Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution

David C. Page*‡, Mary E. Harper†§, Jamie Love† & David Botstein*

* Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
† The Agouron Institute, 505 Coast Boulevard South, La Jolla, California 92037, USA
‡ Present address: Whiteside Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA
§ Present address: Laboratory of Tumor Cell Biology, National Cancer Institute, Bethesda, Maryland 20205, USA

DXYS1, *a site showing greater than 99% DNA sequence homology between the human X and Y chromosomes, maps to the X long arm and to the Y short arm. In great apes, sequences homologous to* DXYS1 *are found only on the X chromosome. These findings suggest an X–Y transposition during human evolution.*

IT is widely held that the mammalian X and Y chromosomes are in part homologous as a consequence of their evolution from a completely homologous pair[1,2]. Terminal pairing of the short arms of the human X and Y chromosomes during meiosis is taken as evidence of such partial homology[1,3–8]. Here we demonstrate that human X–Y sequence homology is not limited to the pairing portions of the sex chromosomes. Moreover, X–Y homology at the locus *DXYS1* appears to be the result of a recent transposition from the X to the Y chromosome.

Chromosomal sublocalization of *DXYS1*

DXYS1 was the first site of single-copy DNA sequence homology to be found between the human X and Y chromosomes[9,10]. A *Taq*I restriction fragment length polymorphism (RFLP)[11] at *DXYS1* consists of allelic X-linked fragments of 11 and 12 kilobases (kb) and a male-specific (Y-specific) 15-kb fragment. The complete association of the 15-kb fragment with the Y chromosome and the strictly X-linked inheritance of the 11- and 12-kb fragments imply that X–Y recombination does not normally occur (or is lethal) at *DXYS1*. Given the long-standing hypothesis that meiotic recombination between the pairing regions of X and Y chromosomes occurs frequently[12], we set out to determine whether *DXYS1* maps to the pairing regions of the sex chromosomes.

To localize *DXYS1* on the X and Y chromosomes, we carried out *in situ* hybridization of a probe defining *DXYS1* [pDP31, a subclone of probe 5 (Fig. 2*a*) into pBR322; ref. 9] to human metaphase chromosomes. As illustrated by the representative cell in Fig. 1*a*, pDP31 showed significant hybridization to the proximal portion of the long arm of the X chromosome and to the short arm of the Y chromosome. Of 72 cells hybridized at 100 ng ml^{-1} probe DNA and exposed for 5–11 days, 47% exhibited grains on the proximal long arm of the X chromosome and/or on the short arm of the Y chromosome. These grains represented 20% of all label observed in the 72 cells. Experiments were then done to refine the sublocalization of *DXYS1* by hybridization of the ^{3}H-labelled probe at a low concentration (12 ng ml^{-1}) followed by exposure for 46 days. Analysis of grains on labelled X and Y chromosomes (Fig. 1*b*) indicated sublocalization of *DXYS1* to Yp and to a region on the X that includes the lower half of band q13 and the upper half of q21 (most likely Xq13.2–q21.2).

To provide a second, independent sublocalization of *DXYS1* on the human X and Y chromosomes, genomic DNAs were

