleic Acids Res. 8*, 1201–1215.

Lewin, B. (1974) *Gene Expression 2: Eukaryotic Chromosomes*, Wiley, New York.

Marsh, L. J., & McCarthy, B. J. (1974) *Biochemistry 13*, 3382.

Martinson, H. G. (1973) *Biochemistry 12*, 145.

McCarthy, B. J., & Farquhar, M. N. (1972) *Brookhaven Symp. Biol. No. 23*, 1–41.

Pearson, W. R., Wu, J.-R., & Bonner, J. (1978) *Biochemistry 17*, 51.

Rubin, C. M., Houck, C. M., Deininger, P. L., Friedmann, T., & Schmid, C. W. (1980) *Nature (London) 284*, 372.

Schmid, C. W., & Deininger, P. L. (1975) *Cell (Cambridge, Mass.) 6*, 345.

Singer, D., & Donehower, L. A. (1979) *J. Mol. Biol. 134*, 835.

Smith, M. J., Britten, R. J., & Davidson, E. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 4805.

Southern, E. M. (1970) *Nature (London) 227*, 794.

Wetmur, J. G., & Davidson, N. (1968) *J. Mol. Biol. 31*, 349.