The EMBO Journal vol.7 no.8 pp.2377-2385, 1988

A physical map of the human pseudoautosomal region

William R.A.Brown

Biochemistry Department, Oxford University, South Parks Road, Oxford OX1 3QU, UK

Communicated by A.P.Bird

A physical map of the human pseudoautosomal region has been constructed using pulsed field gel electrophoresis and the infrequently cutting restriction enzymes *Bss*HII, *Eag*I, *Sst*II, *Not*I, *Mlu*I and *Nru*I. This map extends 2.3 Mbp from the telomere to sex-chromosome-specific DNA, includes at least seven CpG islands and locates four genetically mapped loci. Five of the CpG islands are organized into two clusters. One cluster is adjacent to the telomere, the other extends into sex-chromosome-specific DNA. There is congruence between the genetic and physical maps which implies that the frequency of recombination is approximately uniform throughout the DNA.

Key words: sex chromosomes/long range mapping/telomere

Introduction

Mammalian X and Y chromosomes differ but they pair during male meiosis and exchange DNA sequences (Cooke *et al.*, 1985; Simmler *et al.*, 1985; Harbers *et al.*, 1986; Keitges *et al.*, 1987). The exchange is normally limited to a short region of homology which in humans is located at the tips of the short arms of both chromosomes. Sequences within this stretch of DNA show reduced six linkage or pseudoautosomal behaviour (Burgoyne, 1982). The human pseudoautosomal region is rich in hypervariable minisatellite sequences (Rouyer *et al.*, 1986) and many pseudoautosomal sequences detect informative restriction fragment length polymorphisms. These have been used to analyse the segregation of pseudoautosomal loci in male meioses and to produce a genetic map of the DNA (Figure 1) (Rouyer *et al.*, 1986). This analysis demonstrates that a single recombination event occurs in the pseudoautosomal region in each productive male meiosis and that this recombination event can occur in at least four places. These features are reflected in a gradient of sex linkage with the most centromere proximal pseudoautosomal locus segregating with sex in >90% of meioses and the most telomeric locus showing no sex linkage (Figure 1). The pseudoautosomal region is short. A rate of one recombination event every meiosis is much higher than would be expected in a similar length of autosomal DNA and 10- to 20-fold higher than is observed in the pseudoautosomal region in female meiosis (Rouyer *et al.*, 1986). This elevated rate of recombination makes the pseudoautosomal region a good place in which to identify the sequence changes which accompany such events.

One gene has been mapped to the human pseudoautosomal region (Goodfellow *et al.*, 1986). This is MIC2 which encodes a cell surface antigen. A second pseudoautosomal locus, XGR (Goodfellow *et al.*, 1987), has been postulated to regulate the expression of MIC2 and of the Xg blood group locus. Indirect evidence suggests that there are other pseudoautosomal genes. Women with a 45,X karyotype (Ford *et al.*, 1959) suffer Turner's syndrome (Turner, 1938). The main features of this disorder are short stature, infantilism, streak gonads and a set of characteristic congenital malformations. Monosomy of X and Y homologous loci is thought to be the cause of Turner's syndrome and correlations between phenotype and karyotype (Ferguson-Smith, 1965) indicate that these genes may be in the pseudoautosomal region. The diversity of the features which characterize Turner's syndrome suggests that there are several of these genes.

Progress towards understanding recombination within the pseudoautosomal region and towards locating pseudoautosomal genes requires a physical map of the DNA. With the advent of pulsed field gel electrophoresis (Schwartz and Cantor, 1984) it is now possible to construct such a map. Long-range maps of subchromosomal regions of mammalian DNA are of broad interest in the context of plans to map physically entire mammalian chromosomes. Such maps are particularly valuable if they include biologically significant landmarks. The bulk of mammalian DNA is A+T rich, depleted for the dinucleotide CpG (Swartz *et al.*, 1962) and methylated at cytosine in the sequence CpG. Methylation sensitive restriction endonucleases with recognition sequences composed exclusively of CG interstrand basepairs and including two CpG dinucleotides (here termed CG enzymes) therefore cut mammalian DNA infrequently. Within mammalian DNA are short tracts of sequence, usually 1-2 kb long, which are relatively C+G rich, show no CpG suppression and which are hypomethylated (Bird *et al.*, 1985). These sequence tracts have been termed CpG islands. There are ~30 000 CpG islands in the haploid mammalian genome and many if not all of these are likely to be associated with genes (Bird, 1986; Gardiner-Garden and Frommer, 1987). The sites for CG enzymes are

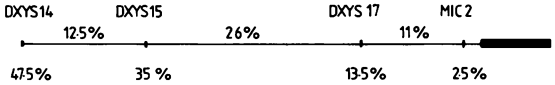

Fig. 1. Genetic map of the human pseudoautosomal region. The line represents the pseudoautosomal region and the bar represents sex-chromosome-specific DNA. Figures above the line are the percentage of male meioses in which a recombination event is observed between individual pairs of loci. Figures below the line are the fraction of male meioses in which a recombination event is observed between an individual locus and the sex-chromosome-specific DNA. This map is a summary of data in Weissenbach *et al.* (1987) and Goodfellow *et al.* (1986).

clustered in CpG islands. Of the total sites for a CG enzyme with a 6-bp recognition sequence 75% are calculated to be in CpG islands, and on average every island will contain one site for each enzyme. CpG islands are thus informative and easily detected landmarks for long-range maps (Brown and Bird, 1986).

This paper describes a long-range map of the human pseudoautosomal region in which the telomere and CpG islands are used as landmarks. The map suggests the position of unidentified pseudoautosomal genes and indicates that recombination sites are not clustered.

Results

Strategy

The pseudoautosomal region includes a telomere which as a natural end enables the DNA to be mapped by a combination of partial restriction endonuclease digestion and indirect end labelling. This is a particularly efficient strategy as it enables many sites to be placed in a single experiment. Putting it into practice required a choice of restriction enzymes and of DNA, suitable probes and a pulsed field gel electrophoresis system. In order to localize CpG islands I mapped sites for the CG enzymes *Bss*HII, *Eag*I, *Sst*II and *Not*I and to provide structural information outside CpG islands I also mapped sites for the enzymes *Mlu*I and *Nru*I. The recognition sequences of all of these enzymes include two CpG dinucleotides and so they will be infrequent in mammalian DNA. However, the recognition sequences of *Mlu*I and *Nru*I also include two AT interstrand basepairs and as a consequence 73% of their sites can be calculated to occur outside CpG islands. Like the CG enzymes these enzymes are susceptible to inhibition of cleavage by cytosine methylation and so many of the sites which lie between islands will not be completely cuttable. Two sequences were used as probes for indirect end labelling. A minisatellite sequence at the DXYS20 locus was used as a probe in almost all of the experiments. This was a particularly sensitive probe as the DXYS20 locus is >20 kb long. The distal end of this sequence is ~30 kb from the telomere although this distance can vary due to polymorphism (Cooke *et al.*, 1985). There are no sites for *Bss*HII, *Not*I, *Mlu*I or *Nru*I distal of the DXYS20 locus. Just telomeric of the DXYS20 locus, however, is a CpG island (Cooke *et al.*, 1985) which includes *Sst*II and *Eag*I sites. This island is unmethylated in some DNA samples and this precluded the use of the DXYS20 probe for the detailed mapping of these DNA samples. For such experiments I used the shorter DXYS14 probe (Cooke *et al.*, 1985) which lies on the telomeric side of this island. The DXYS14 and DXYS20 loci are only 12 kb from one another, have not been separated by recombination and can therefore be regarded as coincident on the genetic map (Figure 1). It seemed possible that the pseudoautosomal region would be polymorphic over long distances. In order to avoid any confusion that such polymorphism would introduce into the interpretation of the mapping experiments I initially mapped restriction sites in the DNA of three individual sex chromosomes; in two Y chromosomes isolated in the hybrid somatic cells 3E7 (Marcus *et al.*, 1976) and 853 (Burk *et al.*, 1985) and an X chromosome in the peripheral blood leukocytes of a woman with Turner's syndrome (45,X karyotype). The restriction site maps of the pseudoautosomal regions in these chromosomes were