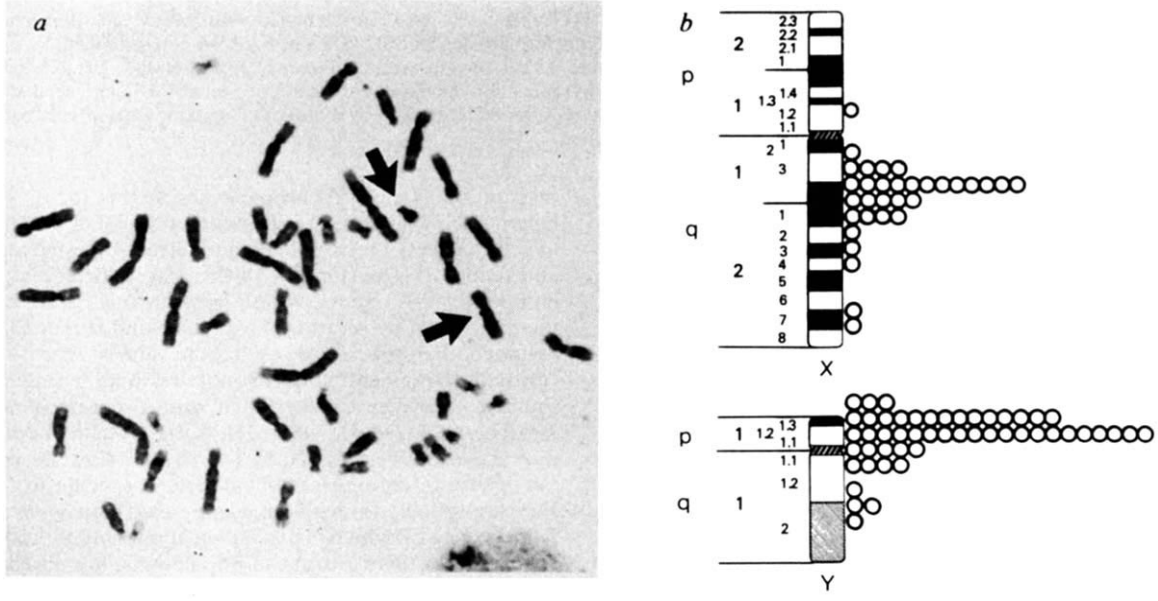

Fig. 1 Sublocalization of *DXYS1* in the human karyotype. *a*, Metaphase cell hybridized with ^{3}H-labelled pDP31 (probe 5 in Fig. 2*a* subcloned into pBR322; ref. 9) illustrating typical labelling of the proximal portion of the long arm of the X chromosome and the short arm of the Y chromosome. *b*, Distribution of grains on metaphase (400-band stage) X and Y chromosomes from 51 labelled cells. pDP31 DNA was ^{3}H-labelled by nick-translation and hybridized at low probe concentrations (12–100 ng ml^{-1}) to male mitotic preparations according to the method previously described[30].

