Organization and Heterogeneity of Sequences within a Repeating Unit of Human Y Chromosome Deoxyribonucleic Acid†

Louis M. Kunkel,‡ Kirby D. Smith,* and Samuel H. Boyer

ABSTRACT: Fragments of 3.4 kilobases (kb) are released from DNA of human males, but not DNA of human females, by cleavage with restriction endonucleases *Hae*III, *Eco*RI, or *Eco*RII. Most, if not all, reiterated DNA which is specific for the Y chromosome (it-Y DNA) is present within these male-specific 3.4-kb molecules. Although such 3.4-kb molecules are themselves localized to the Y chromosome, this is not true for all sequences within them. At least two distinguishable types of reiterated sequences are found within each 3.4-kb molecule. One type consists of at least two families which are highly reiterated and are *not confined* to the Y chromosome. The other type is composed of an estimated minimum of 39 families, each moderately reiterated and localized to the Y chromosome. Y-specific and non-Y-specific sequences are interspersed with one another in the same 3.4-kb molecule. In the average 3.4-kb molecule, three 800 nucleotide lengths of Y-specific sequences alternate with four 250 nucleotide lengths of non-Y-specific sequences. Since the total number of families of Y-specific sequences, calculated on the basis of reiteration frequency and total abundance in a male genome, greatly exceeds the number of Y-specific sequences present in a single 3.4-kb molecule, it necessarily follows that the population of these 3.4-kb molecules is heterogeneous.

The overall sequence organization of human DNA has been shown (Schmid & Deininger, 1975; Ginelli & Corneo, 1976) to be similar to that found for most other eukaryotes (Davidson et al., 1975) insofar as the majority of repetitive sequences are interspersed with nonrepetitive ones. However, details concerning the organization of particular sets of reiterated sequences and the potential role of such sequences in determination of function (Britten & Davidson, 1969) are obscure. One potential way to uncover these details is to limit the variety of sequences to be examined by studying the organization of DNA within a single chromosome. With this in mind, we have analyzed the organization of human Y-chromosome reiterated DNA isolated by two independent methods.

The *first* method of isolation depends upon reassociation between radiolabeled 47,XYY DNA (DNA isolated from a human male with two Y chromosomes) and excess DNA isolated from a woman (female DNA). As described previously (Kunkel et al., 1976), the reiterated radiolabeled DNA that remains unreassociated after two such challenges is human Y-chromosome specific. By using DNA isolated from individuals with variant Y chromosomes, these reiterated Y-chromosome-specific DNA sequences (it-Y) were mapped to the long arm of the human Y chromosome and estimated to represent between 15 and 30 families, each reiterated 300 to 600 times (Kunkel et al., 1977).

The *second* method of isolation is an extension of that first outlined by Cooke (1976). Cleavage of human male DNA with the restriction endonuclease *Hae*III yields two prominent fragments, 3.4 kilobases (kb) and 1.9 kb in length. Neither of these is detected in *Hae*III digests of human female DNA. The resistance of the 3.4-kb fragment to cleavage by a number of other restriction endonucleases facilitates its purification (Cooke, 1976).

Partial nucleotide sequence analysis suggests that at least a portion of 3.4-kb *Hae*III Y DNA consists of variations of a short repetitious oligonucleotide (Cooke, 1976). Although the 3.4-kb fragment is undetectable in females, some sequence homology exists between it and female DNA (Kunkel et al., 1977; Bostock et al., 1978; Cooke & McKay, 1978). Indeed, it has been proposed by Cooke & McKay (1978), largely on the basis of hybridization to DNA digests in agarose gels, that *all* sequences within 3.4-kb *Hae*III Y DNA have some homology to sequences on other chromosomes. They suggest that only the location of restriction enzyme recognition sites is truly male specific. A similar proposal has been made by Bostock et al. (1978).

In this report, we use a combination of methods to examine the nature and organization of DNA sequences within 3.4-kb *Hae*III Y DNA. We show that these molecules contain both Y-specific and non-Y-specific sequences, that the Y-specific sequences are homologous with it-Y DNA, that Y-specific and non-Y-specific sequences coexist and are regularly interspersed with one another in the same 3.4-kb *Hae*III Y DNA molecules, and that these molecules are heterogeneous with respect to the particular sequences within them.

Experimental Procedures

DNA Isolation. DNA was isolated from placentae by methods similar to those described previously (Kunkel et al., 1977). One to two milligrams of DNA per gram of tissue was recovered from placentae frozen immediately after delivery.

Radiolabeling of DNA. Introduction of ^{3}H-labeled nucleotides into purified DNA via nick translation (Kelly et al., 1970) followed an adaptation of conditions previously described (Schachat & Hogness, 1973; Maniatis et al., 1975b). The reaction was carried out in a 100-μL final volume containing the following: 50 mM potassium phosphate buffer (KPB) at a final pH of 7.5 at 25 °C; 30 μM deoxynucleotide triphosphates (unlabeled dATP, dGTP, dCTP, Sigma; [^{3}H]TTP, 40–60 Ci/mmol, Amersham/Searle); 5 mM $MgCl_2$; 1.5 units of DNA polymerase I from *Escherichia coli* (Boehringer Mannheim; 2.9 units/μL); and 10 ng of DNase I from beef pancreas (Worthington). DNase I used in the nick translation reaction had previously been frozen at −20 °C in 10 mM HCl at 1 mg/mL. Prior to use, DNase was thawed and diluted

† From the Howard Hughes Medical Institute Laboratory for Human Biochemical Genetics, Division of Medical Genetics, Department of Medicine, Johns Hopkins University School of Medicine and Hospital, Baltimore, Maryland 21205. *Received February 13, 1979.* This research has been supported in part by grants from the National Institutes of Health.

‡ Present address: University of California, San Francisco, School of Medicine, Department of Biochemistry and Biophysics, San Francisco, CA 94143.

0006-2960/79/0418-3343$01.00/0

1/100 with 0.1% BSA (Miles, 10 mM Tris-HCl and 5 mM $MgCl_2$ at a final pH of 7.5 at 25 °C; thereafter, the diluted DNase was held 2 h at 4 °C and finally further diluted 1/10 with 50 mM KPB and 5 mM $MgCl_2$ (pH 7.5 at 25 °C). Ten microliters of the twice-diluted DNase I was added to the reaction mixture. The complete mixture was incubated at 13.5 °C for 4 h. The reaction was terminated by addition of 0.1 volume of 0.5 M NaEDTA (pH 8.0) and incubation at 70 °C for 10 min. After chilling to 4 °C, unincorporated deoxynucleotide triphosphates were separated from the radiolabeled DNA by a passage through a column (0.75 × 11 cm) of Bio-Gel A-0.5, 100–200 mesh (Bio-Rad Laboratories), equilibrated with 0.1 × SSC (1 × standard saline citrate: 150 mM NaCl, 15 mM sodium citrate) and 5 mM NaEDTA, pH 8.9 at 25 °C. The specific activity of the final DNA product ranged between 5×10^7 and 1×10^8 dpm/μg.

No DNase was used during nick translation of 3.4-kb *Hae*III fragments and incubations were performed for 24 h instead of 4 h. To minimize possible bias arising from base composition, radiolabeled dCTP (18 Ci/mM, Amersham/Searle), dGTP (32 Ci/mM, NEN) and dTTP (46 Ci/mM, Amersham/Searle) were used along with unlabeled dATP.

Preparation of it-Y DNA. ^{3}H-Labeled it-Y DNA was prepared as previously described (Kunkel et al., 1977). Briefly reiterated male DNA, radiolabeled by nick translation, was reassociated with excess unlabeled female DNA at 60 °C in 0.12 M PB. The radiolabeled male DNA which failed to reassociate with female DNA was separated from reassociated molecules by hydroxylapatite (HAP) chromatography.

Shearing and Size Determination of DNA. DNA was sheared to give three different lengths: 300, 600, and 2100 nucleotides. Three hundred nucleotide lengths of DNA were obtained by sonication at 18-W output for 15 min at 4 °C in 1 mL of SSC with a Bronson sonifier (Heat Ultrasonics) fitted with a microtip. Six hundred nucleotide long DNA was obtained by sonication with the same instrument at a setting of 22 W for 5 min at 4 °C in 10 mL of SSC. The 2100 nucleotide long DNA was obtained by shearing in a dual speed commercial Waring Blendor (Model 91-262) at low setting for 20 min at 4 °C in SSC.

Two hundred nucleotide long fragments of 3.4-kb *Hae*III molecules were produced by boiling for 45 min in 0.2 N NaOH. Radiolabeled 1500 nucleotide long DNA was selected from a preparation of 3.4-kb *Hae*III Y DNA molecules nick translated to one-tenth the usual specific activity.

The size of sheared DNA samples was assayed by electrophoresis either as native or denatured molecules. Electrophoresis of native samples was performed in 0.5% agarose and 4% acrylamide disk gels and bands were developed with 2 μg/mL ethidium bromide (Sharp et al., 1973). Following illumination with short-wave ultraviolet, the mean nucleotide length as well as the observable maximum and minimum length was calculated by measurement of distances migrated relative to standards of SV40 DNA (gift of Dr. Thomas Kelly, Johns Hopkins) digested with *Hin*dIII. Three hundred nucleotide long DNA gave a maximum and minimum dispersion of approximately 100 nucleotides on either side of the mean; the 600 and 1500 nucleotide DNA had a larger dispersion of ±200 nucleotides. The fragment size of 2100 nucleotide DNA (dispersion ± 250 nucleotides) was determined by electrophoresis in 1% agarose disk gels and comparison with parallel gel analysis of λ phage DNA (Miles Laboratories) digested with *Hin*dIII.

