Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y–)XY females

Katrin Schiebel1, Martina Winkelmann1, Annellyse Mertz1, Xiaoling Xu2, David C. Page2, Dominique Weil3, Christine Petit3 and Gudrun A. Rappold1,*

1Institute of Human genetics, Ruprecht-Karls-University, Im Neuenheimer Feld 328, D-69120 Heidelberg, Germany, 2Howard Hughes Medical Institute, Whitehead Institute and Department of Biology, MIT, 9 Cambridge Center, Cambridge, MA 02142, USA and 3CNRS URA 1968 at Institut Pasteur, 25, rue du Dr. Roux, F-75724 Paris, Cedex 15, France

Received June 27, 1997; Revised and Accepted July 31, 1997 DDBJ/EMBL/GenBank Accession Nos Y13927–Y13934

XX males and XY females have a sex reversal disorder which can be caused by an abnormal interchange between the X and the Y chromosomes. We have isolated and characterized a novel gene on the Y chromosome, PRKY. This gene is highly homologous to a previously isolated gene from Xp22.3, PRKX, and represents a member of the cAMP-dependent serine threonine protein kinase gene family. Abnormal interchange can occur anywhere on Xp/Yp proximal to SRY. We can show that abnormal interchange happens particularly frequently between PRKX and PRKY. In a collection of 26 XX males and four XY females, between 27 and 35% of the interchanges take place between PRK homologues but at different sites within the gene. PRKY and PRKX are located far from the pseudoautosomal region where XY exchange normally takes place. The unprecedented high sequence identity and identical orientation of PRKY to its homologous partner on the X chromosome, PRKX, explains the high frequency of abnormal pairing and subsequent ectopic recombination, leading to XX males and XY females and to the highest rate of recombination outside the pseudoautosomal region.

INTRODUCTION

The XX male syndrome (1) and Swyer’s syndrome (XY gonadal dysgenesis) (2) are two genetic disorders interfering with normal gonadal development in humans. XX males resemble individuals with Klinefelter’s syndrome (47,XXY) in their general masculine appearance. They are infertile and have small testes. Conversely, individuals with Swyer’s syndrome are phenotypic females with streak gonads, amenorrhea and a male karyotype (46,XY). The XX male syndrome has an incidence of 1:20 000 in the general population, while Swyer’s syndrome occurs more rarely (3). In ~80% of XX males, the male phenotype is caused by the presence of Y-specific DNA, including SRY (4–11), whereas in XY females, at a lower percentage, the phenotype is due to the absence of Y-specific DNA (12–16). The abnormal exchange between the X and Y chromosomes has been characterized in both XX males and XY females as a terminal exchange (10,11,16).

Two hot spots of recombination (HSA and HSB) leading to a high frequency of XX males and XY females have been described previously and positioned in the interval 3F and 3G on the human Y chromosome (17–19). Here we report results on the isolation of a 250 kb PAC and cosmid contig and on the characterization of the Y-chromosomal breakpoints of 26 XX males and four XY females, who were selected because their breakpoints previously were roughly mapped and shown to reside on interval 3 (17–20). We can show that the breakpoints of all tested XX males and XY females either reside within the genomic locus of PRKY or in the direct vicinity of PRKY. This implies that the PRKY-containing subregion exhibits the highest recombination rate on the Y-specific portion of the Y chromosome.

