some cores contained equal amounts of the four core histones (Fig. 1A). We prepared cellular H1-depleted nucleosome cores by solubilizing nucleosomes from cellular chromatin by micrococcal nuclease digestion followed by gel filtration chromatography in the presence of 0.6 M NaCl. Gradient-purified nucleosome cores reconstituted by salt dialysis also resemble cellular nucleosome cores with regard to nuclease digestions. Digestion of the 150-bp PCR fragment nucleosomes, which were labeled throughout during amplification, with DNase I showed a 10-bp ladder between cutting sites. DNase I digestion of end-labeled restriction fragments (which contained the same sequences) after similar reconstitution into nucleosome cores showed a strong 10-bp digestion pattern not apparent on naked DNA, which demonstrates rotational phasing of the DNA on the histone octamer surface (H. Chen and J. L. Workman, unpublished data). The PCR fragments used (150 bp) were close to nucleosome core length (146 bp), which made micrococcal nuclease digestion analysis of these nucleosome cores difficult. However, nucleosome cores reconstituted with 180-bp PCR fragments (from the same plasmid but extended 15 bp on each end) showed a digestion intermediate with micrococcal nuclease of approximately 146 bp (*19*). Thus, the reconstituted nucleosomes resemble cellular nucleosome cores by several criteria.

12. GAL4 derivatives were purified according to Chasman *et al.* (*15*). For the binding reactions shown in Fig. 1B, nucleosome cores that contained approximately 0.3 pmol of DNA (with five GAL4 sites) or the unreconstituted PCR fragment (diluted in the appropriate mixture of the sucrose gradient solutions) in 5 μl was mixed with the indicated amounts of GAL4-AH, which were in 3 μl of BC100 buffer [20% glycerol, 100 mM KCl, 20 mM Hepes (pH 7.9), and 0.2 mM EDTA] with BSA (1 mg/ml), 10 mM 2-mercaptoethanol, and 10 μM $ZnCl_2$ (*8*). Reactions (8 μl total volume) were incubated for 30 min at 30°C and then analyzed on 4% acrylamide (29:1, acrylamide:bis), 0.5× TBE [45 mM tris-borate (pH 8.0) and 1 mM EDTA] gels.
13. For the two-dimensional gels to analyze protein composition (Figs. 1C and 2C), 200 μl of reconstituted nucleosome cores (12 pmol) were incubated for 30 min at 30°C with 180 pmol of GAL4-AH (added as undiluted protein in 3 μl). In the experiment shown in Fig. 2C, 40 μg of nonspecific plasmid DNA was added followed by an additional 30-min incubation at 30°C. After the binding and competition reactions, the samples were separated on native acrylamide gels (4%) as described above. The full-length lanes that contained the samples were cut out of the gel and placed over a short stacking gel (3 cm) above a SDS-polyacrylamide separating gel (15%). Low molecular weight markers (Bio-Rad) were loaded in a marker lane placed directly against the bottom of the native gel slices. The gels were electrophoresed until the bromphenol blue dye (Sigma) ran off the bottom. After electrophoresis, the stacking gels were cut off, and the separating gels were stained with silver [W. Wray, T. Boulikas, V. P. Wray, R. Hancock, *Anal. Biochem.* **118**, 197 (1981)] with the following modifications. The gels were stained in solution C for 30 min and developed until the entire gel became very dark to ensure staining and development of protein in the center of the gels. The gels were cleared (destained) in Kodak Rapid Fix (undiluted) until the background staining was completely removed, allowing visualization of the stained proteins.
14. The experiments shown in Fig. 2, A and B, and in Fig. 3, A and B, were similar to that in lane 8 of Fig. 1B. However, after the binding reactions the indicated amounts of nonspecific competitor DNA or buffer [10 mM tris and 1 mM EDTA (pH 8.0)] were added in a volume of 3 μl. After an additional 30-min incubation at 30°C, the samples were analyzed on acrylamide gels (4%) as described above. The nonspecific plasmid DNA was pNOTM, a plasmid bearing the SV40 replication origin, the major-late promoter of adenovirus, and the tetracycline resistance gene of pBR322 in pUC13. Other nonspecific plasmid DNAs functioned identically. Mobility-shift gels were dried and exposed to film at −80°C. Dried gels were also counted on a Betascope blot analyzer (Betagen) followed by quantitation of the bands.
15. I. Sadowski, J. Ma, S. Triezenburg, M. Ptashne, *Nature* **335**, 563 (1988); D. I. Chasman, J. Leatherwood, M. Carey, M. Ptashne, R. Kornberg, *Mol. Cell. Biol.* **9**, 4746 (1989); M. Carey, H. Kakidani, J. Leatherwood, F. Mostashari, M. Ptashne, *J. Mol. Biol.* **209**, 423 (1989).
16. R. A. Laskey and W. C. Earnshaw, *Nature* **286**, 763 (1980); J. A. Kleinschmidt and W. W. Franke, *Cell* **29**, 799 (1982); S. Smith and B. Stillman, *EMBO J.* **10**, 971 (1991). M. Sapp and A. Worcel, *J. Biol. Chem.* **265**, 9357 (1990).
17. C. L. Peterson and I. Herskowitz, *Cell* **68**, 573 (1992); J. W. Tamkun *et al.*, *ibid.*, p. 561.
18. G. E. Croston, L. A. Kerrigan, L. M. Lira, D. R. Marshak, J. T. Kadonaga, *Science* **251**, 643 (1991); P. J. Laybourn and J. T. Kadonaga, *ibid.* **254**, 238 (1991).
19. L. J. Juan and J. L. Workman, unpublished data.
20. H. M. Blau, *Trends Genet.* **5**, 268 (1989); _____ and D. Baltimore, *J. Cell Biol.* **112**, 781 (1991).
21. We thank I. Taylor and T. Schuetz for valuable discussions, H. Chen for assistance in the preparation of some of the protein fractions, G. Felsenfeld for useful suggestions, and K. Zaret, H. Weintraub, D. Gilmour, S. Abmayr, A. Imbalzano, C. Kara, and O. Papoulas for critical reading of the manuscript. Supported by a grant from Hoechst AG (R.E.K.) and funding from the Eberly College of Science at The Pennsylvania State University (J.L.W.). J.L.W. was a Special Fellow of the Leukemia Society of America.

