Volume 295, number 1,2,3, 22–26 FEBS 10479 December 1991

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 isochores, 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 isochore families H1, H2 and H3.

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

1. 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 isochores, those of the H3 family, exhibit the highest gene concentration; (isochores 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 isochores, which only represent about 3% of the human genome, are characterized by a gene concentration at least 8 times higher than GC-rich isochores, which represent about 31% of the genome, and at least 16 times higher than GC-poor isochores, which represent about 62% of the genome [2]. Very recent investigations have also shown that the GC-richest isochores (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open chromatin structure characterized by DNase sensitivity [8], a wider nucleosome spacing [9] (Aissani and Bernardi, unpublished observation), scarcity of histone H1 and acetylation of histones H3 and H4 [9].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 isochores 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 isochore 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.

Correspondence address: A. De Sario, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France.

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 Cs_2SO_4/BAMD gradient at a ligand/nucleotide molar ratio Rf=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) pTH2Δ is a 390 bp GC-rich (80%) sequence that contains 6 copies of a 29 bp direct repeat;

Published by Elsevier Science Publishers B.V.

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) pTH14Δ contains a 410 bp sequence of 50% of GC; apparently it is a rearranged clone derived from the same human sequence as pTH2Δ [18].

2.5. Restriction enzyme digestion and hybridization

1 μg of each DNA fraction digested either with *Hpa*II or with *Eco*RI 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); cosmid clone Scos146-3 was pre-annealed 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

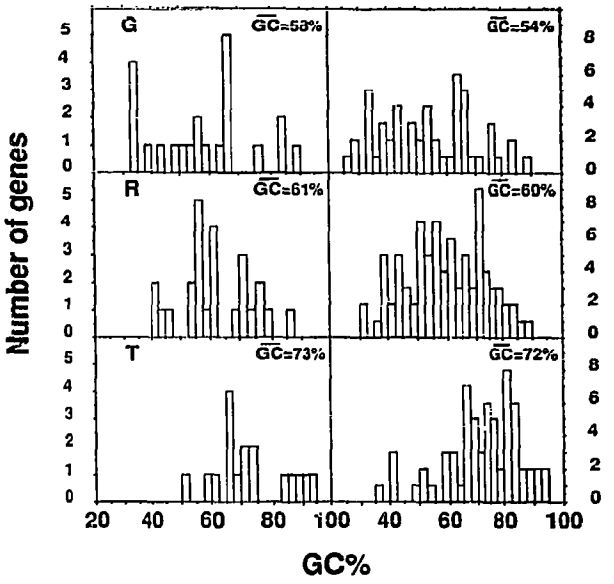

Fig. 1. Histograms of GC levels of third codon positions of human genes localized in G-, R- and T-bands at high resolution (left panels) or low resolution (right panels). Bars correspond to 2.5% GC intervals. Genes from the left panels were included in the right panels using their assignment at high resolution.

Table I List of human genes localized in G-, R- and T-bands used in the present work

G bands

N°	Symbol	Localization	GC III%
1	ACADM	1p31	27.3
2	ALDH1	9q21.1	40.6
3	ALDH2	12q24.2	63.7
4	AMY	1p21	35.3
5	APOC1+2*	19q13.2	66.8
6	APOE*	19q13.2	88.4
7	AR*	Xq12	63.4
8	CALCA*	11p15.4	66.0
9	CD8A*	2p12	83.9
10	CDC2*	10q21.1	34.9
11	CDR	Xq27	54.9
12	CETP*	16q21	75.3
13	CLG4*	16q21	43.3
14	COL2A1*	12q14.3	37.7
15	COL11A1	1p21	27.8
16	CST3*	20p11.22	84.4
17	CYB3*	Xp21.1	50.7
18	DMDL	6q24	49.4
19	EGR2*	10q21.1	65.2
20	EPO	7q21	70.6
21	FGA+B+C	4q28	41.5
22	GLUT5	1p31	78.1
23	GP3A*	17q21.32	68.7
24	GST1	1p31	64.6
25	HGF	7q21.1	30.6
26	IGF1	12q23	53.3
27	IGKC*	2p12	49.5
28	IGKV	2p12	64.1
29	INT1L1	7q31	64.0
30	LAMB2	1q31	54.7
31	LDHB	12p12.2-p12.1	42.7
32	LPL*	8p22	54.8
33	KRAS2*	12p12.1	33.2
34	MET	7q31	49.6
35	MIC2	Xp22.32;Yp11.3	52.1
36	MYCN	2p24	79.9
37	NCA*	19p13.2	50.5
38	NID*	1q43	66.1
39	OTC	Xp21.1	43.7
40	PGY1+3	7q21	44.0
41	PRB1+2+4*	12p13.2	33.5
42	PRH1+2	12p13.2	34.1
43	PRKCG	19q13.4	75.0
44	RAP1	12q14	28.8
45	RB1*	13q14.2	33.7
46	QCPR	4p15.3	84.1
47	SST	3q28	68.4
48	STS	Xp22.32	81.3
49	TCRB*	7q35	57.0
50	TGFB2*	1q41	55.9
51	ZFY	Yp11.3	37.7

