Cell, Vol. 80, 837–845, March 24, 1995, Copyright © 1995 by Cell Press

Mutations in a Putative Global Transcriptional Regulator Cause X-Linked Mental Retardation with α-Thalassemia (ATR-X Syndrome)

Richard J. Gibbons,* David J. Picketts,*
Laurent Villard,† and Douglas R. Higgs*
*Medical Research Council Molecular Haematology Unit
Institute of Molecular Medicine
John Radcliffe Hospital
Headington, Oxford OX3 9DU
England
†Institut National de la Santé et de la Recherche Médicale
Unité 406
Équipe 3
Génome et Développement
13351 Marseilles Cedex 5
France

Summary

The ATR-X syndrome is an X-linked disorder comprising severe psychomotor retardation, characteristic facial features, genital abnormalities, and α-thalassemia. We have shown that ATR-X results from diverse mutations of *XH2*, a member of a subgroup of the helicase superfamily that includes proteins involved in a wide range of cellular functions, including DNA recombination and repair (RAD16, RAD54, and ERCC6) and regulation of transcription (SWI2/SNF2, MOT1, and brahma). The complex ATR-X phenotype suggests that *XH2*, when mutated, down-regulates expression of several genes, including the α-globin genes, indicating that it could be a global transcriptional regulator. In addition to its role in the ATR-X syndrome, *XH2* may be a good candidate for other forms of X-linked mental retardation mapping to Xq13.

Introduction

X-linked mental retardation (XLMR) occurs at a frequency of between 1.8 and 3.4 per 1000 individuals and includes at least 50 distinct syndromes (Neri et al., 1994). Together these represent the most frequent causes of mental retardation in males, and some forms of XLMR provide a significant contribution to mental handicap in manifesting female carriers. The ATR-X syndrome is a severe, nonprogressive type of XLMR characterized by its association with an unusual form of α-thalassemia that can be readily detected by demonstrating hemoglobin H (β4) inclusions in the peripheral red blood cells (Weatherall et al., 1981, cases 2 and 3; Wilkie et al., 1990; Gibbons et al., 1995). Affected boys have a typical facial appearance, genital abnormalities, severe psychomotor retardation, and a wide range of minor congenital abnormalities. Carrier females may have very mild hematologic changes but are otherwise normal. Previous studies localized the ATR-X syndrome to an interval of approximately 11 cM between *DXS106* and *DXYS1X* (Xq12-q21.31) with a peak lod score of 5.42 (*DXS72*) (Gibbons et al., 1992); a significant proportion of all XLMR maps to this region (Neri et al., 1994).

We have previously suggested that the remarkably consistent core of clinical findings in boys with the ATR-X syndrome indicates that the molecular defect probably involves a single gene rather than a group of contiguous genes (Gibbons et al., 1995). The wide spectrum of tissues and systems affected, however, suggests that the mutations within such a gene exert pleiotropic effects throughout development. In this respect, it is of interest that female carriers have a remarkably skewed (greater than 20:1) pattern of X inactivation in tissues representing mesoderm (blood), endoderm (buccal mucosa), and ectoderm (hair root), consistent with a strong selective disadvantage conferred by the mutated gene (Gibbons et al., 1992).

The invariable occurrence of α-thalassemia in this syndrome suggests that the X-linked gene encodes a protein that, when mutated, down-regulates expression of the structurally normal α-globin genes (on chromosome 16) in addition to modifying expression of other (as yet unidentified) genes. Expression of the α-globin genes has been extensively studied: the critical *cis*-acting regulatory sequences have been defined, and some of the proteins that interact with them have been identified (Higgs et al., 1990; Jarman et al., 1991). In the ATR-X syndrome, we have previously shown that although all of these *cis*-acting regulatory sequences are intact and retain their normal patterns of methylation, the levels of α-globin mRNA and α-globin chain synthesis are reduced to 30%–65% of normal (Wilkie et al., 1990). Analysis of α-globin gene expression in this syndrome may thus provide an opportunity to elucidate the mechanism(s) by which mutation of a *trans*-acting factor could perturb the function of target genes with widely differing roles.

To refine the localization of the disease gene responsible for the ATR-X syndrome, we have performed linkage analysis in nine further families and identified key recombinants that reduce the area of interest to 1.4 cM (estimated to be 15 Mb) between *DXS453* and *DXS72*, within Xq13.1-q21.1. Recently, several cDNAs from this area have been isolated, and therefore we have examined these genes as candidates for the ATR-X syndrome. We have further characterized one of these cDNAs, which encodes a gene of the *SNF2* family of putative helicases, and identified two premature in-phase stop mutations, seven missense mutations, and a small deletion that severely down-regulates (to <1%) expression of this gene in ATR-X patients.

Results

Refining the Localization of ATR-X by Linkage Analysis

Previous linkage analysis (Gibbons et al., 1992) in seven pedigrees localized the ATR-X locus to the proximal region of Xq (Xq12-q21.31). In this study we examined a further

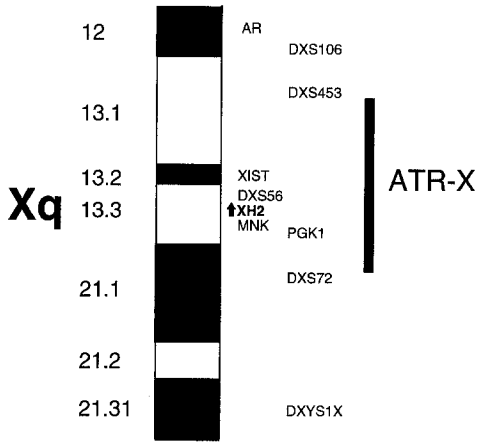

Figure 1. Localization of the ATR-X Locus

The order and cytogenetic positions of loci referred to in the text are shown: the right-hand column refers to markers used for haplotype analysis; the left-hand column refers to other markers of relevance. An arrow shows the direction of transcription of *XH2*. Details of loci and probes are available from the Genome Data Base.

nine families (ATR-X pedigree numbers 12, 14, 17, 20, 21, 22, 23, 24, and 25). These and all other pedigrees used in this study are summarized in Gibbons et al. (1995). Using data from all 16 families, two-point lod scores were calculated for θs between the ATR-X locus and five previously ordered polymorphic loci (Figure 1; Table 1). A peak lod score of 9.56 at $\theta = 0$ was seen at *PGK1*. *DXS453* and *DXS72* are the closest flanking markers that demonstrate recombination with the ATR-X locus. These loci lie approximately 1.4 cM apart (Wang et al., 1994) within the region Xq13.1-q21.1 and thus refine the localization of ATR-X to this segment of the X chromosome (Figure 1).