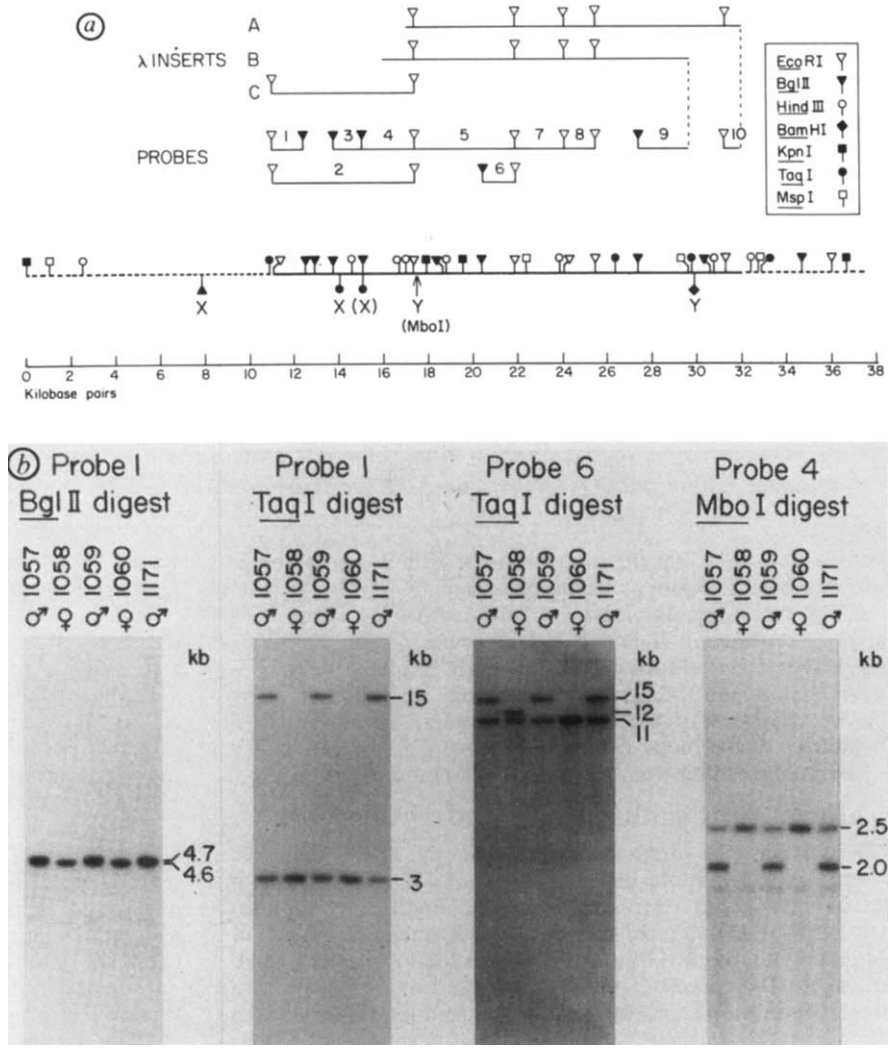

Fig. 2 *a*, Restriction endonuclease mapping of 36 kb of the human X and Y chromosomes at *DXYS1*. At the top are shown the overlapping human DNA inserts of three recombinant λ phage: A, λ-rHs-3-1; B, λ-rHs-4813; and C, λ-rHs-X6.3. The locations of *Eco*RI restriction sites are indicated. Below the phage inserts are 10 restriction fragments purified from these inserts for use as hybridization probes; terminal restriction sites are indicated. Below the probes is shown the restriction map of the X and Y chromosomes at *DXYS1*. Restriction endonuclease cleavage sites common to the X and Y chromosomes are indicated by symbols above the line. Sites specific to either the X or the Y chromosome are indicated below the line. A Y-specific *Mbo*I site (see text) is shown. Sites within the 21-kb region (solid line) spanned by the three λ clones were mapped by direct analysis of those clones and confirmed by blot hybridization using probes 1–10. Sites outside the region (dotted line) were mapped by blot hybridization using probes 1 and 10. Every site was detected or confirmed by blot hybridization of DNAs from at least nine unrelated males and two unrelated females (nine Y chromosomes and 13 X chromosomes). We have previously reported a Y-linked *Msp*I/*Bgl*II restriction fragment length polymorphism at *DXYS1* (ref. 9); this polymorphism is the result of a tandem duplication on the Y chromosome in some males (D.C.P., in preparation). The map of the Y chromosome shown here is that of the nonduplicated form. No new RFLPs were detected during this study. *b*, Differential patterns of hybridization of human male and female DNAs with DNA probes for *DXYS1*. These hybridizations detected restriction sites specific to either the X or the Y chromosome. **Methods:** Recombinant phage λ-rHs-4813, purified from a human genomic library[31], was the original source of hybridization probes for *DXYS1* (ref. 9). Recombinant phages λ-rHs-3-1 and λ-rHs-X6.3 derive from the same genomic library[31] and from an X-specific library[32], respectively. These two phage were identified by plaque hybridization screening[33] using DNA probes purified from the human insert of λ-rHs-4813. Some of these restriction fragments were subcloned into plasmid pBR322 or an abbreviated derivative (pDP322; D.C.P., unpublished) before use as hybridization probes. Human genomic DNAs were prepared from peripheral leukocytes by the method of Kunkel *et al.*[13]. DNA samples were digested with restriction endonuclease, electrophoresed on 0.75% agarose gels, and transferred[34] to AMF Zetapore paper. Restriction fragments purified from phage λ-rHs-3-1, λ-rHs-4813 and λ-rHs-X6.3 were labelled with ^{32}P by nick-translation[35] and hybridized overnight to the filter-bound genomic DNAs at 47 °C in 50% formamide, 5× SSC (1× SSC = 0.15M NaCl, 15 mM Na citrate *p*H 7.4), 1× Denhardt's (0.02% Ficoll 400, 0.02% polyvinyl pyrrolidone, 0.02% bovine serum albumin), 20 mM $NaPO_4$ *p*H 6.6, 50 μg ml^{-1} denatured salmon sperm DNA, and 10% dextran sulphate. Following hybridization, filters were washed three times for 15 min each at 55–60 °C in 0.1× SSC, 0.1% SDS and exposed at −70 °C for 1–7 days with Kodak XAR-5 film backed by a DuPont Lightning-Plus intensifying screen.

obtained from three rodent–human cell hybrids (characterized by G. Bruns, unpublished results) containing portions of the human X chromosome and from three humans whose cells contain portions of the Y chromosome (ref. 13 and K. Smith and collaborators, unpublished results). Gel transfers of *Taq*I digests of these DNAs were hybridized with *DXYS1* probe 5 (see Fig. 2*a*) and scored for the presence or absence of the X- and Y-specific *Taq*I fragments detected by probe 5. The results of these studies map *DXYS1* to Xq1–q22 and to Ycen–pter (data not shown). In addition, *DXYS1* probe 9 (Fig. 2*a*) was applied to the dosage panel used by Kunkel *et al.*[14] to assign DNA segments to regions of the human X chromosome. By this method, *DXYS1* was mapped to Xq13–q24 (data not shown). These results are consistent with the assignment of *DXYS1* to Xq13-q21 and Yp by *in situ* hybridization.

X–Y restriction mapping

The homology between the X and Y chromosomes at *DXYS1* extends for at least 28 kb[9]. In order to quantitate further the X–Y homology at *DXYS1*, we made a comparative restriction map of the X and Y chromosomes at this locus. A series of neighbouring single-copy restriction fragments were hybridized to gel transfers of restriction endonuclease-digested human male and female DNAs (Fig. 2). In this way, a stretch of 36 kb was mapped in this region. Where no differences were detected in the patterns of hybridization with male and female DNA, it was assumed that the X and Y chromosomes share a common restriction fragment. Indeed, most restriction fragments were found to be identical on the X and Y chromosomes. Of 40 *Eco*RI, *Bgl*II, *Hin*dIII, *Bam*HI, *Kpn*I, *Taq*I and *Msp*I restriction sites that we mapped at *DXYS1*, 36 sites are common to the X and Y chromosomes, one site is specific to the Y, two sites are specific to the X, and one (*Taq*I) site gives rise to the X-linked RFLP which is present on about half of the X chromosomes in Northern European populations, but absent from the Y chromosome. These results establish that the X and Y chromosomes are homologous over an expanse of at least 36 kb at *DXYS1*. We have not detected the termination of this X–Y sequence homology.

