J Clin Pathol: Mol Pathol 1996;49:M12–M16

Papers

Complete structural characterisation of the human aryl hydrocarbon receptor gene

P Bennett, D B Ramsden, A C Williams

Abstract

***Aims*—To clone and characterise the complete structural gene for the human aryl hydrocarbon receptor (AhR). This gene, located on chromosome 7, encodes a cytosolic receptor protein which, upon activation by various xenobiotic ligands, translocates to the nucleus, where it acts as a specific transcription factor.**

***Methods*—Primers, based on the AhR cDNA sequence, were used in conjunction with recently developed long range PCR techniques to amplify contiguous sections of the cognate gene. The amplicons produced were then cloned and characterised. A cDNA probe was also used to screen a human P1 library.**

***Results*—Using the cDNA primers, DNA fragments which mapped the entire coding region of the gene were amplified and cloned. All but one of these fragments were amplified directly from human genomic DNA. The remaining fragment was amplified using DNA prepared from a P1 clone as the PCR template. This P1 clone, obtained by screening a human P1 library, also contained the entire Ah locus. Characterisation of amplified and cloned DNA fragments provided sufficient information for the construction of a complete structural map of the gene. This also included 150 base pairs of nucleotide sequence data at all intronic termini.**

***Conclusions*—These data indicate that the human AhR gene is about 50 kilobases long and contains 11 exons. The overall intron/exon structure of the human gene is homologous to that of the previously characterised mouse gene; however, it is probably some 20 kilobases larger. These results demonstrate the need for further characterisation and provide the data to facilitate this.**

(*J Clin Pathol: Mol Pathol* 1996;**49**:M12–M16)

Keywords: PCR, aryl hydrocarbon receptor gene, characterisation, cDNA.

University Department of Medicine, Queen Elizabeth Hospital, Birmingham
P Bennett
D B Ramsden

University Department of Clinical Neurology
A C Williams

Correspondence to:
Mr P Bennett,
University Department of Medicine,
Queen Elizabeth Hospital,
Birmingham B15 2TH.

Accepted for publication
12 October 1995

The human aryl hydrocarbon receptor (AhR) gene locus is located on chromosome 7,[1] and encodes a 96 kDa ligand activated transcription factor.[2] In its transcriptionally inactive state the AhR is found in the cytosol associated with the heat shock protein (hsp) 90.[3] Ligand binding results in the AhR being translocated to the nucleus,[4] where it is believed to undergo transformation to a transcriptionally active state via hetero-dimerisation with the AhR nuclear translocator protein (ARNT) and dissociation from hsp90.[5–7] This complex then mediates increased expression of a number of target genes known collectively as the Ah gene battery. Although the number of genes which comprise this battery is uncertain, the primary role of those characterised to date is xenobiotic metabolism.[8 9] This is unsurprising, because all known ligands for the AhR are xenobiotics. Although endogenous ligands have not yet been identified, it is suspected that they exist and may play an important role in the regulation of cell differentiation or homeostasis.[10 11]

There are several groups of diseases, such as neurodegenerative disorders and certain types of cancer, where xenobiotic insults, or more commonly individual differences in dealing with such insults, have been implicated in the disease aetiology. Disturbances in cell differentiation and homeostasis are also features of these diseases.[12 13] As current data suggest that the AhR system plays a key role in the regulation of such processes, we hypothesise that genetic defects within any part of this complex system may play primary causative roles in some of these disease states. However, before this hypothesis can be fully investigated, a detailed understanding of the genes encoding the various components of the AhR system must be established.

From cloning and sequencing of the human AhR cDNA, previous workers have derived its primary amino acid sequence.[14 15] In this paper we describe use of these cDNA sequences to design specific oligonucleotide primers which permitted complete structural characterisation of the human AhR gene. This was achieved principally by the use of recently developed long range PCR techniques, that have enabled large fragments, previously only obtainable via screening libraries, to be amplified directly from genomic DNA.[16] Amplicons so produced mapped the entire transcribed region of the AhR gene, with the exception of one intron. These amplicons were cloned and characterised by analysis of their nucleotide sequence adjacent to the exon/intron boundaries. We also describe the identification of a P1 clone con-

