

## Gene distribution and nucleotide sequence organization in the human genome

Marino ZERIAL, Julio SALINAS, Jan FILIPSKI and Giorgio BERNARDI

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, Paris

(Received April 2, 1986) – EJB 86 0357

Human DNA was fractionated by centrifugation in  $\text{Cs}_2\text{SO}_4$  density gradients containing 3,6-bis(acetato-mercurimethyl)dioxane (BAMD). Fractions were investigated in their analytical  $\text{CsCl}$  profiles and a number of specific sequences were localized in them. The results so obtained led to an improved understanding of the organization of nucleotide sequences in the human genome, as well as to the discovery that a class of DNA having a very high G + C content and not represented in the mouse genome, is particularly rich in genes and interspersed repetitive sequences.

Although the general background for this work has been discussed in the introduction of the preceding paper [1], it may be worthwhile to recall here some specific points previously made in our laboratory on the human genome as compared with the mouse genome. Even if the human genome could be resolved by density gradient centrifugation in the presence of  $\text{Ag}^+$  into four major components, which were similar in relative amounts, modal buoyant densities and intermolecular heterogeneities to those obtained from the mouse genome [2, 3], some of its properties are strikingly different. First of all, human DNA is characterized [2] by the presence of minor components (defined as components representing less than 3% of total DNA) which show buoyant densities of 1.712, 1.7135, 1.716 and 1.718  $\text{g}/\text{cm}^3$ . In the mouse genome only a minute amount of 1.712  $\text{g}/\text{cm}^3$  component was detected in this buoyant density range. Second, the reassociation properties of the four major components from the human genome [4] are different from those exhibited by the corresponding components from the mouse genome. Differences are particularly striking as far as heavy components are concerned and specially so for the heaviest component. While in the mouse genome 91% of this component is made up by slow-reassociating sequences and 9% by intermediate ones, in the human genome only 46% of this component consists of slow-reassociating sequences, 10% of intermediate sequences and 22% each of foldback and fast sequences.

Other differences between the two genomes were found when families of repeated sequences present in both genomes were studied. The L1 family of repeats is less abundant in the human genome than in the mouse genome (40000 versus 100000 copies per haploid genome) and is present not only in the two light major components as in the mouse genome [5] but also in the first heavy component [6]. In contrast, the *Alu* family of repeats is much more abundant in the human

genome compared to the mouse genome (4% versus 1%; [7, 8]). Its concentration in both cases increases from the light to the heavy components [6].

The present work has been done in parallel with that reported in the preceding paper [1], which it complements in several aspects. The results described here have provided new information on the organization of the genome. One particular discovery reported here concerns the presence in the human genome of a (G + C)-rich fraction representing about 3% of total DNA, which is not represented in the mouse genome. This minor component contains a large number of genes and specific repetitive sequences.

### MATERIALS AND METHODS

#### *DNA preparation*

This was a preparation obtained from human placenta [3]. Its  $s_{20,w}$  value was 32 S corresponding to a relative molecular mass of  $22 \times 10^6$  (33 kb). Major components used in the present work were those obtained and characterized by Cuny et al. [3].

#### *Methods*

Preparative centrifugation of DNA in  $\text{Cs}_2\text{SO}_4$ /BAMD density gradient, analytical  $\text{CsCl}$  density gradient centrifugation, restriction enzyme digestion and hybridization were carried out as described in the preceding paper [1].

#### *Probes*

The localization of *Alu* sequences and of some genes in the fractions from the  $\text{Cs}_2\text{SO}_4$ /BAMD preparative gradients was carried out by hybridization experiments using the following probes: Blur 8 (*Alu* repeats [9]); pH-G1 ( $\alpha$ -globin [10]); the 1.7-kb *Bam*HI restriction fragment of L33 [11] subcloned into pBR322 (*c-sis* [12]); pRyc-7.4 (*c-myc* [13]); pAB (*c-mos* [14]); pHrI (ribosomal DNA [15]).

Correspondence to J. Filipski, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, F-75005 Paris, France

Abbreviations. BAMD, 3,6-bis(acetato-mercurimethyl)dioxane;  $r_t$ , BAMD/nucleotide molar ratio; kb,  $10^3$  bases.

