Compositional properties of telomeric regions from human chromosomes

Albertina De Sario, Brahim Aïssani and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France

Received 23 October 1991

We have investigated the GC levels of third codon position of genes localized in G- (Giemsa), R- (reverse) and T- (telomeric) bands of human metaphase chromosomes, as well as the hybridization of telomeric probes on fractionated human DNA. The first set of results shows much higher GC levels for genes localized in T-bands than in G- or R-bands (the latter being higher than the former). The second set of data shows that telomeric probes corresponding to T-bands hybridize on the GC-richest family (H3) of isochromosomes, whereas telomeric probes corresponding to R-bands hybridize on GC-rich families H1 and H2; in agreement with these findings, the telomeric repeat common to all chromosomes hybridized on isochromosome families H1, H2 and H3.

Human genome; Chromosome; Telomere; T-band; Isochrome

INTRODUCTION

Some years ago it was discovered [1] that the distribution of genes in the human genome is strikingly non-uniform and that the GC-richest isochromosomes, those of the H3 family, exhibit the highest gene concentration; isochromosomes are the long, >300 kb, compositionally homogeneous DNA segments making up the human genome; they belong to a small number of families characterized by different GC levels). Indeed, the GC-richest isochromosomes, which only represent about 3% of the human genome, are characterized by a gene concentration at least 8 times higher than GC-rich isochromosomes, which represent about 31% of the genome, and at least 16 times higher than GC-poor isochromosomes, which represent about 62% of the genome [2]. Very recent investigations have also shown that the GC-richest isochromosomes (i) are the richest ones in CpG doublets and in CpG islands [3,4]; (ii) are preferred integration regions for most retroviruses, and are very actively transcribed [5] (S. Zoubak, A. Rynditch, G. Bernardi, paper in preparation); (iii) are very rich in Alu sequences [6,7]; (iv) are the most recombinogenic ones [8]; and (v) largely correspond to an open chromatin structure characterized by DNase sensitivity [8], a wider nucleosome spacing [9] (Aïssani and Bernardi, unpublished observation), scarcity of histone H1 and acetylation of histones H3 and H4 [9]. Compositional mapping [8,10] of the long arm of human chromosome 21 has shown that the GC-richest isochromosomes correspond to the telomere [10], which is a thermal denaturation resistant band, a T-band [11], and a chromomycin A3-positive, DAPI-negative band [12]. This finding [10] has led to the proposal [8,10] that the GC- and gene-richest isochromosomes, family, H3, corresponds to T-bands [11] and to chromomycin A3-positive bands [12], which are mainly located at about 20 telomeres. This proposal has been tested here by using two different approaches, namely by investigating the GC levels of third codon positions of genes localized in G-, R- and T-bands, and by hybridizing telomeric probes on fractionated human DNA.

2. MATERIALS AND METHODS

2.1. Sequence analysis and chromosomal location of genes

Human genes localized in individual chromosome bands, either at low resolution (400 bands per haploid karyotype) or at high resolution (850 bands), were extracted from HGM10 [13] and HGM11 (Human Gene Mapping Conference, London, August 1991). Gene sequences were obtained from GenBank or EMBL Library. Genes were divided in three classes on the basis of their localization in G-, R- or T-bands, respectively.

2.2. DNA preparation

DNA was extracted from a fresh human placenta as described [7]. The average size of DNA fragments was about 50–100 kb, as determined by gel electrophoresis.

2.3. Preparative centrifugation

DNA was centrifuged in a CsSO4/BAMD gradient at a ligand/ nucleotide molar ratio RI=0.14, as described [7]; BAMD is 3,6-bis(acetato-mercuri-methyl) dioxane. Eleven fractions were collected, dialyzed against 10 mM Tris, 10 mM EDTA, pH 7.5, at room temperature overnight, and against 10 mM Tris, 1 mM EDTA, pH 7.5, at 4°C for 4 days. The fractions were characterized by analytical density gradient ultracentrifugation in CsCl as described [14].

