Chapter 8

PCR Amplification and Sequence Analysis of GC-Rich Sequences: *Aristaless*-Related Homeobox Example

May H. Tan, Jozef Gécz, and Cheryl Shoubridge

Abstract

PCR amplification (followed by mutation scanning or direct sequencing) is a technique widely used in mutation detection and molecular studies of disease-causing genes, such as *ARX*. PCR amplification of high GC-rich regions encounters difficulties using conventional PCR procedures. Here, we present the strategies to amplify and sequence these GC-rich regions for the purposes of mutation screening and other molecular analyses.

Key words Polyalanine tracts, GC-rich, PCR, Sequencing, *ARX*

1 Introduction

Despite GC-rich regions accounting for only ~3 % of the human genome, important regulatory domains including promoters, enhancers, and control elements all have high GC content [1]. GC-rich sequences can be found in promoter regions of most housekeeping genes, tumor suppressor genes, and ~40 % of tissue-specific genes [2]. Repeat-associated disorders caused by expansions of trinucleotide repeats can occur in either noncoding sequences, transcribed but not translated, or within translated sequences for homomeric stretches of either glutamine or alanine amino acids. On the human X chromosome, expansion of GC-repeats in folate-sensitive fragile sites is associated with fragile X syndrome, which is the most common form of familial intellectual disability [3]. A range of inherited human diseases, associated with neurocognitive or neurodegenerative phenotypes, are caused by expansion of GC-rich trinucleotide repeats encoding polyalanine (polyA) or polyglutamine (polyQ) domains (Table 1). Polymerase chain reaction (PCR) amplification of GC-rich sequences can be difficult [4]. Ineffective or failing PCR

Danny M. Hatters and Anthony J. Hannan (eds.), *Tandem Repeats in Genes, Proteins, and Disease: Methods and Protocols*, Methods in Molecular Biology, vol. 1017, DOI 10.1007/978-1-62703-438-8_8,

Table 1
PolyA- or polyQ-codon-containing genes with repetitive GC-rich sequences

Gene (MIM)	Gene name[a]	GC content (%) of ORF	Trinucleotides encoded domains (PolyA or PolyQ)	Disease	References
ARX (300382)	*Aristaless*-related homeobox	72.5	4 polyA domains	NS-XLID; PRTS; ISSX/WS; IEDE; OS	[23]
PHOX2B (603851)	Paired-like homeobox 2b	66.1	2 polyA domains	CCHS	[26]
SOX3 (313430)	SRY (sex determining region Y)-box 3	70	4 polyA domains	XH	[27]
HTT (613004)	Huntingtin	52.6	1 polyQ	HD	[28]
ATXN1 (601556)	Ataxin 1	62.3	1 polyQ	SCA1	[29]

CCHS congenital central hypoventilation syndrome (MIM 209880), *HD* Huntington disease (MIM 143100), *IEDE* infantile epileptic–dyskinetic encephalopathy (MIM 308305), *ISSX(WS)* X-linked infantile spasms (West syndrome) (MIM 308350), *NS-XLID* non-syndromic X-linked intellectual disability (MIM 300419), *OS* Ohtahara syndrome—early infantile epileptic encephalopathy (MIM 308350), *PRTS* Partington syndrome—intellectual disability with dystonic movements, ataxia, and seizures (MIM 309510), *SCA1* spinocerebellar ataxia 1 (MIM 164400), *XH* X-linked hypopituitarism (MIM 300123)

[a]Gene names are official full names provided by HGNC

amplification of GC-rich regions complicates molecular analyses of these regions. The difficulties to PCR amplify these regions in vitro together with a relatively high frequency of errors may simply reflect the in vivo situation when the cells replicate these types of sequences.

The major hindrance to PCR amplification of GC-rich templates is the formation of secondary structures such as hairpin loops of single-stranded GC-rich sequences [4]. Many approaches have been developed to overcome such problems. Addition of organic molecules such as dimethyl sulfoxide (DMSO), glycerol, polyethylene glycerol, betaine, formamide, nonionic detergents, 7-deaza-dGTP, and dUTP into the PCR reaction mixture helps to resolve the complex secondary structure formation, thereby reducing the melting temperature of the primers and the templates [5–17]. Different DNA polymerases [2], template denaturation with NaOH, hot start PCR, stepdown PCR, slow down PCR, and primer modification have also shown to improve PCR amplification of high GC-rich DNA templates [18–22]. In addition, adjustment of magnesium concentration, pH of reaction buffer, PCR cycle temperature, and cycling numbers could have an effect. Incorporating the above-mentioned factors besides being labor and time intensive is only effective in some applications. In this chapter we outline a step-by-step approach to robustly amplify regions that are GC-rich. These optimized PCR conditions are a starting point for mutation screening or to generate or modify such templates for molecular cloning.