Analysis of this region with the restriction enzyme *Mbo*I yields similar results. Of 11 *Mbo*I restriction fragments observed over

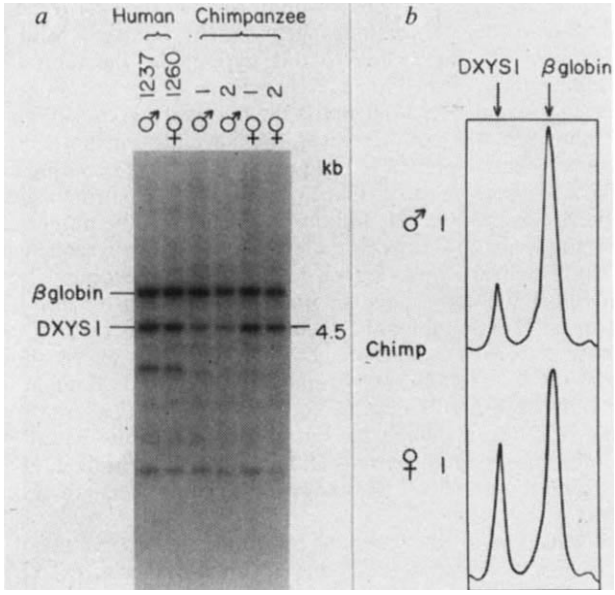

Fig. 3 Male–female dosage at a *DXYS1*-homologous locus in chimpanzee. *a*, Hybridization of *Eco*RI-digested human and chimpanzee male and female DNAs to probes for *DXYS1* and β-globin. Genomic human and chimpanzee DNAs were digested with the restriction endonuclease *Eco*RI, blotted, and hybridized as in Fig. 2*b* with ^{32}P-labelled probe 5 (Fig. 2*a*) and plasmid πSVβ (containing 3.7 kb of human genomic sequence from the β-globin locus; ref. 36). Probe detects a 4.5-kb human genomic *Eco*RI fragment common to the human X and Y chromosomes (Fig. 2*a*). It is also homologous to a 4.5-kb *Eco*RI fragment in chimpanzee. Plasmid πSVβ detects a 5.3–kb *Eco*RI fragment in both human and chimpanzee. In the human, this *Eco*RI fragment contains most of the β-globin gene and derives from chromosome 11. πSBβ also hybridizes to two smaller *Eco*RI fragments in both human and chimpanzee (in man, these are known to correspond to the remainder of the β-globin locus and to the δ-globin locus). *b*, Densitometric tracings of portions of two lanes (chimpanzee male no. 1 and female no. 1) of an autoradiogram produced without an intensifying screen but otherwise identical to that in *a*.

~22 kb at *DXYS1*, nine are common to the X and Y chromosomes, one is specific to the Y chromosome, and one is specific to the X chromosome. Thus, of at least 12 *Mbo*I sites represented by these fragments, one site is specific to the Y (Fig. 2*a*), and all other sites are common to the X and Y chromosomes. The results of several hybridization experiments that revealed restriction sites specific to either the X or the Y chromosome are shown in Fig. 2*b*.

Sequence divergence

The fraction of nucleotides at which a pair of homologous DNAs differ can be estimated by comparing the restriction maps of the two sequences. Given the comparative X–Y restriction map of *DXYS1* (Fig. 2*a*), we set out to determine the fraction of nucleotides at which the X and Y chromosomes differ in this region. This fraction was estimated using the iterative 'nucleotide counting' method of Nei and Tajima[15]. In using this method, we assume that restriction cleavage sites are created or eliminated solely by substitutions of base pairs. The calculation involves only those mapped restriction sites which fall within the 21 kb at *DXYS1* which we have cloned, and which is mapped in its entirety for seven enzymes (Fig. 2*a*). The data on *Mbo*I fragments are not included in the calculation as they do not represent mapped sites. Drawing on data for the seven other enzymes, we estimate that the human X and Y chromosomes differ at 0.83 ± 0.54% of their nucleotides at *DXYS1* (Table 1). It has been argued that CpG nucleotide dimers (such as occur within the recognition sequences of the enzymes *Taq*I and *Msp*I) are remarkably prone to mutation in humans[16]. If the *Taq*I and *Msp*I data are eliminated from the calculation of X–Y sequence divergence, the estimate of that divergence falls to 0.37 ± 0.38%.