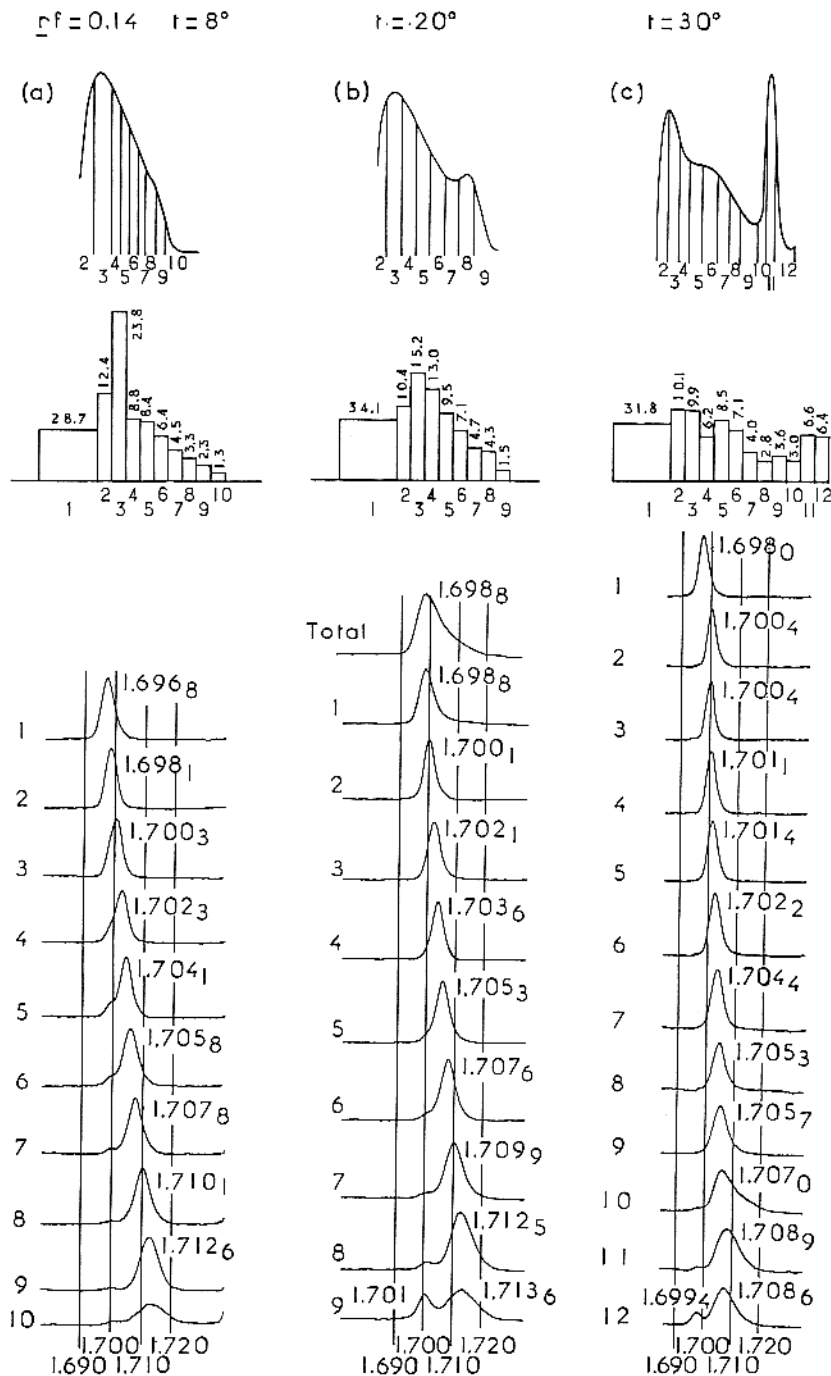


Fig. 1. Fractionation of human DNA by  $\text{Cs}_2\text{SO}_4/\text{BAMD}$  density gradient centrifugation. DNA was centrifuged at the temperature and  $r_f$  values indicated, using the 0.1 M ( $8^\circ\text{C}$ ,  $20^\circ\text{C}$ ) or the 0.4 M ( $30^\circ\text{C}$ ) solvent. The top panel shows the transmission profiles of fractionated DNA as recorded at 253.7 nm. The middle panel shows the histograms of the relative amounts of DNA in the fractions; in all cases, fraction 1 corresponds to the pellet. The bottom panel shows the analytical  $\text{CsCl}$  profiles of pooled fractions. Corresponding fractions from six tubes were pooled in the experiment (b), single fractions were used in the experiments (a) and (c). The analytical  $\text{CsCl}$  profile of total DNA is also shown in experiment (b).

## RESULTS

### Resolution of human DNA by $\text{Cs}_2\text{SO}_4/\text{BAMD}$ density gradient centrifugation

A series of experiments were carried out with the purpose of defining experimental conditions leading to a satisfactory resolution of human DNA by  $\text{Cs}_2\text{SO}_4/\text{BAMD}$  density gradient centrifugation.

The first experiment, presented in Fig. 1a, concerns a preparative  $\text{Cs}_2\text{SO}_4/\text{BAMD}$  density gradient centrifugation at  $8^\circ\text{C}$ ,  $r_f = 0.14$ , 0.1 M solvent. The amount of the pelleted material was 29%. The first three fractions had modal buoyant densities between  $1.697 \text{ g/cm}^3$  and  $1.700 \text{ g/cm}^3$  and represented about 65% of the material. The most (G + C)-rich fractions, 9 and 10, had modal buoyant densities close to  $1.713 \text{ g/cm}^3$ .