13 July 1992; accepted 28 October 1992

Identification of a Second Pseudoautosomal Region Near the Xq and Yq Telomeres

Diha Freije, Cynthia Helms, Michael S. Watson, Helen Donis-Keller*

The telomeres of Xq and Yq have been observed to associate during meiosis, and in rare cases a short synaptonemal complex is present. Molecular cloning of loci from Xqter and Yqter has revealed that their sequence homology extends over 400 kilobases, which suggests the possibility of genetic exchange. This hypothesis was tested by the development of two highly informative microsatellite markers from yeast artificial chromosome clones that carried Xqter sequences and the following of their inheritance in a set of reference pedigrees from the Centre d'Etude du Polymorphisme Humain in Paris, France. From a total of 195 informative male meioses, four recombination events between these loci were observed. In three cases, paternal X alleles were inherited by male offspring, and in one case a female offspring inherited her father's Y allele. These data support the existence of genetic exchange at Xq-Yq, which defines a second pseudoautosomal region between the sex chromosomes.

The human Y chromosome serves two known primary functions. It carries sequences on the short arm coding for a testis-determining factor (*1–3*), which is required to initiate male development. It also provides a pairing partner for the X chromosome during meiosis. Koller and Darlington (*4*) first proposed the existence of differential (sex-specific) and common [pairing, later termed pseudoautosomal (*5*)] segments between the sex chromosomes to explain their cytogenetic association in rat spermatocytes. During the early stages of prophase I, the sex chromosomes condense and associate to form a heterochromatic mass known as the XY body (*6*). The formation of a synaptonemal complex is restricted to the short arm pseudoautosomal region, but the extent of nonhomologous synapsis between the X and the Y chromosomes is variable and could extend into Yq (*7*). A secondary end-to-end association between Xq and Yq is occasionally observed in meiotic spreads, and in rare cases a short synaptonemal complex is present (*8*). The sex chromosomal region that undergoes homologous pairing in humans occupies the most distal segments of the X and Y short arms. The cloning of sequences from the pairing region provided molecular evidence for recombination between distal Xp and Yp (*9*). A gradient of recombination can be observed in males between pseudoautosomal markers and the sex-determining locus, with the most telomeric marker segregating independently of sex (*10*).

