Homologous Ribosomal Protein Genes on the Human X and Y Chromosomes: Escape from X Inactivation and Possible Implications for Turner Syndrome

Elizabeth M. C. Fiehr,$ Peggy Beer-Romero,* Laura G. Brown,* Anne Ridley,* John A. McNeill,‡ Jeanne Bentley Lawrence,‡ Huntington F. Willard,§ Frederick R. Bieber,† and David C. Page*†

*Howard Hughes Research Laboratories
Whitehead Institute for Biomedical Research
Nine Cambridge Center
Cambridge, Massachusetts 02142

and Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

†Department of Cell Biology
University of Massachusetts Medical School
Worcester, Massachusetts 01655

‡Department of Genetics
Stanford University
Stanford, California 94305

§Departments of Pathology
Brigham and Women’s Hospital and
Harvard Medical School
75 Francis Street
Boston, Massachusetts 02115

Summary
We have isolated two genes on the human sex chromosomes, one on the Y and one on the X, that appear to encode isoforms of ribosomal protein S4. These predicted RPS4Y and RPS4X proteins differ at 19 of 263 amino acids. Both genes are widely transcribed in human tissues, suggesting that the ribosomes of human males and females are structurally distinct. Transcription analysis revealed that, unlike most genes on the X chromosome, RPS4X is not dosage compensated. RPS4X maps to the long arm of the X chromosome (Xq), where no other genes are known to escape X inactivation. Curiously, RPS4X maps near the site from which the X-inactivating signal is thought to emanate. On the Y chromosome, RPS4Y maps to a 90 kb segment that has been implicated in Turner syndrome. We consider the possible role of RPS4 haploinsufficiency in the etiology of the Turner phenotype.

Introduction
In human females, each chromosome is paired with a homolog of equivalent size and genetic content. In human males, however, the X and Y chromosomes form a strikingly heteromorphic pair. The banding patterns of the two chromosomes are dissimilar, and the X is 2–3 times as long as the Y. While a few hundred genes have been mapped to the human X chromosome, there is at present good evidence for the existence of only a handful of genes on the Y chromosome (McKusick, 1990).

One might have supposed that genes on the Y would have exclusively male-specific functions. However, of the few Y-chromosomal genes that have been cloned, two (MIC2 and the GM-CSF receptor) have identical counterparts on the X (Goodfellow et al., 1986; Gough et al., 1990), while at least one (ZFY) has a closely related but nonidentical homolog on the X (Schneider-Gädicke et al., 1989a). These observations lend credence to the hypothesis that the human X and Y chromosomes have evolved from what was once a pair of autosomes (Ohno, 1967).

Here we report the discovery of a gene, RPS4Y, which is located on the short arm of the human Y chromosome and which has a closely related but nonidentical homolog, RPS4X, on the X chromosome. Most genes on the human X are dosage compensated via transcriptional silencing of all but one X chromosome per cell, a process called X inactivation (Lyon, 1988). We demonstrate that RPS4X is among a small minority of X-chromosomal genes that escape X inactivation. Also among this small set of noninactivated genes are the X-Y-homologous genes MIC2 and ZFX (Goodfellow et al., 1984; Schneider-Gädicke et al., 1990a). In contrast to other Y-homologous, noninactivated genes, RPS4X maps to the long arm of the X chromosome (Xq). More precisely, RPS4X maps near the X inactivation center (XIC), the Xq locus from which emanates the poorly understood signal that inactivates the entirety of the chromosome (Therman et al., 1979; Mattie et al., 1981). This map location is particularly striking given that no other gene on Xq is known to escape X inactivation.

The eukaryotic ribosome is a massive structure containing four RNA species and about 80 distinct proteins (Wool, 1979). The RPS4X and RPS4Y genes appear to encode isoforms of one of the approximately 33 proteins found in the small subunit of the ribosome. No other mammalian ribosomal protein is known to have isoforms or to be encoded by more than one gene. Since one of the isoforms, RPS4Y, is encoded by the Y chromosome, which is present in males but absent in females, there may be a structural difference between the ribosomes of males and females.

Turner syndrome is a complex human phenotype (Ullrich, 1930; Turner, 1938) most commonly seen in association with a 45,X karyotype (Ford et al., 1959). It has been proposed that the Turner phenotype is the result of monosomy for a gene or genes common to the X and Y chromosomes (Ferguson-Smith, 1965). In the Discussion, we consider the possibility that haploinsufficiency of the RPS4 genes contributes to the Turner phenotype.

Results

An Abundant Transcript from Interval 1A1D of the Human Y Chromosome
Interval 1A, the smallest segment of the human Y chromosome known to confer a completely male phenotype, is
Figure 1. An Abundant Transcript from Interval 1A1B of the Human Y Chromosome

The distal short arm of the Y chromosome is diagrammed above, oriented with respect to the short arm telomere (pter) and centromere (cen). Intervals 1A1A, 1A1B, 1A2, 1B, and 1C are defined by deletion analysis (Page et al., 1967, 1990) and show strictly sex-linked inheritance. The pseudoautosomal region undergoes recombination with the X chromosome during male meiosis. The "DXYS1-like" sequences comprising intervals 1B and 1C are homologous to but do not recombine with sequences on the long arm of the X chromosome. Underneath the map of the chromosome are black bars indicating the portions of the Y chromosome present in XX male LGL203 (case 6 in de la Chapelle et al., 1964; Page et al., 1967) and in X,t(Y;22) female WHT1013 (Page et al., 1967, 1990). Distances from the boundary of the pseudoautosomal region (Ellis et al., 1969) are shown in kilobases. Below is a Northern blot of poly(A)+ RNAs prepared from human cell lines and hybridized with the cDNA insert of plasmid pDP1276. Transcripts of 1.0 kb (abundant) and 10 kb (barely visible) are detected not only in normal males but also in XX male LGL203 and in X,t(Y;22) female WHT1013, who both carry interval 1A1B. Lanes contain RNAs from the following cell lines (left to right; B lymphoblastoid cells unless otherwise indicated): normal male, normal female, LGL201 (father of LGL203; fibroblast), LGL203 (fibroblast), LGL202 (mother of LGL203; fibroblast), WHT101 (father of WHT1013), WHT1013, and WHT101P (mother of WHT1013). Each lane contains approximately 5 μg of poly(A)+ RNA.

defined as the portion of the Y chromosome present in XX male LGL203. Approximately 280 kb in length, interval 1A is located immediately proximal to the pseudoautosomal region, on distal Yp. We previously cloned the entirety of the interval by chromosome walking (Page et al., 1987, 1990).