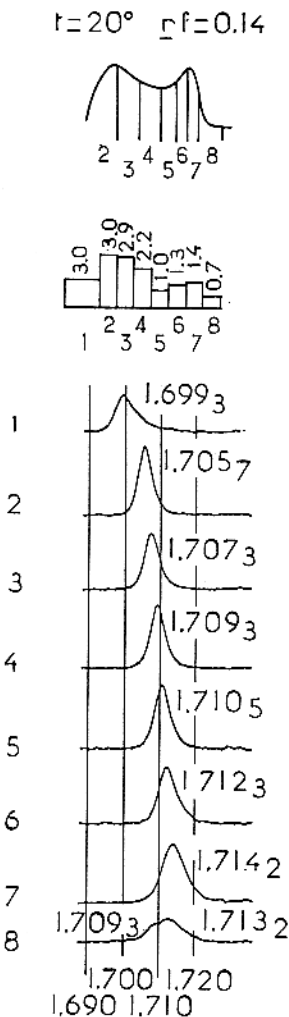


Fig. 2. Recentrifugation of human DNA (G + C)-rich fractions in  $\text{Cs}_2\text{SO}_4/\text{BAMD}$  density gradient. Fractions 10–12 ( $2.5 A_{260}$  units) from four tubes of DNA fractionated in the experiment of Fig. 1c were pooled and recentered at  $20^\circ\text{C}$ ,  $r_f = 0.14$ , using the 0.1 M solvent. Other indications as in Fig. 1

In the second experiment, run under the same conditions but using a temperature of  $20^\circ\text{C}$  (Fig. 1b), the amount of pelleted material was 34%; the first three fractions ranging in modal buoyant density from  $1.698 \text{ g/cm}^3$  to  $1.702 \text{ g/cm}^3$ , represented about 60% of the material. The most (G + C)-rich fraction had a modal buoyant density close to  $1.714 \text{ g/cm}^3$  and was accompanied by a peak centered at  $1.701 \text{ g/cm}^3$ , probably largely corresponding to satellite IV [16].

The third experiment (Fig. 1c) shows that, at  $30^\circ\text{C}$ ,  $r_f = 0.14$ , 0.4 M solvent, the pellet represents 32% of DNA, fractions in the  $1.698 - 1.704 \text{ g/cm}^3$  range are better resolved but heavier fractions are less well resolved. The first five fractions, ranging in modal buoyant density from  $1.698 \text{ g/cm}^3$  to  $1.704 \text{ g/cm}^3$ , represented 67% of DNA. These conditions also led to the resolution of the  $1.6994 \text{ g/cm}^3$  satellite band.

Decreasing or increasing the  $r_f$  values led to less satisfactory resolutions, since a lower  $r_f$  value, 0.12, reduced both the amount of DNA in the pellet and the resolution of light fractions, whereas higher  $r_f$  values (0.16 and 0.19) led to pelleting so much DNA that essentially only the heavy fractions remained in the gradient (not shown).

The last fractions in the gradients of Fig. 1 showed two phenomena (a) the presence of a much broader main peak;

and (b) a lower amount of DNA compared to what was expected from the absorbance measurement of the fractions. Both phenomena have already been commented upon in the preceding article [1].

A better resolution of heavy fractions from human DNA was achieved by recentering at  $20^\circ\text{C}$ ,  $r_f = 0.14$ , fractions 10–12 as obtained from the experiment of Fig. 1c (Fig. 2). A  $\text{CsCl}$  analysis of the fractions so obtained revealed that, under these conditions, the pellet contained the  $1.6994 \text{ g/cm}^3$  satellite already detected in fraction 12 of the experiment of Fig. 1c and that the material present in the gradient could be resolved into fractions ranging in buoyant density from  $1.706 \text{ g/cm}^3$  to  $1.714 \text{ g/cm}^3$ .

#### Localization of specific sequences in the human genome

A number of genes were localized in the fractions of human DNA obtained in the experiments just described.

Two experiments along this line had the aim of comparing hybridization results as obtained on the major DNA components previously isolated [2] and on  $\text{Cs}_2\text{SO}_4/\text{BAMD}$  fractions, in cases where the comparison was of special interest.

In the first experiment the most (G + C)-rich fractions were tested for the presence of a gene, the  $\alpha$ -globin gene, which had been detected in unfractionated human DNA, but in none of its four major components (Fig. 3a). Indeed, the  $\alpha$ -globin probe hybridized on a *Bam*HI fragment having the expected size of 15.8 kb [4], which was localized in fractions 7 and 8 from the second experiment of Fig. 1. These fractions have modal buoyant densities of  $1.7099 \text{ g/cm}^3$  and  $1.7136 \text{ g/cm}^3$ , respectively, a result explaining the absence of the gene from the major components previously isolated.