Single-strand DNA samples were analyzed on gels similar to those described by Maniatis et al. (1975a). Gels were stained overnight with ethidium bromide (2 μg/mL) at 4 °C and analyzed in the same manner as native DNA samples.

For size determination of radiolabeled fragments, gels were sliced by hand into 0.5-cm sections and broken into smaller pieces by forced passage through a 16-gauge hypodermic needle. Samples were removed from the gels by overnight diffusion into 2.5 mL of 0.1 × SSC at 37 °C. Radioactivity in gel slices was determined by liquid scintillation assay of diffused material. Alternatively, radiolabeled samples were applied to 5% acrylamide and 0.2% ethylene diacrylate disk gels. Following electrophoresis, 0.5-cm sections were dissolved in 0.1 N NaOH, and the amount of radioactivity in each gel slice was determined by liquid scintillation assay.

Analytical DNA/DNA Reassociation Assays. Radiolabeled DNA samples were assayed by trace-driver reassociation as described previously (Kunkel et al., 1977). Reassociation incubations shown in Table I were carried out at 60 °C in 0.12 M phosphate buffer (PB) that had a pH of 6.8 at 25 °C. In order to limit breakage of molecules during prolonged incubations at 60 °C, reassociation incubations shown in other tables and figures were performed at 60 °C in a solution containing 165 mM NaCl, 15 mM sodium citrate, 10 mM Tris-HCl, and 1 mM EDTA and having a pH of 7.6 at 25 °C. For each assay point, approximately 150 μg of driver and 1200 dpm of trace DNA were boiled together, incubated at 60 °C to appropriate C_0t and, after dilution (~1/20) in 0.12 M PB, passed across a 1-mL bed volume hydroxylapatite column equilibrated in 0.12 M PB and held at 60 °C. Reassociated duplexes were eluted with 0.3 M PB and the HAP finally dissolved in 6 M HCl. The reassociated molecules (0.3 M PB and HCl fractions) and unreassociated molecules (0.12 M PB fraction) were monitored both radiometrically and optically. The percentage reassociated was calculated as a fraction of the total recovered molecules.

Digestion with Nuclease S_1. Two different methods were used for nuclease S_1 digestion of DNA. When cleavage opposite single-strand nicks was necessary, the procedure described by Efstratiadis et al. (1976) was employed. Briefly, DNA to be digested was either dialyzed or diluted into 0.2 M NaCl, 50 mM sodium acetate, and 1 mM $ZnSO_4$ at a final pH of 4.5 (25 °C). As necessary, wheat germ RNA (Sigma) was added as carrier to obtain a final nucleic acid concentration of 50 μg/mL. Thereafter, nuclease S_1 (Sigma, 1.5 units/μg of nucleic acid) was added and the reaction mixture incubated at 37 °C for 45 min. The reaction was stopped by phenol and chloroform extraction followed by ethanol precipitation. To ensure uniform recovery, carrier RNA was added to give a final nucleic acid concentration of 100 μg/mL prior to ethanol precipitation.

When complete digestion of unreassociated sequences was desired, nuclease S_1 assay was performed in a manner similar to the method of Sutton (1971). DNA was either dialyzed or diluted into 0.2 M NaCl, 0.04 M sodium acetate, and 10^{-5} M $ZnSO_4$ at a final pH of 4.5 (25 °C) and incubated with nuclease S_1 (Sigma, 1 unit/μg of nucleic acid) at 37 °C for 45 min at a nucleic acid concentration of 30 μg/mL. The reaction was stopped either by Cl_3CCOOH precipitation or by phenol and chloroform extraction. Extracted samples were further purified by dialysis vs. 0.12 M PB and subsequent collection of resistant duplexes on HAP.

Digestion with Restriction Endonucleases. Restriction endonucleases were purchased from Bethesda Research Laboratories (BRL) and New England Biolabs. Conditions of use in each instance followed those recommended by the supplier. Before experimental use, each enzyme was tested

for activity on either λ phage DNA or whole genome human DNA substrates. All digestions were incubated for at least 18 h at 37 °C and each was terminated by the addition of 0.1 volume of 0.5 M NaEDTA (pH 8.0) and 2 M KCl. Each sample was ethanol precipitated and the precipitate resuspended in 10 mM NaCl, 10 mM Tris-HCl, and 1 mM NaEDTA at a pH of 7.6 (25 °C).

Isolation of 3.4-kb HaeIII Y DNA from Human Male DNA. Human male DNA (4 mg) purified from placentae was incubated with 650 units of *Hae*III restriction endonuclease (BRL) in a reaction volume of 15 mL containing 6 mM Tris-HCl, 6 mM NaCl, 6 mM $MgCl_2$, and 6 mM 2-mercaptoethanol at a final pH of 7.4 at 25 °C. After incubation for 24 h at 37 °C, the reaction was stopped by ethanol precipitation and the DNA dialyzed into RPC-5 starting buffer (1.5 M sodium acetate and 0.1 M Tris-HCl, pH 7.4, 25 °C). Chromatography on RPC-5 adsorbant (Miles) was performed under conditions similar to those described by Landy et al. (1976) and Hardies & Wells (1976). DNA (3.7 mg) was pump-loaded under pressure, at a concentration of 500 μg/mL, atop a 0.9 × 65 cm high-pressure glass chromatography column (Beckman) previously packed with RPC-5. The packed column was free of entrapped gas and equilibrated with starting buffer. At a pump pressure of ~130 psi, a 2-L linear gradient ranging from 1.5 M sodium acetate and 0.1 M Tris (pH 7.4, 25 °C) to 2 M sodium acetate and 0.1 M Tris (pH 7.4, 25 °C) was passed through the column at 0.8 mL/min at room temperature. The eluate was monitored at 260 nm by continuous flow through a Beckman Model 25 spectrophotometer. Six-milliliter fractions were collected, and 1.0 μg of DNA from each of 40 fractions was assayed by electrophoresis on an analytical 1% agarose gel (Sugden et al., 1975). Those fractions containing 3.4-kb *Hae*III Y DNA were pooled and, after concentration, dialyzed vs. a mixture of 1 volume of *Hin*dIII reaction buffer (10 mM Tris-HCl, final pH 7.4, 25 °C; 7 mM $MgCl_2$; 60 mM NaCl) and 1 volume of *Eco*RII reaction buffer (100 mM Tris-HCl, pH 7.3 at 25 °C; 5 mM $MgCl_2$; 60 mM NaCl). These dialyzed solutions were made 0.01% in autoclaved gelatin (Difco) and 5 mM 2-mercaptoethanol and incubated overnight at 37 °C with a mixture containing 40 units each of the following restriction enzymes: *Hin*dIII (BRL), *Bam*HI (BRL), and *Eco*RII (BRL). The reaction mixture was ethanol precipitated and redissolved DNA was loaded onto a 1.75 × 16 cm cylinder of 1% agarose gel made up in 40 mM Tris base, 13 mM sodium acetate, and 2 mM NaEDTA and adjusted to pH 8 at 25 °C with acetic acid. Following electrophoresis, purified 3.4-kb *Hae*III DNA was cut from the ethidium bromide stained gel and recovered on HAP (Cooke, 1976) with a final yield of approximately 2 μg.

Hybridization to DNA in Agarose Gels. Restriction endonuclease digests of human DNA were assayed by electrophoresis on 1% agarose slab gels (Sugden et al., 1975) and stained with ethidium bromide (Sharp et al., 1973). After photography of the stained gel, the DNA was transferred to a Millipore filter by a method similar to that described by Southern (1975). The filter was then immersed for 24 h at 60 °C in 3 mL of Denhardt buffer (Denhardt, 1966) containing ~0.01 μg of sheared ^{3}H-labeled 3.4-kb *Hae*III Y DNA (specific activity: $\sim 1 \times 10^8$ dpm/μg). After hybridization and washing, the filter was dried and exposed to X-ray film as described by Laskey & Mills (1975).

Verification, via Computer Simulation, of Sequence Organization within 3.4-kb HaeIII Y DNA. Random breakpoints along a 3400 digit line were simulated by a random number generator program by using an HP 9825A computer. Fragment lengths were defined as intervals between two such breakpoints. Lengths of all sizes were acquired and from these 10 000 fragments were selected which corresponded to each of the three DNA lengths shown in Table II: 200 ± 100, 300 ± 100, and 650 ± 200. Since the location of each fragment along the line is defined by the digits with which it begins and ends, the location of fragments along the line can be compared with predicted arrangements. The percentage of fragments in each size class which did not contain designated non-Y-specific regions was taken as a measure of Y-chromosome specificity. Model arrangements involving various numbers and lengths of each of the two types of sequences were tested for their fit to observed data (Table II) for HAP-assayed Y-chromosome specificity.

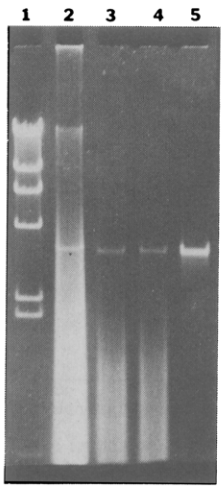

FIGURE 1: Analytical gel electrophoresis of DNA samples at different steps in the purification of the 3.4-kilobase (kb) *Hae*III Y DNA. Analytical 1% agarose slab gel of the following DNA samples: (slot 1) 1.5 μg of *Hin*dIII digested λ phage DNA; (slot 2) 3 μg of 46, male DNA after digestion with *Hae*III; (slots 3 and 4) 1.5 μg of DNA derived from conjoint digestion with *Bam*HI, *Eco*RII, and *Hin*dIII of a pool of RPC-5 column fractions which contained 3.4-kb *Hae*III Y DNA; (slot 5) 0.1 μg of purified 3.4-kb *Hae*III Y DNA after successive RPC-5 chromatography, triple enzyme digest with *Bam*HI, *Eco*RII, and *Hin*dIII, preparative cylindrical agarose gel electrophoresis, and HAP chromatography.