R bands

N°	Symbol	Localization	GC III%
1	ADA*	20q13.11	72.8
2	AMGL	Yq11	64.0
3	APP*	21q21.2	55.5
4	ARG1	6q23	44.3
5	ATP1A1	1p13	58.5
6	BCL2	18q21.3	83.5
7	BLYM	1p32	39.9
8	C1R	12p13	43.9
9	C4BP	1q32	39.1
10	CA2	8q22	54.0
11	CD3D+E+G	11q23	50.2
12	CD9	12p13	71.4
13	CD58	1p13	36.2
14	CR1	1q32	51.2
15	CRH	8q13	81.8
16	CSF1	5q33	71.6
17	CTLA4	2q33	65.0
18	CYP2C	10q24	51.8
19	DAF	1q32	39.3
20	DCP	17q23	81.4
21	DEF1	8p23	55.8
22	EGF*	4q25	54.5
23	ENO2	12p13	65.1
24	ETS1*	11q23.3	60.4
25	F8C*	Xq28	41.5
26	F11	4q35	48.8
27	FABP1*	2p11	69.7
28	FCE	18q21.3	58.0
29	FGF5	4q21	79.5
30	FGFB*	4q25	55.6
31	FLT1	13q12	46.0
32	RARB	10p11.2	38.2
33	FOS*	14q24.3	71.4
34	G6PD*	Xq28	66.2
35	GAA	17q23	64.7
36	GALT	9p13	62.2
37	GAPD	12p13	71.4
38	GCP*	Xq28	76.4
39	GLUT3	12p13.3	57.9
40	GLUT4	17p13	72.5
41	GRP	18q21	63.8
42	GSR*	8p21.1	55.7
43	GST2*	6p12.2	55.7
44	HAR2	1q21	73.1
45	HEXB	5q13	61.5
46	HP*	18q22.1	42.9
47	HPR*	16q22.1	45.5
48	HPRT	Xq26	39.7
49	HSDB3	1p13.1	64.1
50	IFNB1	9p22	50.7
51	IRNG	12q24.1	43.8

T bands

N°	Symbol	Localization	GC III%
52	ITIH2	10p15	53.7
53	L1CAM*	Xq28	41.6
54	LCAT*	16q22.1	78.3
55	MCP	1q32	31.9
56	MGSA	4q21	69.2
57	MOS	8q11	72.9
58	NERL	8p21	75.9
59	NGFB	1p13	60.8
60	NRAS	1p13	45.8
61	PDHA1	Xp22.1	52.4
62	PFC*	Xp11.4	71.9
63	PGK1	Xq13	56.0
64	PKLR	1q21	72.4
65	POLR2*	17p13.1	61.4
66	POMC	2p23	69.2
67	POR	7q11.2	66.9
68	RBP3	10q11.2	79.2
69	RCP*	Xq28	75.9
70	REN	1q32	69.5
71	SOD1	21q22.1	44.5
72	SPN*	16p11.2	61.3
73	SPTA1	1q21	30.8
74	STSP	Yq11	53.5
75	TAT*	16q22.1	58.4
76	TCRA*	14q11.2	53.4
77	TGFA	2p13	71.4
78	TGFB3	14q24	71.6
79	TP53*	17p13.1	62.4
80	TPI1	12p13	65.6
81	TSHB	1p13	43.1
82	UVO*	16q22.1	55.0
83	VIM*	10p13	71.1
84	a)	Xp11.23	65.6
85	b)	1p32	52.0