The second experiment concerned the *c-myc* gene. This was localized in both the  $1.704 \text{ g/cm}^3$  and  $1.708 \text{ g/cm}^3$  major components, the intensity of the hybridization band in the  $1.704 \text{ g/cm}^3$  component being slightly higher (Fig. 3b). When assayed on the fractions from an experiment run under the same conditions as the second experiment of Fig. 1, this gene was detected in fractions 5 ( $1.7049 \text{ g/cm}^3$ ), and 6 ( $1.7064 \text{ g/cm}^3$ ), the gene concentration being higher in the latter fraction.

Two other oncogenes were localized in human DNA fractions. The *c-mos* gene was found (Fig. 4a) in both fraction 3 ( $1.7021 \text{ g/cm}^3$ ) and fraction 4 ( $1.7036 \text{ g/cm}^3$ ) from the second experiment of Fig. 1; the latter fraction showed a stronger hybridization; a very weak hybridization could also be detected in fraction 5 ( $1.7053 \text{ g/cm}^3$ ). The *c-sis* gene was mainly found (Fig. 4c) in fraction 9 ( $1.7126 \text{ g/cm}^3$ ) of the first experiment of Fig. 1, with barely detectable amounts in the two lighter fractions, 7 ( $1.7078 \text{ g/cm}^3$ ) and 8 ( $1.7101 \text{ g/cm}^3$ ).

Fractions having buoyant densities higher than the  $1.708 \text{ g/cm}^3$  component, derived from the second experiment of Fig. 1, were also shown to contain ribosomal genes (Fig. 4b). These were centered in fraction 9 ( $1.7136 \text{ g/cm}^3$ ), but a considerable hybridization was also found in fraction 8 ( $1.7128 \text{ g/cm}^3$ ), which showed some additional bands; a weaker hybridization could also be detected in fractions 7 ( $1.7099 \text{ g/cm}^3$ ) and 10.

As far as interspersed repeated sequences are concerned, the distribution of the most abundant family present in the human genome, the *Alu* family [3], was assessed in the fractions derived from the first experiment of Fig. 1 (Fig. 5a) using the Blur 8 probe. The distribution of the repeats showed a unimodal peak centered at a buoyant density of  $1.708 \text{ g/cm}^3$

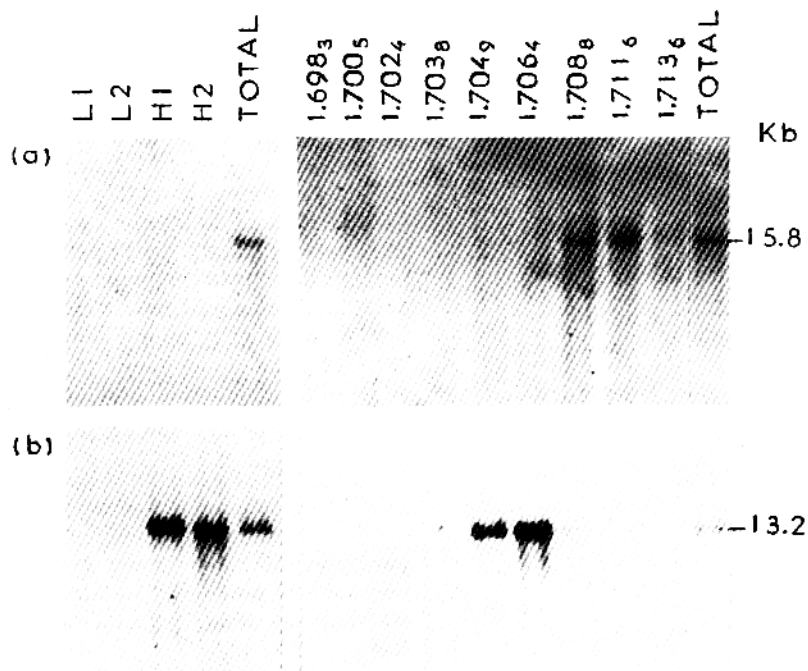


Fig. 3. Localization of the  $\alpha$ -globin and *c-myc* genes in the human genome. 10  $\mu$ g DNA from each fraction of the experiment of Fig. 1b and from the major components [3] of the human genome were digested with *Bam*HI and *Eco*RI, respectively, electrophoresed in 0.8% agarose gels, transferred and hybridized with  $\alpha$ -globin and *c-myc* probes respectively. After hybridization, filters were washed under high-stringency conditions. For details, see Materials and Methods of the preceding paper [1]

(Fig. 5b) and extended to higher densities. Previous results, in which the isolated four major components of the human genome had been used, had revealed an increasing concentration of *Alu* sequences reaching a maximum in the 1.708  $\text{g}/\text{cm}^3$  component [6].