Results

Isolation of 3.4-kb HaeIII Y DNA. Electrophoretic analysis of 3.4-kb *Hae*III Y DNA at successive stages of purification is shown in Figure 1. Visual observation suggests that the fragment comprises more than 90% of the DNA displayed in lane 5.

Reassociation of Radiolabeled 3.4-kb HaeIII Y DNA with Excess Male and Female Human DNA: Identification of Both Y-Specific and Non-Y-Specific Sequences. Purified 3.4-kb *Hae*III Y DNA was radiolabeled by nick translation, sheared under conditions yielding 300 nucleotide lengths, and reassociated with either excess male, female, or *E. coli* DNA. As seen in Figure 2, nearly all ^{3}H-labeled 3.4-kb *Hae*III Y DNA sequences reassociate with male DNA. However, 50–55% also reassociate with female DNA. This means there are at least two types of DNA sequences within the population of 3.4-kb *Hae*III Y DNA molecules: those which are specific to the Y chromosome and those which are not. The curves in this figure were determined by computer-assisted least-squares analysis of the data (Kells & Straus, 1977). Based on a $C_0t_{1/2}$ of 2.2×10^{-1}, the non-Y-specific fraction is homologous to sequences reiterated ~22 700 times in a diploid female genome. The $C_0t_{1/2}$ for the reassociation of these same sequences with male DNA was 1.3×10^{-1}, indicating a re-

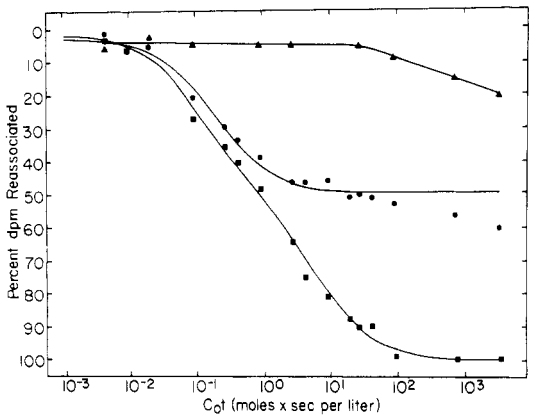

FIGURE 2: Reassociation of radiolabeled 3.4-kb *Hae*III Y DNA with male and female human DNA. Purified 3.4-kb *Hae*III Y DNA was radiolabeled by nick translation (Experimental Procedures), sonicated to 300 nucleotides in length, and then reassociated as trace DNA in a trace-driver reassociation with the following excess DNAs: male (■—■); female (●—●); *E. coli* (▲—▲). Prior to reassociation, sonicated ^{3}H-labeled, 3.4-kb *Hae*III Y DNA was denatured by heating at 100 °C for 5 min in 0.12 M phosphate buffer (PB) and was incubated at 60 °C to a C_0t of 1×10^{-8}, and snapback sequences (~25%) were removed by passage over a 0.5-mL bed volume hydroxylapatite column equilibrated at 60 °C in 0.12 M PB. The curves presented here are an aggregate of points from three sets of assays with excess human male DNA and two with both human female DNA and *E. coli*. In four instances for each curve, plotted values at particular C_0t values represent the mean of duplicate values. All points on the curves were normalized to 100% based on the final reassociation observed for the radiolabeled DNA with male DNA. For most of the determinations, the final observed reassociation of the radiolabeled trace with male DNA was 83.5%. The curves for the reassociation of ^{3}H-labeled, 3.4-kb *Hae*III Y DNA with human male and female DNA were determined by computer-assisted least-squares analysis of the data (Kells & Straus, 1977). The last three C_0t points of the reassociation reaction with female DNA were omitted from the least-squares analysis since they were influenced by the self-reassociation of the trace 3.4-kb *Hae*III molecules as monitored by its reassociation with *E. coli* DNA.

iteration frequency of ~38 500. The final 45% of the reassociation between male driver DNA and ^{3}H-labeled 3.4-kb *Hae*III Y DNA has a $C_0t_{1/2}$ of 8.5. Thus, Y-specific 3.4-kb *Hae*III sequences like it-Y sequences (Kunkel et al., 1977) have a reiteration frequency of ~600.

Although the total number of families of non-Y-specific 3.4-kb *Hae*III Y DNA sequences is not immediately clear, the finding that they reassociate across more than two decades of C_0t suggests (Britten et al., 1974) that at least two families of sequences are present.

Self-Reassociation of 3.4-kb HaeIII Y DNA Fragments and Homology between Y-Specific Sequences and it-Y Sequences. Reassociation between radiolabeled 3.4-kb *Hae*III Y DNA and a known amount (0.033 μg) of unlabeled 3.4-kb *Hae*III Y DNA, both of which were sonicated to 300 nucleotide size, results in the two-component curve shown in Figure 3. As assayed by HAP, the faster component represents ~55% of total reassociable DNA and the slower component ~45%. The addition of 59 μg of unlabeled sonicated female DNA caused a 12-fold acceleration in the reassociation of the faster component but had no effect on the slower component. Comparison of Figures 2 and 3 makes it obvious that the most rapidly reassociating component in Figure 3 is the non-Y-specific portion of 3.4-kb *Hae*III Y DNA. When the $C_0t_{1/2}$ of the faster component is based on the 59 μg of added female DNA, a value of 2×10^{-1} is obtained. This is similar to the $C_0t_{1/2}$ of 2.2×10^{-1} calculated for the reassociation of trace 3.4-kb *Hae*III Y DNA with excess female DNA in Figure 2. Since the reassociation of the slow component in Figure 3 is unaffected by the addition of female DNA, it must be Y specific.

Figures 2 and 3 also demonstrate that it-Y DNA and the Y-specific component of 3.4-kb *Hae*III Y DNA are similar in several respects. *First*, the Y-specific component of 3.4-kb *Hae*III Y DNA reassociates with male DNA (Figure 2) with kinetics similar to those observed for isolated it-Y sequences (Kunkel et al., 1977). *Second*, homology between the Y-specific portion of 3.4-kb *Hae*III Y DNA and it-Y DNA is evident in Figure 3. As indicated by $C_0t_{1/2}$ values in the Figure 3 legend, two preparations of ^{3}H-labeled it-Y DNA reassociated with an excess (0.033 μg) of unlabeled sonicated 3.4-kb *Hae*III Y DNA at rates similar to those observed for the self-reassociation of the Y-specific portion of 3.4-kb *Hae*III Y DNA.

A complication of these analyses is that reassociation between 3.4-kb *Hae*III Y DNA and each of the two ^{3}H-labeled it-Y DNA preparations is incomplete. These incomplete reassociations may, in part, be caused by breakdown of radiolabeled trace material during prolonged incubations rather than by a lack of complete homology between it-Y DNA and Y-specific sequences in 3.4-kb *Hae*III Y DNA. In fact, if the $C_0t_{1/2}$ of these it-Y sequences is estimated on the basis of only 70% homology, it-Y sequences are calculated to be more reiterated in 3.4-kb *Hae*III Y DNA than they are in the whole male genome. In contrast, when $C_0t_{1/2}$ values are estimated after allowance is made for 30% breakdown of it-Y DNA, the reiteration frequencies for it-Y sequences within 3.4-kb *Hae*III Y DNA are quite similar to those found in whole genome male DNA (Table I). We conclude that Y-specific sequences within 3.4-kb *Hae*III Y DNA molecules are identical with it-Y DNA and that most, if not all, it-Y sequences are present in such molecules.

Relationships between Y-Specific and Non-Y-Specific Sequences in 3.4-kb HaeIII Y DNA: Evidence for Linkage within the Same Molecules. Although our studies of sequence homology (Figure 2) and reassociation kinetics (Figure 3) leave no doubt that both Y-specific and non-Y-specific sequences are present in 3.4-kb *Hae*III Y DNA, the relationship between these two kinds of sequences is still unclear. Are these two types of reiterated sequences linked within the same molecule or are there two classes of 3.4-kb *Hae*III molecules: one Y specific, the other non-Y-specific?

In order to answer this question several size classes of tritiated 3.4-kb *Hae*III Y DNA preparations were produced by random fragmentation. Thereafter, the degree of Y-chromosome-specific reassociation obtained with each size class was measured by both HAP and S_1 nuclease assays. As detailed in the Table II legend, Y-chromosome specificity was based on a comparison of 3.4-kb *Hae*III Y DNA reassociation with excess female DNA and excess male DNA. In such experiments, the predicted results for nonlinkage and linkage of Y-specific and non-Y-specific 3.4-kb *Hae*III sequences are quite different.

If there is no physical linkage between Y-specific and non-Y-specific sequences, the results should be straightforward: the proportion of ^{3}H-labeled 3.4-kb *Hae*III Y DNA reassociated with male or female DNA should be independent of fragment length whether assayed by HAP or nuclease S_1.