In addition to the pseudoautosomal segment, the X and Y chromosomes share several regions of homology that are not known to undergo legitimate recombination during male meiosis (*2*, *11*). The isolation of DNA segments corresponding to loci DXYS61 (isolated from a flow-sorted Y library) (*12*) and DXYS64 (iden-

D. Freije, C. Helms, H. Donis-Keller, Division of Human Molecular Genetics, Department of Surgery and Departments of Genetics and Psychiatry, Washington University School of Medicine, 660 South Euclid Street, St. Louis, MO 63110.
M. S. Watson, Edward Mallinckrodt Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110.
*To whom correspondence should be addressed.

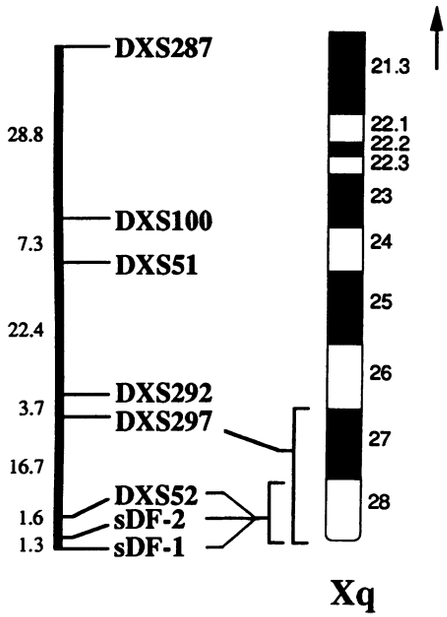

Fig. 1. Genetic linkage map of distal Xq. The female-specific multipoint map was constructed with genotypic data from sDF-1 (DXYS154), sDF-2 (DXS1108), and Xq markers in the CEPH V5 database with the use of CRI-MAP version 2.4 (*20*). Markers were uniquely placed with odds of at least 1000:1. The number of informative meioses for sDF-1 is 430; for sDF-2, 235; and for DXS52, 263. The numbers on the left indicate the distances between loci in centimorgans; the numbers on the right indicate cytogenetic bands. The arrow points in the direction of the centromere on the chromosome.

tified from a random X clone) (*13*) and their localization by in situ hybridization (*14*) in distal Yq and Xq indicated a region of homology near the telomeres. In addition, the pattern of restriction endonuclease sites at these loci predicts greater than 99% sequence identity between the two copies (*12, 13*). Analysis of the DXYS61 locus in higher primates suggests that the observed homology could be the result of its recent transposition to the Y chromosome (*15*). Furthermore, the cloning of human telomeres (*16*) shows that proterminal repetitive sequences (directly proximal to the T_2AG_3 repeats) common to several autosomal telomeres are also present at Xq and Yq. The high degree of sequence homology at Xq and Yq and the cytogenetic association of these telomeres in human meiotic spreads raised the possibility of homologous pairing and genetic exchange.

A 1.6-Mb yeast artificial chromosome (YAC) contig was constructed near the Factor VIII gene (*17*); the contig established that the XY homology region identified by DXYS61 and DXYS64 (which maps 350 kb from the terminal repeats) is the most distal segment in Xq28. This contig provided source DNA to develop polymorphic markers. Here, we provide evidence for genetic exchange between Xq and Yq. We propose that sequence homology is maintained by genetic exchange and that these segments define a second pseudoautosomal region on the human sex chromosomes.

Several cosmid subclones from two overlapping YACs (yWXD364 and yWXD250) that carry the DXYS64 locus were screened for microsatellite polymorphisms. A $(CA)_{18}$ repeat element was identified within cosmid c364-10g, and a polymerase chain reaction (PCR) assay was developed. A highly polymorphic marker, sDF-1 (at locus DXYS154), was identified (70% heterozygosity), and a total of ten alleles (ranging in size from 220 to 244 bp) were observed in the 40 Centre d'Etude du Polymorphisme Humain (CEPH) reference pedigrees studied. Two alleles were amplified from several males, which indicates that the sequence is present in two copies in males as well as females. No male-specific alleles were found.