Sequence and location of PCR primers

Exon	*PCR primer oligonucleotide sequence*	*cDNA position w.r.t. translation start site*
1	AhRF1-5'AGCAGCAGCGCCAACATCACCTACG	7 to 31
2	AhRR4-5'GCAGCAGGCTAGCCAAACGGTCCAA	139 to 163
2	AhRF5-5'GAGTTGGACCGTTTGGCTAGCCTGCTGCCT	136 to 165
3	AhRR9-5'AGTTATCCTGGCCTCCGTTTCTTTCAGTAG	272 to 301
3	AhRF3-5'CCCCTACTGAAAGAAACGGAGGCCA	269 to 293
4	AhRR13-5'AGCATAAAAGACCAAAGCATCTGTAGTG	387 to 414
4	AhRF12-5'CACTACAGATGCTTTGGTCTTTTATGCT	387 to 414
5	AhRR17-5'TTAATGCCCAGTGTAGCTGACGC	513 to 535
5	AhRF10-5'ATTTCAGCGTCAGCTACACTGGGCA	507 to 531
6	AhRR10-5'GAGTTTTCTGGAGGAATCTGGTCTGGG	612 to 638
6	AhRF17-5'TCTACAGAAGCCACTGGTCTCC	572 to 589
7	AhRR12-5'CTATCGCAAACAAAGCCAACTGAGGTG	779 to 805
7	AhRF13-5'CACCTCAGTTGGCTTTGTTTGCGATAG	779 to 805
8	AhRR6-5'TAAACTGATAACCTGAGCCTCTCGTGCACA	944 to 973
8	AhRF4-5'TGAAGCAGAGCTGTGCACGAGAGGC	933 to 957
10	AhRR1-5'GCATCATGGCAGCCAGGAGGGAACT	1360 to 1384
10	AhRF9-5'ATGTTATGCTGGGGCCGTGTCGATGTATCA	2289 to 2318
11	AhRR8-5'AATGCCTCCATGTGAACTTGCTGACGTCCA	2626 to 2655
10	AhRR3-5'ACATCGACACGGCCCCAGCATAACA	2290 to 2314

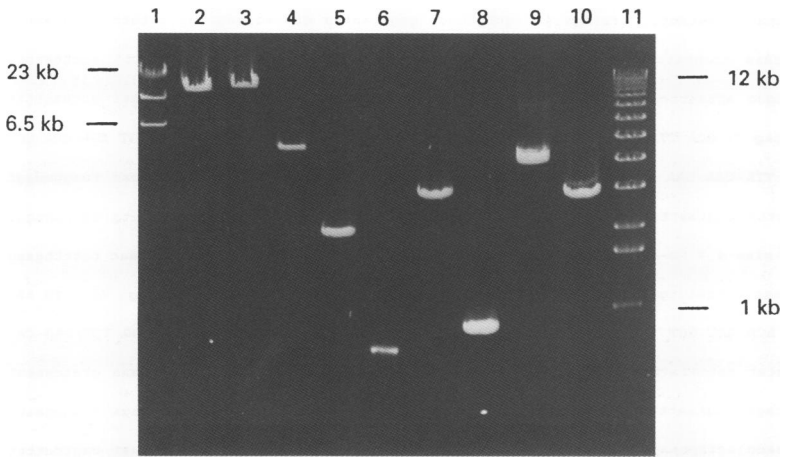

Figure 1 Ethidium bromide stained agarose gel. Lane 1, Hind*III λ DNA digest (TaKaRa); lanes 2–10, PCR products amplified with the following primers: AhRF1/AhRR4, AhRF5/AhRR9, AhRF3/AhRR13, AhRF12/AhRR17, AhRF10/AhRR10, AhRF17/AhRR12, AhRF13/AhRR6, AhRF4/AhRR1, and AhRF9/AhRR8; lane 11, Kb ladder (Gibco). All products were amplified directly from human genomic DNA, except that shown in lane 3, which was amplified from a P1 clone (see text).*

taining the entire AhR locus. Using DNA from this clone, instead of genomic DNA as a PCR template, it was also possible to amplify, clone and characterise the one remaining intron.