If Y-specific and non-Y-specific sequences are linked within the same 3.4-kb molecules, the degree of Y-chromosome specificity should be a function of the length of the fragment used as trace when assayed by HAP. When the fragments are long enough to contain both types of sequences, reassociation of non-Y-specific sequences to female DNA will be

Table I: Comparison of $C_0t_{1/2}$ Values and Reiteration Frequencies for Y-Chromosome Sequences within 3.4-Kilobase (kb) *Hae*III Y DNA and the Y Chromosome

radiolabeled sequences	3.4-kb fragments[a]		Y chromosome[c]	
	$C_0t_{1/2}$	reiteration	$C_0t_{1/2}$	reiteration
3.4-kb non-Y specific	0.0017 0.0016[b]	14 100 15 000	0.13	15 800
3.4-kb Y specific	0.04 0.05[b]	600 480	8.5	590
it-Y	0.04	600	9	570

[a] $C_0t_{1/2}$ values are based on the reassociation of the purified 3.4-kb fragment alone and with it-Y sequences as presented in Figure 3. Reiteration frequencies were calculated from the expected $C_0t_{1/2}$ for the purified 3.4-kb *Hae*III Y DNA divided by the experimentally observed $C_0t_{1/2}$ for each type of reassociated sequences. The expected $C_0t_{1/2}$ for 3.4-kb *Hae*III Y DNA was calculated in the following way. Since (a) a human diploid nucleus has 10.6×10^9 nucleotides as estimated from the mean of 15 determinations of grams of DNA/cell (Sober, 1968) and (b) the 3.4-kb *Hae*III fragment sequences total 0.48% of the male genome (value calculated in text), they represent $(0.0048)(10.6 \times 10^9) = 5.1 \times 10^7$ nucleotides. Knowing that the *E. coli* genome size is 8.5×10^6 nucleotides (Cairns, 1963) and that it reassociates with a $C_0t_{1/2}$ of 4 under conditions used here, the expected $C_0t_{1/2}$ for reassociation of the 3.4-kb *Hae*III fragment is $5.1 \times 10^7 \times 4/8.5 \times 10^6 = 24$. This is the expected $C_0t_{1/2}$ for the total number of nucleotides represented in the 3.4-kb *Hae*III fragment if, like *E. coli* DNA, it consists entirely of single-copy sequences. [b] A second determination of the $C_0t_{1/2}$ for 3.4-kb *Hae*III Y DNA when self-reassociated as given in the Figure 3 legend. [c] $C_0t_{1/2}$ values presented for 3.4-kb *Hae*III sequences are derived from their reassociation with whole genome male DNA shown in Figure 2. The observed $C_0t_{1/2}$ for it-Y sequences was 9. Based on reasoning similar to that just presented, the $C_0t_{1/2}$ of human single copy sequences relative to *E. coli* would be predicted to be $10.6 \times 10^9 \times 4/8.5 \times 10^6 = 5 \times 10^3$. The observed $C_0t_{1/2}$ of human single-copy sequences is 2.4×10^3 or half the expected value of 5×10^3. The difference between observed and expected $C_0t_{1/2}$ results from the fact that the majority of DNA sequences within a diploid nucleus are represented twice, whereas most sequences are represented once in the haploid *E. coli* cell. However, in contrast to autosomes, the Y chromosome is represented once per diploid nucleus and the reiteration frequency of Y-chromosome sequences must therefore be calculated from the predicted $C_0t_{1/2}$ of 5×10^3 for human single copy sequences. To calculate the reiteration frequency of non-Y-specific 3.4-kb *Hae*III sequences in the Y chromosome, an adjustment was made for the presence of homologous sequences on chromosomes other than the Y chromosome. *First*, the reiteration frequency of non-Y-specific sequences was calculated in a human female *diploid* nucleus. In Figure 2, the observed $C_0t_{1/2}$ of the non-Y-specific sequences when reassociated with female DNA was 2.2×10^{-1}. Hence, they are reiterated $5 \times 10^3/2.2 \times 10^{-1} = 22\,700$ times. *Second*, the reiteration frequency of non-Y-specific sequences was calculated in a human male *diploid* nucleus. The observed $C_0t_{1/2}$ of the non-Y-specific sequences when reassociated with male DNA was 1.3×10^{-1}. Hence, they are reiterated $5 \times 10^3/1.3 \times 10^{-1} = 38\,500$ times. The difference between the number of copies found in a male *diploid* genome and those found in a female *diploid* genome or $38\,500 - 22\,700 = 15\,800$ yields the additional number of copies contributed by the Y chromosome.

Table II: Percent of Radiolabeled 3.4-Kilobase (kb) *Hae*III DNA Assayed as Y-Chromosome Specific as a Function of Fragment Length[a]

fragment length of 3.4 kb in nucleotides	DNA	mean % reassociated assayed with		calcd % Y-chromosome specific assayed with	
		HAP	S_1	HAP	S_1
1500	male	93	35		
	female	88	19	6	48
	E. coli	3	2		
650	male	87	70		
	female	63	33	35	67
	E. coli	18	15		
300	male	91	76		
	female	61	38	45	72
	E. coli	24	23		
200	male	81	75		
	female	28	25	68	71
	E. coli	3	5		

[a] The 3.4-kb *Hae*III fragment DNA preparations which had been radiolabeled by nick translation were treated with nuclease S_1 (Efstratiadis et al., 1976; Experimental Procedures) and sized on 4% acrylamide under denaturing conditions. Where necessary, the piece size was reduced by sonication or boiling (Experimental Procedures). Size-selected radiolabeled molecules were reassociated to C_0t 200 with excess *E. coli*, female, or male whole genome DNA (Experimental Procedures). Reactions were assayed in triplicate for percent of radiolabeled molecules reassociated as judged by either HAP analysis or resistance to nuclease S_1 (Sutton, 1971). The percent Y-chromosome specificity was calculated as follows. First, the mean percent reassociation was calculated for each triplicate determination. Second, by using these mean values, the net percent of total 3.4 kb reassociable was calculated by subtracting the reassociation with *E. coli* DNA from the reassociation to male DNA. Third, the net percent of total 3.4-kb reassociation to female DNA was determined by subtracting the reassociation with *E. coli* DNA from the reassociation to female DNA. Fourth, the percent Y-chromosome specificity was calculated by subtracting the net percent reassociated to female DNA from the net percent reassociated to male DNA and dividing by the net percent reassociated to male DNA.

accompanied by the apparent reassociation of Y-specific sequences to which they are linked. Measured Y-chromosome specificity will be lessened accordingly. When short fragments are used, the linkage between non-Y-specific and Y-specific sequences will have been disrupted by fragmentation and, while both kinds of sequences will be able to reassociate to male DNA, only non-Y-specific ones will be able to reassociate to female DNA. Thus, Y-chromosome specificity as assayed by HAP will increase as fragment length is made shorter.

In contrast, when linked sequences are assayed by nuclease S_1, the degree of reassociation of ^{3}H-labeled 3.4-kb *Hae*III fragments to *female* DNA should be independent of fragment length. Unpaired tails containing Y-specific sequences will be digested by nuclease S_1 and the influence of fragment length on the degree of Y-chromosome specificity will be reduced. However, in the case of reassociation to *male* DNA, the influence of fragment length on the extent of reassociation will depend on the internal organization of 3.4-kb *Hae*III Y DNA molecules. If the organization is simple, such that each non-Y-specific sequence is flanked by similar Y-specific sequences, the degree of nuclease S_1 assayed reassociation to male DNA will be independent of length. On the other hand, if there is internal molecular heterogeneity among 3.4-kb *Hae*III Y DNA molecules such that a limited variety of non-Y-specific sequences are flanked by a diversity of Y-specific ones, the extent of nuclease S_1 assayed reassociation to male DNA should be inversely related to fragment length. With short fragments, where linkage between non-Y-specific and Y-specific sequences is minimized, the majority of sequences will be able to reassociate and the influence of nuclease S_1 upon Y-chromosome specificity should be negligible. However, when fragments of 3.4-kb *Hae*III Y DNA molecules are long enough to contain both non-Y-specific and Y-specific sequences, the degree of nuclease S_1 assayed Y-chromosome specificity will be reduced. Because non-Y-specific sequences are more highly reiterated than Y-specific ones (Figures 2 and 3), they reassociate more rapidly. Adjacent but nonhomologous Y-specific sequences will remain unpaired and sus-

ceptible to nuclease S_1 digestion. The longer the fragment length used, the greater will be the fraction of unpaired Y-specific sequences. Thus, if there is internal heterogeneity within 3.4-kb *Hae*III Y DNA molecules, increases in fragment length should be associated with decreases in nuclease S_1 measured Y-chromosome specificity.

As shown in Table II, the degree of HAP-assayed Y-chromosome specificity of ^{3}H-labeled 3.4-kb *Hae*III Y DNA fragments increases from 6% to 68% when the piece size is decreased by fragmentation. This inverse relationship between Y-chromosome specificity and fragment size is not due to intrinsic alteration in reassociability since the sequences which fail to reassociate with female DNA remain reassociable with male DNA. Thus, the evidence from these HAP assays is that Y-specific and non-Y-specific sequences are linked within the same 3.4-kb *Hae*III Y DNA molecules. The remaining data in Table II are entirely consistent with this conclusion. As predicted for the case of intramolecular linkage, the results of nuclease S_1 assays are quite different than those of HAP assays. The extent of S_1-assayed reassociation between *female* DNA and ^{3}H-labeled 3.4-kb *Hae*III Y DNA fragments is similar for all fragment sizes once correction is made for self-reassociation and snapback. With male DNA, the extent of 3.4-kb *Hae*III Y DNA fragment reassociation decreases as the piece size of the fragment increases. This decrease in S_1-measured Y-chromosome specificity is the predicted result for a heterogeneous collection of Y-specific sequences physically linked, within the same molecules, to a more highly reiterated and less complex set of non-Y-specific ones.

With the recognition that Y-specific and non-Y-specific sequences are linked within the same molecules, it becomes apparent that HAP-based estimates of the relative proportions of those two types of sequences may be invalid. The data in Table II indicate that the measured 45% Y-chromosome specificity obtained by HAP assay with 300 nucleotide DNA, like the estimate derived from Figure 2, is a underestimate of the fraction of Y-specific sequences actually present.