An example of an important disease-causing gene with mutations expanding GC-rich sequences is the *Aristaless*-related homeobox (*ARX*) gene (GenBank: NM_139058.2) (MIM 300382). *ARX* is one of the most frequently mutated genes in X-linked intellectual disability. The gene is predominantly expressed in fetal and adult brain, testis, skeletal muscle, and pancreas. This paired-type homeodomain transcription factor is crucial for early embryonic development. Mutations in *ARX* include nonsense and missense mutations and recurrent expansions of the GC-rich regions encoding the first two of four polyalanine tracts [23]. Mutations in these N-terminal polyalanine tracts contribute to ~60 % of all mutations reported in *ARX*. A frequent mutation in the first polyalanine tract (PA1), c.304ins$(GCG)_7$, expands this tract by seven alanine residues, resulting in a 23 alanine tract. The most common mutation in *ARX* occurs in the second polyalanine tract (PA2), c.429_452dup (24 bp) (~40 % of all reported *ARX* mutation), and involves addition of eight alanine residues to the 12 alanine tract. Both polyalanine expansions result in a range of clinical phenotypes (Table 1). Coding sequences of genes with high GC content, such as *ARX*, are not routinely covered by sequence capture followed by massively parallel sequencing, most likely due to the difficulty to sufficiently enrich the GC-rich regions (J.G., unpublished data). We present our data on the *ARX* gene as an

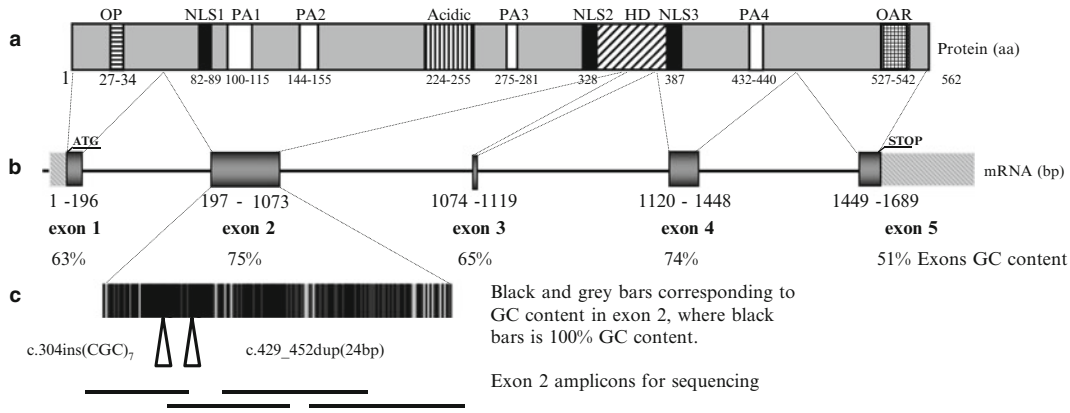


Fig. 1 Schematic diagram of the ARX homeobox transcription factor. (**a**) Known functional domains of ARX protein are highlighted; octapeptide (OP) in *horizontal stripes*, three nuclear localization sequences (NLS) in *black*, four polyalanine tracts (PA) in *white*, acidic domain in *vertical stripes*, homeodomain (HD) in *diagonal stripes*, and *Aristaless* (OAR) domain in *crosshatched*. Amino acid positions of functional domains are indicated below the protein structure. (**b**) Exon–intron structure of *ARX* gene with its five coding exons in *black boxes*; open reading frame in *dark gray*; ATG, STOP codon, and 5′ and 3′ untranslated regions in *diagonal light gray stripes*; base pair sequence of ORF is listed below each exon. (**c**) The percentage of GC content for exon 1–5 is indicated. The positions of *ARX* polyalanine expansion mutations are shown by *triangles* above the amplicons

example of the strategy used to interrogate the most difficult and mutation-prone GC-rich regions in this important disease-causing gene (Fig. 1).

2 Materials

Use autoclaved ultrapure water and analytical grade reagents.

2.1 GC-Rich Polymerase Chain Reaction

1. 100–500 ng of genomic DNA template (*see* **Note 1**).
2. 50 pmol/μl of forward and reverse primers (*see* Table 2).
3. Expand Long Template Enzyme mix (5 U/μl) (Roche) in enzyme storage buffer: 20 mM Tris–HCl, 100 mM KCl, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5 % (v/v) Nonidet P-40, 0.5 % (v/v) Tween 20, 50 % (v/v) glycerol, pH 7.5 (*see* **Note 2**).
4. FailSafe™ PCR 2× PreMix J buffer (Epicentre Technologies): 100 mM Tris–HCl, 100 mM KCl, 400 μM of each dNTP, 3 mM $MgCl_2$, and 8× FailSafe PCR Enhancer, pH 8.3.

2.2 Gel Electrophoresis

1. 5× TBE buffer: 54 g Tris base, 27.5 g of boric acid, 20 ml of 0.5 M EDTA, pH 8.0, and ultrapure water up to a liter.
2. Agarose gel cast and comb.
3. Molecular Biology grade agarose.