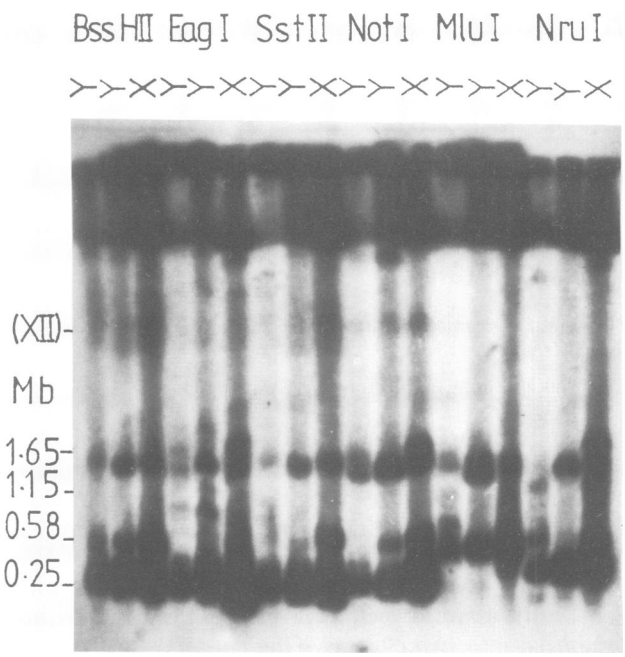

Fig. 2. Low resolution restriction site mapping of the human pseudoautosomal region. High mol. wt DNA isolated from the hybrid somatic cells 3E7 and 853 and from the blood of a woman with Turner's syndrome was partially restricted with *Bss*HII (0.05 U), *Eag*I (1 U), *Sst*I (5 U), *Not*I (2 U), *Mlu*I (1 U) and *Nru*I (1 U). The partial digests were then size fractionated by pulsed field gel electrophoresis, filter transferred and analysed by hybridization with the DXYS20 probe. The electrophoresis was for 14 days in a gel of 0.8% agarose, electric field was 1 V/cm, buffer temperature was 4°C and pulse time was 30 min. The letters Y or X above the illustration refer to the sex chromosome constitution of the DNA analysed in the respective track. DNA in each group of three tracks was respectively from 3E7 cells (Y), 853 cells (Y) or Turner's peripheral blood leukocytes (X). Size markers were the chromosomes of *S.cerevisiae* strain X2180 1B. The XII in brackets to the left of the illustration refers to chromosome XII which has not been accurately sized in this strain.

different but these differences were due to differences in DNA methylation. This observation allowed more thorough mapping of DNA from cells which included two or more sex chromosomes. The original pulsed field gel electrophoresis system gave non-uniform rates of DNA migration across the gel (Schwartz and Cantor, 1984) and is therefore of limited value as a mapping tool. I therefore used the rotating plate gel electrophoresis system (Southern *et al.*, 1987), which gives uniform migration across the width of the gel and allows sizes to be determined accurately and differences between the maps of different chromosomes to be detected with confidence.

Overall organization

The overall organization of the pseudoautosomal region is illustrated by the experiment shown in Figure 2. DNA from the two Y chromosome hybrids and from the woman with Turner's syndrome was partially restricted with each of the CG enzymes and with the other two infrequently cutting restriction enzymes, *Mlu*I and *Nru*I, size fractionated by pulsed field gel electrophoresis so that molecules as big as 2.5 Mb were resolved, filter transferred and analysed with the DXYS20 probe. This experiment suggests that the pseudoautosomal region can be considered in three

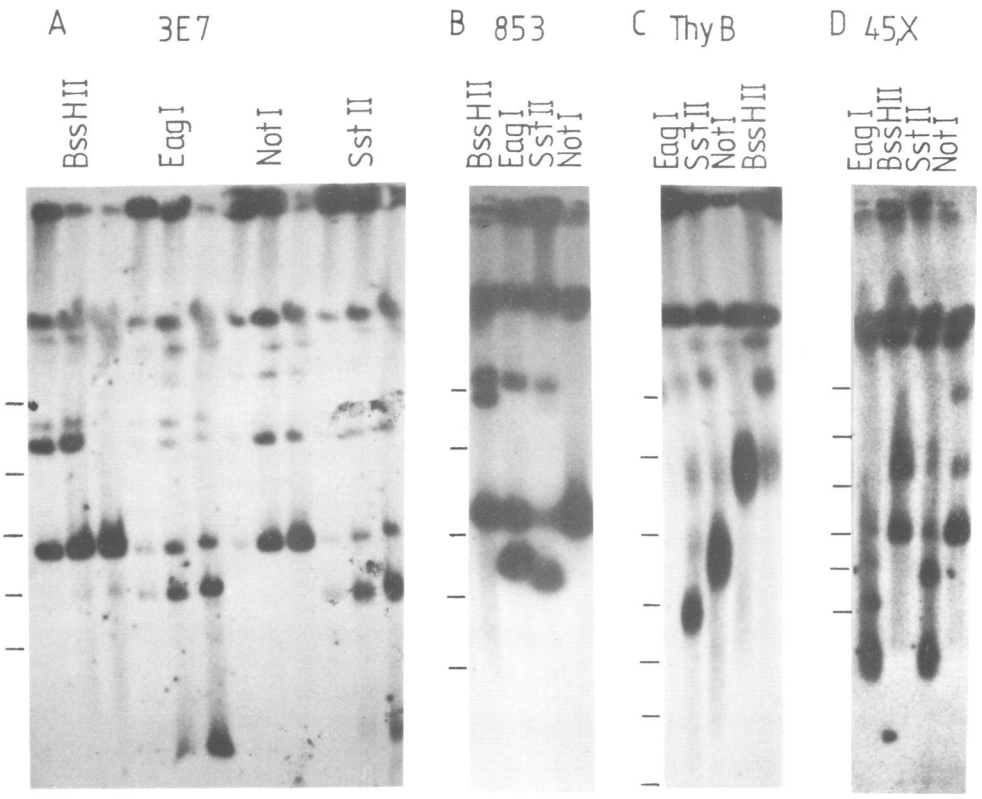

Fig. 3. Sites for *Bss*HII, *Eag*I, *Sst*II and *Not*I in the proterminal subregion. High mol. wt DNA from the hybrid somatic cells 3E7 (**A**), 853 (**B**), ThyB1-33/12 (**C**) or from the blood of a woman with Turner's syndrome (**D**) was partially restricted with the indicated enzyme, size fractionated by pulsed field gel electrophoresis, filter transferred and probed with either the DXYS14 probe (**A** and **D**) or with the DXYS20 probe (**B** and **C**). The 3E7 DNA was partially digested by incubation for 10, 50 or 250 min with either 4 U of *Bss*HII, 5 U of *Eag*I, 12 U of *Sst*II or 10 U of *Not*I. The other DNA samples were incubated for 2 h with either 0.2 U of *Bss*HII, 5 U of *Eag*I, 1.25 U of *Sst*I or 10 U of *Not*I. Gels were run with pulse times and run times of 11 s, 36 h (**A**), 10 s, 36 h (**B** and **D**) or 15 s, 40 h (**C**). Size markers were oligomers of bacteriophage lambda cI857 DNA.