The results just described prompted a search for other repeated sequences in the most (G + C)-rich fractions. This was done by digesting with *Hae*III the fractions from an experiment similar to the first one of Fig. 1 and also the fractions from the recentrifugation experiment of Fig. 2. The first digestion (Fig. 6a) shows the presence of series of sharp bands particularly bright in the last fraction, which contained the 1.6994  $\text{g}/\text{cm}^3$  satellite IV. When the fractions from the experiment of Fig. 2 were digested (Figs 6b and 7), a subset of the bright bands, already seen in the last fraction, were found in the first fraction which contained the satellite. These bands formed a characteristic ladder with a repeat unit of 170 base pairs. The other bands obtained by digestion of the heaviest fractions of the subfractionated DNA did not exhibit any obvious repetitivity pattern.

## DISCUSSION

### *The fractionation of human DNA*

The results of fractionation of human DNA in  $\text{Cs}_2\text{SO}_4$ /BAMD density gradients are essentially similar to the results of experiments done using mouse DNA [1]. The effects of temperature,  $r_f$  and solvent are parallel in both cases. The increase of temperature during centrifugation also results in the formation of a sharp artefactual peak close to the top of the tube composed mainly of the (G + C)-rich DNA. As in the case of mouse satellite DNA, the (A + T)-rich human satellite IV shows anomalously low BAMD binding. Besides, like to mouse satellite DNA it exhibits an anomalous spread

in the gradient under some conditions (Fig. 1) but not under others (Fig. 2).

Two points deserve to be stressed. (a) The conditions used in Fig. 1a and 1b favor the fractionation of (G + C)-rich DNA molecules, whereas those used in the experiment of Fig. 1c favor the fractionation of (G + C)-poor DNA molecules; the poorly resolved (G + C)-rich DNA molecules from the latter experiment could, however, be fractionated by recentrifuging them using the conditions of the experiment of Fig. 1b (see Fig. 2). As in the case of the fractionation of mouse DNA [1], different experimental conditions may, therefore, be used to optimize separations in a chosen range of buoyant densities.

(b) While the light fractions of human DNA are very close in buoyant density to those of mouse DNA, the heaviest ones are definitely higher in buoyant density; recentrifugation of the heavy fractions of human DNA (Fig. 2) showed that as much as 1.3% and 1.4% of it has buoyant densities as high as 1.7123  $\text{g}/\text{cm}^3$  and 1.7142  $\text{g}/\text{cm}^3$  respectively. This result is strikingly different from that obtained in the parallel experiment on mouse DNA (see Fig. 5 of [1]), where no fraction exhibiting modal buoyant densities higher than 1.711  $\text{g}/\text{cm}^3$  could be isolated, and raises the problem of the nature of the very-high-density sequences. This point will be further discussed below.

### *Sequence localization and the organization of the human genome*

The hybridization results of Figs 3–4 complement the information obtained in the preceding article [1].

Except for *c-sis*, practically only detected in a single fraction, the other single-copy genes investigated here, *c-mos*,  $\alpha$ -globin and *c-myc*, were found in comparable amounts in two neighboring fractions. The presence of the DNA molecules

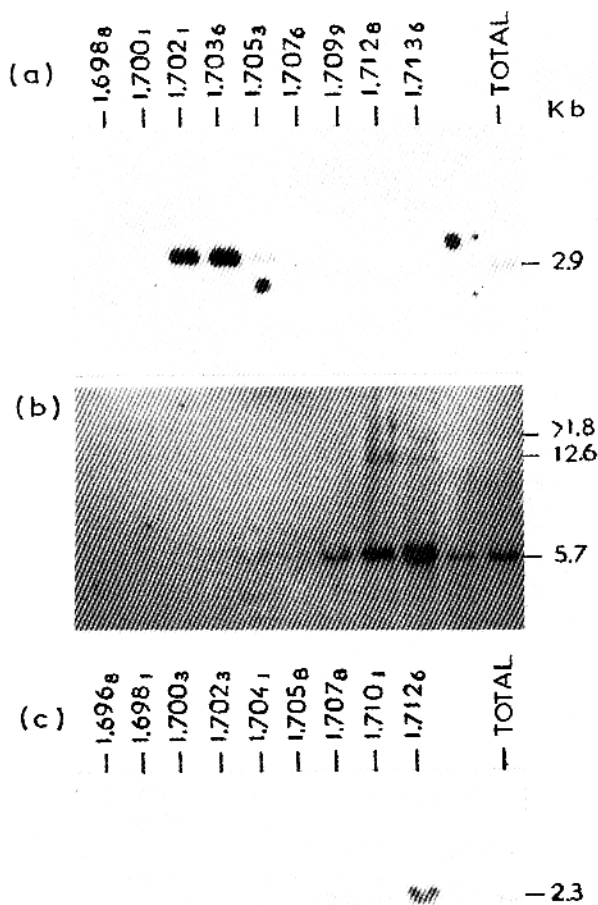


Fig. 4. Localization of *c-mos* and *c-sis* genes and ribosomal sequences in the human genome. 10  $\mu$ g DNA from each fraction of the preparative gradients from the second (*c-mos* and ribosomal) and first experiment (*c-sis*) of Fig. 1 and of total human DNA were digested with *Eco*RI, electrophoresed in 0.8% agarose gels, transferred and hybridized with *c-mos* (a), ribosomal (b) and *c-sis* (c) probes. After hybridization, filters were washed under high-stringency conditions. For details, see the Materials and Methods section of the preceding paper [1]