Portions of interval 1A were found to be deleted in a female (WHT1013) who had a reciprocal translocation between the Y and chromosome 22. As shown in Figure 1, the complex deletion in this X,t(Y;22) female divides 1A into intervals 1A1A, 1A1B, and 1A2. Interval 1A1B is present in the X,t(Y;22) female, while 1A1A and 1A2 are absent (Page et al., 1987, 1990). As argued in the Discussion, 1A1B probably contains one or more Turner genes.

While searching for genomic sequences in interval 1A1B, we discovered that a 1.6 kb segment of single-copy genomic DNA detected transcripts when hybridized to Northern blots of human cellular poly(A)+ RNAs. This probe, the insert of plasmid pDP1241, detected an abundant 1.0 kb transcript and a much less abundant 10 kb transcript, both present in XY male but not XX female fibroblasts and lymphoblastoid cells (data not shown).

To characterize these transcripts, a cDNA library prepared from a human 49,XY,YYY male lymphoblastoid cell line (Sirota et al., 1981) was screened with probe pDP1241. The hybridizing clones were of two distinct classes, differing in the intensity of hybridization with pDP1241. We further analyzed five of the "strong" positives and three of the "weak" positives.

Among the five strongly positive clones, the longest cDNA insert (0.9 kb, subcloned as plasmid pDP1278) was analyzed most thoroughly. First, we hybridized pDP1278 to Northern blots. As shown in Figure 1, this cDNA probe detects the 1.0 and 10 kb transcripts previously seen with genomic probe pDP1241. Again, the transcripts are present in XY male but not XX female cells, consistent with the transcripts deriving from the Y chromosome. Both transcripts are present in XX male LGL203 and in X,t(Y;22) female WHT1013, who have in common interval 1A1B of the Y chromosome. We concluded that the 1.0 and 10 kb transcripts are related and that interval 1A1B contains the entirety of the transcription unit(s) from which they are derived. As demonstrated below, the 0.9 kb insert of cDNA pDP1278 corresponds to the abundant 1.0 kb transcript.

The 1A1B Transcript Encodes a 263 Amino Acid Protein

We determined the nucleotide sequence of cDNA pDP1278 and identified a single long open reading frame (Figure 2). The first ATG in this open reading frame (position 1 in the upper portion of Figure 2) occurs in a sequence context that is highly favorable for initiation of translation (Kozak, 1986). As discussed below, comparisons with homologous genes in other organisms provide further evi-
predicted molecular weight of 29,455. The initiator codon is beginning at this ATG, the open reading frame appears to preceded by a untranslated sequence of 12 nucleotides, evidence that this triplet serves as the initiator codon. Be- and a 3' untranslated sequence of 56 nucleotides is fol- encode a protein 263 amino acids in length, with a pre- nucleotides 5' of the poly(A).

The previously observed 10 kb transcript (Figure 1) appears to be an alternately or incompletely spliced product of the same gene. As expected, genomic sequences defined as introns by cDNA pDP1278 (Figure 3) did not detec- the 1.0 kb transcript when hybridized to Northern blots. However, some of these intron probes did detect the 10 kb transcript. One such probe, pDP1056, was used to screen cDNA libraries selected for inserts of 2 kb or greater (Schneider-Gadicke et al., 1989b). The longest of the cDNA inserts isolated, 0.9 kb in length, was subcloned as plasmid pDP1190. When hybridized to Northern blots, pDP1190 detected both the 1.0 and 10 kb transcripts. We detected a silent base pair difference (an A, as compared with the published G) at nucleotide 492. Pyrimidine-rich stretches are found near the cap sites of mammalian ribosomal protein genes (Harhiran et al., 1989). A pyrimidine-rich stretch (underlined) is present in the 5' UTR of RPS4X, as previously observed in the rat homolog (Devi et al., 1989). A canonical polyadenylation sequence to pDP1278. We concluded that the five "strong positive" cDNA clones were independently derived from a single mRNA species.

The four other strongly hybridizing cDNAs were par- tially sequenced. These clones, each of slightly different length, were entirely contained within and identical in se- quence to pDP1278. We concluded that the five "strong positive" cDNA clones were independently derived from a single mRNA species.

To analyze the structure of the genomic locus, we hy- bridized cDNA pDP1278 to Southern blots of restriction- digested DNAs from recombinant phases spanning inter- val 1A1B. The cDNA hybridized to five HindIII fragments distributed across 30 kb of interval 1A1B. As there are no HindIII sites in the cDNA itself, we concluded that the gene is divided into at least five exons. By hybridizing oligonucleotides (based on the cDNA sequence) to South- ern blots of the 1A1B phases, we crudely mapped the positions of splice sites within the cDNA and the positions of exons within the gene (Figure 3). These results confirmed that the gene, which spans at least 23 kb, contains a mini- mum of five exons.
Figure 3. The Structure of the RPS4Y Gene

The schematic represents the RPS4Y cDNA clone pDP1278, broken into exons. The boxes indicate the minimum number of exons (i.e., each box may contain more than one exon) and are not drawn to scale. Coding regions are shaded. Several splices (dashed vertical lines of boxes) were not precisely defined but were inferred from comparison of the cDNA sequence with the restriction map of the genomic locus. One splice site (after cDNA nucleotide 360) was defined by comparison of cDNA pDP1278 with cDNA pDP1190, which is incompletely or alternatively spliced. Genomic DNA fragments hybridizing with RPS4Y oligonucleotides (numbering, with reference to Figure 2, denotes the tenth nucleotide in these 19- or 20-mer oligonucleotides) are shown as black segments. Each segment contains one or more exons. The genomic locus is scaled in kilobases according to the scheme in Figure 1.