subregions. The proterminal subregion corresponds to a cluster of sites for each of the four CG enzymes and is represented by a strong band of hybridization in the first 12 tracks towards the bottom of the autoradiograph. The centromere distal or more simply the distal subregion lies between the proterminal subregion and the cluster of sites for all six enzymes located ~1.3 Mb from the chromosome end. Beyond this cluster of sites for all six enzymes lies the proximal subregion. The boundary between the proximal subregion and chromosome-specific parts of the X and Y chromosomes is suggested at the top of the resolving part of the gel. A *Not*I site is present in both of the Y chromosomes (it is rather weak in the 3E7 track) but not in the X chromosome at a position just corresponding to the resolution limit of this particular gel. The filter transfer analysed in the experiment shown in Figure 2 corresponds to a gel which was run in such a way as to display the entire pseudoautosomal region. This involved a sacrifice of resolution. I therefore analysed the individual subregions in more detail. This was done in two complimentary ways. The proterminal and distal subregions were mapped by the indirect end-labelling approach but the gels were run with shorter pulse times than the one shown in Figure 2. In order to check these results and in order to define the boundary of the pseudoautosomal region unambiguously, more complete single and double digests were analysed with genetically mapped single copy probes.

The proterminal subregion

The proterminal subregion in each of the three chromosomes described above and in the X chromosome isolated in the hybrid ThyB1-33/12 (Lund *et al.*, 1983) was mapped after partial restriction enzyme digestion with the four CG enzymes, size fractionation and filter hybridization with either the DXYS14 or DXYS20 telomeric probes (Figure 3). The proterminal subregion is rich in sites for these CG enzymes; however, the four DNA samples differ in the number and distribution of cuttable sites. Furthermore the fragments detected in the partial digests of the ThyB1-33/12 DNA differ from those seen in the other digests in that they are smeared. This smearing is not artefactual as other fragments in the same digests can be seen as sharp bands (not shown). The pseudoautosomal telomere, like the telomeres of lower eukaryotes is heterogeneous in size (Cooke *et al.*, 1985). The heterogeneity at the pseudoautosomal telomere in the 3E7, 853 or Turner's DNA is, however, too small to be detectable in the experiments shown in Figure 3, but the heterogeneity at the pseudoautosomal telomere in the ThyB1-33/12 is more extensive. The differences between the four DNA samples in the number and distribution of sites in the proterminal subregion could be caused by differences between the DNA sequences or by differences between the methylation patterns of similar DNA sequences. In order to resolve this point each of the four different types of DNA were partially restricted with *Sca*I and analysed with the

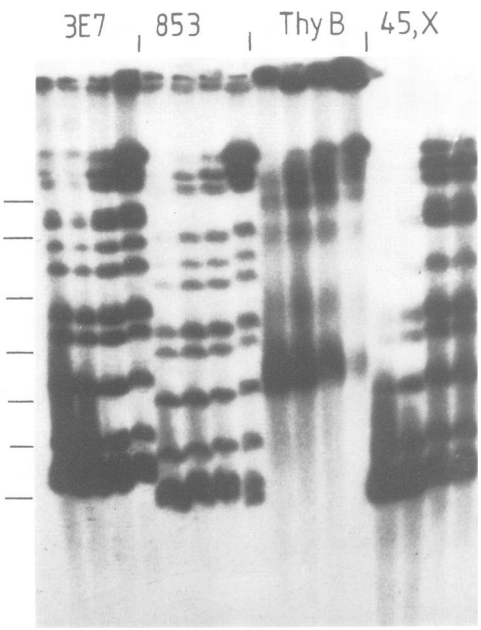

Fig. 4. Partial digestion of 3E7, 853, ThyB1-33/12 and 45,X DNA with *Sca*I. High mol. wt. DNA from the hybrid somatic cells 3E7, 853, ThyB1-33/12 or from the blood of a woman with Turner's syndrome (45,X) was incubated with 8 U, 4 U, 1.6 U or 0.8 U of *Sca*I for 5 h loaded into successive slots, size fractionated by pulsed field gel electrophoresis (14 s pulse time), filter transferred and analysed by hybridization with the DXYS20 probe. Size markers were oligomers of bacteriophage lambda DNA.

DXYS20 probe (Figure 4). *Sca*I is a restriction endonuclease whose recognition sequence does not include CpG and is thus not liable to be inhibited from cleaving mammalian DNA by cytosine methylation. Furthermore and of particular relevance to the experiment shown in Figure 4 there are no *Sca*I sites telomeric of the DXYS20 sequence. The results of this experiment demonstrate that the proterminal subregion in the four DNA samples differ both in their sequence and their pattern of DNA methylation. The *Sca*I fingerprints of the 3E7, 853 and Turner's DNA differ little, however, which suggests that the differences in the distribution of CG enzyme sites between these DNA samples are due to differences in DNA methylation. The small differences between the *Sca*I fingerprints of these DNA samples are consistent with small differences between the amounts of DNA at the very ends of the chromosomes and with the high density of minisatellite sequences in this part of the genome. The *Sca*I fingerprint of the ThyB1-33/12 DNA is hard to interpret because of the terminal heterogeneity. The distribution of fragment sizes, however, suggests that the pseudoautosomal region in the ThyB1-33/12 DNA differs from that in the three other DNA samples both in that an extra 100 kb of DNA is present at the telomere and that there are differences between the methylation patterns of the ThyB1-33/12 DNA and the three other DNA samples.

The distal subregion

The distal subregions of the four individual sex chromosomes were mapped by a combination of partial restriction enzyme

