Hamilton, L., Mahler, I., and Grossman, L. (1974), *Biochemistry 13,* 1886–1896.
Kaplan, J., and Grossman, L. (1971), *Methods Enzymol. 21,* 249.
Kushner, S. R., Kaplan, J. C., Ono, H., and Grossman, L. (1971), *Biochemistry 10,* 3325.
Litman, R. (1968), *J. Biol. Chem. 243,* 6222.
Madden, J. J., Werbin, H., and Denson, J. (1973), *Photochem. Photobiol. 18,* 441.
Marmur, J. (1961), *J. Mol. Biol. 3,* 208.
Minton, K., Durphy, M., Taylor, R., and Friedberg, E. C. (1975), *J. Biol. Chem. 250,* 2823.
Radman, M. (1976), *J. Biol. Chem. 251,* 1438.
Setlow, R. B. (1968), *Prog. Nucleic Acid Res. Mol. Biol. 8,* 257–295.
Simon, T. J., Smith, C. A., and Friedberg, E. C. (1975), *J. Biol. Chem. 250,* 8748.
Sutherland, B. (1973), *J. Biol. Chem. 248,* 4200.

Inverted Repetitive Sequences in the Human Genome†

Peggy J. Dott, Cynthia R. Chuang, and Grady F. Saunders*

ABSTRACT: A specific class of DNA sequences, the inverted repetitive sequences, forms hairpin-like structures in denatured DNA by the folding back of a single linear chain. The reassociation process of these sequences is unimolecular and the rate is extremely fast. Inverted repetitive sequences comprise 6% of the total human genome. They appear to be heterogeneous in length with an overall average length of 190 nucleotides. The inverted sequences are represented in almost all families of repetition frequencies, highly repetitive as well as very few copies per genome. They are not localized at unique sites on metaphase chromosomes. It is estimated that there are approximately 2×10^6 inverted repeats per haploid human genome. The biological function of this class of sequences is unknown.

A sizeable portion of the eukaryotic genome is composed of similar or repetitive sequences which are characterized by their ease of reassociation. The repetitive sequences can be divided into highly repetitive and middle repetitive fractions by their relative reassociation rates. The DNA sequences that appear to occur only once in the genome are called unique or single-copy sequences. These three kinetic classes of DNA can be fractionated on hydroxylapatite (HA)[1] after annealing the denatured, sheared DNA to varying extents (Britten et al., 1974).

The highly repetitive sequences can be divided into three components according to the arrangements of their repeating units: inverted sequences or the so-called "zero-time" DNA fraction, a satellite fraction, and a nonsatellite fraction. The single strands of the satellite and nonsatellite fractions contain the ordinary repetitions of the sequence type ABC . . . ABC and, hence, they reassociate with biomolecular second-order kinetics. Most satellite DNAs are arranged in a long uninterrupted tandem array of simple repeating units (Walker, 1971; Gall and Atherton, 1974). These DNA sequences can be separated from the bulk of the DNA by a variety of equilibrium centrifugation techniques. The highly repetitive sequences of the nonsatellite fraction are more complex than those of the satellite fraction and the repeating units are interspersed with the less repetitive or nonrepetitive sequences (Saunders et al., 1975; Cech et al., 1973). The "zero-time" DNA fraction contains inverted repetitions of the sequence type ABC . . . C′B′A′ which form hairpin-like structures by the folding back of a single linear chain. Thus the reassociation process is unimolecular and the rate is extremely fast. This fraction of DNA can be isolated by reassociating the total DNA to C_0t values of $\leq 10^{-4}$ M s. C_0t is the product of the initial concentration of single-stranded DNA and the time of incubation.

The presence of inverted repetitive sequences in a variety of higher organisms has been reported by several laboratories (Wilson and Thomas, 1974; Davidson et al., 1973; Graham et al., 1974; Schmid et al., 1975). The biological function of this class of sequences is unknown. It has been found that heterogeneous nuclear RNA contains relatively long double-stranded regions which are absent in the cytoplasmic messenger RNA (Ryskov et al., 1972). After denaturation, these RNA sequences corresponding to the double-stranded regions can hybridize efficiently with immobilized cellular DNA, indicating high repetition of the corresponding DNA sequences. It was thus suggested that the inverted repetitive DNA sequences in the genome might correspond to acceptor sites which could interact with regulatory proteins (Jelinek and Darnell, 1972; Ryskov et al., 1973).

The human genome consists of 7×10^9 base pairs, 65% of which are unique sequences (Saunders et al., 1972a). Among the repetitive sequences there exist the inverted repetitive sequences, the simple sequence satellites, the nonsatellite highly repetitive sequences, and the middle repetitive sequences (Saunders et al., 1972b, 1975; Chuang and Saunders, 1974). In this paper, we report the isolation and characterization of the inverted repetitive sequences of human DNA. When DNA with an average single-stranded chain length of 15 000 nu-

† From the Department of Developmental Therapeutics, The University of Texas System Cancer Center, M. D. Anderson Hospital and Tumor Institute, Houston, Texas, and The University of Texas Graduate School of Biomedical Sciences, Houston, Texas 77030. *Received February 17, 1976.* This investigation has been supported by Grant G-267 from the Robert A. Welch Foundation and Grant NP-125 from the American Cancer Society. This is part of the thesis of P.J.D. presented to The University of Texas Graduate School of Biomedical Sciences in partial fulfillment of requirements for the M.S. degree.