RESULTS

PRKX cDNA fragments were used to isolate homologous sequences on the Y chromosome. For this purpose, a Y-sorted Lawrence Livermore cosmid library LLOYNC03 ‘M’ and a human male PAC library (21) were hybridized with PRKX cDNA subfragments (20; see also Materials and Methods), resulting in five overlapping cosmids and two flanking PACs. Sequencing with PRKX-derived primers and PCR analysis with sequence-tagged sites (STSs) DYS263 (sY68), DXYS163Y (HSB), DYS264 (sY69) and DXYS164Y (HSA) and the
Figure 1. Schematic diagram of the PAC/cosmid contig from interval 3G/3F on Yp and the genomic organization of PRKY. Exons were numbered in accordance with PRKX. The position of markers sY68 (DYS263), sY69 (DYS264) (17), HSB (DXYS163) and HSA (DXYS164) (18) are indicated on the map; HSA/HSB: hot spot of recombination A/B. PACs 283F16 and 152F08 were mapped by FISH: both showed strong Yp11.2 signals and weaker signals on the Xp22.3 region (unpublished results). This phenomenon was also seen with the two tested cosmids 67B8 and 67G5 of the region.

hybridization of the amplification products to Southern blots of digested cosmid and PAC DNA allowed the construction of the map depicted in Figure 1. Partial Y-specific cDNAs were derived from fetal brain and bone marrow cDNA libraries at high stringency screening (see Materials and Methods). Y-specific exons were also identified based on their similarity to PRKX (formerly PKX1, 20) using X-specific primers and by sequencing Y-specific cosmids and PACs from the region. The known order of sY markers on the YAC map of Foote et al. (22) revealed that the 5′→3′ orientation of PRKY is towards the telomere. PRKY is composed of eight exons with length ranging from 78 to 6047 bp (Fig. 1). The last exon represents a block of ~6 kb and starts 22 bp downstream of the stop codon in PRKX cDNA, and thus represents the 3′-untranslated region (UTR). Due to a 1.4 kb insertion and several small gaps, this region is 1.3 kb longer in PRKY than in PRKX. The polyadenylation signal sequence, however, is located at the corresponding position (EMBL accession Nos X85545 and Y13934). Sequence comparison with PRKX cDNA revealed that exon 1 of PRKY is composed of 504 bp, with 338 bp representing the 5′ UTR. An exon corresponding to nucleotides 1182–1239 in PRKX cDNA is missing in PRKY. Despite the conservation of start and stop codons at the corresponding positions in both PRKX and PRKY and the conservation of the open reading frame (ORF) in all existing exons, the predicted protein of PRKY is 81 amino acids shorter than that of PRKX. This is due to a frameshift caused by the loss of the smallest 58 bp exon. The deletion leading to the loss of exon 6 seems to be a very recent event during evolution, since no additional frameshift or stop mutations have accumulated in any other exons. In contrast, a pseudogene, PRKXP1, which maps to 15q26, was shown to be truncated not only by a stop codon but also by the insertion of two different repeats (23).