Table 1 Calculations of X–Y sequence divergence at *DXYS1*

Distribution of restriction sites on the X and Y chromosomes

6-base enzymes	m_X	m_Y	m_{XY}	4-base enzymes	m_X	m_Y	m_{XY}
*Eco*RI	6	6	6	*Taq*I	3.5	2	2
*Bgl*II	8	8	8	*Msp*I	2	2	2
*Hin*dIII	6	6	6	Total	5.5	4	4
*Bam*HI	0	1	0				
*Kpn*I	2	2	2	$\bar{m} = 4.75$			
Total	22	23	22				

$\bar{m} = 22.5$

Combined 6- and 4-base enzyme data: $d = 0.0083 \pm 0.0054$

6-base enzyme data only: $d = 0.0037 \pm 0.0038$

The fraction of nucleotides at which the X and Y chromosomes differ at *DXYS1* was estimated using data drawn from the X–Y comparative restriction map of Fig. 2*a* and formulae for 'nucleotide counting' of Nei and Tajima[15]:

$$d = d_1 \frac{\sum_i r_i(\bar{m}_i - m_{XYi})/[\{1-(1-d_1)^{r_i}\}\{2-(1-d_1)^{r_i}\}]}{\sum_i r_i \bar{m}_i/[2-(1-d_1)^{r_i}]}$$

where $d_1 = 1 - \hat{S}^{1/r}$,

$$\hat{S} = \frac{2m_{XY}}{m_X + m_Y},$$

i refers to the ith type of enzyme (in this case there are two types, 4-base and 6-base enzymes),

$$\bar{m} = (m_X + m_Y)/2$$

and where the standard error of d is

$$s_d = \left(\frac{1}{\sum_i \frac{2r_i^2 \bar{m}_i S_i}{(2-S_i)(1-S_i)}} \right)^{1/2}$$

where

$$S_i = e^{-r_i d}$$

where d is the estimated fraction of nucleotides at which X and Y differ; r, the number of base pairs recognized by a given restriction enzyme (in this case, 6 or 4); m_X and m_Y, the number of restriction sites on, respectively, the X and Y chromosomes; m_{XY}, the number of restriction sites shared by the X and Y chromosomes. Note that the polymorphic *Taq*I site on the X chromosome contributes 0.5 to the value of m_X for *Taq*I.

Homologous sequences in great apes

Our estimate of X–Y sequence divergence at *DXYS1* (0.37 ± 0.38%) is lower than estimates of the divergence of the human and chimpanzee genomes as a whole (1–3%)[17–20]. Assuming that the X and Y chromosomes do not evolve more slowly than the autosomes, this suggests one of two possibilities. First, *DXYS1* could represent an evolutionarily ancient homology between the X and Y chromosomes and some mechanism could retard or repair X–Y divergence at that locus. Alternatively, the presence of *DXYS1* sequences on both the X and Y chromosomes might be a phenomenon limited to recent evolutionary times. According to this hypothesis, *DXYS1* sequences have been present on both the X and Y chromosomes only for a period comparable to or shorter than that since humans diverged from chimpanzees.

To distinguish between the two possibilities, genomic DNAs from the great apes were examined for sequences homologous to *DXYS1*. If *DXYS1* represents an ancient homology between the X and Y chromosomes, one would expect to find *DXYS1*-homologous sequences on both the X and Y chromosomes in all the great apes. However, if X–Y homology at *DXYS1* is of recent origin, one would expect to find *DXYS1*-homologous sequences on either the X or Y chromosome, but not on both.

Filter hybridization experiments revealed that the great apes have *DXYS1*-homologous sequences only on their X chromosomes. Gel transfers of *Eco*RI digests of human and ape DNAs were hybridized with probe 5 (Fig. 2*a*) at moderately high stringency as described in Fig. 2 legend. At least one male and one female of each of three ape species (chimpanzee, gorilla

Table 2 Dosage of *DXYS1*-homologous sequences in male and female ape DNAs

	Sex	Autosomal standard: Ratio of areas *DXYS1*/β-globin	X-linked standard: Ratio of areas *DXYS1*/*DXS9*	No. of *DXYS1* copies per cell
Chimpanzee	M	0.30 (0.5)	0.83 (0.9)	1
	M	0.33 (0.6)	0.88 (1.0)	1
	F	0.59 (1.0*)	0.92 (1.0*)	2
	F	0.60 (1.0)	0.89 (1.0)	2
Gorilla	M	0.36 (0.6)	1.06 (1.1)	1
	F	0.62 (1.0*)	0.96 (1.0*)	2
	F	0.75 (1.2)	0.97 (1.0)	2
Orangutan	M	—	1.17 (1.2)	1
	F	—	1.00 (1.0*)	2
Human	M	0.60 (1.1)	1.95 (2.4)	2
	F	0.55 (1.0*)	0.83 (1.0*)	2