carrying these genes in two fractions does not necessarily imply a wider distribution in the gradient compared to those of *c-sis* or of the single-copy mouse genes analyzed in the preceding paper [1]. Indeed, the modal buoyant density of these molecules might simply be located between those of two neighboring fractions; in this case even a very narrow distribution, such as that described for some mouse genes [1], would lead to the results obtained. On the other hand, such results are perfectly compatible with a larger intermolecular heterogeneity, which might be due to the fact that the DNA segment carrying the gene is endowed with a higher degree of compositional heterogeneity compared to those already discussed [1] or to its location near the border of two isochores of widely different base composition. Such could be the case of *c-myc*, a gene which is not only localized on two different fractions but also on two different major components. Finally another factor which certainly has played a role in widening the distribution is the rather low molecular size (33 kb) of the

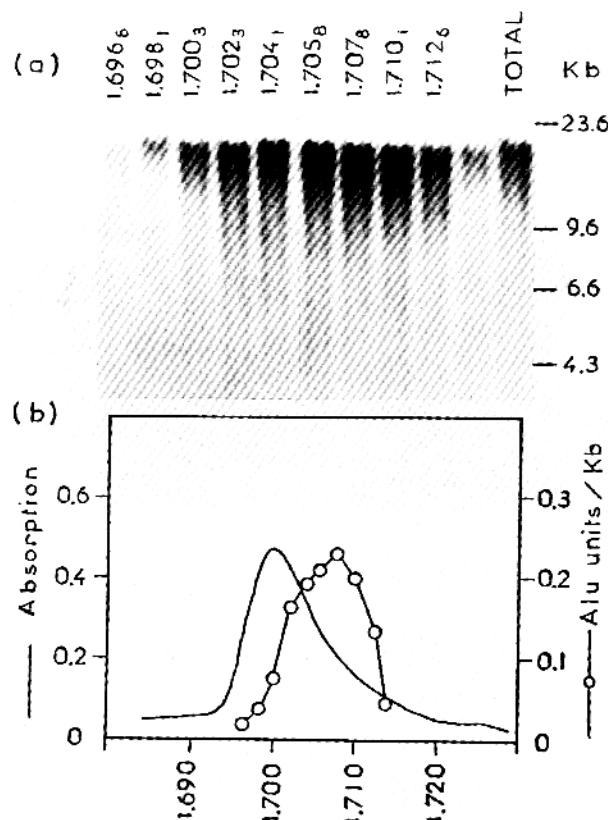


Fig. 5. Localization (a) and distribution (b) of Alu sequences in the human genome. (a) 10  $\mu$ g DNA from each fraction of the experiment of Fig. 1a and of total human DNA were digested with *Kpn*I, electrophoresed in a 0.8% agarose gel, transferred and hybridized with a Blur 8 probe. After hybridization, the filter was washed under low-stringency conditions. For details, see Materials and Methods of the preceding paper [1]. (b) An autoradiogram of (a) was used to calculate the relative amounts of the Alu sequences in the DNA fractions of Fig. 1a. These amounts, expressed as copies/kb DNA, are plotted against the buoyant density values of the DNA in the fractions (O—O) and compared with the analytical CsCl profile of total human DNA (—)

human DNA preparation used, compared to that of mouse DNA (75 kb).

#### The high-density fraction of the human genome

The hybridization experiment localizing the  $\alpha$ -globin gene is also relevant in another context. In previous work we had considered as minor components all the components represented in the genome to an extent of 3% or less [2]. As a consequence, some small fractions higher in density than 1.708 g/cm<sup>3</sup> were considered as minor components [2] and supposed to correspond to some specific sequences, like the clustered ribosomal RNA genes. The first indication that this situation was not that prevailing in the human genome came from our observation (Fig. 3a) that none of the four major components from the human genome showed the presence of the  $\alpha$ -globin genes. The explanation for this result came from an experiment carried out on all fractions from the second gradient of Fig. 1, since this showed the presence of  $\alpha$ -globin in fractions having modal buoyant densities of 1.7088 g/cm<sup>3</sup> and 1.7116 g/cm<sup>3</sup>. This experiment, therefore, provided the evidence for the presence of single-copy genes in what had been previously considered a minor fraction. The localization