Table 2
***ARX*-specific primers used for amplification and sequencing of *ARX* coding regions**

ARX exon	Primer name	5′ to 3′ sequence	Size (bp)	PCR product (bp)	PCR annealing Tm (°C)	Amplicon GC content (%)
1	ARXe1-F	GTC CAC TAC ACT TGT TAC CGC	21	520	60	61
	ARXe1-R	AAT TGA CAA TTC CAG GCC ACT G	22			
2	ARXe2-F1	CTG ATA GCT CTC CCT TGC CC	20	262	60	81
	ARXe2-R1	GCG GCC CCT GCG CCG TCC GGC CGT TC	26			
	ARXe2-F2	CCC CTC CGC CGC CAC CGC CAA C	22	313	60	81
	ARXe2-R2	TCC TCC TCG TCG TCC TCG GTG CCG GT	26			
	ARXe2-F3	GCA AGT CGT ACC GCG AGA ACG	21	371	60	74
	ARXe2-R3	CAG CTC CTC CTT GGG TGA CA	20			
	ARXe2-F4	AAC TGC TGG AGG ACG ACG AGG	21	392	60	68
	ARXe2-R4	TGC GCT CTC TGC CGC TGC GA	20			
3	ARXe3-F	GAA ATA GCT GAG AGG GCA TTG C	22	231	60	61
	ARXe3-R	TCT CTT GGT TTT GTG AAG GGG AT	23			
4	ARXe4-F	GAC GCG TCC GAA AAC AAC CTG AG	23	551	60	72
	ARXe4-R	CCC CAG CCT CTG TGT GTA TG	20			
5	ARXe5-F	ACA GCT CCC GAG GCC ATG AC	20	347	60	71
	ARXe5-R	GAG TGG TGC TGA GTG AGG TGA	21			

4. Ethidium bromide: 50 µg/ml in water.
5. PCR products.
6. DNA marker (e.g., 1 kb plus marker (Invitrogen)).
7. 6× loading buffer containing dyes: 25 mg bromophenol blue, 25 mg xylene cyanol, 3.3 ml 100 % glycerol, and ultrapure water up to 10 ml.
8. Electrophoresis tank and DC power source.

2.3 PCR Purification

1. 100 bp to 10 kb double-stranded PCR product.
2. QIAquick PCR Purification Kit (Qiagen), which contains:
   2.1 Buffer PB (DNA binding buffer).
   2.2 Buffer PE (wash buffer).
   2.3 QIAquick spin columns (DNA-capture columns).
   2.4 Collection tubes.
   2.5 Ethanol (absolute) for dilution of buffer PE.

2.4 Sequencing

1. 800–1,500 ng/µl of double-stranded DNA PCR products or plasmid DNA.
2. BigDye® Terminator v3.1 Cycle Sequencing Kit (PerkinElmer), which contains:
   2.1 BigDye Terminator v.3.1 Sequencing Buffer (5×).
   2.2 BigDye Ready Reaction Mix.
3. Dimethyl sulfoxide (DMSO).
4. 0.2 nM $MgSO_4$ ethanol solution: 70 ml of absolute ethanol, 20 µl of 1 M $MgSO_4$, and ultrapure water up to 100 ml (*see* **Note 3**).
5. 70 % (v/v) ethanol solution.
6. Heating block set at 37 °C.
7. Sequencing analysis software (DNASTAR, Lasergene software package) or CLC Sequence Viewer 6 program (freeware available from http://www.clcbio.com/index.php?id=28).

3 Methods

3.1 PCR of GC-Rich Templates

Purified genomic DNA is used as template for the amplification of coding and flanking noncoding regions of *ARX*.

3.1.1 Oligonucleotides Design

Primer design is a crucial factor for successful PCR amplification. Manual design of each primer pair allowed us to accommodate stretches of difficult GC-rich sequence and achieve approximately

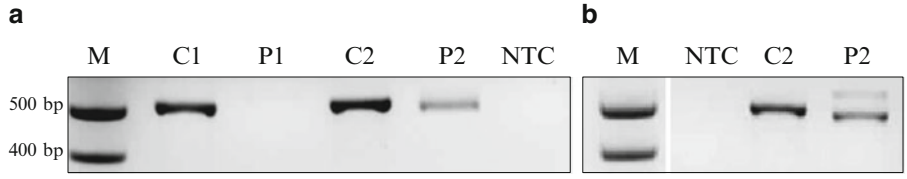
50 % GC content and a melting temperature of at least 60 °C for each primer. The sequences of both forward and reverse primers, size of the amplified gene product, and the GC content of the gene products amplified are listed in Table 2 (*see* **Note 4**).

In order to screen the coding region of the five *ARX* exons, each exon is amplified by a specific set of primers. The exception to this is exon 2, which requires four overlapping amplicons (Fig. 1c) to achieve robust amplification of GC-rich regions coding for polyalanine tract 1–3 (*see* **Note 5**). The overall GC content of exon 2 is 75 %; however, the GC-rich sequences are not evenly distributed along the exon but are instead centered in three regions coding for polyalanine tracts. The GC content of exon 2 is further increased by mutations leading to the expansion of the first two polyalanine tracts. For instance, the c.304ins$(GCG)_7$ expansion mutation in polyalanine tract 1 increases the GC content of amplicon 1 of exon 2 up to 83 % over a 283 bp region. The length of 100 % GC base pairs coding for polyalanine tract 1 increases from 48 to 69 bp.

Previously, only three sets of primer pairs were used to amplify and sequence exon 2 of *ARX* during mutation screening [24, 25]. The amplicon generated by the first primer set of exon 2 consists of polyalanine tract 1 and 2 and has a GC content of 81 % over a 514 bp region. The PCR amplification using this primer set is often inconsistent. Although it is possible to amplify this target from genomic DNA of controls (C1 and C2), the PCR amplification from genomic DNA of a patient with expanded polyalanine tract (P1) has failed (Fig. 2a). When a PCR product is successfully generated from the genomic DNA of patient with polyalanine tract mutation, instead of a strong, clean PCR product, a doublet PCR product is observed (P2) in gel electrophoresis analysis (Fig. 2b). The inconsistency of the PCR amplification, especially from patient's genomic DNA, is most likely due to the presence of complicated secondary structures formed by GC-rich sequences in exon 2.