N°	Symbol	Localization	GC III%
1	AACT	14q32.1	69.2
2	ABL	9q34	54.8
3	ACR	22q13-qter	61.1
4	AK1	9q34.1-q34.2	79.2
5	ALAD	9q34	68.0
6	ALPP	2q37	79.1
7	APRT	16q24	81.2
8	APS	19q13.3	69.0
9	ARP1	15q26	81.2
10	ARSA	22q13.31-qter	75.7
11	ASS	9q34-qter	74.3
12	BCE1*	21q22.3	68.2
13	BCR	22q11	83.0
14	C4A+B	6p21.3	42.2
15	CGB	19q13.3	82.5
16	CHGA	14q32	72.7
17	CK9B	14q32.3	75.3
18	COL4A1+2*	13q34	51.4
19	COL6A1+2*	21q22.3	60.5
20	COL11A2	6p21.3	52.2
21	COMT	22q11.2	83.8
22	CTSD*	11p15.5	87.5
23	CYP21	6p21.3	81.0
24	DBHA+B	9q34	82.3
25	DIA1	22q13.31-qter	79.8
26	END1	1p36	85.5
27	ERG*	21q22.3	65.8
28	ETS2*	21q22.3	66.4
29	F7*	13q34	70.5
30	F10*	13q34	83.9
31	GGT1	22q11.1-q11.2	81.9
32	GLI	12q13	59.2
33	GPI	19q13.1	91.5
34	HBA*	16p13.3	90.7
35	HBB*	11p15.5	85.5
36	HLA-A	6p21.3	80.5
37	HMG14	21q22.3	48.5
38	HRAS	11p15.5	81.0
39	HSPA1	6p21.3	94.5
40	HSTF*	11q13.3	93.7
41	IGF2*	11p15.5	73.2
42	IGHA1+2*	14q32.33	59.0
43	IGHE*	14q32.33	67.3
44	IL2RB	22q13	75.2
45	INS*	11p15.5	70.0
46	INT1	12q13	83.5
47	KLK1	19q13.3	86.8
48	LALBA	12q13	60.2
49	LHB	19q13.3	82.4
50	LNG	19q13.1	85.4
51	MTB1	17q21	67.1

N°	Symbol	Localization	GC III%
52	MUC2	11p15	41.3
53	NAGA	22q13-qter	73.3
54	OAT	10q26	40.9
55	PFKL	21q22.3	71.4
56	PI	14q32.1	67.2
57	PRM1+2*	16p13.3	72.9
58	PTH	11pter-p15.4	35.7
59	RAP2	13q34	75.3
60	S100B	21q22.3	63.4
61	TF	3q21	57.5
62	TGFB1	19q13.1	83.2
63	TH*	11p15.5	86.2
64	TNFA+B	6p21.3	75.0
65	c)	22q13.3	88.7

*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.

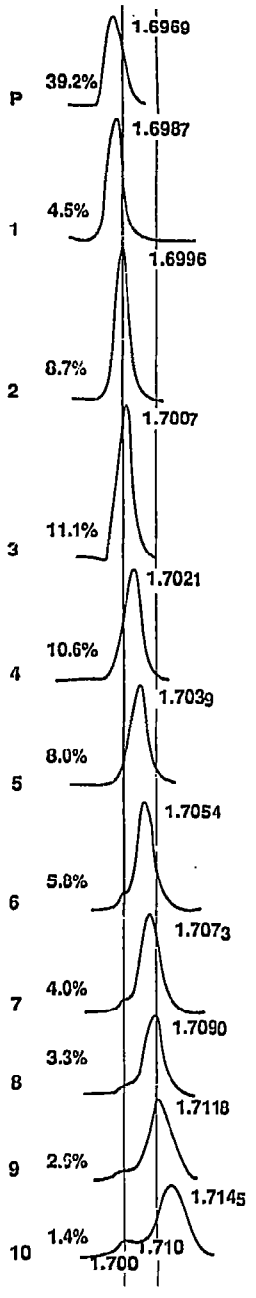

Fig. 2. Analytical CsCl profiles of human DNA fractions. Fractionation was obtained by preparative ultracentrifugation in Cs_2SO_4/BAMD [21,22] at a ligand/nucleotide molar ratio Rf = 0.14. Modal buoyant densities and relative DNA amounts are indicated. P stands for pellet.