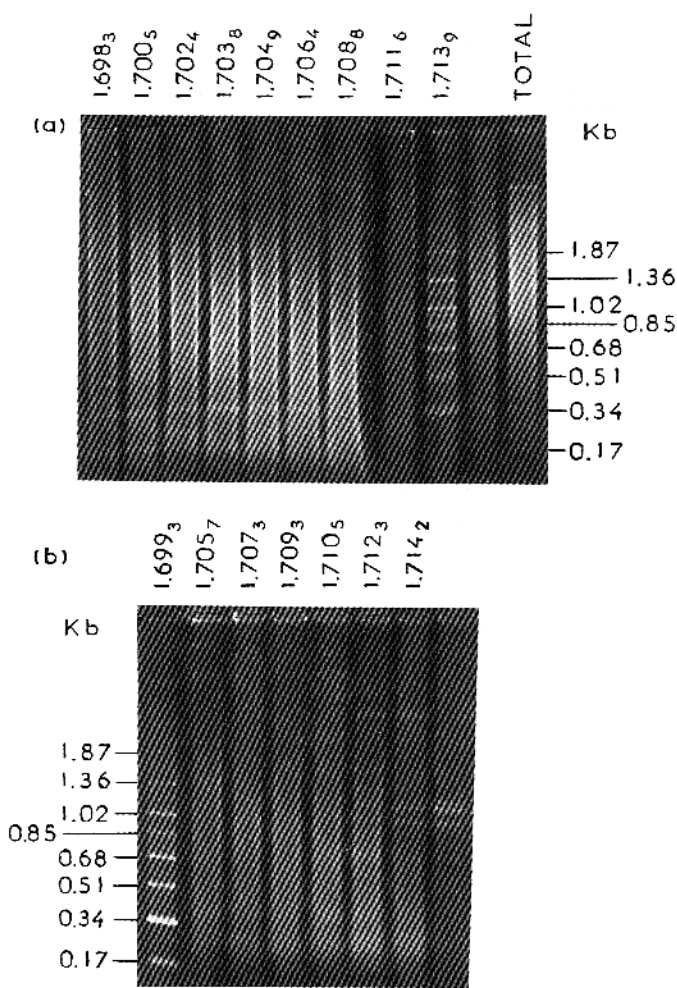


Fig. 6. Gel electrophoresis of *Hae*III digests of human genome fractions. 2  $\mu$ g DNA from each fraction of the experiment of Fig. 1b and of total DNA (a) and of the experiment of Fig. 2 (b) were digested, electrophoresed on 1.4% agarose gel and stained with ethidium bromide. For details, see Materials and Methods of the preceding paper [1]. The sizes indicated correspond to the *Hae*III fragments of human satellite IV [16]

of the *c-sis* gene in a fraction having a density of 1.7126  $\text{g}/\text{cm}^3$  provided a second example of this situation, which was also found for the pre-pro-opiomelanocortin, *c-Ha-ras1* and *c-abl* genes [17]. In contrast, no gene has been localized so far in fractions higher than 1.708  $\text{g}/\text{cm}^3$  in mouse.

These observations raise two important points, namely (a) whether such an unexpected difference between the mouse and the human genome is real and, if so, (b) what is the reason for the difference? As far as the first point is concerned, the recentrifugation experiments of (G + C)-rich fractions, as carried out on both genomes (Fig. 2 of this paper and Fig. 5 of [1]) do prove unambiguously the reality of the difference.

Concerning the second point, two main explanations can be considered. (a) The human genome comprises an additional component characterized by a higher G + C content. This would have arisen by the same general mechanism postulated for the origin of heavy components in warm-blooded vertebrates, namely a regional increase in G + C content [17]. This hypothesis seems to be not very likely in view of the

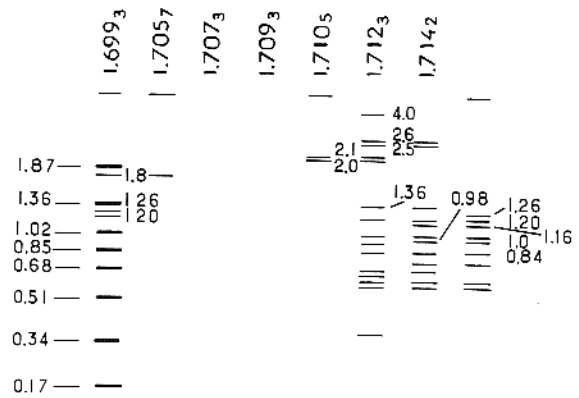


Fig. 7. Schematic representation of the electrophoresis of Fig. 6b. The numbers next to the bands correspond to the molecular sizes of the DNA fragments in kb

apparent conservation of isochores in mammals, since it postulates a rapid increase in G + C in some genome regions. An implication of this hypothesis is that the genes present in this component, which we called H3, should show a higher G + C content than those present in component H2. Along this line there are examples of human genes in which coding sequences are significantly higher in G + C content than the coding sequences of analogous genes in rodents (these are  $\alpha$ -globin, insulin and pre-pro-opiomelanocortin genes). Other 'heavy' genes, however, do not show this difference. Alternatively (b) component H3 can be due to the 'invasion' of part of the H2 component by one or more families of (G + C)-rich repeats. This hypothesis does not imply a point mutation mechanism but rather the duplication and translocation of repeated sequences.