2.4. Probes

The human probes used had been previously localized on one or more telomeric bands either by in situ hybridization or by using somatic hybrids: (i) pHuR93 contains 240 bp of the telomeric tandem repeat [15], and was purchased from ATCC, the American Type Culture Collection; (ii) G2-1H is a single copy sequence localized in telomeric band 4q35 [16]; (iii) Scos146-3 is a cosmid clone which exhibits specific hybridization to 7q36 [17]; (iv) pTH2a is a 390 bp GC-rich (80%) sequence that contains 6 copies of a 29 bp direct repeat,
it is proximal to the telomeric terminal repeat and hybridizes on chromosomes 7, 16, 17, and 21, but not on chromosome 3, as determined by using hybrid cell lines [18]; (v) pTH2 contains a 410 bp sequence of GC; apparently it is a rearranged clone derived from the same human sequence as pTH2d [18].

2.5. Restriction enzyme digestion and hybridization
1 μg of each DNA fraction digested either with HpaII or with EcoRI was loaded on a 0.8% agarose gel. Alkaline DNA transfer was performed onto Hybond-N membrane (Amersham) after partial depurination. Filter hybridization was carried out using probes labeled by the random primer method (Amersham); cosmide clone Scos146-3 was pre-anneled to sonicated DNA from human placenta [19] in order to suppress the effect of highly repetitive sequences. After each hybridization, filters were dehybridized in 0.5% SDS.

3. RESULTS
3.1. Compositional distribution of human genes localized on chromosomal bands
The distribution of GC levels of third codon positions (Fig. 1 and Table I) was done on coding sequences localized on chromosome bands, as determined at high resolution (850 bands per haploid karyotype) or low resolution (400 bands). The former approach could only be applied to 64 coding sequences. The mean GC values

*Genes were localized at high resolution.
for third codon position of coding sequences localized in G-, R- or T-bands were 58, 61 and 73% GC, respectively, the standard deviations being 17, 12 and 12% GC, respectively. While the difference between GC levels of T-band and G- or R-band coding sequences is a large one (12–15% GC), that between coding sequences localized on G- or R-bands is small (3% GC). The possibility of misassignments of genes should, however, be considered. For example, the assignment of APOE to band q13.2 of chromosome 19 is arguable; an alternative possibility would be a localization on the nearest R-band, which is, in fact, a T-band. In such a case, the average GC level of coding sequences localized in G-bands would be 56% instead of 58%. This point is mentioned simply to show how sensitive the mean value is to the removal of even a single (admittedly extreme) gene from such a small sample.

The low resolution approach has the advantage that it can be applied to a larger sample (200 coding sequences in the present case), and the obvious disadvantage of a lower resolution. In this case, mean GC levels for third codon position of genes localized in G-, R- and T-bands were found to be 54, 60 and 72%, respectively, the standard deviations being 16, 14 and 13% GC, respectively. Again, the major difference (12–18% GC) concerns coding sequences which are located in T-bands, whereas that between sequences located in G- and R-bands is definitely much smaller (6% GC). A similar analysis has been published by Ikemura and Wada [20] using a coding sequence sample similar to that used here. In the present work, we have classified genes according to whether they are located in G-, R- (both intercalary and telomeric, but T-negative) and T-bands (both intercalary and telomeric, following Dutrillaux [11] and Ambros and Sumner [12]). In contrast, Ikemura and Wada [20] have used a more complex classification concerning genes located in R-bands, G-bands, telomeric bands (whether T-positive or T-negative), T-type R-bands and intercalary R-bands. There is, therefore, no coincidence between the three classes studied here and the 6 classes investigated by Ikemura and Wada [20] except for G-bands. Another difference concerning the two sets of results concerns the fact that we have considered gene localization not only at low resolution but also at high resolution. The conclusions are, however, largely in agreement.

3.2. Hybridization of telomeric probes on fractionated human DNA

A set of telomeric probes have been hybridized on human DNA compositional fractions to learn about the base composition of telomeres corresponding to either T-positive or T-negative bands. Fig. 2 shows the CsCl profiles of the DNA fractions used.