Though its most 5′ portion remains to be cloned and sequenced, we have no reason to suspect that the 10 kb transcript is translated.

The 1AlB Gene and an X-Linked Homolog Encode Isoforms of Ribosomal Protein S4

As described above, three of the cDNA clones isolated with 1AlB genomic probe pDP1241 had hybridized weakly in comparison with the five other cDNAs. Among the three weakly positive clones, the longest cDNA insert (10 kb, subcloned as plasmid pDP1284) was studied most extensively. As described below, pDP1284 detects an abundant 1.0 kb transcript when hybridized to Northern blots of RNAs from male or female cells.

We completely sequenced pDP1284 and partially sequenced the other two weakly positive cDNAs. Though the three clones differed slightly in length, they were of otherwise identical sequence. We concluded that the three cDNA clones represent a single mRNA species, which, as argued below, is derived from a gene on the X chromosome.

The nucleotide and predicted amino acid sequences for cDNAs pDP1278 (Y derived) and pDP1284 (X derived) are remarkably similar (Figures 2 and 4). Both appear to encode proteins of 263 amino acid residues. (As with pDP1278, the first in-frame ATG in X-derived cDNA pDP1284 [position 1 in the lower portion of Figure 2] occurs in a very favorable context for translation initiation; Kozak, 1988.) These predicted proteins are 93% identical, differing at only 19 residues. The nucleotide sequences of the two cDNAs are 82% identical in the coding region but are quite dissimilar in the 5′ and 3′ untranslated regions.

Figure 4. Comparison of Primary Structures of Human RPS4Y and RPS4X Proteins and Homologs in Other Organisms

Dots represent identity to human RPS4Y, while dashes indicate gaps in one sequence as compared with another. The sequences of human RPS4Y and RPS4X are as predicted from the nucleotide sequences of cDNAs pDP1278 and pDP1284, respectively (Figure 2). The structure of rat RPS4 was determined by direct amino acid sequencing (Wittmann-Liebold et al., 1979) and subsequently by analysis of a full-length cDNA (Devi et al., 1989; as corrected, Y.-L. Chau, personal communication). The nematode sequence, presumably incomplete, was identified by analysis of the potential translation products of the GenBank DNA data base and corresponds to a previously unreported open reading frame 3′ of the deb-1 gene (nucleotides 12,049 through 12,264 in Barstead and Waterston, 1989) on Caenorhabditis elegans chromosome IV. The partial structure of the yeast (Saccharomyces cerevisiae) homolog is based on direct sequencing of the amino terminus of the protein (Obka et al., 1982). The structure of the archaeabacterial protein was predicted (ORF c in Auer et al., 1989) from genomic sequencing of a Methanococcus vannielii transcriptional unit homologous to the E. coli spectinomycin operon.
Searches of computer data bases identified proteins of similar amino acid sequence in organisms as evolutionarily distant as archaeabacteria (Figure 4). The human Y and X-encoded proteins are 33% and 35% identical to an archaeabacterial protein and show even greater similarity to proteins in yeast, nematode, and rat. The significance of the human–archaeabacterial similarity was confirmed by sequence randomization or jumbling (Dayhoff et al., 1983). In the case of archaeabacteria, yeast, and rat, the function of the homologs is known: all are ribosomal proteins. The amino acid sequence of the nematode protein was predicted (Figure 4) from the nucleotide sequence of a previously unrecognized open reading frame in genomic DNA (Barstead and Waterston, 1989).

The most thoroughly studied of these homologs is rat RPS4, ribosomal protein S4 (Devi et al., 1989), one of approximately 33 proteins present in the 40S ribosomal subunit (Sherton and Wool, 1972). Our human Y- and X-derived cDNAs exhibit a striking degree of identity at the nucleotide level (80% and 91%, respectively, in the coding region) and at the amino acid level (93% and 100%, respectively) to rat RPS4. We conclude that human cDNAs pDP1284 and pDP1278 are likely to encode isoforms of RPS4. As these human cDNAs are derived from the X and Y chromosomes, we will refer to the corresponding genes as RPS4X and RPS4Y, respectively.

The nucleotide sequence of RPS4X cDNA pDP1284 differs only slightly (Figure 2) from that of a previously isolated, X-derived human cDNA, SCAR (Wiles et al., 1988). The putative initiator codon in RPS4X (Figure 2) was absent in the published SCAR cDNA sequence, leading Wiles and colleagues to predict a protein slightly shorter at the amino terminus. Nonetheless, the overall similarity of the RPS4X and SCAR nucleotide sequences suggests that they are derived from the same gene.

Mapping RPS4X on the X Chromosome

Mapping and cloning of the RPS4X genomic locus were complicated by the existence of numerous cross-hybridizing sequences in the human genome. When RPS4Y cDNA pDP1278 or RPS4X cDNA pDP1284 was hybridized to Southern blots of human genomic DNAs, a large number of restriction fragments were detected (data not shown). Indeed, the number of fragments detected was far greater than could be accounted for by the RPS4Y and RPS4X loci alone. We suspect that sequences related to RPS4Y and RPS4X are found elsewhere in the genome. In mammals, each ribosomal protein is typically encoded by a single gene, from which a number of silent, processed pseudogenes have been generated (Monk et al., 1981; Dudov and Perry, 1984; Chen and Routa, 1988; Davies et al., 1989). Given this precedent, we speculate that the RPS4X/RPS4Y-related sequences in the human genome may be pseudogenes. This is consistent with the observation that each of the eight cDNA clones we isolated and sequenced pseudogenes. This is consistent with the observation that each of the eight cDNA clones we isolated and sequenced was unambiguously assignable to either RPS4Y or RPS4X.