[1] Abbreviations used: Tris, tris(hydroxymethyl)aminomethane; HA, hydroxylapatite; EDTA, ethylenediaminetetraacetic acid; FUdR, fluorouridine deoxyribose.

cleotides is fractionated after denaturation and neutralization, it is found that 6% of the sequences exist as single-stranded nuclease-resistant duplexes. If the DNA is sheared to a length of 300 nucleotides, it is found that 3% of the sequences can still form duplexes upon denaturation and neutralization. During preparation of this manuscript, Schmid and Deininger (1975) reported similar results on inverted repetitive sequences in the human genome.

Materials and Methods

Preparation of Radioactive HeLa Cell DNA and Placental DNA. HeLa cells were grown in monolayers in 250-ml T flasks at 37 °C until 75% confluency was reached. The cultures were then transferred to 1-l. roller bottles and grown in Eagle's minimal essential medium supplemented with 5% fetal calf serum and antibiotics (penicillin, 100 units/ml; streptomycin, 0.1 mg/ml; neomycin, 0.1 mg/ml). Sixteen hours prior to labeling, cells were partially synchronized at the G1-S boundary by the addition of 0.5 μg of FUdR per ml of medium. The cells were rinsed with fresh medium to remove FUdR and tritiated thymidine ([*methyl*-^{3}H]thymidine, 16 Ci/mmol, Schwarz/Mann) was added to a concentration of 10 μCi/ml of medium. The cells were allowed to grow for 10 additional h. The cells were then harvested, washed with 0.15 M NaCl, 0.1 M EDTA, pH 8.0, and suspended in 0.01 M Tris-HCl, pH 7.9. Sodium dodecyl sulfate was added to a final concentration of 1% to lyse the cells, and the solution density was adjusted to 1.700 g/ml by adding 3 volumes of saturated CsCl. CsCl–DNA solution was centrifuged in a Beckman 60 Ti rotor at 30 000 rpm for 48 h at 20 °C and the radioactive DNA band collected. The specific activity of the [^{3}H]DNA was 4×10^4 cpm/μg.

Human DNA from fresh placental tissue was isolated by a modification of the Marmur (1961) procedure as described previously (Chuang and Saunders, 1974).

Fragmentation and Determination of Size of DNA. DNA was sonicated by the method of Saunders et al. (1972a). The DNA fragments were dialyzed twice against 100 volumes of 0.1 M EDTA and four times against 100 volumes of 0.05 M sodium phosphate buffer (equimolar of NaH_2PO_4 and Na_2HPO_4). The single-strand length of the sonicated DNA was found to be 300 nucleotides as determined by sedimentation velocity measurement (Studier, 1965).

Thermal Denaturation of DNA. DNA solutions in 0.12 M sodium phosphate buffer, pH 6.8, were thermally denatured in Teflon-sealed cuvettes. The change of absorbance at 260 nm was measured in a Zeiss PMQ II recording spectrophotometer equipped with a thermostated cuvette chamber and an automatic sample changer. The temperature was increased at a constant rate of 1 °C/min using a Lauda Linear Temperature Programmer. The absorbance measurements were corrected for thermal expansion of the solution.

Isolation of Inverted Repetitive DNA Sequences. Inverted repetitive sequences were isolated from native and sonicated samples of placenta and HeLa DNA by two methods (schematically shown in Figure 1): (1) Two-step fractionation was performed in which the DNA was denatured with alkali, neutralized, and adjusted to 0.12 M phosphate (Kram et al., 1972), reassociated at 60 °C to a C_0t value of 0.1, and then fractionated on hydroxylapatite (HA). The unbound fraction was eluted with 0.12 M phosphate; then the bound fraction called "C_0t 0.1 HA bound" was eluted with 0.45 M phosphate. The "C_0t 0.1 HA bound" fractions were pooled, concentrated, dialyzed against 0.05 M phosphate buffer, denatured a second time, neutralized, and immediately passed over HA. The bound fractions are termed "C_0t 0.1, zero-time HA bound". (2) A one-step fractionation was carried out in which the DNA was denatured with alkali, neutralized, and immediately passed over HA. These bound fractions are called "zero-time HA bound". Both of these isolation procedures discriminate between reassociation products arising from biomolecular and unimolecular reactions. Zero-time reassociation is equivalent to $C_0t \leq 10^{-3}$ and even the most reiterated sequences should be unable to reassociate. Only those inverted repetitive sequences on a single DNA molecule should be able to anneal or "fold back". The "zero-time HA bound" fraction prepared by the one-step method is treated with S_1 nuclease to remove the single-stranded regions and then extracted with phenol. The S_1 resistant fraction is called "zero-time S_1"