A Search for Mutations of ATR-X

High resolution cytogenetic analysis of all ATR-X patients studied to date has revealed no abnormality (Gibbons et al., 1995). Since this area of the genome has been well characterized (Schlessinger et al., 1993), we examined the X chromosome of up to 26 affected boys from unrelated families by the use of polymerase chain reaction (PCR), pulsed field gel electrophoresis, or conventional Southern blot hybridization using 29 previously described markers distributed across this region (data not shown). Using this approach we found no evidence for allele loss or genomic rearrangement. In a further attempt to identify mutations, we used previously described cDNAs (*DXS1005E* and *DXS1103E*) and six partial cDNA fragments (clones *23.1*, *J15*, *J14*, *E9*, *E4*, and *E13*) that had been isolated by direct cDNA selection (Gecz et al., 1993; L. V. et al., unpublished data) using total yeast artificial chromosome (YAC) DNA from within the candidate region. Using the 84 bp cDNA fragment (*E4*) isolated with YAC4551 (which includes the marker *DXS56*; Figure 1), we noted an absent hybridization signal in one individual (26.3; proband in McPherson et al., 1995) out of 26 ATR-X patients (Figure 2a). This same band was also shown to be missing from 26.3 (data not shown) when the panel of DNA samples was rescreened with the probe *XNP3/4* (1000 bp) obtained by reverse transcriptase (RT)–PCR of cDNA corresponding to the 3′ end of a previously described gene that had been localized to this region (Gecz et al., 1994; Stayton et al., 1994).

Characterization of cDNAs Corresponding to *E4* and *XNP3/4*

Since both *E4* and part of *XNP3/4* were shown to be missing in patient 26.3, it seemed possible that they originated from the same gene. Northern blot analysis using RNA from Epstein–Barr virus (EBV)-transformed lymphoblasts from normal individuals showed that both probes identified indistinguishable, widely expressed 9.5 kb mRNA transcripts (data not shown), and subsequent sequence analysis (see below) confirmed that *E4* lies entirely within *XNP3/4*.

To allow further characterization of this gene in ATR-X patients, we used *XNP3/4* and *E4* to isolate a set of overlapping cDNA clones from human fetal brain and human heart cDNA libraries (Figure 3b). We also generated and cloned cDNA fragments by RT–PCR of cDNA derived from a normal EBV-transformed lymphoblastoid cell line, using the previously described primers *XNP3*, *XNP4*, *XNP7*, and *XNP8* or their reverse complements (Gecz et al., 1994). These clones, which mapped back to YAC4551 (data not shown), were then sequenced. Subsequent analysis demonstrated that the clones we had characterized correspond to the previously described protein and cDNA sequences encoding a widely expressed putative helicase known as *XNP* (Gecz et al., 1994) or *XH2* (Stayton et al., 1994), as set out in Figure 3.

Table 1. ATR-X Syndrome: Two-Point Lod Scores

	Recombination Fraction								
ATR-X versus	0.00	0.01	0.05	0.10	0.20	0.30	0.40	$\hat{\theta}$	Lod_{max}
DXS106	−∞	5.36	6.11	5.85	4.61	3.01	1.32	0.052	6.11
DXS453	−∞	7.68	7.73	7.18	5.60	3.69	1.65	0.027	7.84
PGK1	9.56	9.40	8.71	7.79	5.79	3.64	1.49	0.000	9.56
DXS72	−∞	6.93	7.57	7.19	5.70	3.77	1.65	0.045	7.57
DXYS1X	−∞	4.25	5.66	5.66	4.62	3.06	1.33	0.072	5.74

PGK1 and the dinucleotide tandem repeat sequences PY5-10 and PY2-31 lie on the same 450 kb YAC (Graeber et al., 1992) and were consequently treated as a single locus for the purpose of this analysis. Linkage analysis was performed using the LINKAGE program, version 5. Penetrance figures used for linkage analysis are given in Gibbons et al. (1992).

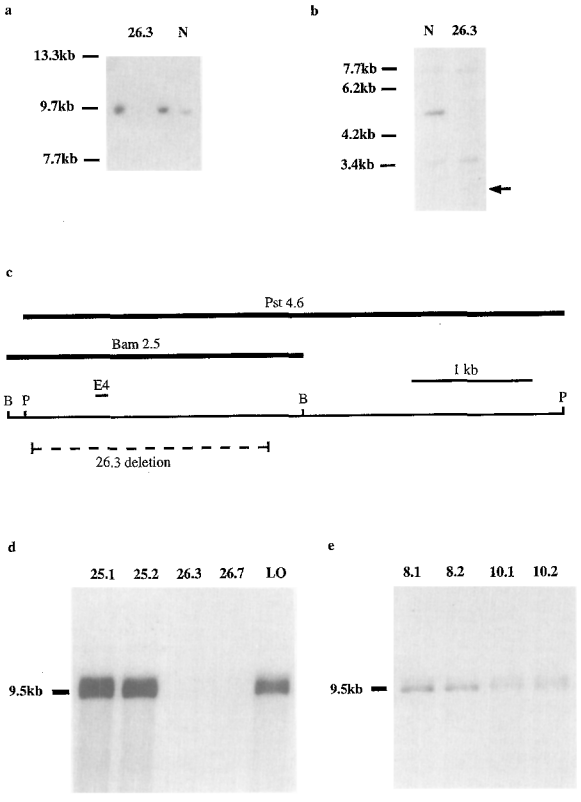

Figure 2. Characterization of a Deletion in Patient 26.3

(a) Genomic DNA (10 μg) from 26 patients and one normal individual was digested with Asp-718 and hybridized to the 84 bp cDNA probe *E4* (representative results from four individuals are shown). The sizes of the DNA markers are indicated in kilobases on the left. Lanes corresponding to patient 26.3 and a normal control (N) are indicated.

(b) PstI digest probed with 2.5 kb BamHI fragment from cosmid ICRFc104A1814 that spans the 26.3 deletion and gives rise to an altered band indicated by the arrow in lane 26.3. Two cross-hybridizing bands are also present.

(c) Map of genomic DNA spanning the deletion in 26.3. Abbreviations: B, BamHI; P, PstI. The 2.5 kb BamHI and 4.6 kb PstI fragments used as probes are shown. The location of the cDNA fragment *E4* is indicated.