To circumvent this problem of processed pseudogenes, polymerase chain reaction (PCR) amplification was carried out using human female genomic DNA as template and oligonucleotides based on the RPS4X cDNA sequence as primers. Assuming that the intron/exon structure of RPS4X might be similar to that of RPS4Y (Figure 3), we chose primers that were likely to flank a small intron in RPS4X. We reasoned that processed pseudogenes would yield PCR products of size identical to those obtained with the cDNA itself, while the true gene would yield a larger product, containing the intron (Davies et al., 1989). We obtained only one such larger product and confirmed by sequencing that it derived from the same gene as the RPS4X cDNAs. When hybridized to Southern blots of EcoRI-digested human genomic DNAs, this intron probe detected a single fragment of 3 kb, twice as intense in females as in males (data not shown), suggesting an X-chromosomal location.

This inference was confirmed and refined when genomic RPS4X clones were labeled with digoxigenin, hybridized in situ to metaphase chromosomes, and visualized by fluorescence. First, a larger portion of the genomic locus was cloned by screening both a recombinant λ phage library of XX female genomic DNA and a library of flow-sorted X chromosome DNA with the PCR-amplified intron. Two recombinant phages were identified, and their inserts were found to overlap. These two genomic phages were labeled and hybridized to metaphase chromosomes of a normal female. An advantage of using larger genomic clones is that signals are highly specific; cross-hybridization to regions of limited sequence similarity is not detected. As illustrated in Figure 5, a single site of hybridization was readily apparent on each of the two homologous chromosomes in essentially all metaphase figures. The signals were clearly on the long arm of a C group chromosome, identified as Xq by DAPI banding and by simultaneous hybridization with a probe for the human dystrophin gene (known to map to Xp). As summarized in the ideogram (Figure 5), analysis of a number of banded chromosome spreads was sufficient to localize the gene unambiguously to Xq13.

This localization of RPS4X was corroborated and further refined by hybridizing the PCR-generated intron probe to Southern blots of EcoRI-digested genomic DNAs from human–rodent somatic cell hybrids retaining intact or partial human X chromosomes (Figure 6). No cross-hybridization to rodent DNA was observed. Positive results were obtained with a hybrid retaining the X as its only human chromosome, verifying an X-chromosomal location. (No hybridization was observed with hybrids retaining the human Y but not the human X chromosome; data not shown.) Positive results were also obtained with all hybrids retaining the entirety of band Xq13.1. Two hybrids with breakpoints in Xq13.1 (and carrying distal portions of Xq) were among those tested. One of these hybrids, A68-2A, yielded a positive result, while the other, W4-1A, yielded a negative result. Previous studies of these two pivotal hybrids with X-linked DNA probes established that the chromosomal breakpoint in A68-2A is proximal to that in W4-1A (H. W. and colleagues, unpublished data). We conclude that RPS4X maps to band Xq13.1 on the long arm of the X chromosome, distal to the breakpoint in A68-2A and proximal to the breakpoint in W4-1A. The results of in situ hybridization (Figure 5) are entirely consistent.
Figure 5. Chromosomal Localization of the RfS4X Gene by In Situ Hybridization

(A) Hybridization of digoxygenin-labeled RPS4X genomic probe to metaphase chromosomes of a normal human female as detected using fluorescein-conjugated anti-digoxygenin antibodies. Both sister chromatids of both X chromosomes are labeled (arrows).

(B) DAPI staining of the same metaphase chromosomes. The Q-banding pattern facilitates identification of chromosomes and specific localization of the hybridization signal. Bright DAPI bands correspond to dark G bands (in ideogram [D]).

(C) Higher magnification of the two X chromosomes showing detail of DAPI banding (left) compared to the fluorescein signal (right). White bars indicate the positions of the centromeres.

(D) Ideogram summarizing the locations of the hybridization signals on 13 banded X chromosomes.

Figure 6. Chromosomal Localization of the RPS4X Gene Using Human–Rodent Hybrids Retaining Partial Human X Chromosomes

A PCR-amplified intron of the human RPS4X gene was radiolabeled and hybridized to EcoRI-digested DNAs from (left to right) a normal human XX female, XXXX female GM1416, mouse, hamster, and the following hybrids, most of which retain one of the two products of a reciprocal translocation (Worton et al., 1984; Myerowitz et al., 1985; Brown et al., 1989; Willard et al., 1990): (A) retaining an entire human X chromosome, C173 (retaining Xpter-p21.2), A2-4 (retaining Xpter-p21.2; the reciprocal of C173), B752a-1b (retaining Xp11.3-qter), A2A1A (retaining Xp11.2-qter), A2-4A (retaining Xp11.2-qter), A2B-2A (retaining Xq13.1-qter), and W4-1A (retaining Xq13.1-qter). The X chromosomal breakpoint in A68-2A is known to be proximal to that in W4-1A (H. F. W. and colleagues, unpublished data).

The most noteworthy results (positive in A68-2A, negative in W4-1A) were confirmed by subsequent control hybridizations (data not shown). This is not surprising given that the nucleotide sequences of the coding regions of the two genes are 82% identical.