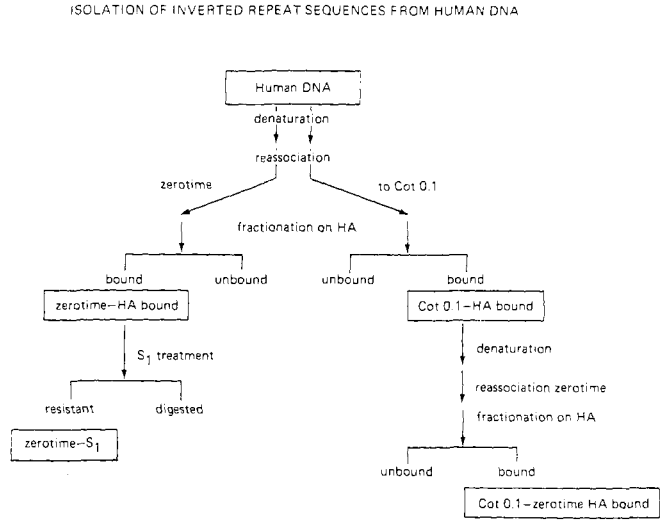

FIGURE 1: Flow diagram of inverted repetitive sequence isolation.

The single-strand-specific nuclease, S_1, was purified from Takadiastase using Sutton's method (1971) as modified by St. John et al. (1974). Before the S_1 treatment, samples of the 0.45 M phosphate eluents were dialyzed against 0.01 M NaCl–0.03 M sodium acetate buffer, pH 4.5. The volumes were recorded after dialysis and 1 M $ZnCl_2$ was added to a final concentration of 3 mM. Twenty microliters of S_1 (at 40 units/ml) per ml of DNA solution was used in each treatment with incubation at 37 °C for 1–2 h to ensure complete digestion of single-stranded regions.

Determination of Double-Strand Content of Reassociated Structures. From the hyperchromicity of each HA bound fraction, the amount of double-stranded DNA within that sample was determined using the following equations:

$$\text{hyperchromicity}_{(\text{sample})} = 1 - \frac{A_{260\ (\text{sample})\ \text{at}\ 60\ ^\circ\text{C}}}{A_{260\ (\text{sample})\ \text{at}\ 93\ ^\circ\text{C}}}$$

$$\text{fraction double strandedness} = \frac{\text{hyperchromicity}_{(\text{sample})}}{\text{hyperchromicity}_{(\text{native DNA})}}$$

The percentage of inverted repetitive sequences in the total genome can be calculated from the product of the percentage of HA bound and the fraction of double strandedness found in each HA bound fraction. Alternatively, the percentage of inverted sequences in the genome can be calculated from the product of the percentage of HA bound and the fraction of S_1 resistant.

Polyacrylamide Gel Electrophoresis. Polyacrylamide gel electrophoresis of purified ^{3}H-labeled inverted repetitive sequences was carried out under denaturing conditions (Maniatis et al., 1975). The samples containing up to 3×10^4 cpm (up to 0.75 μg) were dialyzed in water and then evaporated under

vacuum to dryness. Two hundred microliters of 100% deionized formamide deionized with Bio-Rad Ag 501-x8(D) and 1 μl of bromophenol blue (stock concentration 0.25%) were added. The mixtures were denatured in a boiling water bath for 2 min, quickly cooled, and immediately layered onto prepared gels. The gels contained 4.3% acrylamide and 0.7% bisacrylamide in 100% deionized formamide in 8 × 0.5 cm tubes. Electrophoresis was carried out in a Buchler polyanalyst apparatus at 5 mA per tube for approximately 2 h. After electrophoresis, the gels were sliced into 2-mm sections and incubated at 50 °C overnight in 1 ml of NCS solubilizer (Amersham/Searle)–H_2O(9:1). After cooling to room temperature, 10 ml of a toluene-based scintillation fluid was added and the samples were counted.

After electrophoresis, the gels containing markers were stained with 5 μg/ml ethidium bromide for 30 min and then destained in 5 mM acetic acid–1 mM EDTA, pH 4.5, for 20 min. Restriction fragments were identified by fluoresence.

Annealing Experiments. Trace amounts of S_1-treated radioactive inverted repetitive sequence DNA (specific activity of 4×10^4 cpm/μg) were mixed with a 100-fold excess of sonicated nuclear DNA in either 0.06 M phosphate, 0.12 M phosphate, or 0.48 M phosphate buffer. Each solution was divided into 20-μl aliquots in capillary pipettes whose ends were then flamed sealed. The samples were denatured in a 100 °C water bath for 2 min and incubated at 60 °C to the desired C_0t or equivalent C_0t. A zero-time control point was taken in which the inverted sequence DNA in the absence of the added nuclear DNA driver was immediately diluted into 1 ml of cold 0.12 M phosphate buffer. The C_0t value was determined by the concentration of the DNA and calculated from the following equations:

$$C_0t_{0.12\ \mathrm{M\ PO_4}} = \frac{A_{260}t_{(h)}}{2}$$

$$C_0t_{\mathrm{equivalent}} = \frac{A_{260}t_{(h)}}{2}\gamma$$

where γ is the reassociation rate relative to that which would be obtained in 0.12 M phosphate (Britten et al., 1974). Since temperature was not varied with the salt concentration, DNA reassociation occurred under less stringent conditions for the higher C_0t values due to higher salt concentrations in these reactions.