Sequencing of the PCR products (P1, P2, and C1) in Fig. 2a, b generates a sequencing chromatogram (Fig. 2c). Secondary structures arising from GC-rich sequences may disrupt the sequencing process leading to either a sudden abrupt signal drop-off or attenuation of signal strength after a run of clear sequence signal. This phenomenon is illustrated in PA1 and PA2 regions of the sample from P2, where the sequence signal drops off at the polyalanine tract regions. A reduction in sequencing signal is also seen in PA1 region of C2, even when a good, clean PCR product (C2) was used as a sequencing template. The sequencing results generated from these amplicons are often inconsistent, requiring multiple re-sequencing analyses.

By generating smaller PCR products across this exon, we routinely achieve robust single PCR products of the expected size as observed by gel electrophoresis. Each amplicon designed for exon 2 of *ARX* is less than 400 bp and consists of only one polyalanine tract.



Lanes:		Expected PCR product
M	Marker	-
C1	Control 1	514 bp
C2	Control 2	514 bp
P1	Patient 1 with c.304ins$(GCG)_7$	514 + 21 = 535 bp
P2	Patient 2 with c.429_452dup (24bp)	514 + 24 = 538 bp
NTC	Non-template control	-

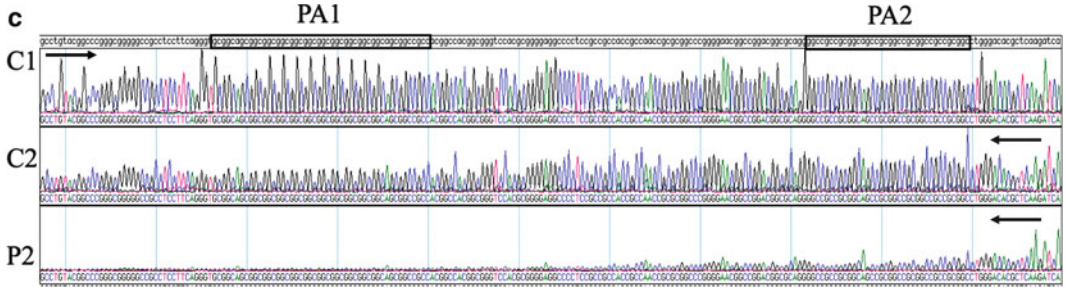
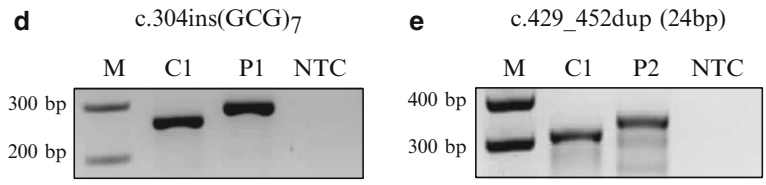


Fig. 2 *ARX* mutation detection using PCR amplification and sequencing. (**a**) *ARX*-specific primers are used to amplify N-terminal region of exon 2 from genomic DNA samples of controls (C1, C2) and patients (P1, P2). The amplicons generated consist of both polyalanine tract 1 and 2 of *ARX*. Strong and clear DNA bands at expected sizes are observed in controls but not in patients. No DNA product is detected in the non-template negative control lane. (**b**) Repeat PCR on sample P2 highlights the inconsistent size of PCR products generated from patients with expanded polyalanine tract mutations in *ARX*. (**c**) DNA sequence chromatogram showing signal drop-off at GC-rich region of *ARX*. An abrupt signal drop-off is observed after initial excellent sequence chromatogram, near the GC-rich region encoding for polyalanine tracts in C2 and P2. Additional sequencing using a forward primer is required to confirm the GC-rich sequence in the signal drop-off region. The *arrows* indicate the direction of sequencing using forward (*left*) or reverse (*right*) primers. Our current protocol uses two sets of *ARX*-specific primers to generate the first (**d**) or second (**e**) amplicon of exon 2 from genomic DNA samples from controls (C1) and patients (P1, P2). DNA bands at expected sizes are observed in both control and patient (listed in *bottom panel*), indicating successful amplification of target sequence. No amplification is present in the non-template negative control lane. The amplicons generated from patients' DNA migrates higher than control, indicating the presence of polyalanine tract expansion. (**f**) Sequence chromatogram of *ARX* exon 2 in control and patient with expanded polyalanine tract mutation c.304ins$(GCG)_7$. Sequence chromatogram from the mutant DNA sample shows 21 additional nucleotides, in the form of seven GCG repeats, in the region coding for the first polyalanine tract. Corresponding amino acids are indicated below the DNA sequences, demonstrating the expansion from 16 alanine residues in the control to 23 residues in the patient. Position of polyalanine tract 1 (PA1) is highlighted in *white box*