Widespread Transcription of the RPS4X and RPS4Y Genes

By Northern analysis, we detected abundant transcription of RPS4Y and RPS4X in a wide variety of male and female cell lines and tissues. RPS4Y cDNA pDP1278 hybridized to a 1.0 kb transcript in all XY male fetal and adult tissues tested (e.g., Figure 7, top) and in a variety of XY male cultured cells, including fibroblasts, B lymphoblastoid, and T cell leukemia cell lines (e.g., Figures 1 and 6a). RPS4X cDNA pDP1284 hybridized to a transcript whose size (1.0 kb) is indistinguishable from that of the abundant RPS4Y transcript. This RPS4X transcript is present in all male and female tissues (e.g., Figure 7, bottom) and cultured cells tested, including fibroblasts, B lymphoblastoid, T cell leukemia, neuroblastoma, retinoblastoma, and cervical carcinoma cell lines (data not shown). (We detected cross-hybridization between the RPS4Y and RPS4X transcripts at reduced stringency [data not shown].) This is not surprising given that the nucleotide sequences of the coding regions of the two genes are 82% identical.)

RPS4X Transcription Is Not Subject to X Inactivation

In mammals, most X-linked genes are dosage compen-
Figure 7. Transcription of RPS4Y and RPS4X in Human Fetal Tissues

Above: The insert of RPS4Y cDNA pDP1276 was hybridized to poly(A)+ RNAs prepared from tissues of 17- to 21-week-old fetuses. Each lane contains approximately 10 μg of poly(A)+ RNA.

Below: After stripping (and reexposure to confirm that the RPS4Y probe had been removed), the blot was rehybridized with the insert of RPS4X cDNA pDP1264.

sated, the result of transcriptional inactivation of all but one X chromosome (Lyon, 1988). The transcription of RPS4X in B lymphoblastoid cells (Figure 1) enabled us to determine, by Northern analysis, whether the gene is dosage compensated. The level of RPS4X transcription was assessed in cell lines derived from individuals with one, two, or four X chromosomes (Figure 8). If RPS4X were dosage compensated, its level of transcription should be independent of the number of X chromosomes. This appears not to be the case. As shown in Figure 8 and Table 1, the intensity of RPS4X hybridization (after correcting for

Table 1. Quantitation of RPS4X and α-Tubulin Transcripts on Northern Blots

<table>
<thead>
<tr>
<th>Number of X Chromosomes</th>
<th>XX</th>
<th>XY</th>
<th>XY</th>
<th>XXXX</th>
<th>XY</th>
<th>XXXXY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized ratio of RPS4X to α-tubulin (Northern 1)</td>
<td>0.089</td>
<td>0.051</td>
<td>0.075</td>
<td>0.396</td>
<td>0.137</td>
<td>0.251</td>
</tr>
<tr>
<td>Normalized ratio of RPS4X to α-tubulin (Northern 2)</td>
<td>0.100</td>
<td>0.071</td>
<td>0.057</td>
<td>0.286</td>
<td>0.173</td>
<td>0.313</td>
</tr>
<tr>
<td>Average of two Normethens</td>
<td>0.095</td>
<td>0.061</td>
<td>0.066</td>
<td>0.341</td>
<td>0.155</td>
<td>0.282</td>
</tr>
</tbody>
</table>

emissions from the RPS4X and α-tubulin bands in each of six lanes on two Northern blots (one of which, "Northern 1," is shown in Figure 8a) were directly quantitated using the Betascope blot analyzer. The absolute numbers of disintegrations were corrected for lane background, and the RPS4X:α-tubulin ratios were calculated. The ratios were normalized so that they sum, for each Northern, to 1.0. The average values of these RPS4X:α-tubulin ratios are plotted in Figure 8b.

Figure 8. The Level of RPS4X Transcription Increases with the Number of X Chromosomes

(a) Northern blot analysis of poly(A)+ RNAs from human cell lines with differing numbers of X and Y chromosomes. The lanes contain RNAs from the following B lymphoblastoid cell lines (left to right): one normal female, two normal males, GM1416 (XXXX female), WHT1002 (XY male), and WHT0706 (XXXXY male). The blot was hybridized sequentially with (above) RPS4X cDNA pDP1284 and (below) hamster α-tubulin cDNA E7 (Elliot et al., 1986). Transcription of α-tubulin was monitored in order to correct for lane-to-lane variation in the amount of RNA.

(b) Ratios of RPS4X to α-tubulin transcription (data from Table 1) plotted versus number of X chromosomes per cell. As explained in Table 1, these data derive from two Northern blots, one of which is shown in (A).
loading differences) increased with the number of X chromosomes. We concluded that RPS4X transcription is proportional to the number of X chromosomes and, therefore, that the gene is not dosage compensated.

To assess directly whether RPS4X is subject to X inactivation, we determined whether the gene is transcribed in three human–rodent hybrid cell lines bearing "inactive" human X chromosomes (Figure 9). Each of these three hybrids contains a genetically distinct inactive human X chromosome. The inactive status of the retained human X chromosome has been thoroughly documented previously (Schneider-Gadicke et al., 1989a, and references therein). In an effort to avoid cross-hybridization to RPS4X-homologous genes in rodents, we used as probe an oligonucleotide derived from the 3' untranslated region of human RPS4X. While this 30-mer detects the abundant 1 kb RPS4X transcript when hybridized to RNAs from human female cell lines, it detects no transcripts in hamster or mouse RNA. In each of the three hybrid cell lines retaining an inactive human X chromosome on a hamster or mouse background, human RPS4X transcripts are readily detected (controls were as previously described; Schneider-Gadicke et al., 1989a). These results demonstrate that the human RPS4X gene is transcribed regardless of whether it is on an active or inactive X chromosome.

Discussion

Isoforms of Ribosomal Protein S4 Encoded by Two Human Genes

Ribosomes are massive ribonucleoprotein complexes that catalyze the synthesis of proteins. In eukaryotes, ribosomes appear to be composed of stochiometric amounts of about 80 distinct proteins and four RNA species (Wool, 1979). The genes encoding the ribosomal proteins are not clustered but instead are distributed throughout the eukaryotic genome (Woolford et al., 1979; D'Eustachio et al., 1981). In mammals, each ribosomal protein gene has typically given rise to a number of processed, silent pseudogenes (Dudov and Perry, 1984). Nonetheless, it appears to be the rule that for each mammalian ribosomal protein, there is only one transcriptionally active gene (Chen and Routa, 1988; Davies et al., 1989).