After the desired incubation period, the sample was diluted into 1 ml of cold 0.12 M phosphate buffer. Samples were kept at 4 °C until all points were collected. They were then adsorbed onto 1.5 ml of HA in 0.12 M phosphate at 60 °C. The unbound material was eluted with five washes of 2 ml of 0.12 M phosphate and then the bound hybrids were eluted with five washes of 2 ml of 0.45 M phosphate buffer. The volume of each wash was recorded. Nuclear DNA reassociation was monitored optically. The absorbance of each fraction was measured at 260 nm at 25 °C and then at 93 °C. At 93 °C, the sample should be fully denatured. DNA reassociation was calculated from the sum of the absorbancies of fully denatured DNA. The extent of the annealing of the inverted repetitive sequences with the nuclear DNA was measured by acid-precipitable radioactivity. The percentage of duplexes formed at zero time in the absence of driver was subtracted from all points. The annealing data were fitted to a second-order kinetic curve by a computer program.

Complementary RNA Synthesis and Chromosomal Localization Studies. High specific activity RNA complementary to the inverted repetitive sequences was synthesized in vitro using purified *Escherichia coli* DNA-dependent RNA polymerase holoenzyme as previously described (Arrighi et al., 1970). Assuming equimolar incorporation of the nucleotide triphosphate precursors and an average molecular weight of 350 for the mononucleotides, the specific activity of the synthesized RNA was calculated to be 4.3×10^6 cpm per μg as counted on filters. Human lymphocyte cultures (continuous cultures) were grown in McCoy's 5a medium (McCoy et al., 1959) supplemented with 20% fetal calf serum. Cells were harvested after 2-h colcemid (0.02 μg/ml) and 10-min hypotonic solution pretreatments and fixed in 40% acetic acid. Squashed preparations on chrome alum-gelatin-coated slides were used for in situ hybridization experiments. In situ hybridization was carried out by the method of Gall and Pardue (1971). Autoradiographs were prepared by using Kodak AR10 stripping film and developed in Kodak D 19-B. The exposure time varied from 2 days to 3 weeks.

TABLE I: Inverted Repetitive Sequences in Human DNA.[a]

Sample	% of Total Human Placental DNA: Sonicated DNA (L = 300 Nucleotides)	% of Total Human Placental DNA: Native DNA (L = 15 000 Nucleotides)
C_0t 0.1 HA bound	16.0 ± 0.9	26.6 ± 2.8
C_0t 0.1 zero-time HA bound	4.7 ± 0.9	19.0 ± 0.7
Zero-time HA bound	5.5 ± 0.9	25.9 ± 3.5
Zero-time S_1	3.2 ± 0.3	6.1 ± 1.1

[a] DNA samples (100 μg/ml) were dialyzed against 0.05 M phosphate buffer, pH 6.8. Eight parts of the DNA solution at room temperature was added to one part of 1.0 M NaOH. After 10 min, 10 parts of water (60 °C) and one part of 2.0 M NaH_2PO_4 (60 °C) were then added to neutralize the solution. The final phosphate concentration at which DNA reassociation occurred was thus 0.12 M. The DNA was incubated at 60 °C (10 s for zero-time reassociation, 15 min for C_0t 0.1) and immediately added to a slurry of HA (1.5 ml) in the same buffer in a 60 °C water bath. The slurry was eluted stepwise seven times with 0.12 and 0.45 M phosphate buffers, each elution consisting of resuspension, centrifugation in a clinical centrifuge at 65 °C, and pipetting of the supernatant. The 0.12 M eluents (HA unbound fraction) and the 0.45 M eluents (HA bound fraction) were assayed by spectrophotometric measurements as described in Material and Methods. L is single-strand chain length.

Results

Interspersion of Inverted Repetitive Sequences. To assess the potential biological role of inverted repetitive sequences, it is first necessary to establish their abundance and arrangement in the genome. The percentages of total DNA retained by HA chromatography using either the one-step or the two-step fractionation methods are summarized in Table I. With increased chain length of the denatured DNA, a greater percentage of the DNA is retained on HA. This indicates interspersion of inverted sequences with other sequence classes.

When the sonicated DNA, with an average length of 300 nucleotides, was reassociated to a C_0t value of 0.1, 16% of the starting material was retained on HA. This bound fraction was carried through the second step, heat denatured, and immediately loaded on HA; 29.5% of the C_0t 0.1 fraction corresponding to 4.7% of the total DNA was rebound at zero-time reassociation. When higher molecular weight placental DNA (15 000 nucleotides) was carried through the two-step fractionation, 19% of the total DNA was reassociated. Four and