Lanes:		Expected PCR product
M	Marker	-
NTC	Non-template control	-
(A) c.304ins(GCG)$_7$		
C1	Control 1	262 bp
P1	Patient 1 with c.304ins(GCG)$_7$	262 + 21 = 283 bp
(B) c.429_452dup (24bp)		
C1	Control 1	313 bp
P2	Patient 2 with c.429_452dup (24bp)	313 + 24 = 337 bp

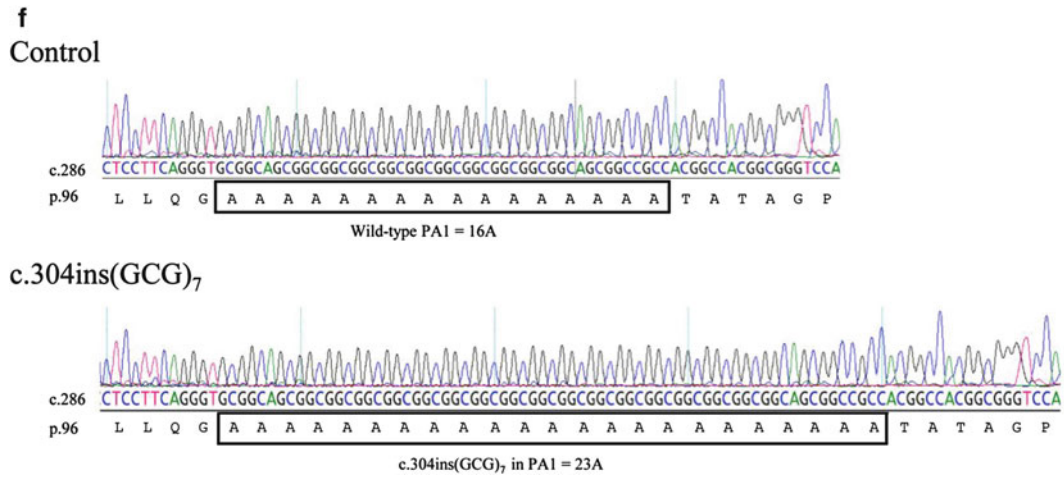


Fig. 2 (continued)

PCR amplification of amplicons 1 and 2 of exon 2 is shown in Fig. 2d, e. Subsequent sequencing outcomes from these products are consistent and reliable, reducing the need for repeat analysis (Fig. 2d–f). The sequencing outcomes from these PCR products (C1 and P1 in Fig. 2d) clearly indicate the difference between a control and a patient with polyalanine expansion mutation (Fig. 2f).

3.1.2 PCR Mixtures

Prepare a 50 µl PCR reaction on ice by adding the following reagents into a 0.2 or 0.5 ml tube:

*x*µl of 100–500 ng genomic DNA template.

25 µl FailSafe™ PCR 2× PreMix J buffer.

0.5 µl of Expand Long Template Enzyme mix.

1 µl of 50 pmol/µl forward primer.

1 µl of 50 pmol/µl reverse primer.

*y*µl of ultrapure water up to 50 µl ($y = 22.5\ \mu l - x$).

Prepare a master mix with all components, but without the template. Add DNA template to each tube before adding an aliquot of master mix. Use of a master mix reduces the error due to differences in pipetting small volumes of multiple components across individual samples. Add DNA template to each tube. Then aliquot the master mix, flick mix, and briefly centrifuge to return sample to the bottom of the tubes before placing in preheated PCR machine. Do not vortex the reaction mixture containing the polymerase/enzyme (*see* **Note 6**).

3.1.3 PCR Conditions

Perform PCR using the following cycling conditions (*see* **Note 7**) in a preheated PCR machine:

Denaturing	94 °C	2 min	
	94 °C	30 s	× 35 cycles
Annealing	60 °C	30 s	
Elongation	68 °C	2 min	
	68 °C	10 min	

3.2 Gel Electrophoresis

The presence and size of the expected PCR amplicons is confirmed by gel electrophoresis. Use a 2 % (w/v) agarose gel as the size of expected PCR product is under 1 kb.

1. Mix 6 g of agarose powder with 300 ml of 1× TBE to make a 2 % (w/v) solution.
2. Heat agarose using a microwave oven to boil gently so the agarose dissolves. Let hot agarose cool down to 55–60 °C before addition of ethidium bromide (*see* **Note 8**).
3. Add 0.5 µg/ml of ethidium bromide into 200 ml warm agarose solution, swirl gently to mix, and pour immediately into a gel cast. Ethidium bromide is added to make DNA visible under UV light (*see* **Note 9**).
4. Whilst still hot, place a comb in the gel cast. This forms wells in the gel for PCR samples loading.
5. Prepare 12 µl PCR samples in 1× loading buffer by adding 2 µl of 6× concentrated loading buffer to 10 µl of PCR product.

Flick mix and briefly centrifuge to return sample to the bottom of the tube before loading to gel.