In contrast—at least in humans—isoforms of the ribosomal protein RPS4 appear to be encoded by two distinct genes, RPS4Y and RPS4X. We have cloned full-length cDNAs from these two homologous genes, one on the Y and one on the X chromosome. There is no evidence that either gene, both of which have introns, is a pseudogene. Nucleotide sequence of the cDNAs suggests that the two genes encode proteins that differ at only 19 of 263 amino acid residues. The coding regions of the two genes differ by 141 nucleotide substitutions, 110 of which are silent, suggesting that stabilizing selective pressure operates at the level of the encoded proteins. Based on the similarity of the predicted amino acid sequences to that of rat RPS4 (to which the human X-encoded protein is absolutely identical), we conclude that both human genes are likely to encode isoforms of RPS4. Given the nearly identical molecular weights of the X- and Y-derived proteins (29,597 and 29,455, respectively; 29,448 and 29,306 if the amino-terminal methionine is cleaved following translation), it is not surprising that the two isoforms were not previously resolved by electrophoresis.

Normal females have no Y chromosome and, therefore, no RPS4Y gene or protein. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X. If there is one RPS4 molecule per ribosome, then the ribosomes of males may be of two types: those containing RPS4Y and those containing RPS4X. All ribosomes in females would contain RPS4X.
A Gene on the Long Arm of the Human X Chromosome Escaping X Inactivation

In mammals, most genes on the X chromosome are subject to X inactivation (Lyon, 1968). However, two lines of evidence indicate that this is not the case for RPS4X. First, quantitation of transcripts in human cells with one, two, or four X chromosomes reveals that the gene is not dosage compensated (Figure 6 and Table 1). Second, hybridization to mRNAs from human rodent cell hybrids retaining RPS4X but also of ZFX (Schneider-Gadicke et al., 1989a) shows no evidence of X-Y cross-hybridization. X-Y sequence similarity has been retained only where divergence would be selectively disadvantageous.

A preliminary analysis of silent nucleotide divergence, corrected for back mutations (Rice, 1987), reveals 4-fold greater divergence between the coding regions of human RPS4X and RPS4Y than between the zinc finger domains of human ZFX and ZFY (K. A. Rice, P. D.-f., and D. C. P., unpublished data). One possible interpretation of this finding is that the RPS4X/Y gene pair has been diverging for a longer period of time than the ZFX/Y gene pair. However, numerous alternative interpretations cannot yet be excluded.

Could RPS4 Haploinsufficiency Cause the Turner Phenotype?

Turner syndrome is a complex human phenotype (Ulrich, 1930; Turner, 1938) that often occurs in association with a 45,X karyotype (Ford et al., 1959). One of the most common of human chromosomal disorders, Turner syndrome affects one in every 5000 liveborn females (Hook and Warren, 1983). Liveborn Turner females are characterized by short stature, gonadal degeneration, and a variety of anatomic abnormalities including webbing of the neck, lymphedema, and aortic coarctation. Turner syndrome is also marked by an extremely high rate of intrauterine lethality (Carr, 1965). It is estimated that 1%-2% of human conceptions are 45,X, and that 99% of such fetuses abort spontaneously (Hassold, 1986).

Virtually all cases of Turner syndrome are associated with sex chromosome anomalies, most commonly monosomy X, but often sex chromosome mosaicism (e.g., 45,X/46,XX) or the presence of a normal X together with a structurally abnormal X or Y chromosome. Such observations led to the proposal that Turner syndrome might be the result of monozygosity for a gene or genes common to the X and Y chromosomes (Ferguson-Smith, 1965).

This hypothesis has been strengthened by studies of two types of human XY females. The first type of XY female has gonadal dysgenesis but no other signs of Turner syndrome. These XY females with "pure gonadal dysgenesis" carry Y chromosomes that are grossly intact as judged by light microscopy and Southern blotting (Cantrell et al., 1989). While some XY pure gonadal dysgenesis females may prove to carry point mutations in Y-chromosomal sex-determining genes, the pattern of inheritance in other cases (Espiner et al., 1979) strongly suggests that the Y chromosome is intact. The second type of XY female not only has gonadal dysgenesis but also has somatic abnormalities characteristic of Turner syndrome.
such XY Turner females studied have deletions of portions of Yp, the short arm (Rosenthal et al., 1979; Magenis et al., 1984; Disteche et al., 1986; Blagowidow et al., 1989; Levilliers et al., 1989; D. C. R. and colleagues, unpublished data). These individuals are female because their Y chromosomes are deleted for one or more sex-determining genes. Phenotypic and genotypic comparison of these two types of XY females suggests that the Yp deletions in the XY Turner females encompass one or more genes whose presence prevents the extragonadal features of Turner syndrome. The Turner phenotype, or at least its extragonadal component, is probably the result of the presence of one rather than two copies of a gene or genes common to the X and Y chromosomes.

Nothing is known about the nature of the Turner gene(s) or the protein(s) it might encode. Attempts have been made to localize the genes by correlating partial deletions of the sex chromosomes with the Turner phenotype. In the case of the Y chromosome, it is agreed that the Turner phenotype maps to interval 1 or 2, on the distal short arm (Rosenthal et al., 1979; Magenis et al., 1984; Disteche et al., 1986; Blagowidow et al., 1989; Levilliers et al., 1989). We refined this localization by additional deletion analysis of XY Turner females and an X[t(Y;22) female who carries most of the Y chromosome but does not exhibit the Turner phenotype (Page et al., 1987, 1990, and unpublished data). The Turner phenotype appears to map either to interval 1A1B or to the region encompassing intervals 1C through 2. Since the latter region is homologous to a portion of the X chromosome that can be deleted without giving rise to the Turner phenotype (Tabor et al., 1983; Page et al., 1984, and unpublished data; Schwartz et al., 1988), we concluded that, on the Y chromosome, one or more Turner genes are probably located in interval 1A1B.