When the probe pHuR93, containing the terminal repeat common to all the chromosomes [15], was hybridized (Fig. 3A), the signals were found on fractions 4–10 corresponding to GC levels ranging from 42.9–55.6%, the latter fraction showing the highest hybridization intensity; 63% of the human genome, corresponding to isochores families L1 and L2, did not show any hybridization.

In order to study the base composition of individual telomeres, different probes specific for a single chromosome or for a group of them were used. The results
obtained showed different GC levels in telomeres corresponding to either T-positive or T-negative bands. Indeed, while the former generally consist of very GC-rich sequences corresponding to the isochores of the H3 family, the latter consist of DNA fragments having a lower GC-content (H1 or H2 isochores).

An example of a probe homologous to sequences located in a T-positive telomere is given in Fig. 3B. pTH2A [18] is a GC-rich minisatellite (80% GC) proximal to the terminal repeat (TTAGGG)n, localized in the telomeres of several chromosomes, such as chromosomes 7, 16, 17 and 21, all comprising at least one T-band, but not of chromosome 3 which has no telomeric T-bands. As expected, this probe hybridized with fractions 9 and 10, which correspond to 52.9% and 55.6% of GC, respectively. Exactly the same pattern was obtained with pTH14A (data not shown), which probably is a rearranged clone derived from the same human sequence [18].

On the contrary, when T-negative telomeres were investigated in their GC levels, lower levels were found. Fig. 3C and D shows two examples: (i) probe G2-1H, specific for 4q35 [16], is clearly located in fraction 5 (44.8% of GC); and (ii) Sos146-3, a cosmid clone specific for 7q36 [17] was localized on several fractions with a hybridization on fractions 3–5 (41.5–44.8% GC). Both bands 4qter and 7q36 are T-negative.

It should be stressed that the hybridization results presented here inform us about the GC levels of DNA segments as large as the size of DNA fragments used, i.e. 50–100 kb.

4. DISCUSSION

The analysis of the third codon positions of coding sequences localized at either low or high resolution on chromosomal bands definitely indicates much higher average GC values for the genes located in T-bands than for those located in either R- or G-bands; the difference among the latter seems to exist, but is much smaller. Since third codon positions above 72% GC correspond to genes located in the H3 family of iso-
chores [2], this finding indicates that this family corresponds to T-bands. Needless to say, this conclusion is of interest if one considers that the H3 family of isochores not only has the highest concentration in genes, but also the highest transcriptional and recombinational activity, as well as a distinct chromatin structure (see Introduction). On the other hand, the smaller differences between sequences located in G- and R-bands may be due to the fact that the latter (at low resolution) comprise a number of thin G-bands; sequences present in such bands would be counted as sequences present in R-bands (as seen at low resolution). Moreover GC-rich isochores belonging to families H1 and H2 cover a relatively broad range of GC levels.

The hybridization results are of interest in that they show that (i) telomeres, as tested with the telomeric tandem repeat [15], practically correspond to isochores H1, H2, H3; (ii) the GC-rich minisatellites present in telomeric T-bands, like those of chromosomes 7, 16, 17 and 21, are located in the two GC-richest fractions; (iii) probes specific for telomeric R-bands (4q, 7q) hybridize on fractions corresponding to GC-rich isochores of the families H1 and H2. There is, therefore, a substantial difference between the four T-bands and the four non-T-bands (4q, 7q and those of chromosome 3) explored.

In conclusion, the present results strongly support the idea that the H3 family of isochores is located in T-bands. Direct evidence on this point has just been obtained from in situ hybridization of biotin-labelled DNA fragments derived from the H3 isochose family (S. Saccone, A. De Sario, G. Della Valle and G. Bernardi, paper in preparation).

Acknowledgements: This work was supported by the French Ministry for Research and Technology (MRT), the Association Francaise contre les Myopathies (AFM) and Association pour la Recherche sur le Cancer (ARC). We thank Titia de Lange, Harold Riehtman and Bernhard Weber for kindly providing us with telomeric probes.

REFERENCES