The most reliable estimates of the relative proportions of Y-specific and non-Y-specific sequences in 3.4-kb *Hae*III Y DNA come from nuclease S_1 measurements of reassociation of 200 and 300 nucleotide fragments. At these shorter lengths, S_1-measured Y-chromosome specificity seems to plateau. Presumably, linkage between the two types of sequences is more-or-less disrupted and, therefore, most sequences are free to form duplexes with their homologues. Under such conditions, S_1-measured Y-chromosome specificity is ~71–72%. Thus, the relative proportions of Y-specific and non-Y-specific sequences within 3.4-kb *Hae*III Y DNA are estimated to be ~71–72% and 28–29%, respectively.

The finding that measured Y-chromosome specificity for 200 nucleotide DNA is slightly less by HAP than nuclease S_1 assay suggests that 3.4-kb *Hae*III molecules of this length contain some Y-specific as well as non-Y-specific sequences. Consequently, the two types of sequences must be interspersed within 3.4-kb *Hae*III Y DNA fragments and not arranged in uninterrupted type-specific blocks.

Estimation of Sequence Length of Reiterated Y-Chromosome-Specific DNA (it-Y) within the Y Chromosome. By utilizing Y chromosome. logic developed to describe the interspersion and length of reiterated DNA within the whole genome (Davidson et al., 1975; Schmid & Deininger, 1975), a two part analysis was performed to estimate the sequence length of it-Y sequences within the Y chromosome.

Table III: The Rate of Reassociation of Y-Specific DNA (^{3}H-Labeled it-Y) with Fractionated Human Male DNA

initial DNA length in nucleotides	treatment of DNA	% of genome[a]	$C_0t_{1/2}$ [^{3}H]it-Y
300	HAP	34.8	3
	HAP + S_1	17.8	3
600	HAP	39.8	3
	HAP + S_1	19.6	3
2100	HAP	58.0	6
	HAP + S_1	20.7	3

[a] The percent of the genome refers to that percentage of total recovered male DNA which was assayed as reiterated either by hydroxylapatite (HAP) or by nuclease S_1 digestion of the reassociated duplexes recovered from HAP. Human male DNA was sheared to three different sizes (300, 600, and 2100 nucleotides in length) as described under Experimental Procedures. Each sheared DNA (~10 mg) was reassociated to C_0t 100, by which point most reiterated DNA is reassociated, in 0.12 M phosphate buffer (PB) at 60 °C and the reiterated portion was collected preparatively on a 50-mL bed volume HAP column equilibrated in 0.12 M PB and held at 60 °C. A portion of each reiterated sample was treated with nuclease S_1 (Sutton, 1971) to remove single-stranded regions. The resulting six DNA fractions were thereafter sonicated under conditions known to yield a size of 300 nucleotides and used as driver DNA in reassociation with trace amounts of ^{3}H-labeled it-Y. Under conditions of reassociation used here (0.12 M PB, 60 °C), the $C_0t_{1/2}$ of ^{3}H-labeled it-Y DNA with *unfractionated* 46,XY DNA was 9.

The *first* part of the analysis involved fractionation of whole genome 46,XY male DNA for use as driver DNA in the second part of the analysis. DNA was sheared to three fragment lengths (300, 600, and 2100 nucleotides), and each fragment size was reassociated to C_0t 100. A value of C_0t 100 was chosen since most human reiterated DNA (~40% of the total genome) has reassociated by this point. Reassociated fragments were collected on HAP. A portion of each reassociated HAP fraction was digested with nuclease S_1 (Sutton, 1971) to remove single-strand DNA in the unpaired tails of the duplexes. Accordingly, six different fractions of male DNA were produced: the three DNA lengths which had been fractionated on HAP and the same three DNA lengths fractionated on HAP and subsequently digested with nuclease S_1. The properties of these fractions are shown in Table III. When assayed by HAP alone, the reassociation of whole genome male DNA increased in proportion to the size of the fragment examined; subsequent digestion with nuclease S_1 minimized this effect. The result is an expected one (Davidson et al., 1975; Schmid & Deininger, 1975); after reassociation to C_0t 100, the amount of DNA actually in duplex structure is largely independent of DNA fragment size and it is only the length of single-strand tails that increases with fragment size.

The *second* part of the analysis involved the following steps: (a) the six fractions of male driver DNA were sonicated under conditions predetermined to yield 300 nucleotide lengths; (b) each sonicated fraction was then assayed for its capacity to reassociate ^{3}H-labeled it-Y DNA in trace-driver reassociations at standard conditions (0.12 M phosphate buffer, 60 °C); and (c) the $C_0t_{1/2}$ was determined for the reassociation of the ^{3}H-labeled it-Y probe with each of the six driver DNA fractions. These $C_0t_{1/2}$ values reflect the relative amounts of it-Y sequences present in each driver fraction.

As seen in Table III, where every entry is the average of duplicate experiments, there is no difference in the $C_0t_{1/2}$ values for ^{3}H-labeled it-Y DNA reassociation with either fraction of 300 and 600 nucleotide driver DNA. Thus, approximately half of driver DNA could be removed from 300 to 600 nu-

cleotide fragments by S_1 digestion without increasing the relative concentration of it-Y sequences. This is the result predicted for the chance reassociation of randomly sheared sequences which are at least as long as the fragment size examined. In such cases, an average of half of all nucleotides within the sequence is expected to exist as unpaired tails. The implication from Table III is that it-Y sequences are at least 600 nucleotides long and are not arranged either as short interspersed sequences or as tandemly repeating sequences. Such arrangements would form duplexes over much of their length and their relative concentration in driver DNA would increase following digestion with nuclease S_1 (Davidson et al., 1975; Schmid & Deininger, 1975).

Reassociation patterns obtained with 2100 nucleotide driver DNA fractions were distinctively different than those obtained with the shorter fragments. As evident in Table III, it-Y DNA reassociates approximately twice as rapidly with driver DNA processed by HAP and nuclease S_1 than with driver DNA processed by HAP alone. Thus, digestion of 2100 nucleotide driver DNA with nuclease S_1 leads to a relative enrichment of it-Y sequences. It-Y sequences must therefore be somewhat longer than 600 nucleotides in length but less than 2100. A closer approximation of their length can be deduced in the following way. *First*, only ~35% of reassociated 2100 nucleotide driver DNA survived digestion with nuclease S_1. Consequently, if the it-Y sequences in this driver DNA had been entirely resistant to S_1, the reassociation rate of ^{3}H-labeled it-Y sequences should have been about threefold greater with DNA processed by HAP and S_1 than it was with DNA processed by HAP alone. *Second*, model building indicates that chance reassociation of a 1000 nucleotide sequence within a 2100 nucleotide fragment produced by random shear of whole genome DNA will result in duplexes where the average 1000 nucleotide sequence is 75% resistant to nuclease S_1. *Third*, a rate difference of 2 (approximately 0.75 × 3) was observed for the reassociation of it-Y DNA with 2100 nucleotide DNA processed by HAP and S_1 as compared with their reassociation with driver DNA processed by HAP alone (Table III). Therefore, the maximum continuous length of it-Y sequences must be on the order of 1000 nucleotides. Although twofold differences in reassociation rate such as seen in Table III are difficult to measure with precision, the result is reproducible. The data in Table III thus strongly suggest that typical it-Y sequences occur in continuous lengths of 600 to 1000 nucleotides. Since it-Y sequences seem to be identical with the Y-specific portions of 3.4-kb *Hae*III Y DNA molecules (Figure 3), Y-specific sequences must also be 600 to 1000 nucleotides long.

Isolation of Non-Y-Specific Sequences from 3.4-kb HaeIII Y DNA and Estimation of Their Sequence Length. Preliminary to the isolation of non-Y-specific 3.4-kb *Hae*III sequences, a population of ^{3}H-labeled 3.4-kb *Hae*III Y DNA fragments exceeding 900 nucleotides in length (mean size ~1650) was obtained by preparative electrophoresis in acrylamide–diacrylate gels. This fraction contained 70% of ^{3}H-labeled 3.4-kb *Hae*III Y DNA applied (Figure 4A). In order to isolate non-Y-specific 3.4-kb sequences, the population of size-selected 3.4-kb *Hae*III Y DNA molecules was reassociated to a C_0t of 1.24×10^{-3} and digested with nuclease S_1. The duplexes resistant to nuclease S_1 digestion (26%) were recovered on HAP. As evident from Figure 3, a substantial portion of non-Y-specific 3.4-kb *Hae*III sequences, but only a few Y-specific ones, is expected to reassociate by a C_0t of 1.24×10^{-3}.