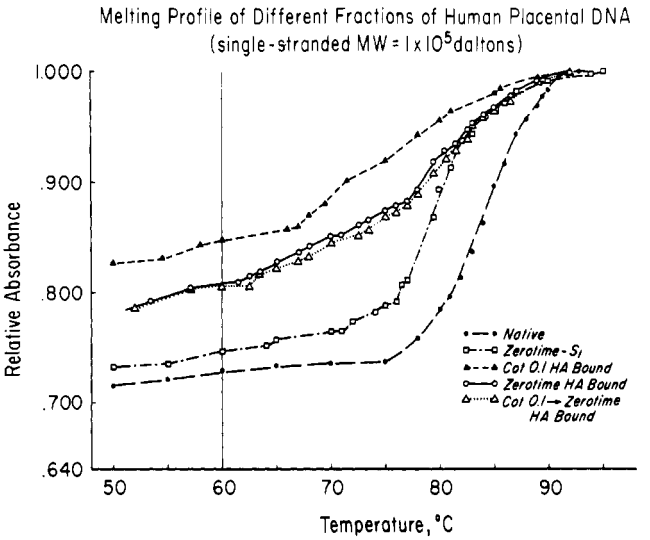

FIGURE 2: Melting profiles of HA-bound DNA fractions from placenta DNA (single-strand molecular weight = 1×10^5). Fractions were isolated by the one- and two-step fractionation methods as shown in Figure 1. HA-bound fractions were concentrated and dialyzed in 0.12 M phosphate buffer. The melting transition was followed spectrophotometrically at 260 nm. The temperature was increased to a constant rate of 1 °C per min.

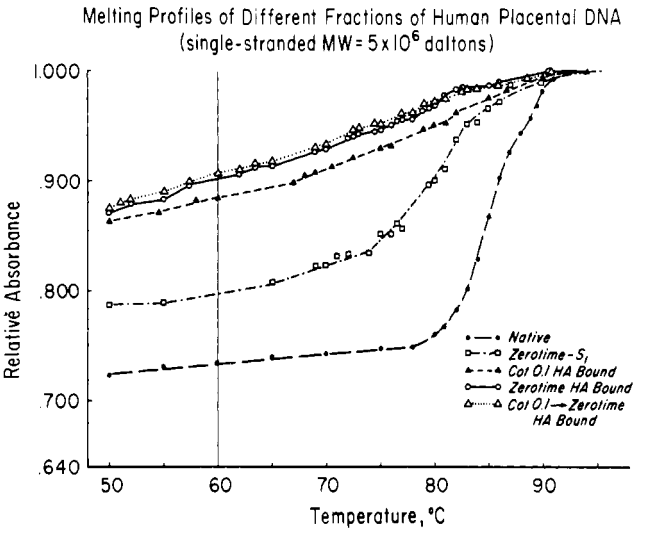

FIGURE 3: Melting profiles of HA-bound DNA fractions from placenta DNA (single-strand molecular weight = 5×10^6). Fractions were isolated by the one- and two-step fractionation methods as shown in Figure 1. HA-bound fractions were concentrated and dialyzed in 0.12 M phosphate buffer. The melting transition was followed spectrophotometrically at 260 nm.

a half times more DNA was bound with the longer fragments (15 000 nucleotides) than with very short fragments (300 nucleotides). If the DNA was fractionated once on HA at zero-time reassociation, then 5.5% of the sonicated DNA was retained as compared with 25.9% of the high molecular weight DNA. Again a four and a half time increase of bound DNA was observed. The increase in bound material can be attributed to the arrangement of DNA sequences in the higher molecular weight samples. The slower annealing portions of the long strands which remain single stranded are covalently linked to the fold-back regions and thus appear in the zero-time reassociated material.

When zero-time HA bound material is treated with S_1 nuclease, the resistant fraction represents the purified repetitive inverted duplexes. The longer fragments contain 6% inverted repetitive sequences while the short fragments contain 3%. Since the length of the inverted repetitive sequence ranges from 100 nucleotides to 900 nucleotides, isolation from 300 nucleotide fragments results in inverted repetitive sequences having a maximal length of 300 nucleotides. Very short inverted repetitive sequences are not detected by the methods used here. The length of duplex required for adsorption onto HA has been estimated at 40–60 nucleotides (Wilson and Thomas, 1973).

Optical Thermal Denaturation. The ordered structure in inverted repetitive sequences was investigated by optical thermal denaturation in 0.12 M sodium phosphate (pH 6.8, 0.18 M Na^+). From thermal stability measurements, the extent of mismatching of base sequences in the reassociated structures can be estimated. The hyperchromicity indicates the amount of true double strandedness. The S_1-treated inverted repetitive sequences isolated from sonicated placental DNA (Figure 2) give a cooperative melting profile which resembles that of the native DNA. Comparison of the hyperchromicity of this sample (25%) with that of native DNA (27%) shows that 94% of the bases are in double-stranded form. Furthermore, the T_m of the sample is 79 °C while that of the native DNA is 84 °C. Since 1 °C lowering of T_m indicates 1% of mismatched base pairs (Bonner et al., 1973), the inverted repetitive sequences isolated by this method have as little as 5% mismatched base pairs. However, if the effect of fragment length on thermal stability is taken into consideration, the T_m of the inverted sequences with an average size of 190 nucleotides (see below) would be reduced 3.4 °C for length out of the observed 5 °C. This is computed using the equation $T_n - T_m = B/L$ where T_n is the melting temperature of long DNA, T_m that of the fragments, L the fragment length in nucleotides, and B = 650 at 0.18 M Na^+ (Britten et al., 1974). This leaves only 1.6 °C difference in T_m which leads one to consider the possibility of cross-links in the DNA since the T_m of cross-linked DNA is approximately the same as that of the native DNA. The cross-linked duplex fragments would be fairly long and that is not observed in our length determination experiments either by gel electrophoresis or by neutral and alkaline sedimentation velocities.

