Gene, 122 (1992) 329–335

GENE 06831

Compositional mapping of the human dystrophin-encoding gene

(*Alu*; isochores; Duchenne Becker muscular dystrophies)

Thomas Bettecken[a,*], Brahim Aissani[b,*], Clemens R. Müller[a] and Giorgio Bernardi[b]

[a] *Institut für Humangenetik der Universität Würzburg, Biozentrum, Am Hubland, D-8700 Würzburg, Germany; and* [b] *Laboratoire de Génétique Moleculaire, Institut Jacques Monod, 75005 Paris, France*

Received by J. Weissenbach: 19 June 1992; Accepted: 4 July 1992; Received at publishers: 30 July 1992

SUMMARY

The genomes of warm-blooded vertebrates are mosaics of long DNA segments (>300 kb, on the average), the isochores, homogeneous in GC levels, which belong to a small number of compositional families. In the present work, the human dystrophin-encoding gene, spanning more than 2.3 Mb in Giemsa band Xp21 (on the short arm of the X chromosome), was analyzed in its isochore organization by hybridizing cDNA probes, corresponding to eight contiguous segments of the coding sequence, on compositional fractions from human DNA. Five DNA regions of uniform (± 0.5%) GC content, separated by compositional discontinuities of about 2% GC, were found, so providing the first high-resolution compositional map obtained for a human genome locus and the first direct estimate of isochore size (360 kb to more than 770 kb, in the locus under consideration). One of the isochores contains 71% and another one 21% of deletion breakpoints found in patients suffering from Duchenne's and Becker's muscular dystrophies.

INTRODUCTION

The genomes of warm-blooded vertebrates are mosaics of long DNA segments (over 300 kb in size, on the average), the isochores, which are compositionally homogeneous and belong to a small number of families characterized by GC levels covering a wide range (see Bernardi et al., 1985; Bernardi, 1989). GC levels of coding sequences (as well as those of the corresponding first, second and third codon positions) are linearly correlated with the GC levels of the isochores that harbor them (Bernardi et al., 1985; Bernardi, 1989; Aïssani et al., 1991).

The human dystrophin gene (see Mandel, 1989, for a review), located in Giemsa band Xp21 on the short arm of the X chromosome (Francke et al., 1985), spans more than 2.3 Mb (Burmeister et al., 1988). Many genomic probes are available and the entire coding sequence of the gene is known (Koenig et al., 1988). A 14-kb mRNA, containing a 11 058-nt coding sequence, corresponding to about 0.5% of the gene, was isolated from a human fetal muscle cDNA library (Koenig et al., 1987). The gene contains more than 75 exons and encodes dystrophin, a protein with 3685 aa residues (Koenig et al., 1988). Intron sizes cover a wide range; the two largest ones were estimated at about 170 and 250 kb (Den Dunnen et al., 1989; Blonden et al., 1991).

The gene is characterized by a very high mutation rate approx. 10^{-4}/gamete/generation). Mutations, mostly consisting of partial gene deletions (about 65% of patients) or partial duplications (up to 10%), are the major cause of

Correspondence to: Dr. G. Bernardi, Laboratoire de Génétique Moleculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France. Tel. (33–1) 43-29-58-24/25 Ext.41-01; Fax (33–1) 44-27-79-77.

* Present addresses: (T.B.) Medizinische Genetik, Kinderspital, Inselspital, CH-3010 Bern, Switzerland. Tel. (41–31) 64 95 45; (B.A.) Mental Health Research Institute, University of Michigan, 205 Zina Pitcher Place, Ann Arbor, MI 48109–0720, USA. Tel. (19–1) 313 936 2089.

Abbreviations: aa, amino acids; BAMD, bis(acetato-mercuri-methyl)dioxane; BMD, Becker muscular dystrophy; bp, base pairs; DMD, Duchenne muscular dystrophy; dystrophin gene, gene encoding dystrophin; GC, % of guanine + cytosine; kb, kilobase(s) or 1000 bp; Mb, megabase(s) or 10^6 bp; nt, nucleotide(s); r_f, molar ligand/nt ratio; TE, Tris·HCl pH 7.4/1 mM EDTA; YAC, yeast artificial chromosome.