(d) Northern blot analysis of RNA from ATR-X patients. Poly(A)$^+$-selected RNA derived from EBV-transformed cell lines (1–2 μg) was loaded in each lane and hybridized to *XNP3/4*. Patients 25.1 and 25.2 are affected maternal cousins; 26.3 and 26.7 are affected maternal cousins, both of whom have a deletion of 1973 bp from *XH2* as described in the text; patient LO has a similar phenotype to ATR-X but without α-thalassemia. Prolonged exposure reveals very faint but faster-migrating bands in 26.3 and 26.7.

(e) The Northern blot was hybridized to *XNP3/4*, identifying a doublet migrating at approximately 9.5 kb. Patients 8.1 and 8.2 are affected maternal cousins; 10.1 and 10.2 are affected sibs.

In addition to the clones corresponding to previously published sequences (GenBank accession numbers L34363 for *XNP* and U09820 for *XH2*), we also isolated and sequenced cDNAs that extend a further 1.5 kb in the 3′ direction. The cDNA fragment *5.1* (Figure 3) contains a putative polyadenylation signal and a poly(A) tail, 14 nt downstream, confirmed by sequence comparison with genomic DNA. From Northern blot analysis (see below), we estimate that the *XH2* mRNA transcripts are approximately 9.5 kb in length; our sequence data (covering 6.5 kb), combined with the previously published sequence of *XH2* (Stayton et al., 1994), cover 7605 bp of contiguous sequence, suggesting that there is an additional segment (2 kb) of uncloned sequence at the 5′ end of the cDNA. We have not ruled out the possibility of an alternative polyadenylation signal further downstream of the signal that we have identified. It is interesting to note that Stayton et al. (1994) have placed the initiation codon at position 53 of the current nucleotide sequence; however, this is somewhat arbitrary since the open reading frame is maintained upstream of this.

ATR-X Associated with a Partial Deletion of *XH2*

Preliminary studies with *E4* and *XNP3/4* suggested that at least part of the *XH2* gene had been deleted in patient 26.3 (see above). Using the small cDNA probe *E4*, we detected no signal in genomic DNA. With the larger *XNP3/4* cDNA fragment, which includes *E4* (see Figure 3b), both normal-sized and missing genomic fragments, corresponding to the missing *E4* fragments, were noted.

To characterize the deletion further, we isolated a cosmid (ICRFc104A1814) from a flow-sorted X chromosome library (the Imperial Cancer Research Fund [ICRF] Reference Library) using *XNP3/4*. A 2.5 kb BamHI fragment from this cosmid, containing the *E4* sequence (see Figure 2c), identified breakpoint fragments in several different digests of genomic DNA from patient 26.3 (e.g., see Figure 2b). Comparison of the restriction map of normal individuals and 26.3 was consistent with a genomic deletion of approximately 2 kb from within the 2.5 kb BamHI fragment, suggesting that the 2.5 kb fragment spans the deletion. This fragment was therefore sequenced, and PCR primers located close to each BamHI site were designed. When DNA from a normal individual was amplified with these primers, a 2.5 kb band was identified, whereas when we used DNA from 26.3, a 0.5 kb band was seen. Comparison of the DNA sequence from the 0.5 kb DNA fragment with the normal sequence confirmed that there is a 1973 bp genomic deletion from this region (data available on request).

To evaluate the effect of this deletion on expression of the *XH2* gene, we analyzed RNA derived from EBV-transformed lymphoblasts. Using primer *XNP3/4*, we observed two closely migrating transcripts of approximately 9.5 kb in RNA from normal individuals. In the proband 26.3 and in a similarly affected maternal cousin (26.7), both 9.5 kb transcripts were missing (see Figure 2d), indicating that *XH2* is down-regulated by this deletion and also showing that both transcripts originate from the same locus. On prolonged exposure, very faint faster-migrating bands were observed. To ensure that *XH2* was the only gene disrupted by this deletion, a Northern blot was probed with a 4.6 kb PstI genomic fragment spanning the deletion (see Figure 2c). Apart from the 9.5 kb doublet previously observed with *XNP3/4*, no other transcripts were seen (data not shown).

These findings therefore provide strong evidence that the genomic deletion observed in 26.3 causes disruption and down-regulation of *XH2*, a candidate gene for the ATR-X syndrome.

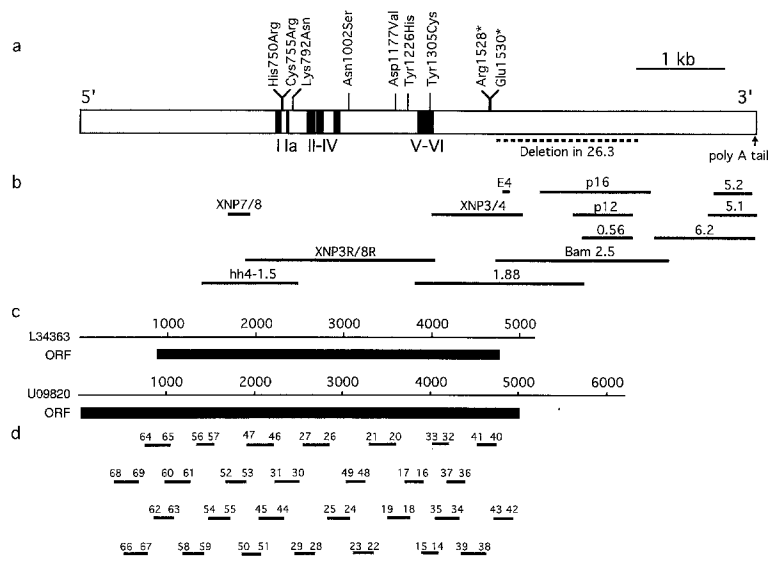

Figure 3. Characterization of *XNP*

(a) Schematic representation of the 7.6 kb cDNA characterized for *XH2* (large open box). The conserved helicase motifs (I–VI as defined by Bork and Koonin, 1993) and the mutations in the ATR-X patients are indicated. All numbering is based on a previously published sequence (GenBank accession number U09820).

(b) The component cDNA clones of the contig established in this study are shown (see Experimental Procedures). *Bam2.5* is part of a 2.5 kb BamHI genomic fragment (see Figure 2c); PCR analysis of genomic DNA and cDNA and sequence analysis revealed that the majority of this fragment (1908 bp) is exonic.

(c) The extent of the cDNA sequences and open reading frames previously determined (GenBank accession numbers L34363 and U09820) is shown.

(d) The locations of the SSCP primer pairs and their products (Table 2) are indicated.

A Search for Other Mutations of *XH2* in the ATR-X Syndrome

We did not detect absent or altered bands on Southern blot analysis in any of the 25 other ATR-X patients studied. However, in view of the findings in 26.3, described above, it seemed likely that point mutations or small deletions might be present in these cases, causing alteration in *XH2* structure or expression.