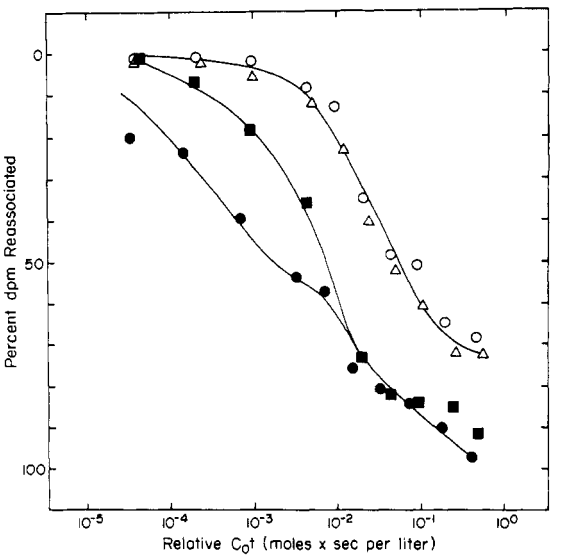

FIGURE 3: Reassociation of 3.4-kb *Hae*III Y DNA and determination of its homology with it-Y sequences. Purified unlabeled 3.4-kb *Hae*III Y DNA was sonicated to give 300 nucleotide long fragments and 0.033 μg was added to each of four reassociation mixtures: (○—○) 12000 dpm of ^{3}H-labeled it-Y + carrier *E. coli* DNA; (△—△) 12000 dpm of a second preparation of ^{3}H-labeled it-Y DNA sequences + carrier *E. coli* DNA; (■—■) 12000 dpm of ^{3}H-labeled 3.4-kb *Hae*III Y DNA (sonicated to give 300 nucleotides and fractionated on HAP to eliminate zero time binding) + carrier *E. coli* DNA; (●—●) 12000 dpm of the same ^{3}H-labeled 3.4-kb DNA used above + 59 μg of sonicated whole genome 46,XX DNA + carrier *E. coli* DNA. Following ethanol precipitation of each mixture from the same volume and the same nucleic acid concentration, pellets were resuspended in 50 μL of reassociation buffer as described under Experimental Procedures. Four microliter aliquots were boiled and incubated under oil at 60 °C for various time periods. Some short time incubations were obtained with 0.1 dilutions of the original resuspended pellets. Following incubation and subsequent 100-fold dilution, the samples were assayed on HAP as described under Experimental Procedures. In parallel with these assays, duplicate determinations were made for the extent of reassociation of each radiolabeled sample with male DNA at C_0t 1000: mean values were 56.7% and 54.3% for ^{3}H-labeled it-Y DNA and 87.0% for the ^{3}H-labeled 3.4-kb fragment. Each curve was adjusted to 100% based on the final reassociation of the appropriate radiolabeled trace to excess male DNA. Recoveries were similar in all assays and ranged between 80 and 90% of input amount. Plotted C_0t values are based on the concentration of unlabeled 3.4-kb *Hae*III Y DNA present in the resuspended ethanol pellets, and this concentration was adjusted according to the recovery of radiolabeled trace DNA. The $C_0t_{1/2}$ values for the it-Y sequences were calculated after assuming that (a) it-Y sequences are completely homologous to the Y-specific portions of 3.4-kb *Hae*III sequences and (b) final reassociation values are not 100% because there has been breakdown of the radiolabeled trace at long incubation times. $C_0t_{1/2}$ values of each of the trace DNAs when reassociated with excess 3.4-kb *Hae*III Y DNA were as follows: ^{3}H-labeled it-Y, 4×10^{-2}; the early component of 3.4-kb *Hae*III Y DNA selfing, 1.7×10^{-3}; early component of 3.4-kb *Hae*III Y DNA with added 59 μg of 46,XX DNA, 1.4×10^{-4}; the later component and the seemingly Y-specific portion of the 3.4-kb *Hae*III Y DNA, 4×10^{-2}. A second self-reassociation curve for 3.4-kb *Hae*III Y DNA (not shown) exhibited a $C_0t_{1/2}$ of 1.6×10^{-3} for the fast component and 5×10^{-2} for the slower Y-specific component.

The non-Y-specific nature of the S-resistant duplexes was confirmed by reassociation with excess male and female DNA. Results in the two cases were the same. As measured by HAP assay after incubation to C_0t 2500, 75% of the S_1-resistant duplexes was reassociated by each driver DNA. However, consistent with data in Figure 2, the rate of reassociation of these nuclease S_1 resistant duplexes was somewhat faster with male DNA than with female DNA. As expected, reassociated duplexes which had been sonicated to 200 nucleotides but *not* digested with nuclease S_1 revealed the presence of Y-specific

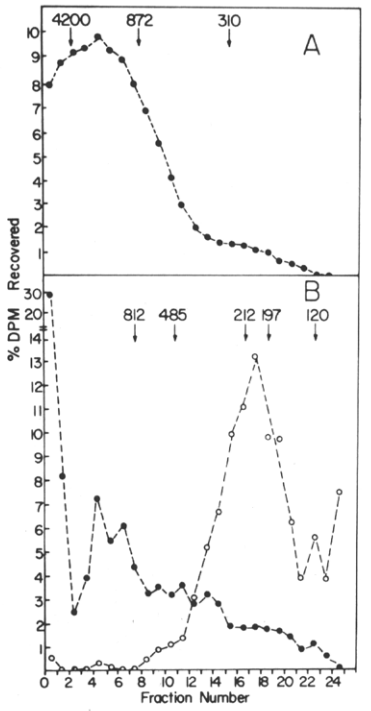

FIGURE 4: Size determination of non-Y-specific 3.4-kb *Hae*III Y DNA. (Panel A) The 3.4-kb *Hae*III Y DNA was radiolabeled by nick translation and its size distribution determined by electrophoresis on 5% acrylamide–0.2% ethylene diacrylate gels. Fragment size was estimated by comparison with DNA fragments of known size released from a *Hin*dIII digest of λ DNA and a *Hae*III digest of ϕ X 174 DNA. Representative nucleotide lengths of the size standards are given over the arrows at the top of the panel. The mean size of nick-translated 3.4-kb *Hae*III Y DNA entering this analysis was ~1650 nucleotides. Seventy percent of this preparation which exceeded 900 nucleotides in length was used for the analysis shown in panel B. (Panel B) Nick-translated 3.4-kb *Hae*III Y DNA exceeding 900 nucleotides in length were self-reassociated to C_0t 1.24×10^{-3} in 0.12 M PB at 60 °C. The reassociated sample was diluted, chilled, and divided into two fractions. The size distribution of one fraction was determined without further treatment. The other fraction was digested with nuclease S_1 and the size of the nuclease S_1 resistant duplexes determined. Estimation of fragment length was obtained by electrophoresis in 5% acrylamide–0.2% ethylene diacrylate gels and comparison with fragments of known length released by *Hae*III digestion of PMB-9 DNA. Representative lengths of the size standards are given over the arrows at the top of the panel. The untreated sample (●—●) had a mean size greater than 6000 nucleotides. The nuclease S_1 digested sample (○—○) had a mean size of 200 ± 100 nucleotides.

sequences: approximately 85% of the ^{3}H-labeled 3.4-kb *Hae*III Y DNA preparation was reassociable by male DNA but only 44% by female DNA.

The nuclease S_1 resistant duplexes obtained after self-reassociation of 3.4-kb *Hae*III Y DNA molecules greater than 900 nucleotides in length were thus measurably non-Y-specific and represent a majority of the non-Y-specific sequences found in unfractionated 3.4-kb *Hae*III Y DNA molecules (Table II). Consequently, measurement of the size of the S_1-resistant duplexes should provide an estimate of the length of non-Y-specific sequences within 3.4-kb *Hae*III Y DNA.

Figure 4B compares the size distribution of reassociated 3.4-kb *Hae*III fragments before and after digestion with nuclease S_1. The mean length of S_1-resistant duplexes measured by agarose gel electrophoresis was 200 ± 100 nucleotides. The same estimate (data not shown) was obtained for denatured, single-strand preparations of the S_1-resistant duplexes. This is a minimum estimate of non-Y-specific 3.4-kb *Hae*III sequence length since imperfect pairing among the reassociated duplexes could result in a fraction of the sequences being sensitive to nuclease S_1.

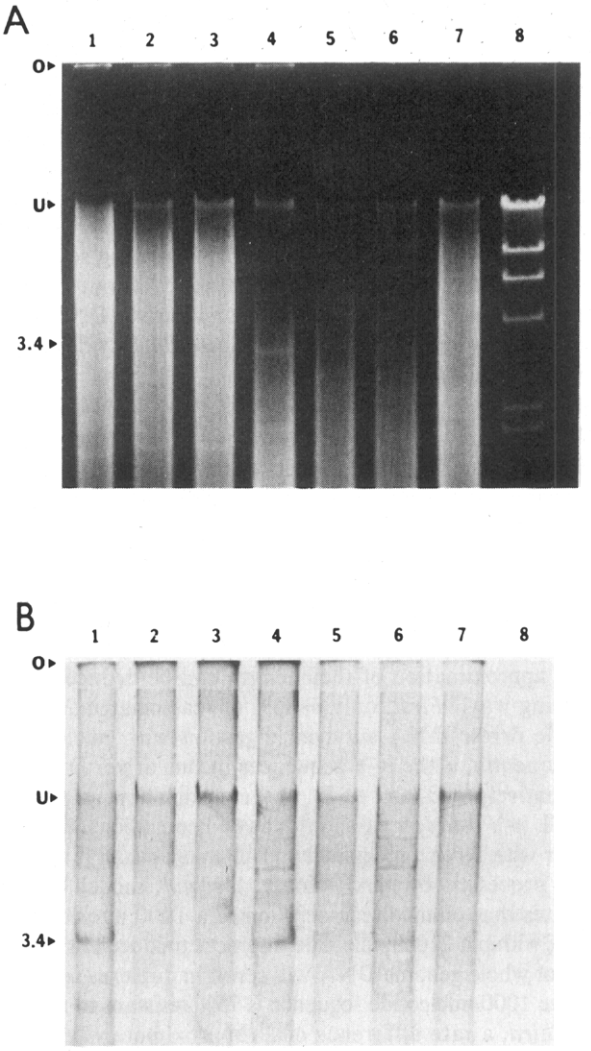

FIGURE 5: Agarose gel electrophoresis of various restriction endonuclease digests of male and female DNA. (Panel A) Ethidium bromide stained 1.0% agarose gel containing various restriction endonuclease digests of 3 μg of DNA from several sources. (Slot 1) *Eco*RI, human male DNA; (slot 2) *Hin*dIII and *Bam*HI, human female DNA; (slot 3) *Hin*dIII and *Bam*HI, human male DNA; (slot 4) *Hae*III, human male DNA; (slot 5) *Hin*dIII, *Bam*HI, and *Eco*RII, human female DNA; (slot 6) *Hin*dIII, *Bam*HI, and *Eco*RII, human male DNA; (slot 7) incomplete digest of human male DNA with *Alu*I; (slot 8) *Hin*dIII, λ DNA. (Panel B) Autoradiograph of the Millipore filter onto which the above gel was transferred by the method of Southern (1975) and then reassociated with ^{3}H-labeled 3.4-kb *Hae*III Y DNA sequences. The arrows preceded by symbols refer to the following: (O) the origin of electrophoresis; (U) DNA not resolved by the gel; (3.4) location of 3.4 kb as calculated from the known lengths (R. Reeder, personal communication) of *Hin*dIII restriction fragments of phage λ DNA.