The one- and two-step zero-time HA bound fractions prepared without S_1 treatment display lower hyperchromicities and broader thermal transitions. The curves are almost superimposable with hyperchromicities of 19%. This indicates that only 73% of bases are in double-stranded form. Broad thermal transitions of inverted repetitive sequences have been reported previously (Britten and Smith, 1970; Schmid et al., 1975).

The S_1-treated inverted repetitive sequences isolated from high molecular weight DNA (Figure 3) also give a cooperative melting profile. The hyperchromicity (21%) reflects that about 80% of the bases are in double-stranded form. The one- and two-step zero-time HA bound material isolated from the high molecular weight DNA, however, gives even lower hyperchromicities (9%), indicating that only one-third of the bases are paired. As mentioned earlier, the longer DNA fragments will contain more of the slow annealing sequences that are covalently linked to the inverted repetitive sequences. At zero-time reassociation, the inverted repetitive sequences fold back to form duplex structures while the other sequences remain single stranded. The entire fragment will bind to HA and thus give higher yields but display lower hyperchromicities.

Size Distribution of Inverted Repetitive Sequences. The inverted repetitive sequences isolated from HeLa [^{3}H]DNA

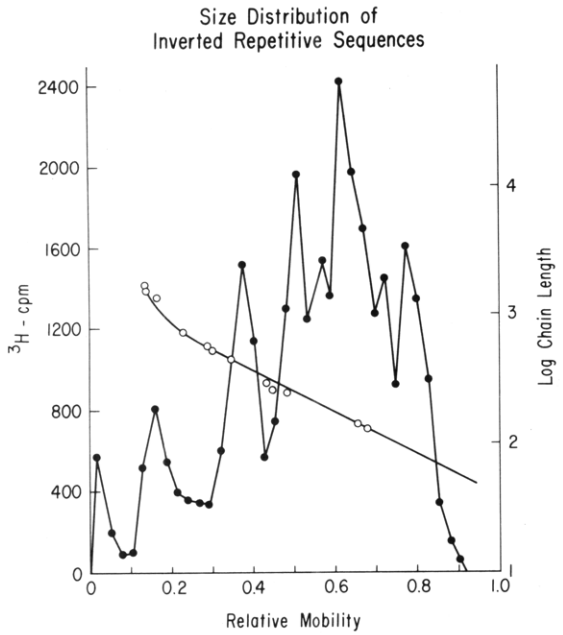

FIGURE 4: Chain-length distribution of inverted sequences on polyacrylamide–formamide gel electrophoresis. Inverted sequences were electrophoresed on denaturing gels composed of 5% polyacrylamide–100% formamide and then sliced and counted. Parallel gels contained a series of fragments of bacteriophage PM2 generated by Hae III restriction enzyme. They had chain lengths of 130, 140, 240, 263, 295, 455, 520, 560, 750, 1280, 1550, and 1650 base pairs.

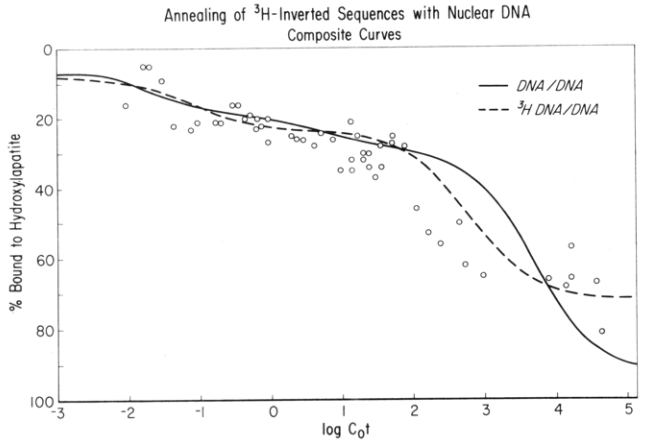

FIGURE 5: Annealing of ^{3}H-labeled inverted repetitive sequences with nuclear DNA. Purified ^{3}H-labeled inverted repetitive sequences from HeLa DNA prepared by the zero-time S_1 method were annealed to a large excess of nuclear placental DNA 300 base pairs long in 0.06 M phosphate, 0.12 M phosphate, or 0.48 M phosphate buffer to the desired C_0t values. After denaturation and incubation at 60 °C to the appropriate C_0t value, the solution (20 μl) was diluted into 1 ml of cold 0.12 M phosphate buffer and fractionated on HA. The reassociation was followed spectrophotometrically at 260 nm and then each fraction was Cl_3CCOOH precipitated and counted. Experimental points from the reassociation and annealing data were best fitted to a second-order kinetic curve.

(18 000 nucleotides) by the zero-time S_1 method were analyzed for size distribution by electrophoresis on 5% polyacrylamide–100% formamide gels (Figure 4). They appear heterogeneous in chain length varying from 100 to 900 nucleotides with an average length of 190 nucleotides. Boundary sedimentation velocity measurement of the purified inverted duplex under neutral conditions gives an average length of 180 base pairs.