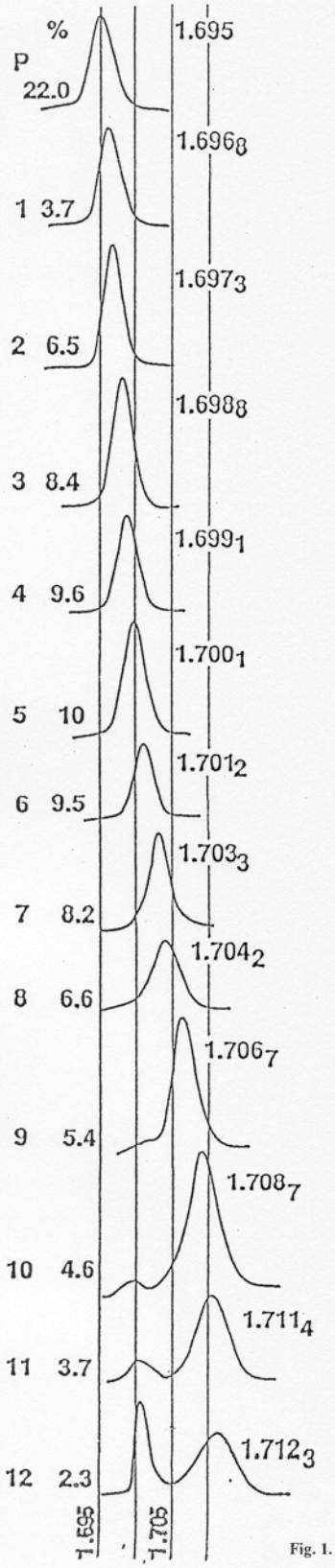

Fig. 1.

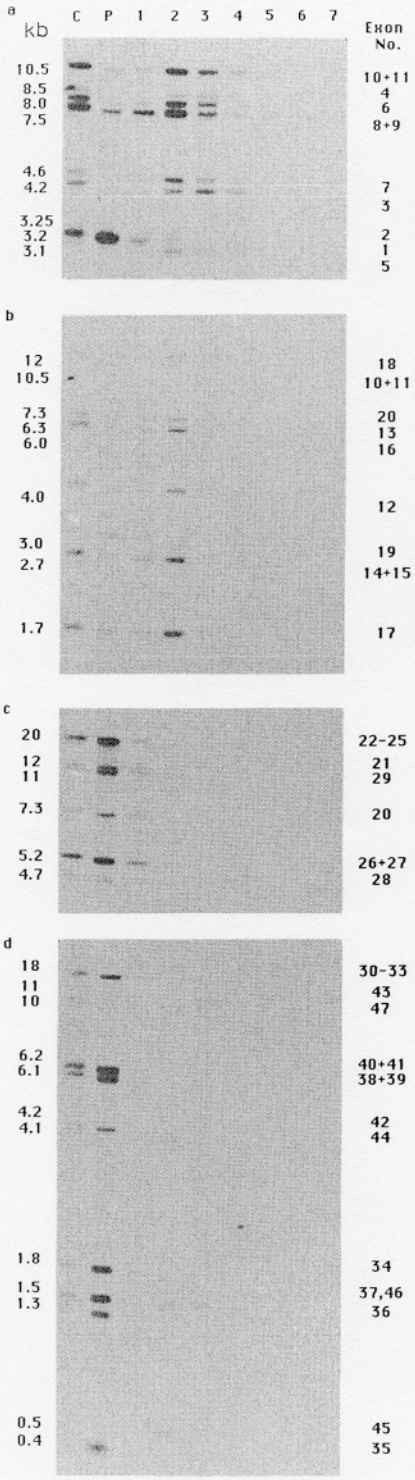

Fig. 2.