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 isochore 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

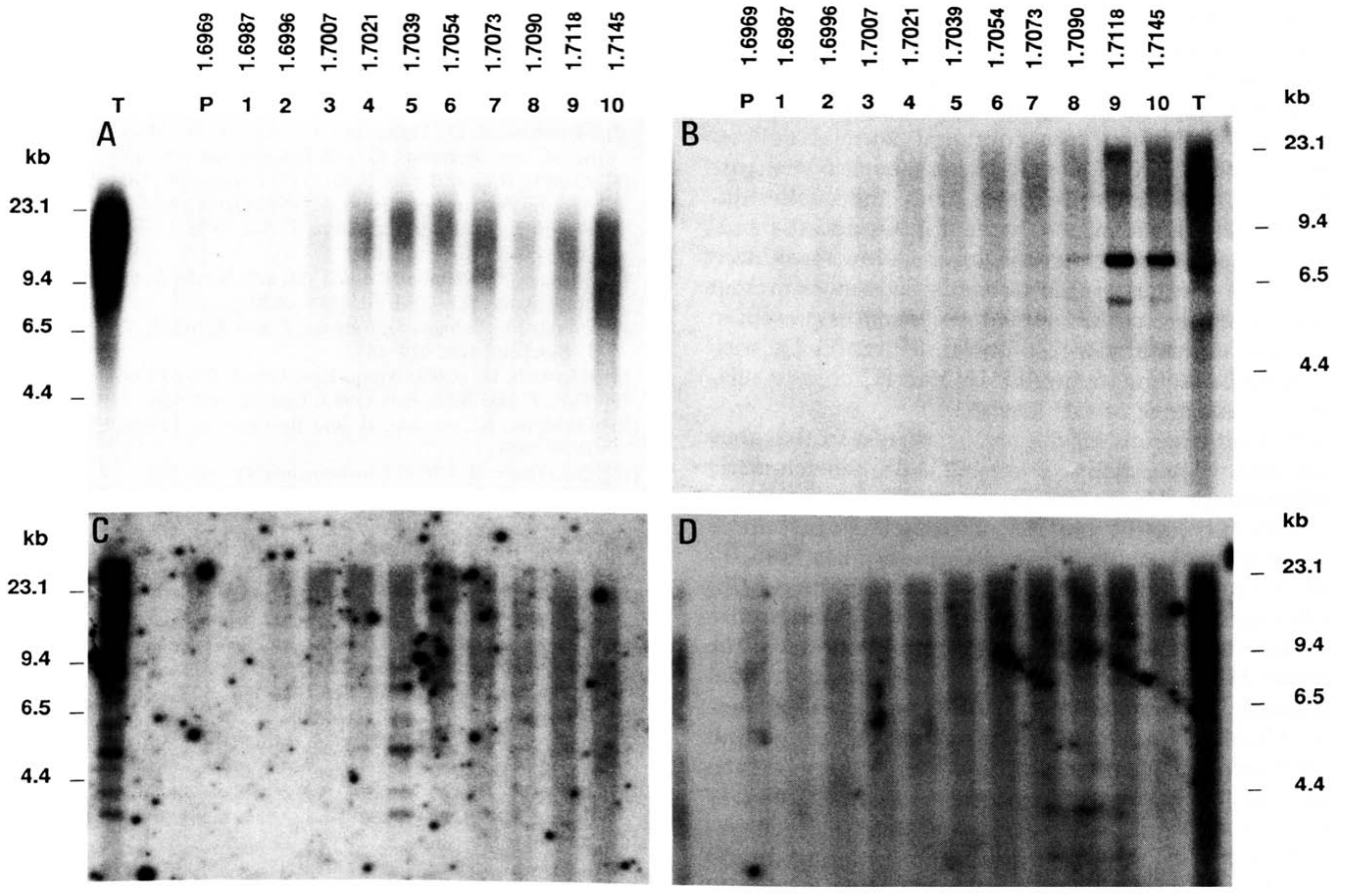

Fig. 3. (A) Hybridization of probe pHuR93, containing a DNA sequence homologous to the human telomeric repeat [15], on equal amounts (1 μg) of human DNA fractions digested with *Hpa*II. (B) Hybridization of probe pTH2Δ, a GC-rich minisatellite located on telomeres of chromosomes 7, 16, 17 and 21, on *Eco*RI-digested human DNA fractions. (C) Hybridization of probe G2-1H, specific for the telomere 4q, on human DNA fractions digested with *Hpa*II. (D) Hybridization of probe Scos146-3, specific for the telomere 7q on human DNA fractions digested with *Eco*RI. In all panels P stands for pellet.

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. pTH2Δ [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 pTH14Δ (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) Scos146-3, 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 isochore 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 Française contre les Myopathies (AFM) and Association pour la Recherche sur le Cancer (ARC). We thank Titia de Lange, Harold Riethman and Bernhard Weber for kindly providing us with telomeric probes.

REFERENCES

[1] Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) Science 228, 953–960.
[2] Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Cautier, C. and Bernardi, G. (1991) Gene 100, 181–187.
[3] Aïssani, B. and Bernardi, G. (1991) Gene 106, 185–195.
[4] Aïssani, B. and Bernardi, G. (1991) Gene 106, 173–183.