The size of reassociated but undigested 3.4-kb *Hae*III Y DNA duplexes (Figure 4B) was considerably larger than the unreassociated sequences from which they were derived (Figure 4A). When analyzed in denaturing gels (data not shown), the undigested single strands had a mean size of about 1200 nucleotides, a value slightly less than the starting material. This result suggests that extensive cross-reaction occurs between the non-Y-specific sequences of a number of individual fragments and indicates that multiple copies of similar non-Y-specific sequences are shared by a majority of 3.4-kb *Hae*III Y DNA molecules.

Interestingly, Bostock et al. (1978) have identified a fragment of approximately 200 nucleotides in *Hae*III digests of human female DNA which is homologous with 3.4-kb *Hae*III Y DNA. The similarity between their results and our

own suggests that the length of non-Y-specific 3.4-kb *Hae*III sequences may be similar in all chromosomes where they are present.

The 3.4-kb Y DNA Sequences in EcoRI and EcoRII Endonuclease Digests. In the course of purifying 3.4-kb *Hae*III Y DNA, a variety of other restriction endonucleases were used to degrade contaminating fragments of similar length. As reported by Cooke & McKay (1978) one of these endonucleases, *Eco*RI, produces a 3.4-kb male-specific fragment which contains sequences homologous to those within 3.4-kb *Hae*III Y DNA. We confirm these observations and, in addition, illustrate in Figure 5 that the same result is achieved when whole genome male DNA is digested with *Eco*RII.

Although electrophoretically separated male-specific bands are discernible by ethidium bromide staining only in *Hae*III digests (Figure 5A), such bands are recognizable in other digests after filter hybridization (Southern, 1975) with ^{3}H-labeled 3.4-kb *Hae*III Y DNA (Figure 5B). As seen, the male-specific fluorographic (Laskey & Mills, 1975) pattern found in *Hae*III digests includes a number of components larger than 3.4 kb. Two of these, 6.7 kb and 10 kb, are approximate multiples of the 3.4-kb length. *Eco*RI digests of male DNA have a pattern of hybridization similar to that seen with *Hae*III and show, in addition, at least one smaller band of ~2 kb. A similar pattern is obtained after combined digestion of male DNA with *Eco*RII, *Bam*HI, and *Hin*dIII endonucleases. We attribute this finding to *Eco*RII since neither *Bam*HI nor *Hin*dIII, by themselves, produces male-specific fragments.

Cooke & McKay (1978) ascribe the similarity of male-specific patterns in *Hae*III and *Eco*RI digests to the juxtaposition of restriction sites in the same DNA molecules rather than to the existence of two homologous but restriction site-specific classes of molecules. Their conclusion was based on semiquantitative gel hybridization reactions and the finding that double digests released no more 3.4-kb male-specific fragments than single enzyme digests. Using an independent approach, we support this conclusion and believe that *Eco*RII sites are clustered with those for *Hae*III and *Eco*RI. If restriction site-specific classes of molecules existed, the total amount of 3.4-kb *Hae*III Y DNA estimated from reassociation kinetics should exceed the amount detected in agarose gels following electrophoresis of *Hae*III digests of male DNA. Since the extent to which any one of the three enzyme digests hybridizes ^{3}H-labeled 3.4-kb *Hae*III Y DNA (Figure 5B) appears to be about the same, estimates based on reassociation kinetics should be threefold greater than those based on densitometry measurements of *Hae*III digests. As we calculate in the following section, the estimates based on reassociation kinetics are quite similar to those based on densitometry. Thus, while a minor subset of 3.4-kb male-specific molecules might contain restriction sites for only one of the three enzymes in question, the majority of molecules must contain clusters for all of them.

Calculation of the Percentage of 3.4-kb HaeIII Y DNA within the Male Genome. Cooke (1976) estimated that 3.4-kb *Hae*III Y DNA represents approximately 0.4% of male DNA. This is probably an underestimate since it omits 3.4-kb *Hae*III Y DNA homologues which are larger than 3.4 kb. These have been described by Cooke & McKay (1978) and are also seen in Figure 5B. In each instance, an estimated 10–25% of silver grains are associated with the fragments exceeding 3.4 kb in length. If these components had been included in Cooke's earlier densitometric analysis (1976), the total fraction of 3.4-kb *Hae*III Y DNA and its male-specific homologues would account for 0.44 to 0.50% of male DNA.

Based on reasoning similar to that of Gelb et al. (1971), it is possible to obtain an estimate of all sequences in the male genome which are homologous with 3.4-kb *Hae*III Y DNA. Figure 3 shows the self-reassociation of 0.033 μg of 3.4-kb *Hae*III Y DNA. Nuclease S_1 analyses (Table II) indicate that non-Y-specific 3.4-kb *Hae*III Y DNA sequences represent 29% of the whole molecule. Therefore, the amount of non-Y-specific 3.4-kb *Hae*III sequences reassociating is 0.0096 μg. Because the non-Y-specific sequences reassociate 12 times faster when 59 μg of female DNA is present, there must be $12 \times 0.0096\ \mu g = 0.12\ \mu g$ of homologous sequences in the female reaction. Thus, non-Y-specific 3.4-kb *Hae*III sequences account for $0.12\ \mu g/59\ \mu g = 0.20\%$ of female DNA. Based on a computer-assisted least-squares analysis (Kells & Strauss, 1977) of the data in Figure 2, non-Y-specific 3.4-kb *Hae*III sequences reassociate more rapidly with male DNA than female DNA by a factor of 1.7 ± 0.2. Accordingly, non-Y-specific 3.4-kb *Hae*III sequences represent approximately $(1.7 \pm 0.2) \times 0.2\% = 0.34 \pm 0.04\%$ of a male genome. By assuming that the male–female difference is attributable solely to the Y chromosome, non-Y-specific 3.4-kb *Hae*III sequences present on the Y chromosome represent $(0.34 \pm 0.04\%) - 0.2\% = 0.14 \pm 0.04\%$ of male DNA. As these sequences account for only ~29% of the entire 3.4-kb *Hae*III Y DNA molecules, the entire molecule represents $(0.14 \pm 0.04)/0.29 = 0.48 \pm 0.14\%$ of male DNA. This value is in close agreement with the densitometric measurements of 3.4-kb *Hae*III Y DNA, suggesting that most if not all non-Y-specific sequences within the Y chromosome are contained within the 3.4-kb molecules and their larger homologues.

Since the male diploid nucleus contains $\sim 5.3 \times 10^9$ nucleotide pairs (Table I legend), there are about 7500 ± 2200 $[(0.48 \pm 0.14\%) \times 5.3 \times 10^9/3400]$ 3.4-kb *Hae*III Y DNA molecules.

As might be expected from the ready detection of a fragment associated with a small chromosome, 3.4-kb *Hae*III molecules represent a considerable fraction of Y-chromosome DNA. By assuming that the 1.2% of diploid chromatin mass associated with the human Y chromosome (Golomb & Bahr, 1971) is also a measure of the fraction of whole genome DNA, 3.4-kb *Hae*III sequences account for $(0.48 \pm 0.14)/1.2 = 40 \pm 12\%$ of all Y-chromosome DNA.

Discussion

Internal Organization of Sequences within 3.4-kb HaeIII Y DNA. The essential features of our findings are that Y-specific and non-Y-specific sequences coexist and are interspersed with one another in the same 3.4-kb *Hae*III Y DNA molecules (Table I). The organizational details within 3.4-kb *Hae*III Y DNA can be deduced by, *first*, determining the average number and nucleotide length of non-Y-specific 3.4-kb *Hae*III sequences most consistent with their observed reassociation kinetics and measured length; *second*, presuming that, since the Y-specific 3.4-kb *Hae*III sequences are homologous with it-Y DNA, their sequence length must be consistent with the data presented in Table III; and, *third*, interspersing the non-Y-specific and Y-specific 3.4-kb *Hae*III sequences to account for the observed discrepancy between HAP and nuclease S_1 assays in the proportion of size-selected 3.4-kb *Hae*III Y DNA fragments estimated to be Y specific (Table II).

Reassociation kinetics in Figures 2 and 3 suggest that the non-Y-specific portions of 3.4-kb *Hae*III Y DNA represent

a *minimum* of two families. As calculated in Table I, which summarizes the reassociation data in Figures 2 and 3, each family is reiterated an average of ~15000 times. Given that there are an estimated 7500 ± 2200 3.4-kb *Hae*III Y DNA molecules, there must be a minimum of $2 \times 15000/(7500 \pm 2200) \simeq 4 \pm 1$ non-Y-specific sequences per 3.4-kb *Hae*III Y DNA molecule. Since 29% of 3.4-kb *Hae*III Y DNA (Table III) is non-Y specific, it occupies a total of approximately 1000 nucleotides. Thus, the *average* tract length is 1000/4 = 250 nucleotides and the possible range is from 200 to about 300 nucleotides. These values are in close agreement with the physical length estimation and non-Y-specific sequences presented in Figure 4.