To verify that sDF-1 is present on the distal long arm of the sex chromosomes, we tested a set of somatic cell hybrids, each carrying one or more human chromosomes and together representing the full human chromosome complement. Amplification was observed only in hybrids that carried X or Y (*18*). Furthermore, fluorescent in situ hybridization with cosmid DNA (from clones c364-10g and c250-9k) detected signals on only Xqter and Yqter (*18*).

A second microsatellite repeat, sDF-2 [$(GT)_5GC(GT)_{13}$, locus DXS1108], was identified in cosmid c250-9k by degenerate cycle sequencing (*19*). Amplification of sDF-2 from 74 unrelated individuals identified seven alleles [75% heterozygosity (*18*)]. All males were found to be hemizygous. Cosmid c250-9k contains an X chromosome–specific sequence (sDF-2) and reveals sequence homology on the Y chromosome by in situ hybridization; therefore, it is likely to contain the physical boundary between X-specific and the XY homology regions.

The markers sDF-1 and sDF-2 were genotyped in 32 informative CEPH families. Two point analyses of these markers against the entire CEPH V5 database did not reveal significant linkage [logarithm of the likelihood ratio for linkage (lod) > 3.0] with any autosomal markers. However, multiple high lod scores were found with markers from the X chromosome, including DXS52, the most distal marker available in Xq28 [for example, for sDF-1–sDF-2, lod = 41, recombination fraction (θ) = 0.01; for sDF-1–DXS52, lod = 23, θ = 0.03; and for sDF-2–DXS52, lod = 27, θ = 0.03]. Figure 1 shows a multipoint linkage map of the

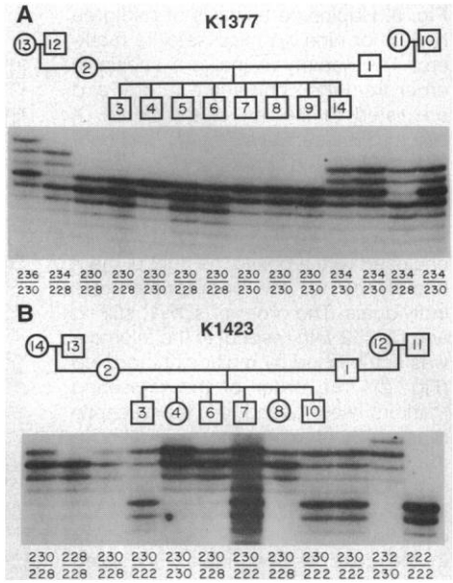

Fig. 2. Segregation of sDF-1 alleles in CEPH pedigrees (**A**) K1377 and (**B**) K1423. Six alleles ranging in size from 222 to 236 bp are evident from the two pedigrees. Genotypes were assayed by PCR amplification with 50 ng of genomic DNA and 2.5 pmol of each sDF-1 primer (5′-GGCCTGAATTCATTTATTATTCTAATAG-3′ and 5′-GAACAGGCAAAGATGCCCACTCTC-3′) in a standard 12.5-μl reaction (*26*). The forward primer was end-labeled with T4 polynucleotide kinase and [γ-^{32}P]adenosine triphosphate, and 22 cycles of 94°C for 1 min, 65°C for 1 min, and 72°C for 1 min were performed in a Perkin-Elmer Cetus thermocycler. An aliquot from each reaction was fractionated by electrophoresis through a 6% denaturing polyacrylamide gel and exposed to x-ray film for 12 to 24 hours. Squares, male family members; circles, female family members. Genotypes (in base pairs) of the individuals are shown beneath the gel lanes.

Xqter region constructed by means of CRI-MAP (*20*) with sDF-1, sDF-2, and Xq genotypic data available from the CEPH database. The map establishes sDF-1 as the most distal marker in Xq28 with sDF-2 located between DXS52 and sDF-1 (with odds for order of at least 1000:1).

