Neuroscience 146 (2007) 236–247

ARISTALESS-RELATED HOMEOBOX GENE, THE GENE RESPONSIBLE FOR WEST SYNDROME AND RELATED DISORDERS, IS A GROUCHO/TRANSDUCIN-LIKE ENHANCER OF SPLIT DEPENDENT TRANSCRIPTIONAL REPRESSOR

O. McKENZIE,[a1] I. PONTE,[b1] M. MANGELSDORF,[a,c] M. FINNIS,[a] G. COLASANTE,[b] C. SHOUBRIDGE,[a] S. STIFANI,[d] J. GÉCZ[a,c*] AND V. BROCCOLI[b]

[a]*Department of Genetic Medicine, Women's and Children's Hospital, 72 King William Road, North Adelaide, Adelaide, South Australia, 5006, Australia*

[b]*Stem Cell Research Institute, DIBIT, San Raffaele Scientific Park, Milan, Italy*

[c]*Departments of Paediatrics and Molecular Biosciences, University of Adelaide, Adelaide, Australia*

[d]*Centre for Neuronal Survival, Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada*

Abstract—*Aristaless*-related homeobox gene (*ARX*) is an important paired-type homeobox gene involved in the development of human brain. The *ARX* gene mutations are a significant contributor to various forms of X-chromosome-linked mental retardation with and without additional features including epilepsy, lissencephaly with abnormal genitalia, hand dystonia or autism. Here we demonstrate that the human ARX protein is a potent transcriptional repressor, which binds to Groucho/transducin-like enhancer of split (TLE) cofactor proteins and the TLE1 in particular through its octapeptide (Engrailed homology repressor domain (eh-1) homology) domain. We show that the transcription repression activity of ARX is modulated by two strong repression domains, one located within the octapeptide domain and the second in the region of the polyalanine tract 4, and one activator domain, the *aristaless* domain. Importantly, we show that the transcription repression activity of ARX is affected by various naturally occurring mutations. The introduction of the c.98T>C (p.L33P) mutation results in the lack of binding to TLE1 protein and relaxed transcription repression. The introduction of the two most frequent ARX polyalanine tract expansion mutations increases the repression activity in a manner dependent on the number of extra alanines. Interestingly, deletions of alanine residues within polyalanine tracts 1 and 2 show low or no effect. In summary we demonstrate that the ARX protein is a strong transcription repressor, we identify novel ARX interacting proteins (TLE) and offer an explanation of a molecular pathogenesis of some *ARX* mutations, including the most frequent *ARX* mutations, the polyalanine tract expansion mutations, c.304ins(GCG)$_7$ and c.428_451dup.

Key words: ARX, X-linked mental retardation, transcription repression, polyalanine tract expansion, mutation, MRX.

Aristaless-related homeobox gene (ARX) belongs to a subset of the *aristaless*-related paired-class homeobox family, which has been structurally and functionally conserved throughout the animal kingdom. The *ARX* gene plays a critical role in vertebrate development. Recently mutations in the X-chromosome-linked *ARX* homeobox gene have been associated with a diverse number of disorders including X-linked lissencephaly with abnormal genitalia (XLAG; MIM 300215) (Kitamura et al., 2002), X-linked infantile spasms (ISSX; MIM 308350) (Strømme et al., 2002) and X-linked mental retardation (MRX) (Bienvenu et al., 2002; Strømme et al., 2002). To date, over 60 different disease causing mutations have been identified in the *ARX* gene, including missense mutations, protein truncations and most commonly, polyalanine tract expansions (Gécz et al., 2006). A study by Kitamura et al. (2002) demonstrated that an *Arx* null mutation caused early perinatal lethality in male mice, diffuse CNS malformations and abnormal testes. These mice also have abnormal differentiation and deficient tangential migration of GABAergic interneurons from their birthplace in the medial ganglionic eminence (Mge) to the cerebral cortex (Kitamura et al., 2002). The *Arx* knockout mice recapitulate many of the clinical and pathological features of XLAG.

Arx has been shown to repress the transcription of Pax4 expression in pancreatic cells, mediating the correct development of α, δ and β endocrine cells (Collombat et al., 2003, 2005). Studies in *Xenopus*, however, suggest the Arx protein is not only a transcriptional repressor but can act as a bifunctional transcriptional regulator in brain development (Seufert et al., 2005). The ARX protein contains a number of conserved domains, providing some insight into its potential role. In addition to the homeodomain, the carboxyl-tail of ARX contains the *aristaless* domain, which is present in more than 40 other members of the paired class homeobox family. The function of this domain is not well understood, although a number of studies have shown it is involved in transcription activation

[1] These authors contributed equally.