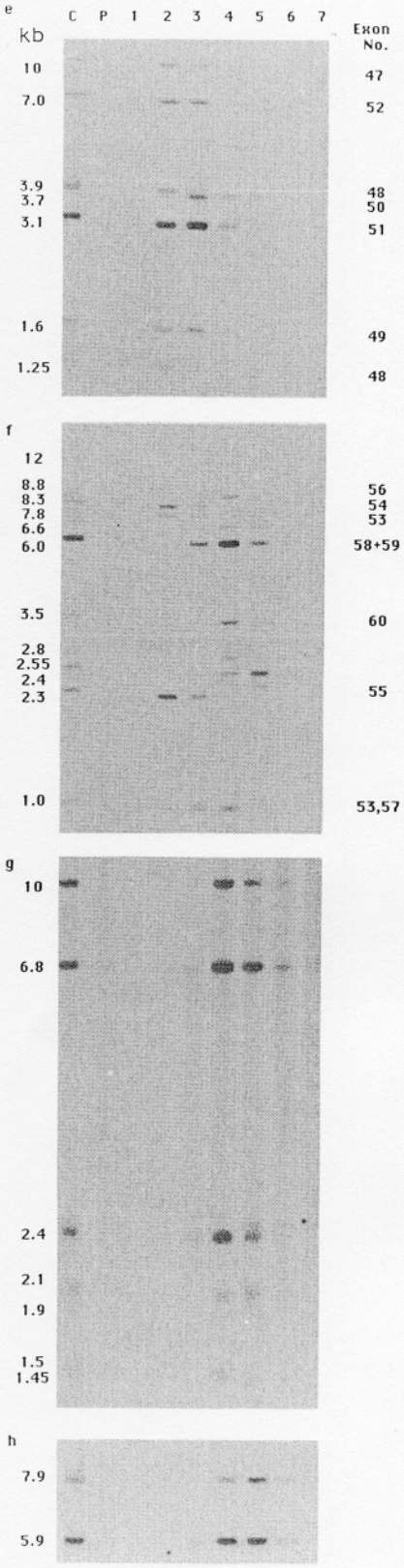

Fig. 1. CsCl profiles, relative amounts and modal buoyant densities of human DNA fractions (as obtained by preparative Cs_2SO_4/BAMD density gradient centrifugation) used to hybridize probes from different regions of the dystrophin gene. The peak appearing at about 1.700 g/ml in the last three fractions corresponds to satellite IV (Macaya et al., 1976). DNA was extracted from peripheral blood leukocytes of a healthy male, essentially as described by Miller et al. (1988). A_{260} = 8.0 units of DNA, having an average size of 50–100 kb, as determined by field inversion gel electrophoresis, were centrifuged in a Cs_2SO_4 density gradient in the presence of BAMD, a sequence-specific DNA ligand which preferentially binds to AT-rich sequences (Bünemann and

DMD and BMD, which are together the most frequent neuromuscular and X-chromosomal diseases in man. In spite of the extremely high variability of mutations with practically every family carrying its own 'private' mutation, there are regions of the gene which are preferentially affected. The mutation mechanism, as well as the reasons for the high deletion frequency and the clustering of breakpoints, are not yet understood. Duplication end points seem to be clustering in the same regions as deletion breakpoints. So far, no common features of local nt sequences around deletion breakpoints were identified. Coincidence of deletion breakpoints from unrelated patients is never exact to the base, clustering being in regions of several hundred kb. Tandem duplications are due to both homologous and non-homologous intrachromosomal recombination (unequal sister chromatid exchanges), the former involving repetitive elements, like *Alu* sequences (Hu et al., 1991).

The availability of so many probes over such a large region make the dystrophin gene an ideal object for studying the compositional organization of a large continuous segment of the X chromosome. Indeed, each exon provides a marker for the determination of its isochore context over the surrounding 100 to 200 kb.

The experimental approach used in the present work is compositional mapping (Bernardi, 1989; Gardiner et al., 1990). Wherever long-range physical maps are available, compositional maps may be constructed by assessing GC levels around landmarks localized on the physical maps that can be probed. This simply requires the hybridization of the probes on DNA fractionated according to base composition. If DNA preparations of about 100 kb in size are used, compositional mapping defines the base composition of DNA stretches of almost twice that size, 200 kb, around the landmark which was probed (Bernardi, 1989). So far, this approach was used to construct a compositional map of the long arm of human chromosome 21 (Gardiner et al., 1990).

RESULTS AND DISCUSSION

(a) The compositional map of the dystrophin gene

The analytical CsCl profiles of DNA fractions from preparative ultracentrifugation are shown in Fig. 1. As expected from the experimental conditions used, a good resolution of GC-poor fractions was obtained.

The compositional environment could be unequivocally determined for practically all exons of the human dystrophin gene. The 14 kb of the dystrophin cDNA are packaged into eight clones. Filter hybridization of *Hin*dIII digested DNA with these clones yielded about 65 bands. Exons were identified by the size of the *Hin*dIII fragments identified by the probes. Since the coding sequence corresponds to only about 0.5% of the gene, this experiment essentially provided information on the GC level of introns. Fig. 2 presents data from autoradiographs together with exon numbers, fragment sizes and numbers of *Hin*dIII fragments recognized by the probes. The 65 landmarks of all cDNA probes scattered over 2.3 Mb provided a considerable amount of overlap of the regions tested, leaving minor gaps only in the largest introns, which are located between exons 1 and 3, 7 and 8, 44 and 45.

The results (Fig. 2) show that all exons are located in DNA regions characterized by low GC levels comprised between 36.6% and 40.9%. Moreover, regions of remarkably uniform composition (± 0.5% GC) could be seen within the gene.