From a total of 195 informative male meioses, four recombination events between sDF-1 and sDF-2 were identified (within four families). Figure 2 shows the inheritance of sDF-1 in two CEPH pedigrees that reveal two of these recombinants. In pedigree K1377, all eight male offspring (individuals 3 through 9 and 14) are expected to inherit the 230-bp Y allele from their father (individual 1) (as inferred by inheritance from their paternal grandfather) and either the 230- or 228-bp alleles from their mother (individual 2). Seven sons inherited the Y allele; however, individual 14 inherited the 234-bp allele present on his father's X chromosome (as

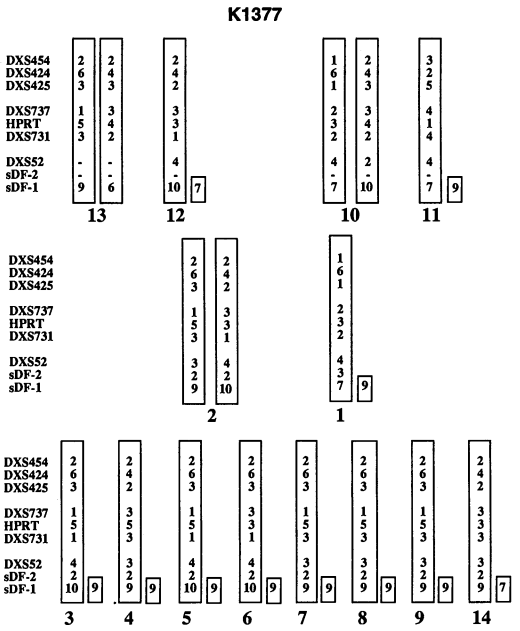

Fig. 3. Haplotype analysis of pedigree K1377 for nine Xq microsatellite markers. The haplotype loci are shown in order from the centromere to qter and are listed to the left of the figure. X chromosome allele numbers are contained within the boxes, and the identification number for each individual is shown in boldface directly under the haplotype. The Y alleles for sDF-1 are enclosed with a box to the right of the X haplotypes for the appropriate male individuals. The order of sDF-1, sDF-2, and DXS52 with respect to the telomere was established by multipoint mapping (Fig. 2). The order of the remaining markers was based on the recently reported X chromosome map (*27*). In the small set of families analyzed, no recombinants were observed between DXS52 and DXS731; therefore, the order with respect to sDF-2 is not known. In the case of nonuniquely localized markers, the order that minimized the number of double recombination events was used.

inferred by inheritance from his paternal grandmother), which indicates that a genetic exchange has occurred that results in the transfer of the 234-bp X allele to the Y chromosome. Similarly, in pedigree K1423 a male offspring (individual 6) inherited the grandmaternal X allele (230 bp) from his father instead of the 222-bp Y allele. The other two recombinants were observed in a son who inherited his father's X allele and in a daughter who inherited her father's Y allele (*21*).

Because the observed recombinants could have resulted from DNA sample errors, it was important to verify first that the recombinant individuals are independent members of the same families and, second, that the reported sex for these individuals is consistent with the information obtained from the DNA analysis. We followed the segregation of 30 highly informative autosomal markers (mapping to chromosomes 2, 6, 7, 8, and 12) in these four families. Mendelian segregation of the alleles consistent with inheritance from the parental DNAs was observed for all individuals analyzed. Variation among the genotypes of the siblings provided verification that they are independent offspring of the two parents. Second, because males are hemizygous for most of the X chromosome, the segregation of seven additional X microsatellite markers (*22*) was followed. All male pedigree members were found to be hemizygous. Figure 3 shows the haplotypes for individuals from pedigree K1377 for a subset of the X chromosome microsatellite markers analyzed. For example, individual 14 inherited a single maternal allele for markers DXS454, DXS425, and DXS731 that differed in size from the paternal allele. This analysis excluded a DNA sample mix-up and verified that the male individual did not inherit other X chromosome sequences from his father.