Is it possible that haploinsufficiency of RPS4Y and RPS4X underlies some or all of the Turner phenotype? Though the evidence is far from definitive, the RPS4 genes conform well to expectations for genes underlying the Turner phenotype. Such genes would most likely have the following characteristics: First, they would exist as X and Y homologs that function interchangeably, or nearly so, in vivo. There are no obvious differences in the patterns of expression of the two genes; both RPS4X and RPS4Y appear to be ubiquitously transcribed. Since the primary structures of the RPS4X and RPS4Y proteins are 93% identical, the isoforms may be interchangeable. Second, they would map to interval 1A1B on the Y chromosome. We have no evidence for the existence of genes other than RPS4Y in this 90 kb segment. Third, they would escape X inactivation. If this were not the case, one would not expect to observe differences in the phenotypes of 45,X and 46,XX females.

How does the mapping of RPS4X to proximal Xq relate to the hypothesized role of this gene in Turner syndrome? In contrast to the Y chromosome, efforts to map the Turner phenotype on the X have failed to yield a precise or even consistent localization. Most cases of Turner syndrome are congruent with a localization to Xp (Ferguson-Smith, 1965: Goldman et al., 1982; Shapiro, 1983). However, as others have pointed out (Epstein, 1988; Thereman and Susman, 1990), the fact that similar Turner phenotypes can be observed with deletions of either Xp or Xq poses a major problem. In an attempt to rationalize such inconsistencies, Thereman and Susman (1990) have proposed that the Turner gene(s) escapes X inactivation on intact X chromosomes but is prone to X inactivation on structurally abnormal chromosomes. If RPS4X is a Turner gene, then some such position effect will be required to explain the occurrence of the Turner phenotype in certain individuals with structurally abnormal X chromosomes (e.g., 46,X(iXq)) and two or more copies of RPS4X. It would be useful to establish whether deletions of RPS4X itself are associated with a Turner phenotype. We know of no individual with a structural abnormality of the X in which RPS4X is deleted.

RPS4Y and RPS4X appear to encode isoforms of a ribosomal protein. Is it reasonable to suppose that reduced levels of a ribosomal protein would produce specific anatomic abnormalities like those seen in Turner syndrome? A precedent is found in Drosophila, where similarly complex but distinctive phenotypes have been shown to result from deficiencies of individual ribosomal protein genes. Flies heterozygous for Minute mutations (usually deficiencies) exhibit delayed larval development, diminished viability, reduced body size, diminished fertility, and specific somatic abnormalities (short, thin bristles and etching of the abdomen; Lindsey and Grell, 1968). As with the human Turner phenotype, these dominant Minute phenotypes have been characterized as haploinsufficiencies. Some of the approximately 50 Minute loci in Drosophila encode ribosomal proteins, and it is thought that the dominant Minute phenotypes are due to reduced levels of individual ribosomal proteins (Kongsuwon et al., 1985). Similarly, we postulate that the Turner phenotype may result, at least in part, from reduced levels of ribosomal protein RPS4. Reduced levels of RPS4 might retard the rate of ribosome assembly, which in turn could reduce the rate of protein synthesis. Rates of protein synthesis vary dramatically among human cell types, tissues, and stages of development. Thus, the consequences of reduced RPS4 levels would likely not be uniform throughout the developing organism.

Ribosomes contain about 80 different proteins (Wool, 1970). Since ribosomal protein genes are not clustered in eukaryotes (Woolford et al., 1979; D'Eustachio et al., 1981), it follows that human chromosomes each carry an average of three or four ribosomal protein genes. In humans, many distinct phenotypes, most including growth retardation, have been associated with monosity for particular portions of autosomes. Perhaps haploinsufficiency of other ribosomal protein genes contributes to the phenotypes observed in some human autosomal monosomies.

Experimental Procedures

DNA Hybridization Probes

Listed in Table 2 are plasmids whose genomic or cDNA inserts were used as hybridization probes in Northern or Southern blot analysis or in screening genomic and cDNA libraries. All inserts are derived from the human X or Y chromosome. Stated are the sizes (and, where appropriate, the restriction site termini) of the inserts, and the plasmid cloning vectors. The Sail site in pDPl241 is not present in the human
Nucleotide Sequencing and Computer Database Searches

Single-stranded and double-stranded DNA templates were sequenced using a modification of the FASTA algorithm (Pearson and Lipman, 1966), and hybridized to Southern or Northern blots at 60°C (with 30-mers) in 6x SSC, 5x Denhardt's, 0.5% SDS, 20 mM sodium chloride (NaCl), 15 mM sodium citrate [pH 7.4], 1x Denhardt's solution (0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 1% SDS, 50 mM NaPO4 (pH 6.5), and 0.01% yeast RNA. Following hybridization, transfer membranes were washed three times for 20 min each at 65°C in 0.1x SSC, 0.1% SDS. Synthetic oligonucleotides were end-labeled with 32P by random-primer synthesis (Feinberg and Vogelstein, 1964). The resulting probes were hybridized overnight to Southern blots, Northern blots, or cDNA libraries at concentrations of 5 x 106 to 3 x 107 cpm/ml at 42°C in 50% formamide, 6x SSC (1x SSC = 0.15 M NaCl, 15 mM sodium citrate [pH 7.4]), 0.1% SDS, transfer membranes were washed three times for 20 min each at 65°C in 0.1x SSC, 0.1% SDS.

Nucleotide Sequencing and Computer Database Searches

Single-stranded and double-stranded DNA templates were sequenced by double-strand termination (Sanger et al., 1977) using Sequenase (US Biochemical Corp.) and synthetic oligonucleotides. Using a modification of the FASTA algorithm (Pearson and Lipman, 1988), the nucleotide sequences (and predicted translation products) of pDP-1278, pDP-1284, and pDP-1190 were compared with the GenBank and EMBL nucleic acid data bases (and their potential translation products).