The promoter region, the first exon (muscle), and, probably, also exon 2 are located in a region of very low GC (36.6%). The exact hybridization maximum of exon 2 could not be determined precisely, owing to the low intensity of the signal. The data suggest, however, that exon 2 could be in a very GC-poor region, even if some signal were also found in GC-richer fractions, probably due to the fact that exon 2 is close to the border of two regions of different GC. Exon 3 showed a hybridization maximum in fraction 3,

Dattagupta, 1973; Cortadas et al., 1977). An r_f value of 0.10 was used to increase the resolution of AT-rich DNA fragments (Zerial et al., 1986). Twelve fractions and the pellet were collected from the gradient, dialyzed overnight against 10 mM Tris·HCl pH 7.4/10 mM EDTA at room temperature and exhaustively against 10 mM TE at 4°C. In another experiment, a r_f value of 0.08 was used to eliminate the formation of a pellet. DNA aliquots from the fractions were centrifuged to equilibrium in a CsCl density gradient in a Beckman Model E analytical ultracentrifuge using conditions previously described (Zerial et al., 1986). GC contents were determined from modal buoyant densities using the equation of Schildkraut et al. (1962). We estimate the accuracy of the determination of the composition to be better than 0.5% GC for exons with sufficiently strong signals. This does not apply to the pellet fraction, due to its heterogeneity.

Fig. 2. Hybridization of dystrophin cDNA probes on Cs_2SO_4/BAMD density gradient centrifugation fractions (1–7); C is unfractionated genomic DNA as a control, P is the pellet fraction. Fragment sizes allowed identification of exons. The same filters were stripped and reprobed, except for the last two clones, on the 3′-end of the gene, in which case two other filters, prepared from the same gradient were used. Proportional amounts of DNA fractions, corresponding to 20 μg of total DNA, were digested with *Hin*dIII (Boehringer Mannheim, Germany), according to the manufacturer's recommendations. After cleavage, DNA was precipitated with ethanol, redissolved in TE and electrophoresed on 1% agarose gels. DNA was then blotted onto Nylon membranes (Hybond N^+, Amersham) by direct alkali transfer and hybridized with ^{32}P-labelled DNA probes. Autoradiography was done for two to ten days. All cDNA probes used were isolated and provided by Kunkel and co-workers (Koenig et al., 1987). Intensities of bands were judged visually. Numbering of exons up to exon 60 is according to Koenig et al. (1987). Panels a–h concern hybridizations with different probes.

corresponding to a region with a GC content of 39.1%. Therefore, there is a transition in GC content of 2.4% between exons 1 and 3. Exons 4 to 18 are located in a region ranging from 38.6–38.1%. A very well defined transition, from 38.1–36.6%, was found between exons 18 and 19. This was followed by a plateau at 36.6% up to and including exon 42. DNA fragments corresponding to exons 19–42 were all found in the pellet. This made this region to appear more uniform in GC level than it really is. Another experiment, carried out at an r_f= 0.08 and avoiding pelleting altogether, indicated, however, that actual values were very close to those reported for pelleted material (points from this experiment are shown as diamonds in Fig. 3A). Exon 43 is located in a region of much higher GC content, reaching the same level as exon 3, namely 39.1%. A plateau at about 38.3–38.1% GC is followed by another increase to 39.0% near exon 49 and 39.2% at exon 50. The GC decrease to 38.4% from exons 51 to 55 is followed by an increase to about 39.9% at exons 56 to 60. On the 3′ side of exon 60 no more exon numbers have been assigned, but exons are indicated by the size of corresponding *Hin*dIII fragments. Fragments of 2.4 and 1.45 kb are located in a region of 40.9% GC, dropping again to 40.1% for several fragments, and ending at 40.5–40.9% GC. The exact number and order of these exons is still debated (Den Dunnen et al., 1989; Koenig et al., 1987). For the present work, this problem is not a serious one, however, since no compositional transitions are located in the controversial areas.

Fig. 3A provides a compositional map of the dystrophin locus. Positions of exons are taken from Den Dunnen et al. (1989) and may still be subject to refinement. This graph allows the identification of five regions, as judged by uniformity of GC content and the location of unequivocal transitions:

(1) The 5′ region of the gene, including exon 1 and prob-

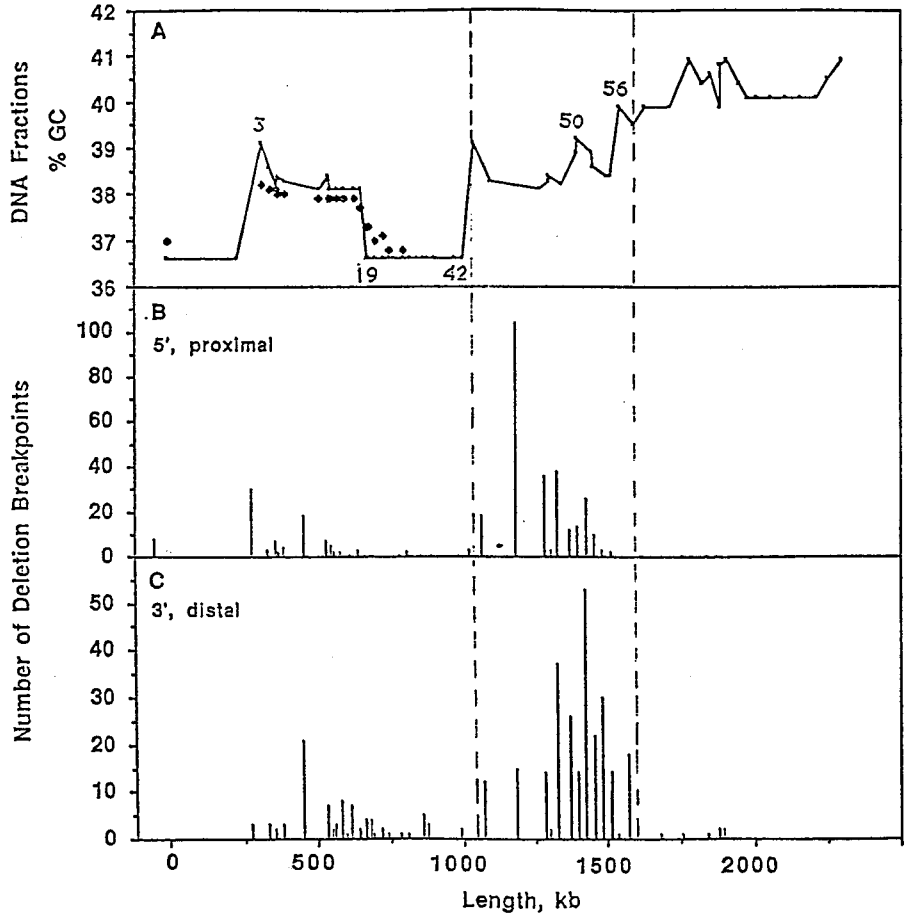

Fig. 3. Compositional map and histograms of deletion end points in human dystrophin gene. (**A**) Compositional map. Points connected by a line correspond to exons (diamonds concern data from another experiment run at r_f = 0.08, in which no pellet was present; see section **a**). The positions of exons (some of them are numbered) on the physical map of the gene are estimated according to data of Den Dunnen et al. (1989). GC content of exon 2 could not be exactly determined, and is shown as identical to that of exon 1 (see section **a**). Possible gaps exist between exons 1–3 and 44–45 due to intron size being larger than the area covered by the experimental approach (about 100–200 kb). (**B** and **C**) Histogram of proximal (5′; centromere-oriented) and distal (3′; telomere-oriented) deletion breakpoints, respectively, as determined by missing and present exons. Data points are centered between the last nondeleted exon and the first deleted exon. Deletions are from Koenig et al. (1989) and Den Dunnen et al. (1989).

ably exon 2 is undetermined in size, because no discontinuity was reached on its 5′ end; the minimal size of the region is 240 kb; (*2*) The region between exons 3 and 18 (ca. 390 kb); (*3*) The region between exons 19 and 42 (ca. 360 kb); (*4*) The region between exons 43 and 55 (ca. 510 kb); (*5*) The region starting at exon 56 and extending beyond the 3′ end of the gene (over 770 kb).

(b) A computer analysis of the dystrophin cDNAs

A computer analysis of cDNA sequences from the dystrophin gene indicates a 44.4% GC level for the coding part of the cDNA, and a 45.7% GC level for third-codon positions. Remarkably, the GC level in the noncoding part 5′ of exon 1 (before nt 209) is 38%, whereas that beyond nt position 11266 (end of translated region) until the end of the cDNA (nt position 13957) is only 33.7%. In fact, exons stand out in GC level, not only relative to these noncoding regions characterized by extremely low GC level, but also to introns (see Table I and below).

The data for average GC, exon GC, and GC in third-codon position concerning the regions defined by the compositional transitions are shown in Table I. Exon GC was systematically higher than the corresponding intron GC, as expected (Aïssani et al., 1991; D'Onofrio et al., 1991). On the other hand, third-codon position GC was very close to exon GC for low GC exons, but was higher for higher GC exons, also as expected (D'Onofrio and Bernardi, 1992). Interestingly, the only significant change to be seen this way, is the transition between exons 55 and 56 which was characterized by an exon GC change of 5% and a third-codon position GC change of about 11%.