To evaluate expression, RNA from EBV-transformed lymphoblasts derived from 24 ATR-X patients was examined by Northern blot analysis using *XNP3/4*. RNA was detected in all patients studied; however, in some there was a marked alteration in the relative intensities of the two *XH2* transcripts, and the ratio was identically altered in other affected members of the same family (e.g., see Figure 2e). These findings suggested that mutations in these families were causing some perturbation in *XH2* mRNA expression or stability.

We next analyzed by single strand conformation polymorphism (SSCP) analysis the structure of *XH2* cDNA de-

Table 2. SSCP Primer Pairs

Primer A		Primer B		Product Size	Annealing Temperature (Celsius)
14	TCATAGTAAAATGACGCTCCACCTG	15	TTATATTCGACGCTTCTTGGAATCC	201	52.9
16	TGGATGTCATAAGATGGATTCCAAG	17	AAAGGTGAGGGGAAGTGGCTTC	224	52.1
18	TTCCTTGACTGTGCAGTAGTGGAA	19	GCTGAGGTTTTAGAGCATTCTGGG	257	53.4
20	GGACTTTATCCCCAATTTCCTCTG	21	TAGCTCAAGTGGAAGTGGCAGTG	290	53.8
22	AACATCATTGTCACTGCCACTTCC	23	GGACTCATCCTTGGTGTTTGCAG	204	52
24	ACCTGCCTTTCCTCTTCCACCTTC	25	CCAATTCAAAATGGTCAGTGTGCAG	261	53.2
26	TGCACACTGACCATTTTGAATTGG	27	TGCTCAAGGAAGGAATGTGAAGAG	302	51.8
28	GCCTTCATCACAAACAACAAATCAG	29	ACGTCCTCAGGAGAGAAGCTACATG	182	53.3
30	ACCACCATCTTCTTGCCACCTCTG	31	TGCATTCTTGCCCACTGTATGG	264	54.6
32	AGTATGGTATCCTTTGGCAGCATG	33	TCGGCAAGTAACTAAGCAGTCACTG	180	52.8
34	CAGCCCAAGCTGCTTTTCTTTC	35	GCAGGTGGAGCGTCATTTTACTATG	251	53.5
36	TTTGAGAGTTGAAACTGACAGGGG	37	GCCAAAGGATACCATACTTGCAGAG	224	53.7
38	TCCTGGCTGGCTTGTCTACTTAATC	39	CGTTTCAACATACCAACTGGGACC	281	56.6
40	GCTGCTGGAGCCTTCTGTTCATAAG	41	GCAGTGAGGATTCAACCTCTTGAGG	220	54.6
42	TTTTGCTTCTCATTGGGGGTG	43	AGGCTCCAGCAGCAGTACAATCAG	211	55.1
44	CGTGCTGAAATCCAGTTTGTCACAC	45	CACCCACCAAGTGTCCAATAACAAC	284	54.1
46	ACAGACTCACAGCAGCAATCCCAC	47	CGGAAGATTCTTAAAGATGATAAA	295	59
48	CGGAGCTTAAACTCATGGAGGTTTC	49	TGTGGGCAATAATAGTGAAGGTGG	201	52.6
50	GGAGGAGGAAGATGAAAATGATG	51	ACACTTGGTGGGTGAAGCATC	211	52.6
52	GCCCAGAACAAGGTCTGCAAAG	53	TGCCTTTTCCAGGAGACTTGG	227	53.7
54	CAGCATAGGCTTTTCGGCAC	55	TTCCAACTCTGCTTTCTTTGCAG	231	53.1
56	TGGATCTTCAGATGATGAGCCAG	57	AGATTCTCCGAAACTCAAAAGCC	187	51.7
58	TGGATTTTGCCAATCTTCAGG	59	TTTGCTTCCTCATCTCCTGGG	234	52.2
60	AAAGAAGGCAGTCATTGTCAAGG	61	TTGTCGTCATCATCATCATCCAC	270	52.6
62	TCAACACGCATAGAATTGAGGG	63	TGAGGATGTAATGTCAGCTTGCTTC	225	51.1
64	GCAAGTTGCTTGAAAGAGTTCATGG	65	TCTCCTTGACAATGACTGCCTTC	270	52
66	TGATGGCAACTGAAAAGTTACCTG	67	CGTGTCTCTTCCATGAACTCTTTC	277	51.2
68	AAAACGCAGAGCAAAAACGG	69	TCCCTGTTGACTTCTCAGCATAATC	268	50.5

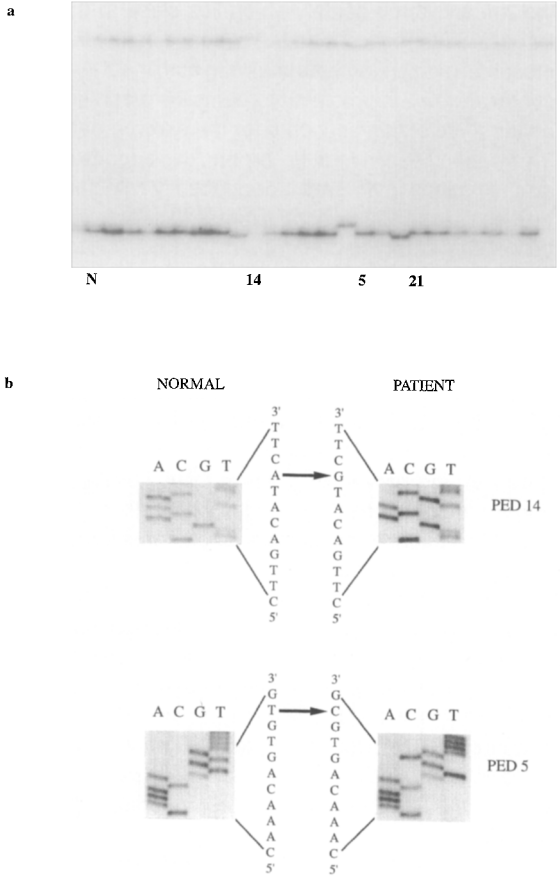

Figure 4. Mutation Analysis in Patients with ATR-X

(a) SSCP analysis of a panel of ATR-X patients and a normal individual (N), utilizing the primer pair *XNP30/31*. Altered bands were noted in pedigrees 14, 5, and 21 and were subjected to sequence analysis. (b) Sequence analysis of the sense strand from pedigrees 14 and 5 showing an A to G transition at nucleotide 2302 and a T to C transition at nucleotide 2316, respectively.

rived from EBV-transformed lymphoblasts from 24 ATR-X patients, using primers designed to generate overlapping 150–300 bp fragments (Figure 3d; Table 2). In this way, 4.5 kb of the 9.5 kb mRNA transcript, including most of the currently known coding sequence, was analyzed. In addition, amplified fragments generated from primers 7 and 8, which contain an imperfect trinucleotide repeat (Gecz et al., 1994), were examined and found to be of normal size in 24 patients. By SSCP analysis we detected 11 altered bands corresponding to cDNA fragments from nine unrelated pedigrees. In two cases, SSCP changes were present in adjacent fragments where a mutation occurred in the region of overlap between PCR fragments (Table 3).