Genomic human, chimpanzee, gorilla and orangutan DNAS were digested with *Eco*RI, blotted and hybridized with probes for *DXYS1* and β-globin or *DXYS1* and *DXS9* (RC8, ref. 21) as described in Fig. 3 legend. To ensure a linear relationship of filter-bound radioactivity to absorbance on the autoradiogram, no intensifying screen was used. As in Fig. 3*b*, densitometric scans of the autoradiograms were made. The areas under the peaks corresponding to *DXYS1* and β-globin or *DXYS1* and *DXS9* were integrated, and, for each species, the ratios of these areas were normalized to the female, indicated by an asterisk. (*DXYS1*/β-globin ratios could not be calculated for the orangutans because of poor separation of the *DXYS1* and β-globin peaks.) The number of copies of *DXYS1* per cell was calculated assuming that females have two copies per cell of both β-globin and *DXS9* and that males have two copies per cell of β-globin and one of *DXS9*.

and orangutan) were examined. In humans, probe 5 detects a 4.5-kb *Eco*RI fragment that is common to the X and Y chromosomes. Thus there are two copies per cell in both males and females. All ape DNAs contained sequences homologous to *DXYS1* (for example, Fig. 3*a*). In the gorilla and chimpanzee, probe 5 detects a 4.5-kb *Eco*RI fragment, and in the orangutan, probe 5 detects two *Eco*RI fragments of 3.4 and 5.3 kb. Unlike the human, however, in all three ape species the dosage of *DXYS1*-homologous sequences is a function of sex. As judged using an autosomal (β-globin) or X-linked (*DXS9*, probe RC8; ref. 21) reference standard, in each of the three ape species, females contain twice as many copies of the *DXYS1*-homologous sequence as do males (Table 2, Fig. 3); no male-specific *Eco*RI (or *Taq*I, data not shown) fragment was found in any of the apes. The most economical interpretation of these findings is that *DXYS1*-homologous sequences occur only on the X chromosome among the great apes. Thus, it seems that X–Y homology at *DXYS1* is of recent origin and, among extant species, is found only in man.

Conclusions

Homology of single-copy DNA sequences on mammalian X and Y chromosomes was first reported at the human locus *DXYS1* (ref. 9), and there is now increasing evidence of substantial homology of single-copy sequences of no known function on the human X and Y chromosomes[22]. However, although a monoclonal antibody elicited by an antigen (12E7) encoded by the human X chromosome cross-reacts with an antigen coded for by the human Y chromosome[23], there is little rigorous evidence of homologous functional genes on the X and Y chromosomes of any mammal.

Our mapping of *DXYS1* to Xq13–q21 is consistent with the findings of other workers. J. Weissenbach (personal communication) has, by means of a different panel of rodent–human cell hybrids containing portions of the human X chromosome, assigned *DXYS1* to Xq12–q22. Drayna *et al.*[24] have recently demonstrated a genetic distance of about 12 centimorgans between *DXYS1* and *DXS17* (S21), which in turn has been mapped to Xq21–q22.

The findings reported here are not consistent with a model that limits X–Y sequence homology to the short arms of the X and Y chromosomes. Terminal portions of the short arms of the X and Y chromosomes pair during meiosis in human spermatogenesis[3–7], and this observation provided the basis for the prevailing model of homology between the human X and Y chromosomes[1,8]. According to this hypothesis, the terminal, pairing portions of the X and Y chromosomes are homologous at the DNA sequence level, while the non-pairing, or 'differential', portions are not. However, we have demonstrated the presence of single-copy DNA sequences (*DXYS1*) homologous to the Y chromosome at Xq13–q21, far outside the pairing region of the X chromosome. In addition, there seem to be numerous other single- or low-copy number[22] and repetitive[25] sequences on Xq which are homologous to the Y chromosome. Thus, although single-copy sequence homology between Xp and Yp may exist, X-Y homologous sequences are clearly not limited to those regions. Furthermore, *DXYS1* sequences on the short arm of the Y chromosome, virtually all of which is thought to pair with the X (ref. 26), have no homologue on the short arm of the X. Thus, although the pairing regions of the X and Y chromosomes may be partially homologous, it is unlikely that there is continuous X–Y homology throughout the length of the pairing region.

We have previously reported[9] our failure to observe meiotic recombination between the X and Y chromosomes at *DXYS1*. Given the location of *DXYS1* at Xq13–q21, it is now clear that single reciprocal cross-overs between the X and Y at *DXYS1* would destroy the integrity of both the X and Y chromosomes.