Breakpoint analysis of XX males and XY females

The two largest collections of XX males and XY females with known breakpoints in interval 3F and 3G on Yp were investigated in order to define the breakpoint regions with respect to PRKY (17,19,24). Using PRKY exon-specific primers and primers flanking the gene on either side we show in Figure 2 that 13/30 deletion breakpoints occur intragenically (within the 110 kb genomic locus of PRKY), two of the deletion breakpoints reside close to exon 9 (WHT1209 and WHT1870) and 15 deletion breakpoints are upstream to exon 1, 12 of which reside in a distance of 30 kb in the previously described hot spot HSA (18). The distribution of translocation breakpoints reveals that five different intervals appear to be particularly prone to this type of rearrangement: translocation breakpoints within PRKY occur between exons 1 and 2 (patient 693), between exons 3 and 5 (LGL105, GA001, WHT304, LGL850/WHT438 and WHT1327), between exons 5 and 8 (CHM347 and 651), between 9A and 9B in exon 9 in the 3′ UTR (WHT1856, GA/WHT1123, TA, NAN194 and PAR089), between exon 1 and HSA (WHT1010) and in HSA ~30 kb upstream of exon 1 (Figs 1 and 2). The breakpoint regions of two patients WHT950 and WHT2127 were narrowed down to a 650 bp fragment using an
Figure 2. Deletion map of the PRKY gene region on the short arm of the Y chromosome. Listed on the left are those 26 XX male and four XY female individuals analysed, who carry deletion breakpoints between sY68 and AMELY. One normal male and one normal female were used as controls. WHT1209, WHT950, WHT2127, WHT745, WHT1615 and WHT1327 are from the series of Volfrath et al. (17). XX male WB from the original series (17) was omitted, as his breakpoint could be redefined distal to sY68 (data not shown) and was substituted by WHT1870, WHT1856 and WHT304. TA, NAN194, PAR089, CHM347, GA001, CHM005, CHL071, PAR475, PAR221, MAR320, AG, RH, DR, PAR090 and LGL850 are from the patient series of Wang et al. (19); GA/WHT1123, LGL105, LGL115 and LGL850/WHT438 are contained in both collections; 651 and 693 have been described in Klink et al. (20). Across the top, one hybridization probe and 12 STSs are listed, eight of which represent ESTs from PRKY (exons were numbered in accordance with PRKX). Eleven STSs (sY68–TranD) lie in a genomic interval of <140 kb ∼7 Mbp from the Yp–telomere (Yp–tel distance is based on available Y AC contigs) (33). The short arm telomere (TEL) is to the left and the centromere (CEN) to the right. HSA and HSB are recombination hot spots (18). The presence or absence of loci was detected by PCR. The experimentally demonstrated presence of a locus is indicated by +, its absence by a minus and the inferred absence/presence of a locus is not marked at all. Patients with an asterisk: DYS7, TranD and TranC loci have been tested previously by hybridization (19). Note that TranC11orf is an X-specific primer used for the amplification of the X-Y junction fragment (see Materials and Methods).

Figure 3. Comparison of DNA sequences of patients and hot spot A sequence (18). The sequence starts at the BamHI site and is numbered according to HSTRANXYC (accession No. X70412; breakpoint region sequenced from the X chromosome). DNA sequences of the patients are indicated by symbols: – nucleotide identical in X,Y and patient DNA; x nucleotide identical in X and patient DNA, y nucleotide identical in Y and patient DNA, p polymorphic nucleotide (in patient DNA but neither in X nor in Y DNA); * nucleotide of breakpoint region (identical in X and Y DNA).

In summary, the previously described recombination hot spots A and B have a distance of maximally 140 kb on the Y chromosome and enclose the PRKY gene. The extensive XY homology of PRKX and PRKY genes leads to abnormal XY interchange in at least 33 (17/52 based on ref. 19) to 35% (8/23 based on ref. 17) of all (Y+)XX males and in 27% (3/11 based on ref. 17) of all (Y−)XY females.