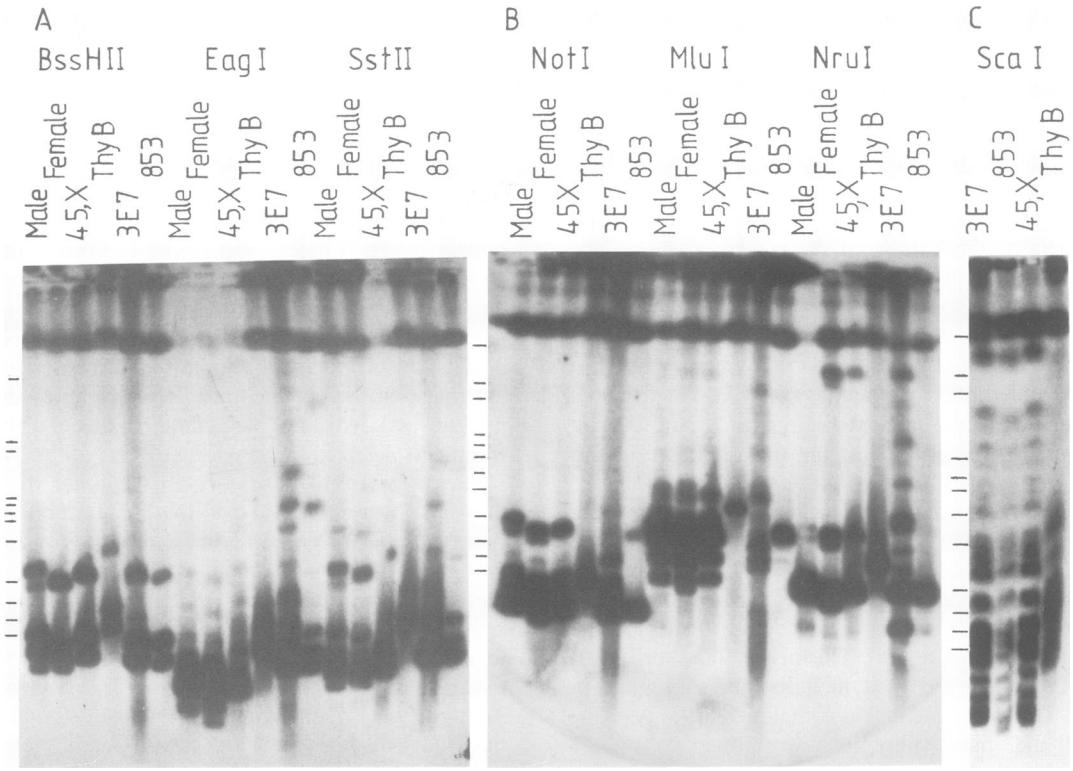

Fig. 5. Restriction site mapping of the distal subregion. (**A**) and (**B**) High mol. wt DNA from the blood of a man, a woman and a patient with Turner's syndrome or from cells of the hybrid somatic lines 3E7, 853 and ThyB1-33/12 containing single human sex chromosomes was partially restricted with *Bss*HII (0.1 U), *Eag*I (1 U), *Sst*II (5 U), *Not*I (5 U), *Mlu*I (5 U) or *Nru*I (2 U), size fractionated by pulsed field gel electrophoresis (75 s pulse time), filter transferred and analysed by hybridization with a DXYS20 probe. (**C**) High mol. wt DNA from the blood of a woman with Turner's syndrome or from the hybrid somatic cells 3E7, 853 or ThyB1-33/12 was partially restricted with *Sca*I (2 U), size fractionated by pulsed field gel electrophoresis (60 s pulse time) and analysed by filter hybridization with the DXYS20 probe. Size markers were chromosomes of *S.cerevisiae* strain X2180 1B.

digestion and indirect end labelling (Figures 5A, B and 8A). This experimental approach was also applied to DNA isolated from the blood of a normal man or a normal woman. The patterns of fragments in the six samples in any one set of partial digests shown in Figure 5 differ from one another. The differences between the male, female or Turner's DNA are small and suggest that the pseudoautosomal region does not show extensive long-range hypervariability. The fragments in the partial *Mlu*I or *Nru*I digests of the Turner's DNA, however, differed markedly from those in the partial *Mlu*I or *Nru*I digests of the DNA from the three hybrid somatic cells which also differed from one another. It seemed probable that these differences were due mainly to differences between the methylation of the peripheral blood leukocyte DNA and of the hybrid somatic cell DNA. This interpretation was confirmed by the analysis of *Sca*I partial digests (Figure 5C) with the DXYS20 sequence. The *Sca*I fingerprints of the 3E7, 853 and Turner's DNA were very similar. The fingerprint of the ThyB1-33/12 DNA was difficult to interpret due to terminal heterogeneity but was consistent with those of the other three cell types.

The locus DXYS15 recombines with the DXYS14 locus in 13.5% of meiosis (Figure 1). This corresponds to about a quarter of the genetic length of the pseudoautosomal region and suggested that the DXYS15 locus might be in the distal subregion. Hybridization analysis of limit digests with a DXYS15 sequence confirmed this (Figure 6). The 3E7, 853 and Turner's DNA all gave cognate fragments with sizes

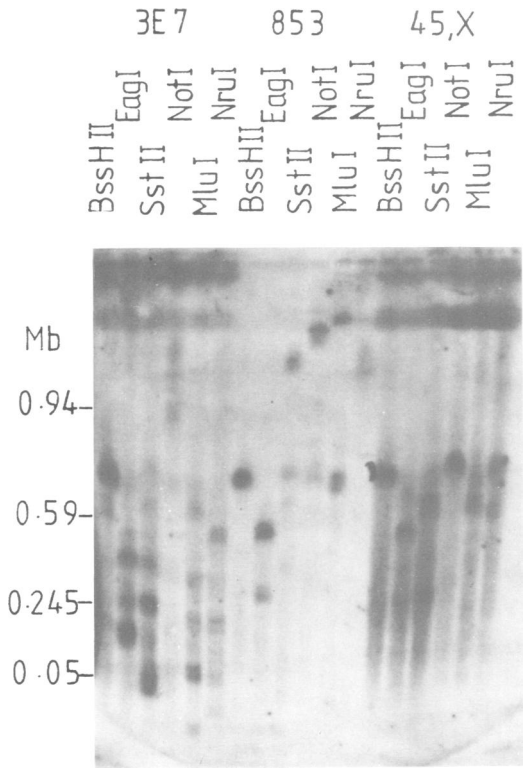

Fig. 6. Positioning the DXYS15 locus in the centromere distal subregion. DNA from either 3E7 cells, 853 cells or from the blood of a woman with Turner's syndrome was restricted with *Bss*HII (4 U), *Eag*I (15 U), *Sst*II (15 U), *Not*I (15 U), *Mlu*I (16 U) or *Nru*I (16 U), size fractionated by pulsed field gel electrophoresis (65 s pulse time), filter transferred and analysed by hybridization with the DXYS15 probe 113D. Size markers were chromosomes of *S.cerevisiae* strain X2180 1B.

which, on the whole, were consistent with the maps of the distal subregion predicted by partial digestion and indirect end labelling. There were some inconsistencies between the two sets of data. One example is the presence of small cognate *Sst*II fragments and a 0.76-Mb cognate *Bss*HII fragment in the 3E7 DNA analysed with the DXYS15 probe. The *Bss*HII fragment is presumed to extend from ~0.51 to 1.23 Mb (Figure 8A) but there are no closely spaced *Sst*II sites in this stretch of DNA. This and other similar inconsistencies could be resolved by a re-examination of the partial digest data. These revealed very weakly cut *Sst*II sites at several places between 0.51 and 1.23 Mb in the 3E7 DNA. Sites which are extensively cleaved in limit digests but poorly cleaved in partial digests are to be anticipated given the extensive kinetic heterogeneity of CG enzyme sites (New England Biolabs, 1986). Despite this sort of problem the sizes of the cognate fragments in Figure 6 suggested that the DXYS15 locus was 0.65 Mb from the telomere.

Two additional experiments (not shown) confirmed that the maps shown in Figure 8A were correct and that the DNA in peripheral blood showed no long-range restriction site hypervariability. Firstly four additional sequences were isolated from the distal subregion and used to analyse complete digests of 3E7, 853 and Turner's DNA. They revealed a collection of fragments consistent with the maps shown in Figure 8A. All of the sequences were located in the vicinity of the DXYS15 locus. Secondly DNA samples from nine indivduals were digested to completion with *Bss*HII and analysed by filter hybridization with one of these four sequences. All contained a single cognate fragment of 0.76 Mb.