[5] Rynditch, A., Kadi, F., Geryk, J., Svoboda, J. and Bernardi, G. (1991) Gene 106, 165–172.
[6] Soriano, P., Meunier-Rotival, M. and Bernardi, G. (1983) Proc. Natl. Acad. Sci. USA 80, 1816–1820.
[7] Zerial, M., Salinas, J., Filipski, J. and Bernardi, G. (1986) Eur. J. Biochem. 160, 479–485.
[8] Bernardi, G. (1989) Annu. Rev. Genet. 23, 637–661.
[9] Tazi, J. and Bird, A.P. (1990) Cell 60, 909–920.
[10] Gardiner, K., Aïssani, B. and Bernardi, G. (1990) EMBO J. 9, 1853–1858.
[11] Dutrillaux, B. (1973) Chromosoma 41, 395–402.
[12] Ambros, P.F. and Sumner, A.T. (1987) Cytogenet. Cell Genet. 44, 223–228.
[13] Human Gene Mapping 10 (1989) Cytogenet. Cell Genet. 51, 1–1148.
[14] Thiery, J.P., Macaya, G. and Bernardi, G. (1976) J. Mol. Biol. 108, 219–235.
[15] Moyzis, R.K., Buckingham, J.M., Scott Cram, L., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) Proc. Natl. Acad. Sci. USA 85, 6622–6626.
[16] Weber, B., Collins, C., Robbins, C., Magenis, R.L., Delaney, A.D., Gray, W.J. and Hayden, M.R. (1990) Nucleic Acids Res. 11, 3353–3361.
[17] Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T. and Olson, M.V. (1989) Proc. Natl. Acad. Sci. USA 86, 6240–6244.
[18] de Lange, T., Shiue, L., Myers, R.M., Cox, D.R., Naylor, S.L., Killery, A.M. and Varmus, H.E. (1990) Mol. Cell. Biol. 10, 518–527.
[19] Sealey, P.G., Whittaker, P.A. and Southern, E.M. (1985) Nucleic Acids Res. 13, 1915–1922.
[20] Ikemura, T. and Wada, K. (1991) Nucleic Acids Res. 19, 4333–4339.
[21] Bünemann, M. and Dattagupta, N. (1973) Biochim. Biophys. Acta 331, 341–348.
[22] Cortadas, J., Macaya, G. and Bernardi, G. (1977) Eur. J. Biochem. 76, 13–19.