It is widely accepted that the mammalian X and Y chromosomes evolved from a completely homologous pair[2]. It has been assumed, as a corollary, that if limited homology of the human X and Y chromosomes exists, the homology should be ancient in an evolutionary sense[1]. The findings reported here demonstrate that this need not be the case. Among extant species, *DXYS1*-homologous sequences apparently occur on both the X and Y chromosomes only in man; in chimpanzee, gorilla and orangutan, *DXYS1*-homologous sequences are found only on the X chromosome. Thus, the simplest explanation is that *DXYS1*-homologous sequences were transposed from the X chromosome to the Y chromosome after the human line diverged. (To argue that the common ancestor of all four species had *DXYS1*-homologous sequences on both X and Y chromosomes, one must appeal to the loss of *DXYS1*-homologous sequences from the Y chromosome three times during evolution.) This hypothesis is consistent with the finding that the divergence of human X and Y DNA sequences at *DXYS1* is somewhat less than the divergence of the human genome from that of its closest relative, the chimpanzee. It is estimated that the human and chimpanzee lines diverged between 4 and 8 Myr ago[20,27,28]. Thus, *DXYS1* sequences have been present on both the X and Y chromosomes for less than—perhaps much less than—8 Myr.

The full extent of the X-to-Y transposition that occurred at *DXYS1* during human evolution is unknown, but it is clear that at least 36 kb of DNA was involved. To our knowledge, this X-to-Y transposition has not been detected by comparative light microscopic examination of banded human and ape chromosomes[29]. The recent finding[22,25] of other human X–Y homologous sequences on Xq suggests either that the transposition was much larger than 36 kb or that it was in some way directed by preexisting X–Y homologies.

We thank Tom Maniatis for a human genomic DNA library; Kay Davies for a human X chromosome library and for phage RC8; Gail Bruns for cell hybrid DNAs; Kirby Smith for ape and human DNAs; Lou Kunkel for human DNAs; Stu Orkin for plasmid πSVβ; and Arlene Wyman and Linda Wolfe for preparing some of the human DNAs and Southern blots. We particularly thank Masatoshi Nei for assistance with the calculations of nucleotide divergence, and we appreciate the advice of Luca Cavalli-Sforza. We thank Connie Holm, Arlene Wyman and Linda Wolfe for useful criticisms of the manuscript. This work was aided by grants from the NIH (to D.B., Arlene Wyman and Melvin Simon) and from the Wills Foundation (to Arlene Wyman). D.C.P. received research stipends from Harvard Medical School, and M.E.H. received a NIH National Research Service Award.

Received 2 April; accepted 2 July 1984.

1. Polani, P. E. *Hum. Genet.* **60**, 207–211 (1982).
2. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer, New York, 1967).
3. Koller, P. C. *Proc. R. Soc. Edinb.* **B57**, 194–214 (1937).
4. Ford, C. E. & Hamerton, J. L. *Nature* **178**, 1020–1023 (1956).
5. Solari, A. J. & Tres, L. L. *J. Cell Biol.* **45**, 43–53 (1970).
6. Pearson, P. L. & Bobrow, M. *Nature* **226**, 959–961 (1970).
7. Moses, M. J., Counce, S. J. & Paulson, D. F. *Science* **187**, 363–365 (1975).
8. Burgoyne, P. S. *Hum. Genet.* **61**, 85–90 (1982).
9. Page, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 5352–5356 (1982).
10. Page, D., Wyman, A. & Botstein, D. in *Banbury Rep. 14: Recombinant DNA Applications to Human Disease* (eds Caskey, C. & White, R.) 267–273 (Cold Spring Harbor, New York, 1983).
11. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 317–331 (1980).
12. Darlington, C. D., Haldane, J. B. S. & Koller, P. C. *Nature* **133**, 417 (1934).
13. Kunkel, L. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **74**, 1245–1249 (1977).
14. Kunkel, L. M., Tantravahi, U., Eisenhard, M. & Latt, S. A. *Nucleic Acids Res.* **10**, 1557–1578 (1982).
15. Nei, M. & Tajima, F. *Genetics* **105**, 207–217 (1983).
16. Barker, D., Schafer, M. & White, R. *Cell* **36**, 131–138 (1984).
17. King, M.-C. & Wilson, A. C. *Science* **188**, 107–116 (1975).
18. Benveniste, R. E. & Todaro, G. J. *Nature* **261**, 101–108 (1976).
19. Deininger, P. L. & Schmid, C. W. *Science* **194**, 846–848 (1976).
20. Sibley, C. & Allquist, J. E. *J. molec. Evol.* **20**, 2–15 (1984).
21. Murray, J. M. *et al. Nature* **300**, 69–71 (1982).
22. Weissenbach, J., Geldwerth, D., Guellan, G., Fellous, M. & Bishop, C. *Cytogenet. Cell Genet.* **37**, 604 (1984).
23. Goodfellow, P. *et al. Nature* **302**, 346–349 (1983).
24. Drayna, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2836–2839 (1984).
25. Mueller, C. R., Davies, K. E. & Ropers, H. H. *Cytogenet. Cell Genet.* **37**, 545 (1984).
26. Solari, A. J. *Chromosoma* **81**, 315–337 (1980).
27. Sarich, V. M. & Cronin, J. E. in *Molecular Anthropology* (eds Goodman, M. & Tashian, R. E.) 141–170 (Plenum, New York, 1976).
28. Johanson, D. C. & White, T. D. *Science* **203**, 321–330 (1979).
29. Yunis, J. J. & Prakash, O. *Science* **215**, 1525–1530 (1982).
30. Harper, M. E. & Saunders, G. F. *Chromosoma* **83**, 431–439 (1981).
31. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. *Cell* **15**, 1157–1174 (1978).
32. Davies, K. E., Young, B. D., Elles, R. G., Hill, M. E. & Williamson, R. *Nature* **293**, 374–376 (1981).
33. Benton, W. & Davis, R. *Science* **196**, 180–182 (1977).
34. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
35. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
36. Triesman, R., Orkin, S. H. & Maniatis, T. *Nature* **302**, 591–596 (1983).