The proximal subregion and the boundary with sex-chromosome-specific DNA

The proximal subregion lies furthest from the telomere and so was not amenable to detailed mapping by partial digestion and indirect end labelling. However, the data in Figure 2 suggested a map of the proximal subregion and this was checked by analysing single and double restriction endonuclease digests with sequences detecting genetically mapped loci. Two such loci were important in this analysis. They were the MIC2 gene which lies ~100 kb from the boundary with sex-chromosome-specific DNA (Pritchard *et al.*, 1987) and the more distal DXYS17 locus (Rouyer *et al.*, 1986) which recombines with the MIC2 locus in ~11% of meioses (Figure 1). The DNA samples analysed were isolated from the blood of a normal man, a normal woman, a woman with Turner's syndrome, from the lymphoblastoid lines OXEN (karyotype 49,XYYYY) and GM1416 (karyotype 48,XXXX) and from the hybrid somatic cell line 853. Chromosome-specific fragments were assigned on the basis of dosage and segregation in the Turner's and 853 DNA. Figure 2 suggests that the proximal subregion lacks *Nru*I sites and includes a single *Not*I site. In agreement with this prediction the MIC2 probe detected sex-chromosome-specific *Not*I and *Nru*I fragments. The Y chromosome *Not*I fragment is 0.3 Mb long, the X chromosome *Not*I fragment is 1.2 Mb long and the X chromosome *Nru*I fragment is 1.2 Mb long (Figure 7A). The MIC2 probe detected a Y-chromosome-specific 1.0-Mb *Nru*I fragment in the male and in the OXEN DNA but not in the 853 DNA. This is consistent with the data of Figure 2 which indicate the presence of an *Nru*I site in the sex-chromosome-specific

part of the 3E7 DNA but not in the 853 DNA. It seems likely that this site is cuttable in the male, OXEN and 3E7 DNA but is refractory to cleavage in the 853 DNA due to methylation of the enzyme recognition site. The suggestion that the 853 DNA is hypermethylated at this *Nru*I site is consistent with the map of this DNA elsewhere in the pseudoautosomal region. The DXYS17 probe (Figure 7B) evidently recognizes some sequences which are repeated in the genome but as predicted the locus-specific *Nru*I fragments are the same size as those detected by the MIC2 probe while the locus-specific *Not*I fragments are common to the sex chromosomes. These fragments were 0.8 Mb long (Figure 7B). This result places the DXYS17 locus distal of the *Not*I site in the proximal subregion. Analysis of double digests with each of the two probes located *Not*I and *Nru*I sites with respect to one another and yielded a map (Figure 8B) consistent with the data of Figure 2. In the analysis shown in Figure 7B no locus-specific fragments are detectable in the tracks containing 853 DNA. This was observed in two other experiments and is possibly a consequence of the short size of the DXYS17 probe.

The partial digests with *Bss*HII, *Eag*I and *Sst*II analysed

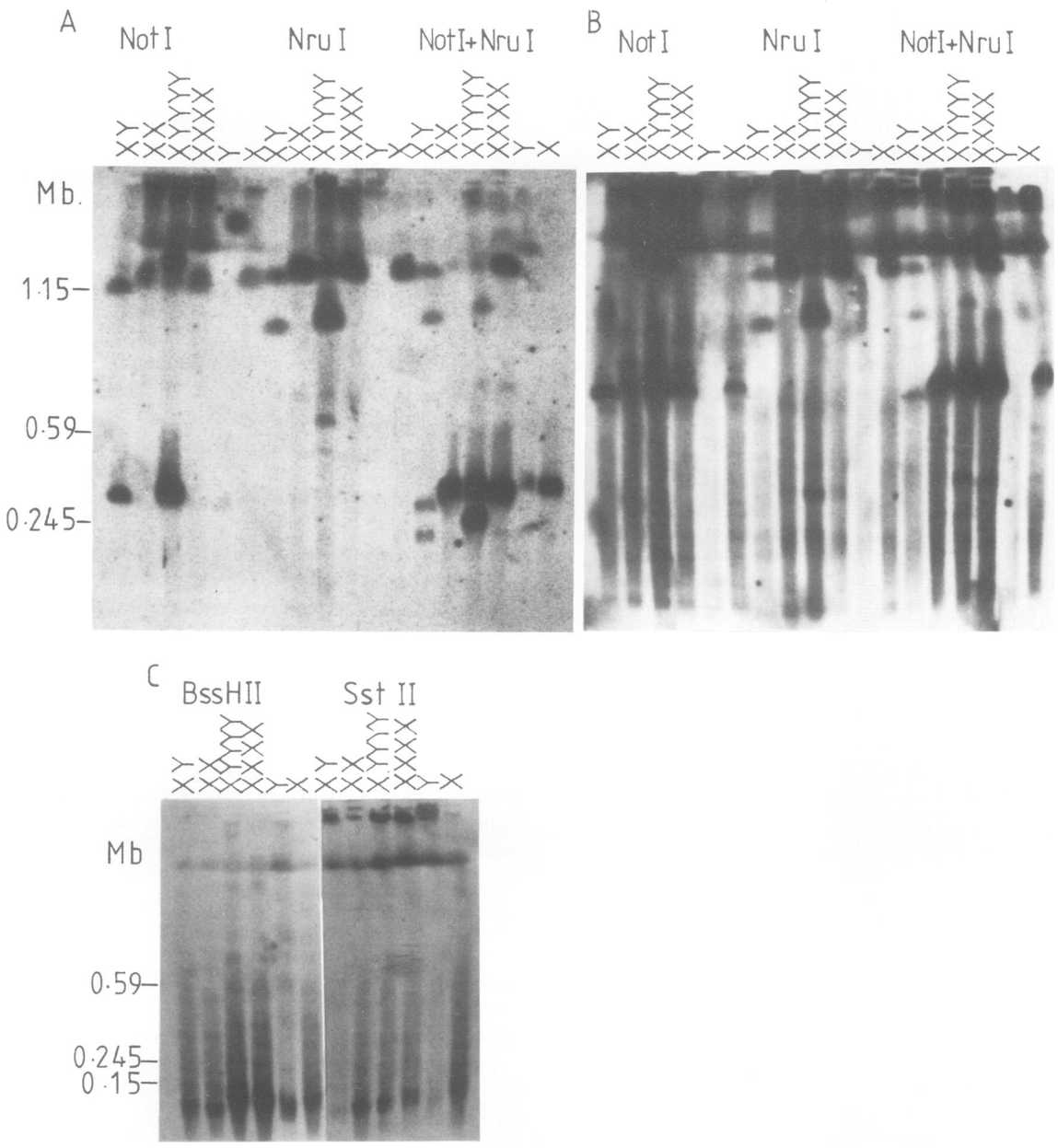

Fig. 7. Mapping *Not*I and *Nru*I sites and CpG islands in the centromere proximal subregion and in sex-chromosome-specific DNA. (**A**) and **B**) DNA from the blood of a man, a woman, a woman with Turner's syndrome, from cells of the lines OXEN (karyotype 49,XYYYY), GM1416 (karyotype 48,XXXX), 3E7 or 853 was digested for 5 h with 15 U *Not*I, 15 U *Nru*I or with 15 U of each enzyme sequentially. Digests were size fractionated by pulsed field gel electrophoresis (70 s pulse time), filter transferred and analysed by hybridization with the MIC2 probe p19B (**A**) or the DXYS17 probe 601 (**B**). The same filter was used and stripped of probe between the two experiments. Size markers were *S.cerevisiae* strain X2180 1B chromosomes. (**C**) DNA from the same six sources was digested for 5 h with *Bss*HII (4 U) and *Sst*II (20 U), size fractionated by pulsed field gel electrophoresis (60 s pulse time) and analysed with the 2.5-kb *hin*dIII fragment from the p44A subclone. This sequence lies immediately distal of the CpG island at the 5′ end of the MIC2 gene. A further round of filter hybridization with the p19B probe established that the *Bss*HII and *Sst*II sites in the MIC2 island had been completely cleaved. The letters above the panels indicate the sex chromosome constitution of the various DNA samples.