These data provide evidence for recombination between the most distal segments of the long arms of the human X and Y chromosomes. Two mechanisms could explain the molecular basis of the recombination events observed: gene conversion and reciprocal exchange. Gene conversion refers to the unidirectional transfer of information from one copy of a given sequence to another. Homologous recombination results in the reciprocal exchange of sequence information between the two alleles. The generation of more distal markers could in principle establish the length of the DNA segment involved in each genetic event. However, because of the relatively small size of the XY homology segment (less than 500 kb) and the inability to analyze all four products of a meiotic event, it may not be possible to determine with certainty which mechanism is responsible for the genetic exchange seen at Xq and Yq.

The two microsatellite markers used in this study are separated by at most 100 kb (*23*). Therefore, the rate of recombination in male meiosis (4 out of 195 or 2%) is relatively high compared to that in female meiosis, where a single crossover was observed in 238 informative meioses. The observed rate of recombination is sufficient to maintain sequence identity between the two copies, assuming a neutral mutation rate and a lack of selective pressure favoring the divergence of the sequences (*24*).

Because we observed recombination between genetic markers from distal Xq and Yq, it is interesting to consider the possibility that homologous pairing might occur in this region during male meiosis. On the basis of our data, synapsis between Xq and Yq can be inferred for at least 4% of meiotic events. Before this, Xq and Yq telomeres were not known to pair homologously during meiosis. Our data raise two possibilities with regard to recombination. It is conceivable that the acquisition of a pairing segment at Xq and Yq in higher primates is a result of the nonhomologous pairing that occasionally occurs between chromosomal ends during the early stages of meiosis and does not reflect an essential biological role for this DNA segment (*25*). Alternatively, pairing at Xq and Yq may be required for proper synapsis and segregation of chromosomes in male meiosis. The role of Xq-Yq pairing, if any, in male meiosis remains to be elucidated. Such a role could be investigated in males monosomic for this DNA segment, assuming that deletions of one copy of these sequences do not affect the viability of the individuals.

REFERENCES AND NOTES

1. M. Andersson, D. C. Page, A. de la Chapelle, *Science* **233**, 786 (1986).
2. D. C. Page *et al.*, *Cell* **51**, 1091 (1987).
3. A. H. Sinclair *et al.*, *Nature* **346**, 240 (1990).
4. P. C. Koller and C. D. Darlington, *J. Genet.* **29**, 159 (1934).
5. P. S. Burgoyne, *Hum. Genet.* **61**, 85 (1982).
6. A. J. Solari, *Chromosoma* **81**, 315 (1980).
7. A. C. Chandley, P. Goetz, T. B. Hargreave, A. M. Joseph, R. M. Speed, *Cytogenet. Cell Genet.* **38**, 241 (1984).
8. R. M. Speed and A. C. Chandley, *Hum. Genet.* **84**, 547 (1990).
9. H. J. Cooke, W. R. A. Brown, G. A. Rappold, *Nature* **317**, 687 (1985); M.-C. Simmler *et al.*, *ibid.*, p. 692.
10. F. Rouyer *et al.*, *ibid.* **319**, 291 (1986); D. C. Page *et al.*, *Genomics* **1**, 243 (1987).
11. N. A. Affara *et al.*, *Nucleic Acids Res.* **14**, 5353 (1986); D. C. Page, M. E. Harper, J. Love, D. Botstein, *Nature* **311**, 119 (1984); A. Schneider-Gadicke, P. Beer-Romero, L. G. Brown, R. Nussbaum, D. C. Page, *Cell* **57**, 1247 (1989).
12. H. J. Cooke, W. A. R. Brown, G. A. Rappold, *Nature* **311**, 259 (1984).
13. B. Arveiler, A. Vincent, J.-L. Mandel, *Genomics* **4**, 460 (1989).
14. A. Pedicini, G. Camerino, R. Avarello, S. Guioli, O. Zuffardi, *ibid.* **11**, 482 (1991).
15. W. A. Bickmore and H. J. Cooke, *Nucleic Acids Res.* **15**, 6261 (1987).
16. W. R. A. Brown *et al.*, *Cell* **63**, 119 (1990); A. O. M. Wilkie *et al.*, *ibid.* **64**, 595 (1991).
17. D. Freije and D. Schlessinger, *Am. J. Hum. Genet.* **51**, 66 (1992).
18. For chromosome localization, a panel of human-rodent somatic cell hybrids that carried one or a small number of human chromosomes and that represented the entire human chromosome complement was obtained from the National Institute of General Medical Sciences Mutant Cell Repository at the Coriell Institute (Camden, NJ). Genomic DNA was amplified with the use of the sDF-1 primers and the conditions specified in Fig. 2 except that the reactions were performed for 35 cycles without radioisotope labeling of primers and the products were visualized by ethidium bromide staining after electrophoresis through 7.5% polyacrylamide gels. Cell lines GM06317