6. When gel solidifies, remove the comb gently from the gel.
7. Place the gel in cast into the electrophoresis tank, and submerge in 1× TBE buffer.
8. Load the DNA marker and PCR samples into individual wells of the gel. The DNA standard is used as a reference to estimate the size of PCR products migrating in the agarose.
9. Close the lid of the electrophoresis tank and connect the tank to electric current. Ensure the negative (black) end and positive (red) end of the current are placed correctly.
10. Turn on the electric current from 90 to 110 V. The DNA strands move away from negative end of the tank. The shorter DNA products move through the gel quicker than the longer DNA strands.
11. Turn off the electric current when the bromophenol blue DNA tracking dye has migrated past halfway down the gel (*see* **Note 10**) (approximately an hour on a 2 % (w/v) agarose gel at 100 V). Remove the gel from the tank.
12. Observe the DNA bands of DNA standard and PCR samples using UV light (*see* **Note 11**).

3.3 PCR Purification

1. Add 5 volumes of DNA binding buffer, PB buffer into per volume of PCR product.
2. Transfer resulting mixture solution into a DNA-capture QIAquick column in a collection tube and centrifuge for a minute.
3. Discard the flow-through and wash the column with 750 μl of wash buffer, PE buffer.
4. Centrifuge for a minute at 17,900 × *g* and discard the wash.
5. Centrifuge for an additional minute to remove residual ethanol from the column (*see* **Note 12**).
6. Elute DNA into a clean 1.5 ml labelled microcentrifuge tube by adding 30 μl of ultrapure water to the center of membrane.
7. Let the column stand at room temperature for a minute.
8. Centrifuge for a minute to obtain purified DNA.
9. Determine the concentration of purified DNA using a spectrophotometer such as NanoDrop spectrophotometer ND-1000 (Thermo Scientific). Absorbance at 260 nm is multiplied by a factor of 50 for double-stranded DNA to get DNA concentration (ng/μl).
10. Store the purified DNA at −20 °C (*see* **Note 13**).

Table 3
***ARX*-specific primers used for sequencing of *ARX* from plasmid**

ARX exon	Primer name	5′ to 3′ sequence	Size (bp)
1–2	ARXex1/2F	TGC AAG GCT CCC CTA AGA GCA	21
2	ARXex2-R2	TCC TCC TCG TCG TCC TCG GTG CCG GT	26
3–4	ARXex3/4F	CCG AGT CCA GGT CTG GTT CCA	21
4	ARXcR1	CAG TCC AAG CGG AGT CGA GCG	21

3.4 Bidirectional Sequencing of GC-Rich Template

3.4.1 Oligonucleotides

The primers used for PCR amplification (Table 2) are also used to sequence the purified PCR products. We routinely use a different set of primers to sequence plasmid DNA containing the open reading frame of *ARX* (no intron and untranslated region). These primers are listed in Table 3.

3.4.2 Sequencing Mixtures

Prepare a 20 µl sequencing reaction on ice by adding the following reagents into a 0.2 or 0.5 ml tube:

*x*µl of 800–1,500 ng of PCR product or plasmid DNA.

3 µl of BigDye Terminator v.3.1 Sequencing Buffer (5×).

1 µl of BigDye Ready Reaction Mix.

1 µl of DMSO (5 % (v/v) final concentration) (*see* **Note 14**).

1 µl of 3.2 pmol/µl of forward OR reverse primer (*see* **Note 15**).

*y*µl of ultrapure water up to 20 µl ($y = 14\ \mu l - x$).

Make a master mix with all components minus the template. Add DNA template to each tube before adding an aliquot of master mix. Mix (do not vortex) and pulse the samples by brief centrifugation to return samples to the bottom of the tube before placing in preheated PCR machine.

3.4.3 Bidirectional Sequencing Conditions

Perform sequencing amplification under the following cycling conditions:

Denaturing	96 °C	1 min	×25 cycles
	96 °C	10 s	
Annealing	50 °C	5 s	
Elongation	60 °C	4 min	

3.4.4 Precipitation of Sequencing Products

1. Allow tubes to equilibrate at room temperature.
2. Add 75 µl of 0.2 nM $MgSO_4$ ethanol solution to each sample.
3. Mix samples by vortexing and incubate samples for 15 min at room temperature (keep samples in the dark where possible).

4. Centrifuge samples for 15 min at 17,900 × *g*.
5. Remove tubes from microcentrifuge and gently invert the tubes over a paper towel for 1–2 min.
6. Add 100 µl of 70 % ethanol to wash the pellet and centrifuge again for 15 min at 17,900 × *g*. Do not vortex the tubes.
7. Allow tubes to air-dry in a 37 °C heating block for 5 min (*see* **Note 16**). To avoid samples being dislodged, do not aspirate remaining solution.
8. Submit dry samples to DNA sequencing service facility for capillary electrophoresis for size separation, detection, and recording of incorporated dye fluorescence and data output as fluorescent peak trace chromatograms.
9. Analyze the sequencing results using commercial software such as EditSeq and SeqMan programs (DNASTAR, Lasergene software package) or CLC Sequence Viewer 6 program (freeware available from http://www.clcbio.com/index.php?id=28).