Variant SSCP fragments were amplified and sequenced, revealing nine different mutations (Figure 4; Table 3); in one case, the same mutation was seen in two similarly affected individuals from the same family (21.1 and 21.2). In pedigrees 13 and 14, fetuses with the "affected" haplotype were found to have the same mutation as the affected proband. The locations of these mutations in *XH2* are shown in Figure 3a. Two (C4635T and G4641T) give rise to premature in-phase stop codons. The remainder are missense mutations that fall in the central portion of the molecule, a region that is known to be highly conserved among members of the superfamily of DNA helicases to which *XH2* belongs (Bork and Koonin, 1993). Although the mutations did not fall precisely into the previously proposed motifs (I–VI), based on protein sequence alignment, with one exception (N1002S), they occurred in areas that are well conserved among members of the helicase superfamily (Figure 5). It is interesting that none of the mutations involves a highly convserved amino acid, which may mean that such mutations are lethal. Given the relatively conservative amino acid change in N1002S and its nonconserved location, we cannot be certain that this is the disease-causing mutation in family 20. However, the nature of the other eight mutations identified here, their occurrence in similarly affected individuals from the same family, and their locations in highly conserved regions of *XH2* suggest that they cause significant disruption of its structure and function, thereby leading to the ATR-X phenotype.

Discussion

We have shown that the ATR-X syndrome results from mutations involving *XH2*, a gene that lies in the region Xq13.3. A wide variety of mutations distributed throughout the central portion of the XH2 protein give rise to a surprisingly uniform phenotype that includes severe psychomotor retardation, a similar pattern of facial dysmorphism, and α-thalassemia. Consistent with the complex ATR-X phenotype, studies of *XH2* and its murine homolog *Mxnp* (Gecz et al., 1994; Stayton et al., 1994; unpublished data)

Table 3. Summary of ATR-X Mutations Identified by SSCP Analysis

Pedigree Number	SSCP Primer Pair	cDNA Nucleotide Change	Amino Acid Change	Restriction Enzyme Site
14 (1, 2)	30/31, 44/45	A2302G	H750R	Creates RsaI
5	30/31	T2316C	C755R	
21 (1, 2)	30/31	G2429T	K792N	Disrupts HindIII
20	24/25	A3058G	N1002S	
23	18/19	A3583T	D1177V	Creates MseI
3	16/17, 18/19	T3729C	Y1226H	Creates HphI
13 (1, 2)	14/15	A3967G	Y1305C	
16	40/41	C4635T	R1528*	
27	40/41	G4641T	E1530*	Disrupts EarI

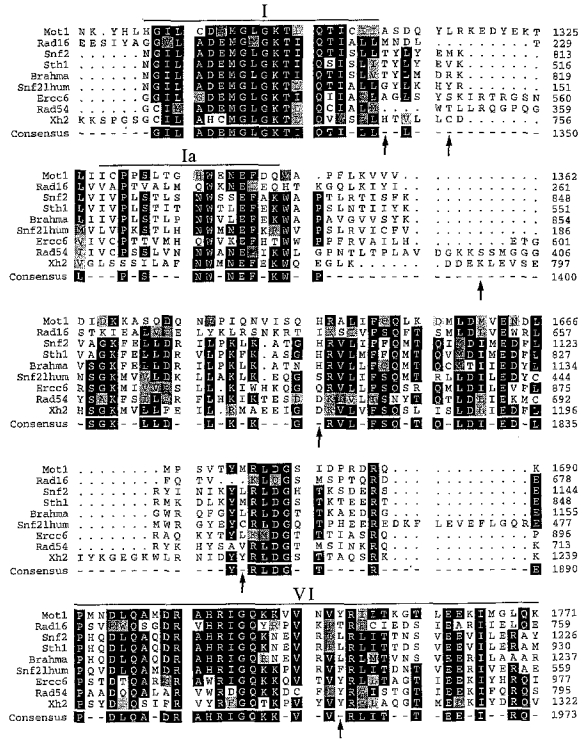

Figure 5. Location of Six ATR-X Mutations in Relation to Conserved Domains

Sequence alignment of various members of the helicase superfamily (indicated on the left). The locations of six mutations identified in ATR-X patients (arrows) are shown in relation to conserved regions both within and outside the helicase domains (overlined). The N1002S missense mutation that occurs in a nonconserved region is not shown. The amino acids are divided into blocks of ten, with the residue number for each family member given on the right.

have shown it to be a widely expressed gene that might play an important role in development of the central nervous system. At present, the role of *XH2* in development of the urogenital system is not known but, since 96% of ATR-X patients also have genital abnormalities (Gibbons et al., 1995), it seems likely that *XH2* will also be involved in this pathway. The most extreme urogenital abnormalities, causing severe hypogenitalism, male pseudohermaphroditism, and in some cases female gender assignment, were caused by a partial deletion of *XH2*, premature in-phase stop codons (R1528* and E1530*), and a missense mutation (C755R). We have previously shown that severe genital abnormalities appear to "breed true" with the disease (McPherson et al., 1995; Reardon et al., 1995). It will therefore be interesting to see whether specific types of mutation, perhaps those producing no XH2 protein, are associated with this particular phenotype.

Previous studies have shown that *XH2* is subject to X inactivation and is exclusively transcribed from the active X chromosome (Stayton et al., 1994). Female carriers of ATR-X have a markedly skewed pattern of X inactivation in all tissues that have been studied, and we previously considered whether this might be due to selection against the affected chromosome or to the direct involvement of the X inactivation center, which lies close to the ATR-X locus (*XIST* in Figure 1). In the absence of any large chromosomal rearrangements involving both the X inactivation center and *XH2* loci, selection seems the most likely explanation. If selection accounts for the skewed pattern of X inactivation seen in female carriers, this suggests that unless mutations in *XH2* are independently selected against in each tissue, *XH2* must be expressed prior to differentiation of the embryonic endoderm, ectoderm, and mesoderm in humans. It is therefore of interest that *Mxnp* is already expressed in the primitive streak at the earliest stage of mouse development (7.0 days postcoitum) that has yet been studied (Stayton et al., 1994).