(containing human Y chromosome) and GM06318B (containing human X chromosome) gave the expected product size (approximately 230 bp). sDF-2 was amplified under the same conditions used for sDF-1. The sequences of the sDF-2 primers are 5′-ACTAGGCGACTAATACAGTGGTGC-3′ and 5′-GTGAATTCATCATATGTGATTTCC-3′. The product size ranges between 163 and 177 bp. For in situ hybridization, approximately 500 ng of DNA from cosmids c364-10g and c250-9k was digested with Rsa I or Hinf I, labeled with biotinylated deoxyuridine 5′-triphosphate by random priming, and hybridized independently to metaphase spreads of normal human lymphocytes (46, XY). Signal detection was done with the use of an in situ hybridization kit from Oncor (Gaithersburg, MD) according to the manufacturer's recommendation.

19. Cycle sequencing was performed as described [B. R. Krishnan, R. W. Blakesley, D. E. Berg, *Nucleic Acids Res.* **19**, 1153 (1991)]. Fifteen cycles of 95°C for 1 min, ramp to 65°C for 5 min, and 75°C for 30 s were performed in a Perkin-Elmer Cetus thermocycler with 50 ng of cosmid DNA and 5 pmol of each end-labeled primer mix [$(AC)_{11}N$ (A, T, or G) or $(CA)_{11}N$(T, G, or C)] in a separate reaction. Gel separation and autoradiography were done according to standard protocols.
20. P. Green, CRI-MAP version 2.4, unpublished material.
21. In pedigree K1362, individual 4, a female offspring, inherited the 228-bp allele from her father's Y chromosome. In pedigree K1332, individual 6, a male offspring, inherited a 222-bp allele from his father's X chromosome.
22. J. A. Luty *et al.*, *Am. J. Hum. Genet.* **46**, 776 (1990); T. J. Hudson *et al.*, *Genomics* **13**, 622 (1992); A. Edwards, A. Civitello, H. A. Hammond, C. Caskey, *Am. J. Hum. Genet.* **49**, 746 (1991).
23. The physical distance between sDF-1 and sDF-2 was estimated from PCR amplification of these markers from overlapping YACs reported in (*17*). The smallest region of overlap among the YACs that contained both sDF-1 and sDF-2 is approximately 100 kb. For example, both markers were present in YAC yWXD630 but absent in YACs yWXD501 and yWXD37. sDF-2 (but not sDF-1) was present in yWXD250, and sDF-1 (but not sDF-2) was present in yWXD364 and yWXD427.
24. B. O. Bengtsson and P. N. Goodfellow, *Ann. Hum. Genet.* **51**, 57 (1987).
25. R. M. Speed, *Hum. Genet.* **78**, 260 (1988).
26. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985).
27. NIH/CEPH Collaborative Mapping Group, *ibid.* **258**, 67 (1992).
28. Supported in part by NIH grants HG00100 and HG00201 (H.D.-K.). We thank R. Veile for assistance with in situ hybridization studies, T. Steinbrueck and R. Weaver for assistance with linkage map construction, and H. F. Willard for helpful comments on the manuscript.