*Correspondence to: J. Gécz, Department of Genetic Medicine, Women's and Children's Hospital, 72 King William Road, North Adelaide, Adelaide, South Australia, 5006, Australia. Tel: +61-8-8161-6339; fax: +61-8-8161-7342.
E-mail address: jozef.Gecz@adelaide.edu.au (J. Gécz).

Abbreviations: ARX, *aristaless*-related homeobox gene; CAT, chloramphenicol acetyltransferase; Cge, caudal ganglionic eminence; DMEM, Dulbecco's modified Eagle's medium; eh-1, Engrailed homology repressor domain; Et, eminentia thalami; Hy, hypothalamus; ISSX, X-linked infantile spasms; Lge, lateral ganglionic eminence; LUC, luciferase; Mge, medial ganglionic eminence; MRX, prefix used for families with non-syndromic X-linked mental retardation; ORF, open reading frame; *TLE*, transducin-like enhancer of split; Vt, ventral thalamus; XLAG, X-linked lissencephaly with ambiguous genitalia.

0306-4522/07$30.00+0.00 © 2007 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2007.01.038

(Collombat et al., 2003, 2005; Norris et al., 2000; Norris and Kern, 2001). The ARX protein contains four hydrophobic polyalanine tracts, suggested to be involved in protein–protein and protein–DNA interactions and as such to stabilize the interactions between transcription regulators and/or DNA (Brown and Brown, 2004). It has also been shown that expansions of the first polyalanine tract lead to protein aggregation and subsequent cell death (Nasrallah et al., 2004). ARX protein contains a highly conserved octapeptide domain located near the N-terminus. This sequence shares high similarity with the Engrailed homology repressor domain (eh-1) known to be involved in transcriptional repression both *in vitro* and *in vivo*. This domain recruits Groucho/transducin-like enhancer of split (TLE) co-factor proteins (TLE1-4), which modulate transcription repression activity.

In this study we investigate the function of the ARX protein and show it acts as a strong transcriptional repressor. We identify 95 amino acids within ARX that include the fourth polyalanine tract and show this region is largely responsible for the transcription repression activity of ARX. We investigate the function of the *aristaless* domain in transcriptional regulation and show this domain alone increases gene expression and as such acts as transcription activation domain. We identify TLE proteins as ARX interacting proteins, which bind specifically to the octapeptide domain and function to mediate its transcriptional repression. We show this interaction is abolished when a known ARX mutation (c.98T>C; causing non-syndromic mental retardation) is introduced in the eh-1 domain. We also look at the transcriptional effect of the two most frequent mutations of the *ARX* gene, the expanded polyalanine tract mutations, which are known to cause a range of clinical presentations. We show that increasing the length of alanines within polyalanine tracts of the ARX protein increases repression, while deletions do not seem to have an effect. We show for the first time how two different polyalanine tract expansions cause a measurable effect on the function of the ARX protein.

EXPERIMENTAL PROCEDURES

Construction of full length and partial ARX cDNAs

Full length ARX (ARX^{Wt}), c.428_451dup mutation (ARX^{dup24}) and c.304ins$(GCG)_7$ ($ARX^{GCG10+7}$) were constructed by first cloning three overlapping fragments which were later ligated together to generate a full length cDNA clone. As the polyalanine tracts were often found to be deleted in generated clones, all propagation was carried out in the recombination deficient bacterial strain SURE® 2 (Stratagene, La Jolla, CA, USA). For the first and third parts, fragments were amplified from normal brain cDNA (Clontech, Mountain View, CA, USA), using the following oligonucleotides; A F-5′ga**agatct**atgagcaatcagtaccagg and A R-5′tgccgcaccctgaaggaggcggccc; C F-5′aactgctggaggacgacgagg and C R-5′cc**gaattc**ttagcacacctccttgccc. Restriction sites *Bgl*II and *Eco*RI (in bold and underlined) were introduced at the 5′ and 3′ ends of A and C respectively in order to facilitate cloning into the cloning vector phrGFP-N1 (Stratagene). Next, a region from exon 2 containing the polyalanine tracts was PCR amplified using normal and patient genomic DNA as template, using the following oligonucleotides; B F-5′ccg**gaattc**cagcagcgccccgttcgagg and B R-5′gcg**ggatcc**cgggtagtgcgtcttctgg. This product overlapped with fragments A and C such that restriction sites naturally occurring in the ARX open reading frame (ORF) could be used for ligating A to B (*Eco*NI) and B to C (*Pst*I). All three PCR products generated were cleaned using QIAquick PCR purification kit (Qiagen, Doncaster, VIC, Australia) and were ligated into the PCR cloning vector pGEM-T (Promega Corp., Madison, WI, USA). The three fragments were then excised from pGEM-T using the restriction enzymes mentioned (New England Biolabs, Ipswich, MA, USA) and then ligated together into pUC19 (New England Biolabs). All ligations were performed using standard conditions except *Eco*NI which was ligated using excess ligase (2000 U, New England Biolabs) with the addition of 10% (v/v) PEG_{8000}. All constructs were sequence verified before proceeding.