Sequence analysis has shown that *XH2* belongs to the SNF2 subgroup (Kolstø et al., 1993) of a superfamily of proteins with similar DNA-dependent ATPase (I and II in Figure 3a) and proven or putative DNA helicase (III–VI in Figure 3a) domains (Bork and Koonin, 1993). Regions outside the helicase domains also show considerable conservation among members of this SNF2 subgroup (Matson et al., 1994). These proteins are involved in a wide variety of cellular functions, including mitotic chromosome segregation (lodestar), DNA recombination and repair (RAD16, RAD54, and ERCC6), and global regulation of transcription (SWI2/SNF2, MOT1, brahma) (reviewed by Carlson and Laurent, 1994). Using comparative analysis, we found no additional protein motifs associated with particular subgroups of the SNF2 family (e.g., bromodomains [Haynes et al., 1992], chromodomains [Delmas et al., 1993], DNA-binding or activation motifs [Stayton et al., 1994], kinases [Tsuchiya et al., 1992], and TPR boxes [Davis et al., 1992]). Furthermore, recent evidence has shown that although some proteins in this superfamily are predominantly involved with either repair (e.g., RAD16) or transcriptional control (e.g., SWI2/SNF2), the two processes are closely connected (Drapkin et al., 1994). Therefore, at present, the best guide to the normal function of *XH2* comes from an evaluation of the phenotypic effects of mutation. In contrast with other human genetic diseases (xerodema pigmentosum [Hoeijmakers, 1993], trichothiodystrophy [Broughton et al., 1994], and Cockayne's syndrome [Troelstra et al., 1992]) that result from mutations in genes of the helicase superfamily, in the ATR-X syndrome there is no clinical evidence for ultraviolet sensitivity or the premature development of malignancy. Furthermore, cytogenetic analysis has not demonstrated any evidence of abnormal chromosome breakage or segregation. However, the consistent association of ATR-X, due to a wide range of independent mutations, with down-regulation of α-globin expression suggests that *XH2* is involved in the regulation of gene expression.

Mutations of *XH2* might affect any stage of gene expression, but the *SNF2* family to which it belongs is distinct from the group of DEAD box RNA helicases, some members of which have been implicated in RNA processing, stability, or translation (Pain, 1986; Strauss and Guthrie, 1991; Iost and Dreyfus, 1994). From a functional point of view, perturbation of *XH2* appears to resemble most closely the effects that would be expected from disruption of a global transcriptional regulator, the best-studied examples being

brahma in Drosophila (Tamkun et al., 1992) and the *SNF/SWI* family in Saccharomyces cerevisiae (Winston and Carlson, 1992). Although the precise mechanism by which these proteins exert their effect is not known, they appear to regulate transcription of restricted classes of genes via interactions with gene-specific activators, most likely by altering chromatin structure to relieve repression (Carlson and Laurent, 1994). In this respect, it is most interesting that mutations in *XH2* down-regulate expression of the α-globin genes but do not appear to affect the closely related β-globin genes. Activation of the α- and β-globin genes appears to involve a common group of lineage-specific (GATA-1 and NF-E2) and ubiquitous (CACC box) DNA-binding proteins. However, despite their common ancestry, human α- and β-globin clusters have been previously shown to lie in remarkably different chromosomal environments (reviewed by Vyas et al., 1992). The α genes are associated with CpG-rich islands located in a constitutively "open," transcriptionally active chromatin domain that is early replicating in all cell types, and their expression is regulated by a remote tissue-specific enhancer. By contrast, the β cluster is contained within a segment of chromatin that opens and becomes early replicating in a tissue-specific manner under the influence of a remote locus control region (Grosveld et al., 1987; Forrester et al., 1990). In addition, unlike the α cluster, the β cluster contains several matrix attachment regions (Jarman and Higgs, 1988) that may act as histone H1 nucleation sites (Zhao et al., 1993) and thereby influence long-range chromatin structure. The specific effect of *XH2* mutations on α- rather than β-globin expression may therefore indicate that *XH2* is a transcriptional activator that, like other similar proteins (e.g., SNF/SWI and brahma), works via an interaction with chromatin. It is interesting to note that all the diverse mutations of *XH2*, including a deletion that severely reduces the level of *XH2* mRNA, cause α-thalassemia, implying that rather than the α genes being affected as "innocent bystanders" in these mutations, *XH2* is normally necessary for their correct regulation. Further analysis of the role of *XH2* in the control of gene expression is therefore likely to elucidate the normal mechanisms of α-globin regulation.

It is of interest that many forms of XLMR and complex human genetic disorders map to Xq13-q21 (Schlessinger et al., 1993). Although the patients that we have studied have been highly selected for a specific phenotype, they have mutations distributed throughout the *XH2* gene. Although this might suggest that any perturbation of *XH2* expression will cause ATR-X, mutations should be sought in other forms of XLMR that map to this region. We have identified several pedigrees in which affected individuals have a phenotype very similar to ATR-X but in whom we cannot detect hemoglobin H inclusions; these include a pedigree with the Juburg–Marsidi syndrome (Saugier-Veber et al., 1995). Analysis of *XH2* in all of these families is currently underway.

Experimental Procedures

DNA/RNA Sources and Analysis

EBV-transformed lymphoblastoid patient cell lines (Gibbons et al., 1995) were used as a source of both genomic DNA and RNA. cDNA was prepared from cell line RNA by RT–PCR as described by Brown et al. (1990). YAC4551 (provided by Dr. A. Monaco, ICRF, Oxford, England) was used for mapping back cDNA probes. Cosmids covering the 3′ end of the *XH2* gene were obtained by screening the X chromosome flow-sorted ICRF Reference Cosmid Library (Lehrach et al., 1990) with probe *XNP3/4*. The 2.5 kb BamHI and the 4.6 kb PstI fragments from cosmid ICRFc104A1814 were subcloned into pBluescript II. Northern and Southern blots were performed as described in Sambrook et al. (1989). Final washing conditions were performed at either room temperature or at 65°C in 0.1 × SSC and 0.1% SDS for 5 min.

cDNA Isolation and Characterization

Initially, a human fetal brain library in λ phage (from Dr. P. C. Harris, Institute of Molecular Medicine, Oxford, England) was screened with *E4*. One cDNA was isolated that contained two EcoRI fragments (1.88 and 0.56) that were cloned into pBluescript II for further analysis. Three additional cDNAs (*P12*, *P16*, and data not shown) were isolated from this library after screening with the 0.56 fragment. Five additional cDNA fragments (*hh4*, *5.1*, *5.2*, *6.2*, and data not shown) were isolated from a human heart λ phage cDNA library (Stratagene Cloning Systems) or from a human fetal brain plasmid cDNA library (from Dr. D. J. Blake, Institute of Molecular Medicine, Oxford, England) prepared in the vector pcDNAII (Invitrogen Corporation), probing with either *XNP3R/8R*, the BamHI 2.5 kb genomic fragment, or the 2.1 kb BamHI–PstI genomic fragment (Figure 2c). cDNA fragments *XNP3/4* and *XNP3R/8R* were amplified from cDNA from a normal individual and cloned into pUC18 at the SmaI site.