By exclusion, the remaining 2400 nucleotides of 3.4-kb *Hae*III Y DNA are assignable to Y-specific sequences. Since the majority of these Y-specific sequences are synonymous with it-Y sequences (Figure 3) and since it-Y sequences are each between 600 and 1000 nucleotides long (Table III), there must be three Y-specific sequences in a typical 3.4-kb *Hae*III Y DNA molecule.

The results of linkage analysis given in Table II indicate that Y-specific sequences alternate with non-Y-specific ones in the typical 3.4-kb *Hae*III Y DNA molecule. Such interspersion of 600–1000 nucleotide Y-specific sequences with 200–300 nucleotide non-Y-specific sequences predicts that 800 nucleotide fragments produced by random shear should all contain at least some non-Y-specific sequences. As seen in Table II, there is a sharp decline in HAP-assayed Y-chromosome specificity in the interval between 650 and 1500 nucleotide fragments. At the other extreme, it is expected from this model that, when the average fragment length falls below 250 nucleotides, Y-specific sequences will be substantially freed of linkage with non-Y-specific sequences and Y-chromosome specificity should reach a maximum. Table II shows that HAP-assayed Y-chromosome specificity reaches a maximum with 200 nucleotide fragments and closely resembles the value obtained by nuclease S_1 assay.

An arrangement of 600–1000 nucleotide Y-specific sequences alternating with 200–300 nucleotide sequences within the *average* 3.4-kb *Hae*III Y DNA molecules is compatible with our linkage data and estimates of the abundance and length of each sequence component. The general validity of this model can be tested by computer-based simulation. We used simulated random shear of a 3.4-kb length with various arrangements of the two kinds of sequence components to determine the predicted degree of Y-chromosome specificity. The predicted values for Y-chromosome specificity of an alternating arrangement of 250 and 800 nucleotide sequences approximated those actually found (Table II). When such 3.4-kb lengths are reduced by simulated random shear to 200 nucleotide lengths, the predicted Y-chromosome specificity is ~66%. This value closely approximates the estimate of 68% found for HAP assays of 200 nucleotide-length fragments. Similarly, simulated random shear to mean lengths of 300 and 650 nucleotides led to values of ~48% and ~34%, respectively. These resemble the corresponding observed estimates of 45% and 35% given in Table II. The fit between computer-based predictions and observed values was much less good for other model arrangements.

While an alternating arrangement of 800 nucleotide Y-specific sequences and 250 non-Y-specific sequences is consistent with all our analyses, the fact that it is based on a population of molecules necessarily means that it is no more than an average representation of any one molecule. Considerable variation in sequence length and arrangement among individual 3.4-kb molecules is possible. Nonetheless, this general arrangement allows us to explain several otherwise puzzling observations. For example, it-Y DNA formed only ~0.018% of total male DNA when isolated from starting DNA fragments which averaged 500–600 nucleotides in length as compared with the 0.16% calculated for preparations which began with ~180 nucleotide long DNA (Kunkel et al., 1976). This disparity in the absolute amount of it-Y DNA is predictable from the arrangement of sequences within 3.4-kb *Hae*III Y DNA proposed here. The longer the DNA, the more likely it is that any one it-Y sequence will be attached to a non-Y-specific one and thereby lost during challenge reassociation with female DNA.

The interspersion of two types of sequences within the same molecules also explains the apparent homology of all 3.4-kb *Hae*III Y DNA with female DNA. This conclusion depended on hybridization between nick-translated 3.4-kb *Hae*III DNA and electrophoretically separated components released from male and female DNA either by *Hae*III digestion of satellite III DNA (Bostock et al., 1978) or by combined digestion with *Hae*III and *Mbo*I or with *Eco*RI and *Mbo*I (Cooke & McKay, 1978). As evident from Figure 5A, almost all radiolabeled molecules prepared by nick translation are long enough to contain non-Y-specific sequences. Consequently, radioactivity associated with the Y-specific portion of the probe will be bound by female DNA fragments through linkage with non-Y-specific sequences.

The intramolecular heterogeneity described here is not inconsistent with the partial sequence data (Cooke, 1976), suggesting that a large fraction of 3.4-kb *Hae*III Y DNA consists of simple variations of a short oligonucleotide repeating many times in each molecule. Because non-Y-specific sequences are highly reiterated and consist of a small number of families present within most 3.4-kb *Hae*III molecules, they are the ones most likely to be revealed by sequence analysis. The more abundant Y-specific sequences, being only moderately reiterated and highly variable between 3.4-kb *Hae*III molecules, would not be expected to yield a distinguishable sequence when the entire population of 3.4-kb *Hae*III molecules is examined.

Intermolecular Heterogeneity of 3.4-kb HaeIII Y DNA. A distinctive feature of 3.4-kb *Hae*III Y DNA is that it is heterogenous. Aside from the internal heterogeneity of interspersed Y-specific and non-Y-specific sequences, the data presented here demonstrate considerable intermolecular heterogeneity as well. The number of kinds of Y-specific sequences in the overall population of 3.4-kb *Hae*III molecules greatly exceeds the maximum number which can be accommodated by any one 3.4-kb *Hae*III molecule. The evidence that there are ~3 Y-specific sequences per molecule and ~7500 molecules per Y chromosome means that there should be $3 \times 7500 = 22500$ copies of Y-specific sequences. This estimate is ~39 times greater than the average reiteration frequency of 570 calculated for Y-specific sequences from reassociation kinetics in Table I. The implication is that there are at least 39 families of reiterated Y-specific sequences. This value is about twice the number previously estimated from the reiteration frequency and abundance of it-Y DNA (Kunkel et al., 1977). Since no single 3.4-kb *Hae*III Y DNA molecule can encompass more than about three Y-specific sequences, there must be considerable heterogeneity among 3.4-kb *Hae*III molecules.

Independent evidence of intermolecular heterogeneity among 3.4-kb *Hae*III Y DNA molecules has been reported by Cooke & McKay (1978). They found that a subset of these molecules

could be cleaved by restriction endonuclease *Mbo*I. Our data would suggest that the regions of *Mbo*I restriction site variability are most likely to occur within Y-specific portions of 3.4-kb *Hae*III Y DNA molecules.

Since the fragments produced by *Hae*III digestion represent ~40% of human Y chromosome DNA and since this fraction is further augmented by another 20% from seemingly nonhomologous 1.9-kb *Hae*III Y DNA molecules (Cooke, 1976), it seems possible that refinements of the methods described here might be used to construct an organizational model of the entire human Y chromosome.

Acknowledgments

We thank N. S. Kunkel for advice and assistance with gel transfer techniques, Dr. P. C. Wensink for introducing us to the technique of nick translation, C. Conover Talbot for assistance with computer analysis, and Drs. B. J. Schmeckpeper and A. F. Scott for helpful discussion of this work.

References

Bostock, C. J., Gosden, J. R., & Mitchell, A. R. (1978) *Nature (London) 272*, 324.

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

Britten, R. J., Graham, D., & Neufeld, B. R. (1974) *Methods Enzymol. 29E*, 363.

Cairns, J. (1963) *Cold Spring Harbor Symp. Quant. Biol. 28*, 43.

Cooke, H. (1976) *Nature (London) 262*, 182.

Cooke, H. J., & McKay, R. D. G. (1978) *Cell 13*, 453.

Davidson, E. H., Galau, G. A., Angerer, R. C., & Britten, R. J. (1975) *Chromosoma 51*, 253.

Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun. 23*, 641.

Efstratiadis, A., Kafatos, F. C., Maxam, A. M., & Maniatis, T. (1976) *Cell 7*, 279.

Gelb, L. D., Kohne, D. E., & Martin, M. A. (1971) *J. Mol. Biol. 57*, 129.

Ginelli, E., & Corneo, G. (1976) *Chromosoma 56*, 55.

Golomb, H. M., & Bahr, G. F. (1971) *Exp. Cell Res. 68*, 65.

Hardies, S. C., & Wells, R. D. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 3117.

Kells, D. I. C., & Straus, N. A. (1977) *Anal. Biochem. 80*, 344.

Kelly, R. B., Cozzarelli, N. R., Deutscher, M. P., Lehman, I. R., & Kornberg, A. (1970) *J. Biol. Chem. 245*, 39.

Kunkel, L. M., Smith, K. D., & Boyer, S. H. (1976) *Science 191*, 1189.

Kunkel, L. M., Smith, K. D., Boyer, S. H., Borgaonkar, D. S., Wachtel, S. S., Miller, O. J., Breg, W. R., Jones, H. W., & Rary, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 1245.

Landy, A., Foeller, C., Reszelbach, R., & Dudock, B. (1976) *Nucleic Acid Res. 3*, 2575.

Laskey, R. A., & Mills, A. D. (1975) *Eur. J. Biochem. 56*, 335.

Maniatis, T., Jeffrey, A., & Van de Sande, H. (1975a) *Biochemistry 14*, 3787.

Maniatis, T., Jeffrey, A., & Kleid, D. G. (1975b) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184.

Schachat, F. H., & Hogness, D. S. (1973) *Cold Spring Harbor Symp. Quant. Biol. 38*, 371.

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

Sharp, P., Sugden, B., & Sambrook J. (1973) *Biochemistry 12*, 3055.

Sober, H. A., Ed. (1968) *Handbook of Biochemistry Selected Data for Molecular Biology*, pp H-52–H-61, Chemical Rubber Co., Cleveland, OH.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503.

Sugden, B., Detroy, B., Roberts, R. J., & Sambrook, J. (1975) *Anal. Biochem. 68*, 36.

Sutton, W. D. (1971) *Biochim. Biophys. Acta 240*, 522.