Constructs for transcription studies

The complete ORFs of ARX^{Wt}, ARX^{dup24} and $ARX^{GCG10+7}$ were cut from pUC19 using *Eco*RI and *Hind*III and were fused in-frame to the gal4 DNA binding domain of the *Eco*RI and *Hind*III digested expression vector, pM (Clontech). To generate various segments of the ARX protein, two methods were used. For constructs ARX^{1-281} (*Eco*RI and *Pst*I), ARX^{1-495} (*Eco*RI and *Sal*I), $ARX^{Ac\&pA3}$ (amino acids 210–281) (*Xma*I and *Pst*I) and ARX^{pA4} (amino acids 400–495) (*Xma*I and *Sal*I) the relevant regions were cut directly from full length *ARX* in pM vector using the specific restriction endonucleases shown in brackets. These digested products were inserted in frame into matching sites of pM. Synthesis of the remaining domain constructs of *ARX* was by PCR amplification; ARX^{Hom} (amino acids 306–431) (F-5′ca**gaattc**aaggacggcgaggacagc and R-5′gcg**ctcgag**tccaagcggagtcgagcg), ARX^{Arist} (amino acids 472–562) (F-5′tt**gagctc**cgacacccagctttcatc and R-5′gcg**ctcgag**ttagcacacctccttgcc) and ARX^{Oct} (amino acids 1–66) (F-5′tt**gagctc**atgagcaatcagtaccag and R-5′ccg**ctcgag**gccttgcaccgcctttc). The PCR products of ARX^{Arist} and ARX^{Oct} domains were restriction endonuclease digested with *Sac*I and *Xho*I and the ARX^{Hom} product digested with *Eco*RI and *Xho*I. These digested products were inserted into matching restriction sites of the digested vector, pHybLex/Zeo (Invitrogen Corp., Carlsbad, CA, USA). The products were then re-cut from pHybLex/Zeo using *Eco*RI and *Pst*I and inserted in frame into *Eco*RI and *Pst*I digested pM. The polyalanine-myc-tagged constructs were prepared by PCR amplification of bases 171–523 (amino acids 61–175) of ARX^{Wt}, ARX^{dup24} and $ARX^{GCG10+7}$ ORFs using F-5′g**gaattc**-gaaaaggccgtgcaaggctc and R-5′aa**ctgcag**cagcaggtcctcctcgctgatcagcttctgctccgacttgctgcggctgatg primers, which contained a myc tag. The PCR products were restriction endonuclease digested with *Eco*RI and *Pst*I and cloned into these sites in the digested pM vector. During this cloning process a number of interesting mutated constructs were generated. These contained insertions and deletions within the polyalanine tracts and were kept for analysis. All constructs were verified by sequencing.

A positive control expression plasmid pMNK10 contained KRAB domain of mouse NK10 protein (amino acids 1–112) fused to the *gal4* DNA-binding domain of the pM vector (provided by M. Kochetkova, Adelaide University). The chloramphenicol acetyltransferase (CAT) reporter vector, *gal4-TK-CAT*, contained five copies of the GAL4 DNA-binding site directly upstream of the thymidine kinase promoter (provided by M. Kochetkova). The *pSVβ-galactosidase* plasmid (Promega) expressed b-galactosidase and was used as an internal control of transfection efficiency.

The ARX^{OCT} and ARX^{PA4} domains (as above) were cloned into the C-terminus of pM-ALL1 construct (Hillman and Gécz, 2001) in frame with the ALL1 activator domain to test whether the chimeric protein of GAL4-ALL1-ARX repressed transcription of CAT by ALL1. The pM vector contains a GAL4 DNA binding domain flanked by the simian virus 40 early promoter (SV40) and polyA transcription termination signal. pG5CAT (Clontech) reporter vector was used in co-transfections with pM-ALL1 con-