in the experiment shown in Figure 2 suggest that there are several sites for each of these enzymes in the proximal subregion close to the boundary with sex-chromosome-specific DNA. These sites can be seen as a smear toward the top of the resolving part of most of the tracks containing these partial digests. This smear is not evident in the tracks containing the *Eag*I and *Sst*II partial digests of 3E7 DNA because of underloading. Three quarters of the total number of sites for each of these CG enzymes can be calculated to occur in CpG islands (Brown and Bird, 1986) and the smear seen in these tracks raised the possibility that there were CpG islands distributed in this part of the pseudoautosomal region. In light of the association between genes and CpG islands it was of interest to map these islands. The MIC2 gene has a CpG island at its 5′ end. This is the last island before sex-chromosome-specific DNA (Pritchard *et al.*, 1987). The MIC2 CpG island includes a *Bss*HII site and could be positioned 100 kb distal of the Y-chromosome-specific *Nru*I site (Figure 8B) immediately beyond the boundary with sex-chromosome-specific DNA by hybridization analysis of *Bss*HII and *Nru*I double digests with a sequence from the proximal side of the MIC2 CpG island (data not shown).

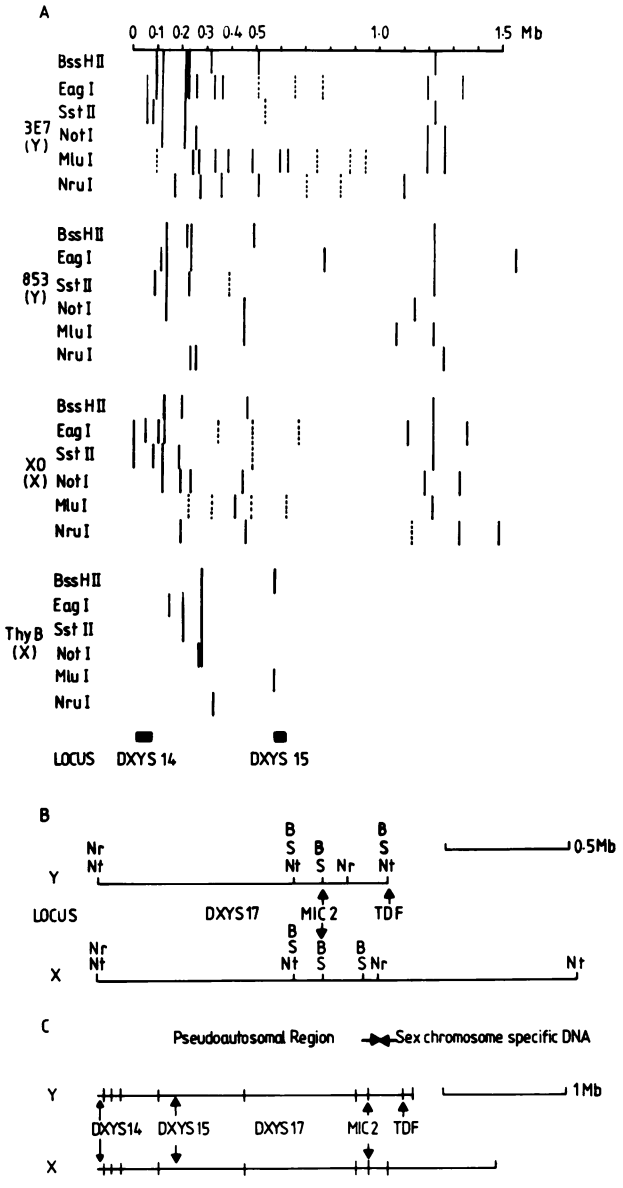

Fig. 8. Restriction site maps and overall organization of the human pseudoautosomal region. (**A**) *Bss*HI, *Eag*I, *Sst*II, *Not*I, *Mlu*I and *Nru*I sites in the distal and proterminal subregions of the pseudoautosomal region. Maps of the DNA in the blood of a Turner's syndrome patient or in the cells of the hybrid somatic lines 3E7, 853 or ThyB1-33/12 were constructed by analysis of the data in Figures 3 and 5 and other similar experiments. Figure 5 suggesets that the map of the DNA in the blood of the Turner's syndrome patient is similar to that in the blood of a normal man or woman. The broken bars indicate sites which appear as weakly cut in the partial digests. The positions of the genetically mapped loci DXYS14 and DXYS15 are indicated below the restriction site map. (**B**) *Bss*HII, *Sst*II, *Not*I and *Nru*I sites in the proximal subregion and in the adjacent sex-chromosome-specific DNA. Maps were constructed by analysis of the data in Figure 7 and on the basis of the data in Page *et al.* (1987) and Pritchard *et al.* (1987) which indicate the positions of *Bss*HII and *Sst*II sites in sex-chromosome-specific DNA. Only the pseudoautosomal *Bss*HII and *Sst*II sites which lie adjacent to the MIC2 CpG island could be mapped. The positions of the probes used in the analysis are indicated. (**C**) CpG islands in the pseudoautosomal region and in adjacent sex-chromosome-specific DNA. (A) and (B) are summarized to indicate the positions of CpG islands. These are assumed to lie at the position of any coincident CG enzyme sites in the map of the DNA in the peripheral blood leukocytes of the Turner's patient or in the maps of sex-chromosome-specific DNA in Page *et al.* (1987) or Pritchard *et al.* (1987). The DNA is indicated by the horizontal line, the CpG islands are indicated by the vertical bars which cross the horizontal line and the *Not*I site which bounds the X chromosome map is indicated by the vertical bar above the horizontal line.

The extent of the smear of hybridization seen in the partial *Bss*HII, *Eag*I and *Sst*II digests and discussed above suggested that there were other CpG islands distal to the MIC2 island. In order to confirm this a set of complete *Bss*HII and *Sst*II digests was analysed with a probe immediately distal to the MIC2 CpG island. This experiment revealed a set of common fragments of 100 kb in length (Figure 8C). The MIC2 CpG island contains *Sst*II sites in addition to *Bss*HII sites and so this result places *Bss*HII and *Sst*II sites together with the previously identified *Not*I site in a cluster 100 kb distal to the MIC2 island. This cluster suggests the presence of a second CpG island in this region. Although the resolution of the gel in Figure 2 is poor the extent of the smear of hybridization extends over several hundred kbp. The disparity between this figure and the short distance between the MIC2 island and its distal neighbour suggests that there may be other CpG islands in this part of the pseudoautosomal region. Identifying these additional islands will require further molecular cloning experiments.