Several families of repetitive sequences were found to be interspersed in the human genome. The most abundant, the *Alu* family [3], are distributed in the human genome in such a way that they form a maximum centered at 1.708  $\text{g}/\text{cm}^3$  as shown in Fig. 5. Thus the *Alu* sequences alone (at least those which are detected by the Blur 8 probe) cannot be directly responsible for the high buoyant density of molecules found in the H3 component of the human genome. Good candidates for invading repetitive sequences, partially responsible for the high average G + C content of DNA in the H3 component, are the 'minisatellite' sequences. These interspersed repetitive sequences, composed of tandem repeats of short (G + C)-rich units, have been found in the proximity of coding sequences of several (G + C)-rich human genes [18]. Another family of interspersed tandem repeats has been recently described [19], which also has the same type of organization as that of minisatellite sequences. A number of bands, possibly corresponding to unknown interspersed repeats, are found in fractions having densities of 1.710  $\text{g}/\text{cm}^3$  and higher (Figs 6b and 7). The presence of these and other repetitive sequences in the heaviest fractions of human DNA would thus explain the high amount of fast-renaturing DNA in (G + C)-rich components of human DNA as compared with the mouse DNA.

At the present state of knowledge it is difficult to decide whether the last mechanism [the invasion of component H2 by (G + C)-rich repetitive sequences] was the only one which contributed to the formation of component H3 in the human genome. In this connection it would be interesting to examine the DNAs of other mammals, like cat, rabbit and dog, which also contain very high (G + C)-rich components [2].

Two of us (J.S. and M.Z.) thank the *Ministerio de Educacion y Ciencia* (Madrid, Spain) and the *Associazione Italiana per la Ricerca sul Cancro* (Milan, Italy) respectively, for financial support. Thanks are due to Dr H. Bünemann (Düsseldorf) for his gift of BAMD, and Drs C. Croce (Philadelphia), W. R. Jelinek (New York), M. Mottes (Milan) and C. Weissmann (Zürich) for their gifts of cloned DNA probes. Finally, we wish to thank Martine Brient for typing this manuscript and Philippe Breton for the artwork.

## REFERENCES

- Salinas, J., Zerial, M., Filipinski, J. & Bernardi, G. (1986) *Eur. J. Biochem.* **160**, 469–478.
- Thiery, J. P., Macaya, G. & Bernardi, G. (1976) *J. Mol. Biol.* **108**, 219–235.
- Cuny, G., Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 227–234.
- Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 235–239.
- Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. & Bernardi, G. (1982) *Proc. Natl Acad. Sci. USA* **79**, 355–359.
- Soriano, P., Meunier-Rotival, M. & Bernardi, G. (1983) *Proc. Natl Acad. Sci. USA* **80**, 1816–1820.
- Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. L. & Saunders, G. F. (1981) *Proc. Natl Acad. Sci. USA* **78**, 1508–1513.
- Georgiev, G. P., Ilyin, Y. V., Chmelianskaite, Ryskov, V. G., Kramerov, A. P., Skryabin, D. A., Krayev, K. G., Lukanidin, E. M. & Grigorian, M. S. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 641–654.
- Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. P., Chondary, P. V., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) *Proc. Natl Acad. Sci. USA* **77**, 1398–1402.
- Dierks, P., Van Ooyen, A., Mantei, N. & Weissmann, C. (1981) *Proc. Natl Acad. Sci. USA* **78**, 1411–1415.
- Della Favera, R., Gelmann, E. P., Gallo, R. C. & Wong-Staal, F. (1981) *Nature (Lond.)* **292**, 31–35.
- Della Favera, R., Gallo, R. C., Giallongo, A. & Croce, C. M. (1982) *Science (Wash. DC)* **218**, 686–688.
- Stanton, L. W., Wah, R. & Marcu, K. B. (1983) *Nature (Lond.)* **303**, 401–406.
- Watson, R., Oskarsson, M. & Vande Woude G. F. (1982) *Proc. Natl Acad. Sci. USA* **79**, 4078–4082.
- Mottes, M., Tsai Lai, S.-A., Montoya, J. & Attardi, G. (1984) *Gene* **27**, 109–113.
- Frommer, M., Prosser, J., Tkachuk, D., Reisner, A. H. & Vincent, P. C. (1982) *Nucleic Acids Res.* **10**, 547–563.
- Bernardi, G., Olofsson, B., Filipinski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985) *Science (Wash. DC)* **228**, 953–958.
- Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (Lond.)* **314**, 67–73.
- Meneveri, R., Agresti, A., Della Valle, G., Talarico, D., Siccardi, A. G. & Ginelli, E. (1985) *J. Mol. Biol.* **186**, 483–489.