**DISCUSSION**

Recombination between X and Y chromosomes is normally restricted to the pseudoautosomal regions located at both ends of the sex chromosomes, where pairing of X and Y chromosomes also initiates. An obligate crossover in the pseudoautosomal region (PAR1) during male meiosis seems to be necessary for male fertility (25). The sequence identity between X and Y chromosomes found within the PAR1 is interrupted abruptly at the pseudoautosomal boundary where X- and Y-specific regions begin. On the Y chromosome, this region harbours the testis-determining gene, SRY (26). Recombination occurring outside the pseudoautosomal region (PAR1) during male meiosis seems to be necessary for male fertility (25). The sequence identity between X and Y chromosomes found within the PAR1 is interrupted abruptly at the pseudoautosomal boundary where X- and Y-specific regions begin. On the Y chromosome, this region harbours the testis-determining gene, SRY (26). Recombination occurring outside the pseudoautosomal region in the Xp–Yp homologous regions results in a translocation of the SRY gene from the Y to the X chromosome, and consequently to the development of (Y+)XX males or (Y−)XY females. Illegitimate recombination involving Xp–Yq homologous regions has also been described as the cause of Xp22.3–Yq11 translocations, but these seem to be extremely rare (27,28). We show that abnormal interchange happens particularly frequently between two small X–Y homologous regions harbouring the PRKX/PRKY genomic loci.
PRKY and PRKX have a high overall sequence similarity of 94% and encode proteins with an intact ATP-binding domain and a catalytic domain with high homology to protein kinases. PRKY and PRKX are expressed at different levels (data not shown), and we do not know if both proteins are functional. Twelve differences in amino acids between PRKY and PRKX fall in the first exon, and the putative promoter region shows only 89% sequence similarity, suggesting that potential differences in transcription activity and functional relevance probably reside in the respective 5′ portions of the genes. The shortening of the putative PRKY protein results in the loss of a highly conserved arginine (named Arg280), which was shown to interact with a glutamic acid (Glu208) and seems to play an important role in the formation of the three-dimensional structure of all protein kinases investigated so far (29). On the other hand, several members of the protein kinase family are known to have short C-termini similar to the PRKY product (30). Further biochemical investigations are needed to determine whether both transcripts are translated and have similar or different functions. What is the reason for this frequent occurrence of abnormal XY interchange intragenically or directly adjacent to PRKY, ~7 Mbp proximal to PAR1 on Yp (31)? To our knowledge, the two other known Xp–Yp gene pairs, ZFX/ZFY and AMELX/AMELY, with sequence similarities of 92 and 91%, respectively, are not prone to ectopic recombination. Whereas in AMELX/AMELY high similarities between exons and the flanking intron regions have also been noticed, gene loci are short (10 kb) and the orientation of both genes is unknown (32). In comparison, ZFX/ZFY are in the same orientation on X and Y, cover larger genomic loci (70 and 50 kb), yet sequence homology seems to be restricted to the exons and the CpG island region at the 5′ ends (33). Thus, it appears that the biology of the abnormal interchange is implicated by a combination of factors such as the high sequence similarity of PRKY to PRKX (94% homology) extending into the introns, the unusually long untranscribed last exon (6 kb/4.7 kb) with 93% identity, the large genomic locus of PRKY (~110 kb) and the identical orientation of both genes on the sex chromosomes. The high incidence of ectopic recombination in the PRKY/PRKX gene regions in both XX males and XY females also demonstrates that (Y–)XY females represent true mirror images of (Y+)XX males, as originally hypothesized by Ferguson-Smith in 1966 (4).

MATERIALS AND METHODS

Patients

XX males and XY females patients are from two large series described by Vollrath et al. (17) and Wang et al. (19). Patients 693 and 651 have been described by Klink et al. (20).

cDNA isolation

Two fetal brain cDNA libraries (Stratagene 937227; 1.2×10^6 clones and Clontech HL3003b; 0.9×10^6 clones; mixed male and female tissue), a placenta (Stratagene 936203; 0.75×10^6 clones; male tissue), a bone marrow (Clontech HL1058b; 0.75×10^6 clones; male tissue) and a testis cDNA library (Clontech HL3024b; 1.2×10^6 clones; male tissue) of human origin were screened with PRKY cDNA fragments according to standard techniques. Final stringencies of washes were 1% SDS, 20 mM Na2HPO4 at 60°C. Pre-selection of Y-specific cDNA clones was done by PCR amplification of plaques with primer 166cyfor: CCA AGC ATT TCT TCG CCC and 166cyrev: CGT GAG CTT GAT GTG ACC (annealing at 55°C) followed by a PstI digestion. The 356 bp amplification products from Y chromosomal transcripts cannot be cleaved with PstI to 248 and 108 bp fragments, in contrast to the X-derived transcript.

Three Y-specific partial cDNA clones were derived from the Clontech fetal brain and the bone marrow cDNA libraries extending from 577 to 834 bp ( exon 2 + partial exon 3 ), 639 to 894 bp ( partial exon 2 + partial exon 3 ) and 577 to 966 ( exon 2 + exon 3 ). Localization refers to the PRKY cDNA sequence (accession No. X85545).