Discussion

There is considerable current interest in long-range structural studies of human chromosomes. This paper demonstrates firstly a general strategy for long-range mapping around telomeres of human chromosomes and secondly how to approach several of the problems inherent in any long-range mapping project. I started to map the human pseudoautosomal region in DNA from the hybrid somatic cells 3E7 (Marcus *et al.*, 1976), 853 (Burk *et al.*, 1985) and in DNA from the blood of a woman with Turner's syndrome. It is well known that the pattern of DNA methylation often differs in cells from different tissues and is disturbed in cells in culture and it was therefore to be expected that these maps, made with methylation-sensitive restriction endonucleases, would differ. The results bear this out. It was a little surprising that no major long-range DNA polymorphism was detected in the mapping studies since many

anonymous pseudoautosomal DNA sequences detect informative restriction-fragment-length polymorphisms. These polymorphisms often arise because of minisatellite-like sequences and typically involve only a few kbp. They are apparently not reflected in long-range differences between the maps.

A main reason for interest in the pseudoautosomal region is that it is the site of a high rate of recombination in male meioses. The physical map (Figure 8C) includes the positions of four genetically mapped loci: DXYS14, DXYS15, DXYS17 and MIC2. The DXYS14, DXYS15 and MIC2 loci are positioned exactly, while there is an 0.8 Mb uncertainty in the whereabouts of the DXYS17 locus. The congruence between the physical and genetic distances (Figure 1) suggests that the recombination sites are not markedly clustered within the pseudoautosomal region. This conclusion is qualified because only three loci are located precisely. More loci need to be physically and genetically mapped before it becomes possible to exclude the presence of recombination hotspots. This will become relatively straightforward when a series of molecular clones have been isolated which span the pseudoautosomal region. The physical map will facilitate isolation of these clones.

The pseudoautosomal region may contain unidentified genes of which some may be involved in the aetiology of Turner's syndrome. Many genes have been found to be associated with CpG islands (Gardiner-Garden and Frommer, 1987) and conversely several random CpG islands are associated with transcripts (Lindsay and Bird, 1987). CpG islands are short tracts of G+C-rich DNA which show no CpG suppression and which are unmethylated (except on the inactive X chromosome). CG enzyme sites are clustered in such sequence tracts and these enzymes can be used to diagnose the presence of a CpG island (Brown and Bird, 1986). This paper shows that there exist stretches of sequence which are normally methylated but can be demethylated to reveal clusters of sites for CG enzymes. Such stretches of DNA are evident in the map of 3E7 DNA. They are not CpG islands because they are methylated in the Turner's DNA. However, if a cluster of sites for two or more CG enzymes in the Turner's DNA is considered to correspond to a CpG island then at least seven CpG islands can be located within the pseudoautosomal region (Figure 8C). This figure is possibly an underestimate because the size resolution of the map is nowhere better than 10 kb. These CpG islands are candidate gene loci. It is of course possible that some of the CpG islands indicated on the map are not associated with genes.

It is striking that many of the CpG islands within the pseudoautosomal region occur in clusters (Figure 8C). There is one cluster of three islands which starts 20 kb from the telomere and extends for 200 kb. A second set of clusters spans the boundary between the pseudoautosomal region and the X or Y chromosomes. The Y-chromosome-associated cluster includes at least two islands from the pseudoautosomal region and two islands from the testis-determining region (Page *et al.*, 1987; Pritchard *et al.*, 1987). The X-chromosome-associated cluster includes the two or more pseudoautosomal islands and at least one island from X-chromosome-specific DNA (Pritchard *et al.*, 1987). Clusters of CpG islands have been seen by others and may correspond to C+G-rich isochores (Bernardi *et al.*, 1985). The human α globin gene complex lies in a stretch of C+G-rich DNA which spans at least 300 kb and which includes a stretch of 30 kb which includes five CpG islands (Fischel-Ghodsian *et al.*, 1987). Similarly an unexpectedly large number of cosmids, isolated at random from chromosome three, included two or more CpG islands (Smith *et al.*, 1987). Conversely many of the genes contained within C+G-rich isochores (Bernardi *et al.*, 1985) are associated with CpG islands (Gardiner-Garden and Frommer, 1987). Little is known about CpG island clusters. It is unclear whether sequences which include such tightly packed islands as the α globin cluster can be considered as being the same class of sequence as the pseudoautosomal region island clusters. The original experiments suggested that the isochores were at least 250 kb long but were unable to put an upper limit on their size or to define the number of isochores. If the CpG island clusters correspond to the C+G-rich isochores then it will be possible to answer these questions by further chromosome mapping studies. Such mapping studies will also give information about the chromosome distribution of the CpG island clusters and their association with cytologically detected aspects of chromosome structure such as banding.

Materials and methods

Cells and DNA

DNA was prepared from three hybrid somatic cell lines, two lymphoblastoid lines and from the peripheral blood leukocytes of a woman with Turner's syndrome, a normal woman and a normal man. The hybrid somatic cells each contained a human sex chromosome as the only cytologically detectable human constituent. The Y-chromosome-containing hybrids were 3E7 (Marcus *et al.*, 1987) and 853 (originally described as 7631, Burk *et al.*, 1985). The X-chromosome-containing hybrid was ThyB1-33/12 (Lund *et al.*, 1983). The lymphoblastoid lines were OXEN (karyotype 49,XYYYY) and GM1416 (karyotype 48,XXXX). The woman with Turner's syndrome had a 45,X karyotype and showed no signs of mosaicism. Peripheral blood leukocytes were isolated from whole blood after lysing erythrocytes in three volumes of 0.155 M, NH_4Cl, 0.01 M $KHCO_3$, 0.01 mM ethylene diamine tetra acetic acid (EDTA), pH 7.4.

High mol. wt DNA was prepared from cells encapsulated in low-gelling-temperature agarose plugs. A suspension of single cells at either 10^7 or 2×10^7/ml in phosphate-buffered saline (PBS) was held at 42°C and mixed with an equal volume of molten 1% (w/v) low-gelling-temperature agarose (Betheseda Research Laboratories Ultrapure) in PBS also at 42°C. The mixture was dispensed into many small moulds each $2 \times 6 \times 9$ mm and allowed to set into plugs. The plugs were then extruded into 0.5 M EDTA, 0.01 M Tris−HCl, 1% sodium lauroyl sarcosine pH 9.5 (NDS) and incubated at 55°C for 1 h. The NDS was then decanted, replaced with a solution of proteinase K (Boehringer) at 1 mg/ml in NDS. The plugs were incubated at 55°C in this solution for 2 days and were then stored indefinitely at 4°C without additional buffer changes.

Restriction endonuclease digestion

Plugs were washed extensively with 10 mM Tris−HCl, 1 mM EDTA, pH 7.4 and residual proteinase K activity was destroyed by adding phenylmethylsulphonylfluoride to 0.1 mM. The plugs were then rinsed in distilled water, divided into three across the length of the plugs and equilibrated with restriction enzyme buffer as recommended by the supplier. Excess buffer was then removed, gelatin added to 0.5 mg/ml, dithiothreitol added to 1 mM and Triton X-100 added to 0.01% where appropriate. Complete digests were incubated for between 4 and 6 h at the recommended temperature. Partial digests, with the exception of those shown in Figures 3A and 4, were carried out by incubating the samples with limited amounts of enzyme for 1 h on ice and then for 2 h at the recommended temperature. For analysis of the partial digests the slices of agarose were divided into two 1-mm-wide pieces of which one was analysed. Restriction enzymes were from New England Biolabs with the exceptions of *Sca*I, which was from Anglian Biotechnology, and *Sst*II, which was from Bethesda Research Ltd.