Table 2. DNA Hybridization Probes

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Origin</th>
<th>X or Y</th>
<th>Size (kb)</th>
<th>Enzyme(s)</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDP1056</td>
<td>Genomic Y</td>
<td>0.4</td>
<td>HindIII</td>
<td>Bluescript</td>
<td></td>
</tr>
<tr>
<td>pDP1190</td>
<td>cDNA</td>
<td>Y</td>
<td>6.9</td>
<td>--</td>
<td>Bluescript</td>
</tr>
<tr>
<td>pDP1241</td>
<td>Genomic Y</td>
<td>1.6</td>
<td>HindIII/BalI</td>
<td>--</td>
<td>Bluescript</td>
</tr>
<tr>
<td>pDP1278</td>
<td>cDNA</td>
<td>Y</td>
<td>0.9</td>
<td>--</td>
<td>Bluescript</td>
</tr>
<tr>
<td>pDP1394</td>
<td>cDNA</td>
<td>X</td>
<td>0.9</td>
<td>--</td>
<td>Bluescript</td>
</tr>
</tbody>
</table>

Cloning of the RPS4X Genomic Locus

The strategy employed was essentially that previously used to clone the rat L19 ribosomal protein gene (Davies et al., 1969). In the present case, an intron of the RPS4X gene was amplified by PCR (Salik et al., 1989) using human female genomic DNA as template and the oligonucleotides based on the RPS4X cDNA sequence, TGAAGTCAAAGT GAAAGATGC and AAAGATCTGACCATGTCACCC, as primers. Thirty-five cycles of amplification were carried out using Taq polymerase and a DNA thermocycler (Cetus); annealing, extension, and melting temperatures were 52°C, 72°C, and 94°C, respectively. Products of approximately 200 and 750 bp were observed. The 750 bp product, also observed using RPS4X cDNA pDP1284 as template, probably derives from one or more processed pseudogenes in the genome. Partial nucleotide sequencing of the 750 bp product, which contained about 200 bp of coding sequence, confirmed that it corresponded to the same gene as cDNA pDP1284 and suggested that an intron is located between cDNA nucleotides 690 and 691. We therefore screened a second PCR reaction using the 750 bp product as template and internal sequences, GACTGAGTCCTAAGAT and CUGGGTTTITGGCTGCT, as primers. As expected, a 550 bp product, presumably containing the entire of the intron but little of the flanking exons, was obtained. This 550 bp PCR product was 32P-labeled by random-primer synthesis and used to screen recombinant λ phage libraries of human genomic DNA. In this manner, clone λBar226 was isolated from a library of human female DNA (a gift of Stuart Orkin), while clone λLaol was isolated from a library of flow-sorted X chromosomes (λAXNLOl, constructed at the Los Alamos National Laboratory and obtained from the American Type Culture Collection). The human genomic inserts of the two phages were restriction mapped and found to overlap. The 32P-labeled 550 bp product served as hybridization probe in mapping the RPS4X gene using rodent-human hybrids (Figure 6).

Chromosome In Situ Hybridization

Standard metaphase chromosome spreads were prepared using peripheral blood lymphocytes from a normal human female. Fluorescence in situ hybridization was carried out as previously described (Lawrence et al., 1988, 1990). Briefly, recombinant λ phage clones λBar226 and λLaol were labeled by nick translation using digoxigenin-11-dUTP (Boehringer Mannheim) and hybridized to denatured metaphase chromosomes at a concentration of 5 µg/ml at 37°C in 50% formamide, 2x SSC. Excess unlabelled total human DNA was added to the hybridization mixture to compete for hybridization to repetitive sequences. Hybridization was detected using fluorescein-conjugated anti-digoxigenin antibody. Chromosomes were identified by simultaneous banding with DAPI (diaminophenylindole), enhanced by prior incorporation of 5-bromodeoxyuridine into chromosomal DNA. Photographs taken at 1000 x magnification on a Zeiss Axioplan fluorescence microscope were analyzed independently by two investigators, whose scorings were concordant.

Acknowledgments

We thank Alan Buckler, Albert de la Chapelle, David Housman, Barbara McGillivray, and George Mutter for tissue samples and human cell lines; Robert Nussbaum and T. K. Mohandas for human-rodent hybrid cell lines; Stuart Orkin for a human female genomic library; Graeme Mardon for a human XYYYY male cDNA library; Anebert Schneider-Gädicke for human testis RNA; Yuen-Ling Chan and Ira Wool for providing rat RPS4 sequence data; Katsuyuki Suzuki and Tat-suo Tanaka for sharing their ribosomal protein database; and Tracy Pittirsch for exemplary library work. We especially thank Jasper Rees for extensive data base searches and consistent help and encouragement with UNA sequence analysis. We are grateful to Albert de la Chapelle, Terry Hassold, Tom RajBhandary, Ruth Lehmann, Barbara McGillivray, Robert Nussbaum, Ken Rice, Ira Wool, and John Woolford for formative discussions and to Jonathan Bogan, Simon Foote, Eric Lander, Rebecca Mosher, Douglas Voldreth, and Andrew Zinn for criticism of the manuscript. E. M. C. F. was supported by fellowships from EMBO and the Royal Society. A. R. was supported by an EMBO fellowship. This study was supported in part by grants from the NIH, the DuPont Center of Molecular Genetics at the Whitehead Institute, the Lucille P. Markey Charitable Trust, the Muscular Dystrophy Association, and the Searle Scholars Program/Chicago Community Trust.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received September 7, 1990; revised October 10, 1990.

References


Nygard, O., and Nika, H. (1962). Identification by RNA-protein cross-linking study on 


**GenBank Accession Number**

The accession numbers for the RPS4Y and RPS4X sequences reported in this paper are M58469 and M58468, respectively.