(c) The breakpoints of dystrophin deletions

Figs. 3B and 3C present statistical compilations of 364 out of 388 breakpoints of individual partial gene deletions, leading to absent or defective dystrophin molecules and, as a consequence, to DMD and BMD. The data are compiled from Koenig et al. (1989) and Den Dunnen et al. (1989). All breakpoints were contracted into one data point which was positioned in the middle of the intron. Considering the orientation of the gene, breakpoints can be divided into two groups (Meuth, 1990). Breakpoints located on the 5′ ends of the deleted area are closer to the centromere (proximal), while breakpoints on the 3′ side of the deleted region are closer to the telomere (distal).

TABLE I

GC levels of different regions in the dystrophin gene[a]

Regions (exons)	% GC (average)	% GC (exons)	% GC (3rd codon)
3–18	38.5	42.4	43.6
19–42	36.6	42.7	43.1
43–55	38.5	44.4	42.5
56–end	40.5	49.3	53.9

[a] Computer calculations were done on nt sequences from GenBank and EMBL databanks using the ACNUC retrieval system and processing software available in the Bisance system (Deverieux et al., 1984; Dessen et al., 1990).

TABLE II

Deletion breakpoints in the dystrophin gene regions

Regions (exons)	Size (kb)	Number of breakpoints[a]		
		5′ (per kb)	3′ (per kb)	Total (%)
1–2	240	–	–	
3–18	390	85 (0.22)	66 (0.17)	151 (21)
19–42	360	9 (0.03)	20 (0.06)	29 (4)
43–55	510	269 (0.53)	250 (0.49)	519 (71)
56–end	770	4 (0.005)	25 (0.03)	29 (4)

The numbers of breakpoints in the regions defined by the compositional map, as well as their concentration per kb, are listed in Table II. The region bordered by exons 43 and 55 does have the highest concentration (71%) of breakpoints; this practically reaches 75% if the breakpoints between exons 56 and 57 are included (this region is marked by vertical broken lines in Fig. 3). This region is followed in breakpoint frequency by that between exons 3 and 18 which comprises 21% of deletions.

(d) The isochores of the dystrophin gene

The human dystrophin gene is located in a DNA region of very low GC content which corresponds to the isochore family L1, representing about 30% of human genomic DNA (Macaya et al., 1976; Cuny et al., 1981) and containing very few genes (Mouchiroud et al., 1991). The GC levels of the gene vary from 36.6% to 40.9%. Five regions of remarkably uniform (± 0.5%) GC content can be recognized; the central ones are about 390, 360 and 510 kb in size, respectively. Considering earlier work (Bernardi et al., 1985; Bernardi, 1989), these regions separated by 2% GC discontinuities can be identified with isochores. In fact, the present data provide the first direct estimate for the size of isochores, which is, in fact, well within the approximate range originally estimated (> 300 kb, on the average, in the mouse genome; Macaya et al., 1976). Moreover, these isochores provide a first example of the range of different GC levels that can be found within an isochore family. Interestingly, while the well-defined compositional transitions between exons 18 and 19 and between exons 42 and 43 do not correspond to GC differences in coding sequences, one compositional transition, between exons 55 and 56, is also reflected by the GC levels and appears in the GC levels of exons and third-codon positions (Table I). The analysis of coding sequences of mouse and chicken dystrophin genes

revealed corresponding changes in codon usage and GC content (data not shown). In this connection, it should be mentioned that Bar et al. (1990) identified an additional 6.5-kb mRNA from this region of the dystrophin gene, which is expressed in numerous tissues other than muscle and brain, in contrast to the total dystrophin mRNA.

Contrary with the dystrophin gene, a recent analysis of five YACs, ranging from 180–380 kb and covering almost 1.2 Mb around the cystic fibrosis-controlling gene (Krane et al., 1991), suggests that this 42.5% GC region is uniform (± 0.5% GC) in base composition. The presence of compositional discontinuities, although unlikely, is, however, not altogether ruled out at this point. Indeed, one should keep in mind that in the present work the genomic environments of 65 landmarks spanning 2.3 Mb were analyzed, against five regions spanning 1.2 Mb in the cystic fibrosis locus. In any case, the possible existence of isochores as long as, or longer than, 1 Mb is also suggested by the last isochore of the dystrophin gene which is larger than 770 kb. The possibility should also be considered that average isochore size could be different for GC-poor, GC-rich and GC-richest isochores.