DNA Sequence Analysis

Double-stranded sequencing of genomic or cloned cDNA fragments (Chen and Seeburg, 1985) was performed with Sequenase T7 polymerase (United States Biochemicals) using sequential truncations (Henikoff, 1984) or selected oligonucleotide primers (see Table 2). ATR-X patients showing an altered SSCP band and normal control samples were directly sequenced after PCR amplification as described in Thein and Hinton (1991).

PCR Amplification and SSCP Analysis

XNP3/4 and *XNP7/8* were amplified from cDNA derived from a normal EBV-transformed lymphoblastoid line utilizing primers *XNP3*, *XNP4*, *XNP7*, and *XNP8* using previously described conditions (Gecz et al., 1994). Similarly, *XNP3R/8R* was amplified using primers *XNP3R* and *XNP8R*, the reverse complements of *XNP3* and *XNP8*, respectively. To characterize the 26.3 breakpoint, we employed primers *Bam2.5f* (5′-GATCCACCAAACCTCACAAATTCTG-3′) and *Bam2.5r* (5′-GCCAGCCTTCTGCCTTATACTTG-3′). Cycles were as follows: once at 94°C for 5 min, at 52.7°C for 1 min, and at 72°C for 2.5 min; 28 times at 94°C for 1 min, at 52.7°C for 1 min, and at 72°C for 2.5 min; once at 94°C for 1 min, at 52.7°C for 1 min, and at 72°C for 10 min. Typically, all PCRs were performed in PCR buffer (50 mM KCl,10 mM Tris–HCl [pH 8.8], 0.01% gelatin) containing 0.5–2.0 mM $MgCl_2$, 200 μM dNTPs, 0.5 μM of each primer, 2.5 U/100 μl AmpliTaq (Perkin-Elmer Cetus), 100 ng of genomic DNA or 1 μl of cDNA prepared as described above, in a final volume of 50 μl. For SSCP analysis, PCRs were performed in 10 μl with 1 μl of a 5-fold dilution of cDNA template and the addition of 0.1 μl of [α-^{32}P]dCTP (3000 Ci/mmol; Amersham). Annealing temperature was adjusted for each primer pair (Table 2). Cycles were as follows: once at 94°C for 5 min, annealing for 1 min, and then at 72°C for 1 min; 28 times at 94°C 1 for min, annealing for 1 min, and then at 72°C for 1 min; once at 94°C for 1 min, annealing for 1 min, and then at 72°C for 4 min. Analysis of labeled SSCP products was performed according to the method of Glavac and Dean (1993).

Acknowledgments

Correspondence should be addressed to D. R. H. We are grateful to Prof. Sir D. Weatherall for initiating this project and for his continued help and support throughout. We also thank the families for their cooperation and the clinicians who referred patients. We are grateful to Dr. W. G. Wood and Dr. A. O. M. Wilkie for help, for ideas, and for careful reading of the manuscript. J. Sloane-Stanley and H. Ayyub are thanked for their excellent technical assistance. We thank Dr. M.

Fontés for permitting the use of probe *E4*. We thank Dr. A. Monaco for providing many of the probes and YACs used in this study and Dr. D. Blake for providing a cDNA library. We thank L. Rose for preparation of the manuscript. R. J. G. was funded by Action Research, D. J. P. by the Canadian Medical Research Council, and L. V. by the French Muscular Dystrophy Association.

Received December 2, 1994; revised January 16, 1995.

References

Bork, P., and Koonin, E. V. (1993). An expanding family of helicases within the 'DEAD/H' superfamily. Nucl. Acids Res. *21*, 751–752.

Broughton, B. C., Steingrimsdottir, H., and Weber, C. A. (1994). Mutations in the xeroderma pigmentosum group D DNA repair/transcription gene in patients with trichothiodystrophy. Nature Genet. *7*, 189–194.

Brown, C. J., Flenniken, A. M., Williams, B. R. G., and Willard, H. F. (1990). X chromosome inactivation of the human TIMP gene. Nucl. Acids Res. *18*, 4191–4195.

Carlson, M., and Laurent, B. C. (1994). The SNF/SWI family of global transcriptional activators. Curr. Opin. Cell Biol. *6*, 396–402.

Chen, E. J., and Seeburg, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA *4*, 165–170.

Davis, J. L., Kunisawa, R., and Thorner, J. (1992). A presumptive helicase (*MOT1* gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. *12*, 1879–1892.

Delmas, V., Stokes, D. G., and Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SW12-like helicase domain. Proc. Natl. Acad. Sci. USA *90*, 2414–2418.

Drapkin, R., Sancar, A., and Reinberg, D. (1994). Where transcription meets repair. Cell *77*, 9–12.

Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus. Genes Dev. *4*, 1637–1649.

Gecz, J., Villard, L., Lossi, A. M., Millasseau, P., Djabali, M., and Fontes, M. (1993). Physical and transcriptional mapping of DXS56–PGK1 1 Mb region: identification of three new transcripts. Hum. Mol. Genet. *2*, 1389–1396.

Gecz, J., Pollard, H., Conzalez, G., Villard, L., Stayton, C., Millasseau, P., Khrestchatisky, M., and Fontes, M. (1994). Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to *PGK1* in Xq13.3. Hum. Mol. Genet. *3*, 39–44.

Gibbons, R. J., Suthers, G. K., Wilkie, A. O. M., Buckle, V. J., and Higgs, D. R. (1992). X-linked α thalassemia/mental retardation (ATR-X) syndrome: localisation to Xq12-21.31 by X-inactivation and linkage analysis. Am. J. Hum. Genet. *51*, 1136–1149.