4 Notes

1. The quality and purity of template DNA is important and will affect the outcome of PCR amplification.
2. Expand Long Template Enzyme mix is a unique blend of *Taq* DNA polymerase and a thermostable high fidelity *Tgo* DNA polymerase with proofreading activity. Together, these DNA polymerases synergistically generate PCR products with greater yield and three times higher fidelity than *Taq* DNA polymerase alone. GC-rich regions with high melting temperature domains can act as permanent termination sites and have profound effects in the amplification mimicking competitive PCR; when *Taq* DNA polymerase is used in the presence of a proofreading enzyme, this effect may be reduced or eliminated [4].
3. Prepare 0.2 nM $MgSO_4$ stock solution on a weekly basis. Store stock solution at room temperature, do not refrigerate or freeze the solution.
4. To ensure the primer set detects only the desired *ARX* exon (or exon region) when screening genomic DNA, each primer set is submitted for analysis by BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). BLAST will find sequences in the database, which are similar to the sequences submitted in the query. The results will be displayed in (a) a graph showing the hits found, (b) a table showing sequence identifiers for each hit (scoring related data), (c) alignments for the sequence of interest, and (d) corresponding BLAST scores for each hit

obtained. An ideal primer set will have a 100 % best match to desired target and with no cross homology.

5. Due to high GC base composition at the N-terminal of exon 2, overlapping amplicons with GC content of 69–82 % are designed across exon 2. These overlapping amplicons enable robust PCR efficiency and sequencing coverage. This design is particularly helpful in the case of PCR amplification and sequencing of templates from affected individuals with expanded polyalanine tract mutations. The products generated will have higher GC content than the amplicons generated from controls.
6. Mixing by vortex of PCR reactions containing the enzyme blend will damage the enzyme integrity and cause PCR failure.
7. Optimum initial denaturation temperature for human genomic DNA is 93–95 °C. Expand Long Template Enzyme mix is use for efficient amplification of GC-rich DNA template. Instead of standard PCR elongation temperature at 72 °C, the optimal primer extension temperature is reduced to 68 °C to avoid enzyme loss during prolonged extension times.
8. Direct skin contact with hot agarose can cause severe burns. Always wear protective gloves, goggles, and a lab coat while preparing and casting agarose gels. Swirling of hot agarose leads to boil over. Let hot molten agarose cool down to 55–60 °C before addition of ethidium bromide to prevent boil over.
9. Ethidium bromide is a suspected carcinogen and at high concentration is irritating to the eyes, skin, mucous membranes, and upper respiratory tract. Handle ethidium bromide with extreme care. Always wear protective gloves, goggles, and a lab coat during handling. Gels and solutions containing ethidium bromide should be disposed appropriately as hazardous waste. Refer to Material Safety Data Sheet (MSDS) for more information.
10. On a 1 % agarose gel, xylene cyanol and bromophenol blue typically migrate about the same as a 4,000 bp and 200–400 bp DNA fragment, respectively.
11. Overexposure to ultraviolet (UV) radiation can lead to different stages of erythema (sunburn) and photokeratitis (eye inflammation due to lesions on the cornea and conjunctiva). These symptoms are not immediate and normally appear 4–24 h post exposure. Exposure to UV radiation can be minimized by having adequate skin and eye protection. Always wear a UV-resistant full face shield, goggles with side shields, nitrile gloves, and lab coats when working with UV radiation.

12. Residual ethanol from wash buffer will not be totally removed if flow-through from previous centrifugation is not discarded before this step (as recommended by Qiagen).
13. For long-term storage, store DNA at −20 °C as DNA may degrade in water in the absence of a buffering agent.
14. DMSO can improve the outcome of sequencing reactions from difficult templates such as those with GC-rich regions by relaxing complex secondary structure formations such as hairpins.
15. Primer for the sequencing reaction is freshly prepared at the required dilution each time it is required as DNA primers at low concentration are more prone to degradation through freeze-thaw cycles. Sequencing reactions are conducted with only one primer in each reaction.
16. Do not allow samples to overdry in 37 °C heating block as samples will be difficult to resuspend for subsequent analysis.

Acknowledgments

This work was supported by grants from the National Health and Medical Research Council of Australia (Project Grant 1002732 to C.S.; Principle Research Fellowship 508043 to J.G.) and MS McLeod Foundation Fellowship to C.S. and postgraduate scholarship to M.H.T.

References

1. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW (1997) Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc Natl Acad Sci USA 94(7):3195–3199
2. Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, Subramanian A, Viljoen HJ (2008) A fundamental study of the PCR amplification of GC-rich DNA templates. Comput Biol Chem 32(6):452–457. doi:S1476-9271(08)00088-1 [pii] 10.1016/j.compbiolchem.2008.07.021
3. Sutherland GR, Richards RI (1995) The molecular basis of fragile sites in human chromosomes. Curr Opin Genet Dev 5(3):323–327
4. McDowell DG, Burns NA, Parkes HC (1998) Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. Nucleic Acids Res 26(14):3340–3347. doi:gkb479 [pii]
5. Musso M, Bocciardi R, Parodi S, Ravazzolo R, Ceccherini I (2006) Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. J Mol Diagn 8(5):544–550. doi:S1525-1578(10)60343-1[pii] 10.2353/jmoldx.2006.060058
6. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S (2006) An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun 347(3):747–751. doi:S0006-291X(06)01470-7 [pii] 10.1016/j.bbrc.2006.06.151
7. Baskaran N, Kandpal RP, Bhargava AK, Glynn MW, Bale A, Weissman SM (1996) Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content. Genome Res 6(7):633–638
8. Chakrabarti R, Schutt CE (2001) The enhancement of PCR amplification by low molecular-weight sulfones. Gene 274(1–2):293–298. doi:S0378111901006217 [pii]
9. Guerrini R, Moro F, Kato M, Barkovich AJ, Shiihara T, McShane MA, Hurst J, Loi M, Tohyama J, Norci V, Hayasaka K, Kang UJ, Das S, Dobyns WB (2007) Expansion of the first PolyA tract of ARX causes infantile spasms and status dystonicus. Neurology 69(5):427–433.