18 August 1992; accepted 26 October 1992

DNA Sequencing by Primer Walking with Strings of Contiguous Hexamers

Jan Kieleczawa, John J. Dunn, F. William Studier

When template DNA is saturated with a single-stranded DNA binding protein (SSB), strings of three or four contiguous hexanucleotides (hexamers) can cooperate through base-stacking interactions to prime DNA synthesis specifically from the 3′ end of the string. Under the same conditions, priming by individual hexamers is suppressed. Strings of three or four hexamers representing more than 200 of the 4096 possible hexamers primed easily readable sequence ladders at more than 75 different sites in single-stranded or denatured double-stranded templates 6.4 kilobases to 40 kilobase pairs long, with a success rate of 60 to 90 percent. A synthesis of 1 micromole of hexamer supplies enough material for thousands of primings, so multiple libraries of all 4096 hexamers could be distributed at a reasonable cost. Such libraries would allow rapid and economical sequencing. Automating this strategy could increase the speed and efficiency of large-scale DNA sequencing by at least an order of magnitude.

Improvement in the efficiency of nucleotide sequencing is needed if the entire human genome is to be sequenced in the next 15 years at reasonable cost (*1*). A potentially efficient method is enzymatic sequencing by primer walking. A primer within a segment of known sequence (such as a vector sequence) is used to extend the sequence into an unknown region. The sequence thus determined is in turn used to select a primer to extend the sequence further, and this process is repeated until the sequence of the entire molecule has been determined. Advantages of primer walking are that the entire sequence can be determined on a single preparation of template DNA without subcloning and that the sequence can be determined with a minimum number of sequencing reactions.

A disadvantage of primer walking has been the inconvenience and expense of having to synthesize a primer for each sequencing reaction. A proposed solution is to use primers that are short enough that a manageable library of primers would allow any DNA molecule to be sequenced entirely with primers selected from the library (*2*). A potential way to obtain specific priming with a small library of primers is to generate longer, more specific primers from combinations of shorter, less specific ones. One proposal for doing this is by template-directed ligation (*3*), which would be compatible with current sequencing procedures. Efficient ligation requires that oligonucleotides pair at adjacent sites in the template DNA (*4*). We refer to sets of contiguous oligonucleotides as strings.

In trying to optimize conditions for specific priming by short oligonucleotides and to develop conditions for template-directed ligation, we discovered that saturating the template DNA with SSB stimulated strings of three or more unligated hexamers to prime specifically at the position of the string and at the same time suppressed priming by individual hexamers or by many pairs of contiguous hexamers. This stimulation of priming required that the hexamers be contiguous: priming by sets of three or four hexamers whose complements were separated by one or two nucleotides was suppressed by added SSB.

Conditions for priming by hexamer strings. The DNA polymerase used for sequencing was modified T7 DNA polymerase (*5*). The SSB of *Escherichia coli*, the gene *32* protein of bacteriophage T4, and the gene *2.5* protein of bacteriophage T7 (*6*) each stimulated good sequence ladders, but the *E. coli* SSB gave the best initial results and was used for the experiments described here. We used single-stranded M13 DNA (6407 nucleotides) or M13mp18 DNA (7250 nucleotides) (*7*) and one string of four hexamers (Table 1, A4–A1) to establish optimal conditions for priming by hexamer strings (*8*).

Standard reaction conditions were derived from the two-step Sequenase labeling and termination protocols for sequencing with ^{35}S label (U.S. Biochemical). An equilibration reaction (10 μl) typically contained 0.7 μg (33 nM) of M13 DNA, 3 μg of SSB (16 μM SSB monomer), and 50 pmol (5 μM) of each hexamer in 40 mM tris-HCl (pH 7.5), 50 mM NaCl, and 10 mM $MgCl_2$, plus 5% glycerol contributed by the stock solution of SSB. After equilibration for at least 5 min at 0°C, labeling was for 5 min at 0°C, followed by termination for 5 min at 37°C (*9*). We added SDS (0.1%) to the stop solution to prevent SSB from interfering with electrophoresis of the DNA on sequencing gels.

Labeling intensity was reduced at 5 or 20 mM $MgCl_2$ relative to 10 mM but was insensitive to NaCl concentration between 40 and 100 mM. The priming complex appears to reach equilibrium in 2.5 min at 0°C, whether SSB is added to an equilibrated mixture of DNA and hexamers or hexamers are added to an equilibrated mixture of DNA and SSB, as judged by intensity of labeling at different times after the last addition.

SSB concentration. *Escherichia coli* SSB

Biology Department, Brookhaven National Laboratory, Upton, NY 11973.