Sequence Arrangement of Inverted DNA. In order to know the sequence arrangement within the repeated inverted region, reassociation experiments of inverted repetitive sequences were

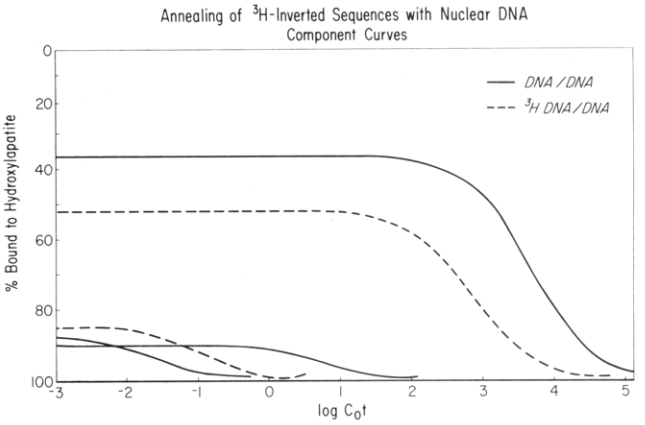

FIGURE 6: The best fitted curves of the nuclear DNA reassociation and the ^{3}H-labeled inverted sequence annealing shown in Figure 5 were resolved by the same computer program into their component curves, respectively.

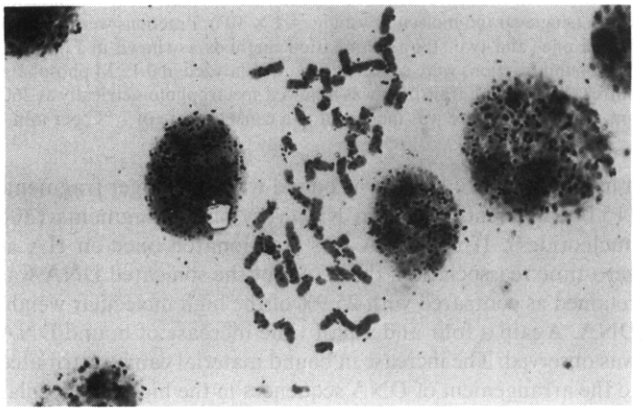

FIGURE 7: Autoradiographs of in situ hybrids between human lymphocyte chromosomes and ^{3}H-labeled RNA complementary to human inverted repetitive sequences. Exposure time was 3 weeks.

carried out in the absence of nuclear DNA. Purified inverted repetitive sequences obtained by zero-time S_1 treatment were heat denatured and reassociated; 10% of the material reassociates at zero time. It appears that 10% of the sample was not S_1-nicked during the purification process. Two models are considered to explain this result. One model proposes that a short single-strand loop may occur at one end of the inverted repetitive duplex and this single-strand loop is too short to be sensitive to S_1 digestion. In the other model, which may account for the majority of the population, each inverted repeat itself contains a shorter internal inverted repeat. During the isolation process, S_1 digests the rather long loop and the single-stranded tails giving a well-paired duplex; however, after further denaturation each strand can fold back on its internal inverted repeat. These models have been previously discussed by Schmid et al. (1975).

In order to know whether inverted sequences form a single class of repetitive sequences in the human genome, reassociation experiments were carried out. The purified ^{3}H-labeled inverted repetitive sequences were incubated with a 100-fold excess of nuclear DNA to various C_0t values and analyzed. Results depicted in Figure 5 show that the inverted repetitive sequences are contained in almost all kinetic classes of human DNA. The nuclear DNA reassociation curve and the ^{3}H-labeled inverted sequence annealing curve were resolved into their component curves, respectively, by a computer program (Figure 6). The slowest reassociating component of nuclear DNA comprises 64% of the total and reassociates with a $C_0t_{1/2}$

of 5000. The slowest annealing component of the ^{3}H-labeled inverted sequences comprises 48% and anneals with a $C_0t_{1/2}$ of 640. When corrected for purity of each component, the $C_0t_{1/2}$ values become 3200 and 308, respectively. The ratio of these $C_0t_{1/2}$ values shows that the slowest annealing component in the inverted sequences is repeated about ten times per haploid genome.

Distribution of Inverted Repetitive Sequences on Metaphase Chromosomes. The distribution of the inverted repetitive sequences on human metaphase chromosomes was studied by in situ hybridization experiments (Figure 7). ^{3}H-Labeled RNA complementary to the inverted repetitive DNA was hybridized to chromosome preparations. One observes that grains are scattered throughout the chromosomes. There is no localization on any specific chromosomes or clustering on interphase nuclei as was observed with human satellite DNAs. Thus inverted repetitive sequences appear to be distributed throughout the chromosomes.

Discussion

Inverted repetitive sequences comprise 3–6% of the total human DNA (Wilson and Thomas, 1974; Schmid and Deininger, 1975; this report). These sequences are heterogeneous in chain length with an average of 190 nucleotides and reassociate with unimolecular kinetics. The single-strand nuclease-resistant regions contain sequences representative of almost all repetition frequencies, highly repetitive as well as very few copies per genome, and they are widely distributed throughout the metaphase chromosomes.

The properties of the inverted repetitive sequences in human DNA found in our work were similar to those found by other workers (Wilson and Thomas, 1974; Schmid and Deininger, 1975). Wilson and Thomas (1974) found inverted repetitive sequences in HeLa DNA to be about 300–1200 nucleotides in length. Our methods for measuring the average chain length of 190 nucleotides of these sequences are more sensitive but may have selected against detection of very long inverted repetitive sequences.