doi:69/5/427 [pii] 10.1212/01.wnl.0000266594.16202.c1

10. Henke W, Herdel K, Jung K, Schnorr D, Loening SA (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res 25(19):3957–3958. doi:gka606 [pii]
11. Kang J, Lee MS, Gorenstein DG (2005) The enhancement of PCR amplification of a random sequence DNA library by DMSO and betaine: application to in vitro combinatorial selection of aptamers. J Biochem Biophys Methods 64(2):147–151. doi:S0165-022X(05)00082-5 [pii] 10.1016/j.jbbm.2005.06.003
12. Mutter GL, Boynton KA (1995) PCR bias in amplification of androgen receptor alleles, a trinucleotide repeat marker used in clonality studies. Nucleic Acids Res 23(8):1411–1418. doi:4s0766 [pii]
13. Pomp D, Medrano JF (1991) Organic solvents as facilitators of polymerase chain reaction. Biotechniques 10(1):58–59
14. Sidhu MK, Liao MJ, Rashidbaigi A (1996) Dimethyl sulfoxide improves RNA amplification. Biotechniques 21(1):44–47
15. Sun Y, Hegamyer G, Colburn NH (1993) PCR-direct sequencing of a GC-rich region by inclusion of 10 % DMSO: application to mouse c-jun. Biotechniques 15(3):372–374
16. Turner SL, Jenkins FJ (1995) Use of deoxyinosine in PCR to improve amplification of GC-rich DNA. Biotechniques 19(1):48–52
17. Weissensteiner T, Lanchbury JS (1996) Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. Biotechniques 21(6):1102–1108
18. Agarwal RK, Perl A (1993) PCR amplification of highly GC-rich DNA template after denaturation by NaOH. Nucleic Acids Res 21(22):5283–5284
19. Bachmann HS, Siffert W, Frey UH (2003) Successful amplification of extremely GC-rich promoter regions using a novel 'slowdown PCR' technique. Pharmacogenetics 13(12):759–766. doi:10.1097/01.fpc.0000054140.14659.61
20. Li LB, Yu Z, Teng X, Bonini NM (2008) RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature 453(7198):1107–1111. doi:nature06909 [pii] 10.1038/nature06909
21. Frey UH, Bachmann HS, Peters J, Siffert W (2008) PCR-amplification of GC-rich regions: 'slowdown PCR'. Nat Protoc 3(8):1312–1317. doi:nprot.2008.112 [pii] 10.1038/nprot.2008.112
22. Sahdev S, Saini S, Tiwari P, Saxena S, Singh Saini K (2007) Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. Mol Cell Probes 21(4):303–307. doi:S0890-8508(07)00028-X [pii] 10.1016/j.mcp.2007.03.004
23. Shoubridge C, Fullston T, Gecz J (2010) ARX spectrum disorders: making inroads into the molecular pathology. Hum Mutat 31(8):889–900. doi:10.1002/humu.21288
24. Kato M, Das S, Petras K, Kitamura K, Morohashi K, Abuelo DN, Barr M, Bonneau D, Brady AF, Carpenter NJ, Cipero KL, Frisone F, Fukuda T, Guerrini R, Iida E, Itoh M, Lewanda AF, Nanba Y, Oka A, Proud VK, Saugier-Veber P, Schelley SL, Selicorni A, Shaner R, Silengo M, Stewart F, Sugiyama N, Toyama J, Toutain A, Vargas AL, Yanazawa M, Zackai EH, Dobyns WB (2004) Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation. Hum Mutat 23(2):147–159. doi:10.1002/humu.10310
25. Fullston T, Finnis M, Hackett A, Hodgson B, Brueton L, Baynam G, Norman A, Reish O, Shoubridge C, Gecz J (2011) Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene. Clin Genet 80(6):510–522. doi:10.1111/j.1399-0004.2011.01685.x
26. Matera I, Bachetti T, Puppo F, Di Duca M, Morandi F, Casiraghi GM, Cilio MR, Hennekam R, Hofstra R, Schober JG, Ravazzolo R, Ottonello G, Ceccherini I (2004) PHOX2B mutations and polyalanine expansions correlate with the severity of the respiratory phenotype and associated symptoms in both congenital and late onset Central Hypoventilation syndrome. J Med Genet 41(5):373–380
27. Laumonnier F, Ronce N, Hamel BC, Thomas P, Lespinasse J, Raynaud M, Paringaux C, Van Bokhoven H, Kalscheuer V, Fryns JP, Chelly J, Moraine C, Briault S (2002) Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. Am J Hum Genet 71(6):1450–1455
28. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Huntington's Disease Collaborative Research Group (1993). Cell 72(6):971–983
29. Banfi S, Servadio A, Chung MY, Kwiatkowski TJ Jr, McCall AE, Duvick LA, Shen Y, Roth EJ, Orr HT, Zoghbi HY (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat Genet 7(4):513–520. doi:10.1038/ng0894-513