Pulsed field gel electrophoresis

Agarose plugs containing restricted DNA were rinsed in TE pH 7.4 and loaded into the slots of an agarose gel in 0.5 × TAE (1 × TAE is 0.04 M Tris-acetate, 0.001 M EDTA). The gel concentration was 1.5% except in the experiment shown in Figure 2 where it was 0.8%. The electrophoresis was in a rotating plate device (Southern *et al.*, 1987). Pulse times were as described in the figure legends, run times were 33 h except where stated otherwise in the figure legends, buffer temperature was 18°C except where stated in the figure legends, electric field was 5 V/cm except for Figure 2 where it was 1 V/cm. Size markers were oligomers of bacteriophage lambda cI857 DNA or the chromosomes of *Saccharomyces cerevisiae* strain X21801B (Anand, 1986).

Probes and filter hybridization

The following sequences were used as probes, 29c1 (Cooke *et al.*, 1985) at the locus DXYS14, a 3.04-kb *Hin*dIII fragment from cY29 (Cooke *et al.*, 1985) at the locus DXYS20, 113D at the locus DXYS15 (Simmler *et al.*, 1985), 601 at the locus DXYS17 (Rouyer *et al.*, 1986), p19B centromere proximal to the CpG island at the MIC2 locus (Pritchard *et al.*, 1987) and a 2.5-kb *Hin*dIII fragment from subclone p44A centromere distal to the CpG island at the MIC2 locus. This probe was a gift from Peter Goodfellow and Paul Goodfellow, both of the Imperial Cancer Research Fund, London. The 2.5.-kb *Hin*dIII fragment from p44A contained some repeated sequences which were removed by prereassociation to a Cot value of 15.

DNA in agarose gels was depurinated (Wahl *et al.*, 1979) and filter transferred (Southern, 1975) to either nylon (Genescreen, New England Nuclear) or nitrocellulose (Schleicher and Schuell). Nylon filters were baked dry and UV irradiated with a bacteriocidal lamp. Nylon or nitrocellulose filters were hybridized with random hexamer-primed probes (Feinberg and Vogelstein, 1983) in either 7% sodium dodecyl sulphate, 0.5 M sodium phosphate, 1% bovine serum albumin pH 7.4 or as described in Brown and Bird (1986). Filters were washed to 0.2 × SSC at 65°C with either 1% (nylon) or 0.1% (nitrocellulose) sodium dodecyl sulphate except for those probed with the DXYS20 probe which were washed to 0.5 × SSC at 65°C with sodium dodecyl sulphate. Autoradiographs were made using Kodak XAR-5 film.

Acknowledgements

Thanks to Mark Ross, Ian Craig and Chris Tyler-Smith for comments on a draft manuscript, Peter Goodfellow and Jean Weissenbach for probes, Martin Johnson for building and maintaining the gel machine, Lesley Taylor for growing cells, Frank Caddick for illustrations, Michael Lardelli for enthusiastic help, Dr R.Lindenbaum for advice about the Turner's patient and M.G. for generously donating blood samples. This work was supported by the Cancer Research Campaign, London.

References

Anand,R. (1986) *Trends Genet.*, **2**, 278–283.
Bernardi,G., Olofson,B., Filipski,J., Zerial,M., Salinas,J., Cuny,G., Meunier-Rotival,M. and Rotier,F. (1985) *Science*, **228**, 953–958.
Bird,A.P. (1986) *Nature*, **321**, 209–213.
Bird,A.P., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) *Cell*, **40**, 91–99.
Brown,W.R.A. and Bird,A.P. (1986) *Nature*, **322**, 477–481.
Burgoyne,P.S. (1982) *Hum. Genet.*, **61**, 85–90.
Burk,R.D., Ma,P. and Smith,K.D. (1985) *Mol. Cell. Biol.*, **5**, 576–581.
Cooke,H.J., Brown,W.R.A. and Rappold,G.A. (1985) *Nature*, **317**, 687–692.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Ferguson-Smith,M. (1965) *J. Med. Gen.*, **2**, 142–155.
Fischel-Ghodsian,N., Nicholls,R.D. and Higgs,D.R. (1987) *Nucleic Acids Res.*, **15**, 9215–9227.
Ford,C.E., Jones,K.W., Polani,P.E., de Almeida,J.C. and Briggs,J.H. (1959) *Lancet*, **ii**, 711–713.
Gardiner-Garden,M. and Frommer,M. (1987) *J. Mol. Biol.*, **196**, 261–282.
Goodfellow,P.J., Darling,S.M., Thomas,N.S. and Goodfellow,P.N. (1986) *Science*, **234**, 740–743.
Goodfellow,P.J., Pritchard,C., Tippet,P. and Goodfellow,P.N. (1987) *Ann. Hum. Genet.*, **51**, 161–167.
Harbers,K., Soriano,P., Muller,U. and Jaenisch,R. (1986) *Nature*, **324**, 682–685.
Keitges,E.A., Schordet,D.F. and Gartler,S.M. (1987) *Genetics*, **116**, 465–468.
Linday,S. and Bird,A.P. (1987) *Nature*, **327**, 336–338.
Lund,E., Bostock,C., Robertson,M., Christie,S., Mitchen,J.L. and Dahlberg,J.E. (1983) *Mol. Cell. Biol.*, **3**, 2211–2220.
Marcus,M., Tantravahi,R., Dev,V.G., Miller,D.A. and Miller,O.J. (1976) *Nature*, **262**, 63–65.
New England Biolabs (1986) *New England Biolabs Catalog (1986/87)*, p. 121.
Page,D.C., Mosher,R., Simpson,E.M., Fisher,E.M.C., Mardon,G., Pollack,J., McGillivray,B., de la Chapelle,A. and Brown,L.G. (1987) *Cell*, **51**, 1091–1104.
Pritchard,C.A., Goodfellow,P.J. and Goodfellow,P.N. (1987) *Nature*, **328**, 273–275.
Rouyer,F., Simmler,M.-C., Johnson,C., Vergnaud,G., Cooke,H.J. and Weissenbach,J. (1986) *Nature*, **319**, 291–295.
Schwartz,D.C. and Cantor,C. (1984) *Cell*, **37**, 67–75.
Simmler,M.-C., Rouyer,F., Vergnaud,G., Nystrom-Lahti,M., Ngo,K.Y., de la Chapelle,A. and Weissenbach,J. (1985) *Nature*, **317**, 692–697.
Smith,D.I., Golembieski,W., Gilbert,J.D., Kizyam,L. and Miller,O.J. (1987) *Nucleic Acids Res.*, **15**, 1173–1183.
Southern,E.M. (1975) *J. Mol. Biol.*, **237**, 503–517.
Southern,E.M., Anand,R., Brown,W.R.A. and Fletcher,D.S. (1987) *Nucleic Acids Res.*, **15**, 5925–5943.
Swartz,M.N., Trautner,T.A. and Kornberg,A. (1962) *J. Biol. Chem.*, **237**, 1961–1967.
Turner,H.H. (1938) *Endocrinology*, **23**, 566–578.
Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **83**, 5611–5615.
Weissenbach,J., Levilliers,J., Petit,C., Rouyer,F. and Simmler,M.-C. (1987) *Development* (Suppl.), **101**, 67–74.

Received on March 23, 1988; revised on May 17, 1988