The number of inverted repeats per haploid genome can be estimated from the mean length of 190 base pairs (Figure 4) and the haploid genome size of 7×10^9 base pairs. Assuming inverted repeats comprise 6% of the DNA, then 4×10^8 base pairs are involved in fold-back structures or 2×10^6 such inverted repeats occur per haploid genome.

The human genome is remarkable in that it contains such a wide variety of sequence classes. These include several simple sequence satellite DNAs, a variety of highly repetitive and middle repetitive sequences as well as long stretches of single-copy sequences (Schmid and Deininger, 1975). Added to this array of sequence classes are the inverted repeated sequences which themselves are present in a variety of multiplicities in the human genome.

From a teleological point of view, one can argue that inverted repetitive sequences must be biologically important. They may serve as binding sites for regulatory proteins (Gilbert and Maxam, 1973), cleavage sites for restriction enzymes (Kelly and Smith, 1970), or they may be involved in the processing of heterogeneous nuclear RNA, ultimately generating messenger RNA (Ryskov et al., 1972, 1973; Jelinek and Darnell, 1972). It now appears that all of these functions may require DNA sequences with a twofold axis of rotational symmetry, a property characteristic of inverted repetitive sequences.

Acknowledgments

We are very grateful to Dr. M. T. Kuo for assistance in carrying out the chromosome localization experiments and to Drs. William R. Crain and John Monahan for helpful discussions.

References

Arrighi, F. E., Hsu, T. C., Saunders, P., and Saunders, G. F. (1970), *Chromosoma 32,* 224–236.

Bonner, T. I., Brenner, D. J., Neufeld, B. R., and Britten, R. J. (1973), *J. Mol. Biol. 81,* 123–135.

Britten, R. J., Graham, D. E., and Neufeld, B. R. (1974), *Methods Enzymol. 29*E, 363–418.

Britten, R. J., and Smith, J. (1970), *Carnegie Inst. Washington, Yearb. 68,* 378–386.

Cech, T. R., Rosenfeld, A., and Hearst, J. E. (1973), *J. Mol. Biol. 81,* 299–325.

Chuang, C. R., and Saunders, G. F. (1974), *Biochem. Biophys. Res. Commun. 57,* 1221–1230.

Davidson, E., Hough, B., Amenson, C., and Britten, R. J. (1973), *J. Mol. Biol. 77,* 1–23.

Gall, J. G., and Atherton, D. (1974), *J. Mol. Biol. 85,* 633–664.

Gall, J. G., and Pardue, M. L. (1971), *Methods Enzymol. 21D,* 470–480.

Gilbert, W., and Maxam, A. (1973), *Proc. Natl. Acad. Sci. U.S.A. 70,* 3581–3584.

Graham, D., Neufeld, B., Davidson, E., and Britten, R. J. (1974), *Cell 1,* 127–137.

Jelinek, W., and Darnell, J. E. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 2537–2541.

Kelly, T. J., and Smith, H. O. (1970), *J. Mol. Biol. 51,* 393–409.

Kram, R., Botchan, M., and Hearst, J. E. (1972), *J. Mol. Biol. 64,* 103–117.

Maniatis, T., Jeffrey, A., and Van deSande, H. (1975), *Biochemistry 14,* 3787–3794.

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

McCoy, T. A., Maxwell, M., and Kruse, P. F. (1959), *Proc. Soc. Exp. Biol. Med. 100,* 115–118.

Ryskov, A. P., Farashyan, V. R., and Georgiev, G. P. (1972), *Biochim. Biophys. Acta 262,* 568–572.

Ryskov, A. P., Saunders, G. F., Farashyan, V. R., and Georgiev, G. P. (1973), *Biochim. Biophys. Acta 312,* 152–164.

Saunders, G. F., Chuang, C. R., and Sawada, H. (1975), *Acta Haematol. 54,* 227–233.

Saunders, G. F., Hsu, T. C., Getz, M. J., Simes, F. L., and Arrighi, F. E. (1972b), *Nature (London), New Biol. 236,* 244–246.

Saunders, G. F., Shirakawa, S., Saunders, P. P., Arrighi, F. E., and Hsu, T. C. (1972a), *J. Mol. Biol. 63,* 323–334.

Schmid, C., Manning, J., and Davidson, N. (1975), *Cell 5,* 159–172.

Schmid, C. W., and Deininger, P. L. (1975), *Cell 6,* 345–358.

St. John, T., Johnson, J. D., and Bonner, J. (1974), *Biochem. Biophys. Res. Commun. 57,* 240–247.

Studier, F. W. (1965), *J. Mol. Biol. 11,* 373–390.

Sutton, W. D. (1971), *Biochim. Biophys. Acta 240,* 522–531.

Walker, P. M. B. (1971), *Prog. Biophys. Mol. Biol. 23,* 145–190.

Wilson, D. A., and Thomas, C. A., Jr. (1973), *Biochim. Biophys. Acta 331,* 333–340.

Wilson, D. A., and Thomas, C. A., Jr. (1974), *J. Mol. Biol. 84,* 115–144.