Gibbons, R. J., Brueton, L., Buckle, V. J., Burn, J., Clayton-Smith, J., Davison, B. C. C., Gardner, R. J. M., Homfray, T., Kearney, L., Kingston, H. M., Newbury-Ecob, R., Porteous, M. E. P., Wilkie, A. O. M., and Higgs, D. R. (1995). The clinical and hematological features of the X-linked α thalassemia/mental retardation syndrome (ATR-X). Am. J. Med. Genet., in press.

Glavac, D., and Dean, M. (1993). Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. Hum. Mutat. *2*, 404–414.

Graeber, M. B., Kupke, K. G., and Müller, U. (1992). Delineation of the dystonia–Parkinsonism syndrome locus in Xq13. Proc. Natl. Acad. Sci. USA *89*, 8245–8248.

Grosveld, F., Blom van Assendelft, G., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell *51*, 975–985.

Haynes, S. R., Dollard, C., Winston, F., Beck, S., Trowsdale, J., and Dawid, I. B. (1992). The bromodomain: a conserved sequence found in human, *Drosophila* and yeast proteins. Nucl. Acids Res. *20*, 2603.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene *28*, 351–359.

Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I.-M., and Ayyub, H. (1990). A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Dev. *4*, 1588–1601.

Hoeijmakers, J. H. J. (1993). Nucleotide excision repair II: from yeast to mammals. Trends Genet. *9*, 211–217.

Iost, I., and Dreyfus, M. (1994). mRNAs can be stabilized by DEAD-box proteins. Nature *372*, 193–196.

Jarman, A. P., and Higgs, D. R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. EMBO J. *7*, 3337–3344.

Jarman, A. P., Wood, W. G., Sharpe, J. A., Gourdon, G., Ayyub, H., and Higgs, D. R. (1991). Characterization of the major regulatory element upstream of the human α-globin gene cluster. Mol. Cell. Biol. *11*, 4679–4689.

Kolstø, A.-B., Bork, P., Kvaløy, K., Lindback, T., Grønstadt, A., Kristensen, T., and Sander, C. (1993). Prokaryotic members of a new family of putative helicases with similarity to transcription activator SNF2. J. Mol. Biol. *230*, 684–688.

Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A. P., Nizetic, D., Zehetner, G., and Poutska, A. (1990). Hybridization fingerprinting in genome mapping and sequencing. In Genome Analysis, Volume 1: Genetic and Physical Mapping, K. E. Davies and S. M. Tilghman, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 39–81.

Matson, S. W., Bean, D. W., and George, J. W. (1994). DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. Bioessays *16*, 13–22.

McPherson, E., Clemens, M., Gibbons, R. J., and Higgs, D. R. (1995). X-linked α thalassemia/mental retardation (ATR-X) syndrome: a new kindred with severe genital anomalies and mild hematologic expression. Am. J. Med. Genet., in press.

Neri, G., Chiurazzi, P., Arena, J. F., and Lubs, H. A. (1994). XLMR genes: update 1994. Am. J. Med. Genet. *51*, 542–549.

Pain, V. M. (1986). Initiation of protein synthesis in mammalian cells. Biochem. J. *235*, 625–637.

Reardon, W., Gibbons, R. J., Winter, R. M., and Baraitser, M. (1995). Sex reversal in the α thalassemia/mental retardation (ATR-X) syndrome: a further case. Am. J. Med. Genet., in press.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Saugier-Veber, P., Gibbons, R. J., Munnich, A., and Lyonnet, S. (1995). Lumping Juberg–Marsidi syndrome and X-linked α-thalassemia/mental retardation syndrome? Am. J. Med. Genet., in press.

Schlessinger, D., Mandel, J. L., Monaco, A. P., Nelson, D. L., and Willard, H. F. (1993). Report of the Fourth International Workshop on Human X Chromosome Mapping 1993. Cytogenet. Cell Genet. *64*, 148–169.

Stayton, C. L., Dabovic, B., Gulisano, M., Gecz, J., Broccoli, V., Giovanazzi, S., Bossolasco, M., Monaco, L., Rastan, S., Boncinelli, E., Bianchi, M. E., and Consalez, G. G. (1994). Cloning and characterization of a new human Xq13 gene, encoding a putative helicase. Hum. Mol. Genet. *3*, 1957–1964.

Strauss, E. J., and Guthrie, C. (1991). A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. Genes Dev. *5*, 629–641.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992). *brahma*: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SW12. Cell *68*, 561–572.

Thein, S. L., and Hinton, J. (1991). A simple and rapid method of direct sequencing using Dynabeads. Br. J. Haemat. *79*, 113–115.

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. J. (1992). *ERCC6*, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell *71*, 939–953.

Tsuchiya, E., Uno, M., Kiguchi, A., Masuoka, K., Kanemori, Y., Okabe,

S., and Mikayawa, T. (1992). The *Saccharomyces cerevisiae NPS1* gene, a novel *CDC* gene which encodes a 160 kDa nuclear protein involved in G_2 phase control. EMBO J. *11*, 4017–4026.

Vyas, P., Vickers, M. A., Simmons, D. L., Ayyub, H., Craddock, C. F., and Higgs, D. R. (1992). Cis-acting sequences regulating expression of the human α-globin cluster lie within constitutively open chromatin. Cell *69*, 781–793.

Wang, L. H., Collins, A., Lawrence, S., Keats, B. J., and Morton, N. E. (1994). Integration of gene maps: chromosome X. Genomics *22*, 590–604.

Weatherall, D. J., Higgs, D. R., Bunch, C., Old, J. M., Hunt, D. M., Pressley, L., Clegg, J. B., Bethlenfalvay, N. C., Sjolin, S., Koler, R. D., Magenis, E., Francis, J. L., and Bebbington, D. (1981). Hemoglobin H disease and mental retardation: a new syndrome or a remarkable coincidence? N. Engl. J. Med. *305*, 607–612.

Wilkie, A. O. M., Zeitlin, H. C., Lindenbaum, R. H., Buckle, V. J., Fischel-Ghodsian, N., Chui, D. H. K., Gardner-Medwin, D., MacGillivray, M. H., Weatherall, D. J., and Higgs, D. R. (1990). Clinical features and molecular analysis of the α thalassemia/mental retardation syndromes. II. Cases without detectable abnormality of the α globin complex. Am. J. Hum. Genet. *46*, 1127–1140.

Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. Trends Genet. *8*, 387–391.

Zhao, K., Kas, E., Gonzalez, E., and Laemmli, U. K. (1993). SAR-dependent mobilization of histone H1 by HMG-I/Y *in vitro*: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J. *12*, 3237–3247.

GenBank Accession Number

The accession number for the sequence reported in this paper is X83753.