The variety of gene defects was one of the many reasons that prompted detailed investigations of numerous DMD and BMD patients. A total of 364 well characterized deletions were defined (Koenig et al., 1989; Den Dunnen et al., 1989). The compilation of Fig. 3,B and C, summed up in Table II, shows that the concentration of deletion breakpoints is highest in two isochores. Exons are preferentially deleted from the same regions. Interestingly, only very few deletions described so far cross the compositional discontinuity located between exons 42 and 43. It should be pointed out that in the histograms of Fig. 3,B and C, breakpoints are shown in the center of introns, irrespective of their actual location. One should, therefore, not be misled by the very high concentration of breakpoints in the center of intron 44. It has, in fact, been shown (Blonden et al., 1991) that these breakpoints are well distributed over the entire intron. However, exon 45 is much more frequently deleted than exon 44, as indicated by the fact that only 13% of deletion breakpoints in intron 44 are 3′ breakpoints.

The correlation between the isochores of the dystrophin gene and chromosomal bands is not clear. It has been reported (Human Gene Mapping 11, 1991) that the gene extends over parts of two large Giemsa-positive bands (Xp21.1 and Xp21.3) separated by a thin Giemsa-negative band (Xp21.2). The latter, as well as exons 1–28, have been reported to be lacking together with several million nt 5′ of the dystrophin gene (Francke et al., 1985; Burmeister et al., 1988) in patient BB, who was instrumental in locating the gene. It is now known that, while Giemsa-positive bands appear to be homogeneous in composition and consist essentially of GC-poor isochores, Giemsa-negative bands are heterogeneous in composition and comprise both GC-poor and GC-rich isochores (Gardiner et al., 1990; G. Pilia, R. Little, B. A., G. B. and D. Schlessinger, submitted). The most heat-denaturation-resistant R-bands, the T-bands of Dutrillaux (1973) correspond to the GC-richest isochores (Gardiner et al., 1990; De Sario et al., 1991; Saccone et al., 1992).

The observed isochore pattern does not seem to be correlated with the described structural features of dystrophin (actin-binding region, triple-helical part and cysteine-rich terminus; Koenig et al., 1988). Isochores do not coincide with any of these features and GC transitions are not close to their borders.

(e) Conclusions

The compositional mapping reported here provides evidence that the dystrophin gene comprises several isochores, establishes their sizes and their GC levels, their compositional uniformity, and the range of GC levels exhibited by isochores belonging to the same family. Moreover, it quantifies the nonrandom distribution of deletion breakpoints. An analysis of the density of repetitive sequences and chromatin structure in these isochores may be useful for a better understanding of the high density of deletion breakpoints and deleted areas in these regions and, more generally, of the so-called contiguous deletion syndromes (Ballabio, 1991).

ACKNOWLEDGEMENTS

This work was made possible through grants from the Association Française contre les Maladies Musculaires (Association Française contre les Myopathies), and from the Fritz-Thyssen Stiftung. B.A. thanks Association Française contre les Myopathies for a two-year Postdoctoral Fellowship.

REFERENCES

Aïssani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K., Gautier, C. and Bernardi, G.: The compositional properties of human genes. J. Mol. Evol. 32 (1991) 497–503.

Ballabio, A.: The contiguous deletion syndrome. Curr. Opin. Genet. Dev. 1 (1991) 25–29.

Bar, S., Barnea, E., Levy, Z., Neuman, S., Yaffe, D. and Nudel, U.: A novel product of the Duchenne muscular dystrophy gene which greatly differs from the known isoforms in its structure and tissue distribution. Biochem. J. 272 (1990) 557–560.

Bernardi, G.: The isochore organization of the human genome. Annu. Rev. Genet. 23 (1989) 637–661.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F.: The mosaic genome of warm-blooded vertebrates. Science 228 (1985) 953–958.

Blonden, L.A., Grootscholten, P.M., Den Dunnen, J.T., Baker, E., Abbs, S., Bobrow, M., Boehm, C., Van Broeckhoven, C., Baumbach, L. and Chamberlain, J.: 242 breakpoints in the 200-kb deletion-prone P20 region of the DMD gene are widely spread. Genomics 10 (1991) 631–639.

Bünemann, M. and Dattagupta, N.: On the binding and specificity of 3,6-bis-(acetylo-mercurimethyl) dioxane to DNAs of different base compositions. Biochim. Biophys. Acta 331 (1973) 341–348.

Burmeister, M., Monaco, A.P., Gillard, E.F., Van Ommen, G.J.B., Ferguson-Smith, M.A., Kunkel, L.M. and Lehrach, H.: A 10-megabase physical map of human Xp21, including the Duchenne muscular dystrophy gene. Genomics 2 (1988) 189–202.