Correlation between segmental mobility and the location of antigenic determinants in proteins

E. Westhof*, D. Altschuh*, D. Moras*, A. C. Bloomer†, A. Mondragon†, A. Klug† & M. H. V. Van Regenmortel*

* Institut de Biologie Moleculaire et Cellulaire, CNRS, 15 rue Descartes, 67000 Strasbourg, France
† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Most continuous antigenic determinants of tobacco mosaic virus protein (TMVP), myoglobin and lysozyme correspond to those surface regions in the protein structure, as determined by X-ray crystallography, which possess a run of high-temperature factors along the polypeptide backbone, that is, a high segmental mobility. The mobility of an antigenic determinant may make it easier to adjust to a pre-existing antibody site not fashioned to fit the exact geometry of a protein. The correlation found between temperature factors and antigenicity is better than that between hydrophilicity and antigenicity.

THE antigenic reactivity of proteins resides in restricted parts of the molecule known as antigenic determinants or epitopes. Antigenic determinants represent the accessible patches on the surface of a native protein that interact with the binding sites of antibody molecules. Epitopes made up of a continuous sequence of amino acid residues are called 'continuous determinants' while 'discontinuous determinants' consist of residues that are not contiguous in the primary structure but which are brought together by the folding of the polypeptide chain[1]. The continuous epitopes comprise five to eight residues and their uniform size in different proteins reflects the constant size of the complementary antibody combining sites.

The localization of continuous determinants in proteins is based on measurements of the antigenic reactivity of natural or synthetic peptides of the molecule[2,3]. Usually, the fragments are tested for their ability to inhibit the reaction between antibodies and the whole antigen, and very high molar ratios of peptide to intact antigen are often required to achieve significant inhibition; this is usually explained by assuming that peptides exist in solution in a variety of random conformations and that only the rare conformations which approximate to that found in the native protein will bind to the antibody[4,5]. Both the N- and C-termini of different proteins often show higher than average antigenic activity[6–14], presumably because they are located at the surface and have a high relative flexibility[15].

In the last few years, the importance of the internal dynamics of protein molecules in relation to function has been emphasized[16–18]. This appreciation prompted us to examine whether there is a more general link between the local mobility of short segments of proteins and the antigenic structure. For this study we selected three proteins: the coat protein of tobacco mosaic virus (TMVP), myoglobin and lysozyme, which have been studied extensively by X-ray crystallography[19–21]. Present refinement methods[22,23] give not only precise atomic coordinates but also atomic temperature factors (B values). The temperature factor represents the mean-square displacement of each atom and, when plotted against residue number, provides a graphic image of the degree of mobility existing along the polypeptide chain. Such plots were used to examine the possible correlation between the mobile segments of these proteins and the locations of their epitopes.

Tobacco mosaic virus protein

The antigenic properties of the coat protein of tobacco mosaic virus have been extensively studied by several groups[24]. Milton and Van Regenmortel[25] originally reported that TMVP possesses five continuous antigenic determinants situated in tryptic peptides I, IV, VIII and XII. Peptide I contains two epitopes located in residues 1–10 and 34–39. The epitope in peptide VIII has been studied extensively by Benjamini[26], who located it in residues 105–112, although some residual activity is also found in the shorter pentapeptide, 108–112. Two additional antigenic determinants were recently identified[10] in residues 55–61 and 80–90, by comparing the inhibitory activity of tryptic peptides II, III (42–61) and VI (72–90) from three strains of the virus. Three peptides corresponding to residues 72–77, 129–134 and 142–147 that were synthesized because they correspond to accessible regions on the surface of the TMVP molecule were found to have no significant (that is, above the 10% level) antigenic activity in enzyme-linked immunosorbent assay (ELISA) inhibition tests.

Thus, the many studies of the antigenic structure of TMV[10,26,27] have so far revealed only seven continuous epitopes in the protein. The immunoassays were carried out at very low