Cosmid and PAC screening

PRKY cDNA fragments (0.77 and 0.69 kb SacI subfragments of FB166) and a genomic 4.5 kb BamHI–SacII fragment enclosing the 3′ UTR (20) were hybridized to gridded filters of the Y-sorted Lawrence Livermore cosmid library LLOYNC03 ‘M’ (constructed by J. Games and P. de Jong) and to a human male PAC library (21). Final wash conditions were 1% SDS, 20 mM Na2HPO4. The hybridization and wash temperature was reduced from 65 to 58°C for exon 1. Localization of PACs and several cosmids was confirmed by fluorescence in situ hybridization which was carried out as previously described (20).

Sequencing

DNA double-strand sequencing of PCR products cloned in pMOS-vector (Amersham) was performed with an ALF express automated sequencer (Pharmacia).

Primers and PCR conditions for breakpoint analysis

PCR products were performed in 50 μl volumes containing 20–100 ng of genomic template DNA, 50 pmol of each primer, 200 μM dNTP, 5% dimethylsulphoxide (DMSO), buffer and MgCl2 as supplied with Taq polymerase (Eurogentec). Cycling was carried out in a Thermocycler 60 (Biomed): 1 min at 94°C, 30 s at the annealing temperature (see below), 30 s at 72°C.

Primer sequences, annealing temperatures and fragment length were as follows: sY68, sY69 and AMELY see ref. 17, annealing at 62, 64 and 62°C, respectively; exon 9b: FB1Yfor, GAC CTT TTC TTC AGC TGA C; FB1Yrev, AAA ACA GAC AAC ATA AAA TTA CA, 58°C annealing, 578 bp; exon 9a: 166A2Yfor, CAA GAC TTT TTT CTC CAC C; 166A2Yrev, CAT TCT CCT TGA CAT TTT GC, 58°C annealing, 364 bp; exon 8: pky8for, TCG TGC CCA CGA TGA CGT GCA; pkyrev, TCT TCC AGA TGT GAG FCT C . GTC C, 67°C annealing, 121 bp; exon 7: pky7for, ACA ACT TCA ATG TGT GGA GAA GAA; pky7rev, AAT CCA AAT GTC TGG GGA AAT ATA G, 60°C annealing, 228 bp; exon 6: 166cy3for, CTC ATG GAG TAT GTG CCG GGT; 166cy2rev, GTC CGT GAG TAT GTG ACC, 65°C annealing, 197 bp; exon 5: 166cy2for, GGA AGC AGG AGC AGC AGC TG; pky12Brev, GAG TGC TGC GTC GGA AGA GCC C, 65°C annealing, 108 bp; exon 1: pkyE1for, ACT CCC GAG AGG TGA CCG; pkyrev, CAT GGT GAC CAG CCG GTC GC, 62°C annealing, 121 bp; TranD: TranDfor, CCT GCC TTT TTT AGT TTC CAG CA, TranDrev, TAC GTG GAT AGG TAG AAT AAT GGC, 63°C annealing, 292 bp; TranC: TranCX1for, GGG CAC AGT GGT TCA CAC TG; TranDrev, see above, annealing, 63°C, 649 bp. Bases in bold indicate Y-specific bases (versus XY homologous bases); dots mark
missing bases in the Y copy. Note that all primers are Y-specific except for TranCX1 for, which is an X-specific primer used for the amplification of the X--Y junction fragment (here bases in bold indicate X-specific bases).

ACKNOWLEDGEMENTS

We thank Pieter de Jong for the PAC and cosmid libraries (constructed at the HGC, LLNL, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by U.S. DOE). We also thank Jeff Craig, Anne Jordan, Albrecht Klink, Birgit Schechinger, Gerd Scherer, Laura Brown and Raaja Alapaggan for help and critical comments and C.R. Bartram for his support. The research was supported by the Deutsche Forschungsgemeinschaft RA 380/9-1 and the Medizinische Fakultät, Universität Heidelberg.

REFERENCES