Cortadas, J., Macaya, G. and Bernardi, G.: An analysis of the bovine genome by density gradient centrifugation: fractionation in Cs_2SO_4/3,6 bis-(acetato-mercurimethyl) dioxane density gradient. Eur. J. Biochem. 76 (1977) 13–19.

Cuny, G., Soriano, P., Macaya, G. and Bernardi, G.: The major components of the mouse and human genomes: preparation, basic properties and compositional heterogeneity. Eur. J. Biochem. 111 (1981) 227–233.

Den Dunnen, J.T., Grootscholten, P.M., Baker, E., Blonden, L.A.J., Ginjaar, H.B., Wapenaar, M.C., Van Paassen, H.M.B., Van Broeckhoven, C., Pearson, P.L. and Van Ommen, G.J.B.: Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. Am. J. Hum. Genet. 45 (1989) 835–847.

De Sario A., Aïssani, B. and Bernardi, G.: Compositional properties of telomeric regions from human chromosomes. FEBS Lett. 295 (1991) 22–26.

Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C.: Bisance: a French service for access to biomolecular sequence data bases. Cabios 6 (1990) 355–356.

Devereux, J., Haeberli, P. and Smithies, O.: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 (1984) 387–395.

D'Onofrio, G., Mouchiroud, D., Aïssani, B., Gautier, C. and Bernardi, G.: Correlations between the compositional properties of human genes, codon usage and aminoacid composition of proteins. J. Mol. Evol. 32 (1991) 504–510.

D'Onofrio, G. and Bernardi, G.: A universal compositional correlation among codon positions. Gene 110 (1992) 81–88.

Francke, U., Ochs, H.D., de Martinville, B., Giacalone, J., Lindgren, V., Disteche, C., Pagon, R.A., Hofker, M.H., Van Ommen, G.J.B., Pearson, P.L. and Wedgwood, R.J.: Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. Am. J. Hum. Genet. 37 (1985) 250–267.

Gardiner, K., Aissani, B. and Bernardi, G.: A compositional map of human chromosome 21. EMBO J. 9 (1990) 1853–1858.

Hu, X., Ray, P.N. and Worton, R.G.: Mechanisms of tandem duplication in the Duchenne muscular dystrophy gene include both homologous and non-homologous intrachromosomal recombination. EMBO J. 10 (1991) 2471–2477.

Human Gene Mapping 11. Cytogenet. Cell Genet. 58 (1991) 1–2000.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C. and Kunkel, L.M.: Complete cloning of the Duchenne Muscular Dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50 (1987) 509–517.

Koenig, M., Monaco, A.P. and Kunkel, L.M.: The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell 53 (1988) 219–228.

Koenig, M. and 35 co-authors: The molecular basis for Duchenne *versus* Becker muscular dystrophy: correlation of severity with type of deletion. Am. J. Hum. Genet. 45 (1989) 498–506.

Krane, D.E., Hartl, D.L. and Ochman, H.: Rapid determination of nucleotide content and its application to the study of genome structure. Nucleic Acids Res. 19 (1991) 5181–5185.

Macaya, G., Thiery, J.P. and Bernardi, G.: An approach to the organization of eukaryotic genomes at a macromolecular level. J. Mol. Biol. 108 (1976) 237–254.

Macaya, G., Thiery, J.P. and Bernardi, G.: In: Yunis, J.J. (Ed.), Molecular Structure of Human Chromosomes. Academic Press, New York, 1977, pp. 35–58.

Mandel, J.-L.: The gene and its product. Nature 339 (1989) 584–586.

Meuth, M.: The structure of mutation in mammalian cells. Biochim. Biophys. Acta 1032 (1990) 1–17.

Miller, S.A., Dykes, D.D. and Polesky, H.F.: A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16 (1988) 1215.

Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Gautier, C. and Bernardi, G.: The distribution of genes in the human genome. Gene 100 (1991) 181–187.

Saccone, S., De Sario, A., Della Valle, G. and Bernardi, G.: The highest gene concentrations in the human genome are in T-bands of metaphase chromosomes. Proc. Natl. Acad. Sci. USA 89 (1992) 4913–4917.

Schildkraut, C.L., Marmur, J. and Doty, P.: Determination of the base composition of DNA from its buoyant density in CsCl. J. Mol. Biol. 4 (1962) 430–443.

Zérial M., Salinas J., Filipski J. and Bernardi G.: Gene distribution and nucleotide sequence organization in the human genome. Eur. J. Biochem. 160 (1986) 479–485.