Methods

The primers listed in the table 1 were designed according to published human AhR cDNA nucleotide sequences.[14 15] Adjacent primer pairs were then used in PCR reactions containing 100 ng of human genomic DNA, prepared by the method of Sykes.[17] Recently developed Long and Accurate (LA) PCR technology was used to amplify DNA. The thermostable DNA polymerase mixture used was Ex-taq from TaKaRa Ltd (European distributors: ITC Biotechnology, Heidelberg, Germany); other reaction components were as suggested in the TaKaRa supplied protocol. Thermal cycling was performed using a Hybaid Omni-Gene thermocycler. Initial cycling parameters were 30 cycles, with denaturation at 98°C for 20 seconds, and annealing/extension at 68°C for 10 minutes. For primer pairs generating amplicons less that 5 kilobases, the annealing/extension time was reduced to five minutes in subsequent reactions. Amplicons generated by these PCR reactions were analysed by electrophoresis using 0·5% agarose $\frac{1}{2}\times$ TBE (0·045 M Tris-borate, 0·001 M EDTA) gels; DNA was visualised by staining with ethidium bromide. Size estimation was by comparison with known size markers (KB ladder, Gibco BRL, Paisley, UK; λ *Hind*III digest, TaKaRa).

PCR products were purified using the Promega Wizard DNA Clean Up System and concentrated by ethanol precipitation where necessary. These purified products were then ligated into pMOSBlue T-vector (Amersham, Little Chalfont, UK) and transformed into competent XL-blue 1 MRF' *Escherichia coli* (Stratagene, Cambridge, UK). The presence of the correct insert in recombinant clones was verified by PCR and size estimation of restriction digested clone DNA. The partial nucleotide sequences of such clones were then determined using an ABI 373 automated sequencer (Applied Biosystems, Warrington, UK).

Filters containing a human genomic P1 library were supplied by the Imperial Cancer Research Fund (ICRF) Genome Research Laboratories, London, UK. These filters were probed with a ^{32}P labelled AhR cDNA probe, generated from human liver RNA by reverse transcription PCR using primers AhRF1 and AhRR3 (table). The coordinates of potentially positive clones were then used to obtain live stab cultures from the ICRF. The procedures for amplifying and cloning fragments from P1 DNA were as described above for genomic DNA, except only 1 ng of template and 25 PCR cycles were used.

Results

PCR amplifications using genomic DNA template generated single high yield amplicons for all adjacent primer pairs, except AhRF5 and AhRR9 (fig 1) (note: in fig 1, a PCR product is shown to be generated by AhRF5/AhRR9; as described below this was amplified using a P1 clone instead of genomic DNA as the PCR template). In each case the amplicon was larger than that which would be generated using cDNA as template, thus suggesting that the primers were positioned in separate exons. PCR reactions with primers AhRF5 and AhRR9 generated no visible products; this was also the

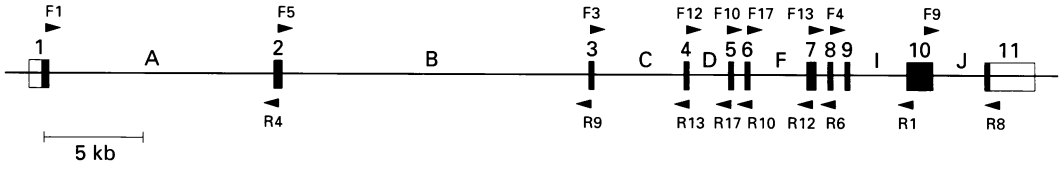

Figure 2 Structural organisation of the human AhR gene. Exons are denoted by numbered boxes; translated regions are shaded. The approximate positions of primers used are indicated by labelled arrowheads.

+1 *

ATTCAGCCGG TGCGCGCGGC GGCGGGAGGC AGTGGCTGGG GAGTCCCGTC GACGCTCTGT TCCGAGAGCG TGCCCCGGAC CGCCAGCTCA GAACAGGGGC
AGCCGTGTAG CCGAACGGAA GCTGGGAGCA GCCGGGACTG GTGGCCCGCG CCCGAGCTCC GCAGGCGGGA AGCACCCTGG ATTTGGGAAG TCCCGGGAGC
AGCGCGGCGG CACCTCCCTC ACCCAAGGGG CCGCGGCGAC GGTCACGGGG CGCGGCGCCA CCGTGAGCGA CCCAGAGCCA GGATTCTAAA TAGACGGCCC
AGGCTCCTCC TCCGCCCGGG CCGCCTCACC TGCGGGCATT GCCGCGCCGC CTCCGCCGGT GTAGACGGCA CCTGCGCCGC CTTGCTCGCG GGTCTCCGCC
CCTCGCCCAC CCTCACTGCG CCAGGCCCAG GCAGCTCACC TGTACTGGCG CGGGCTGCGG AAGCCTGCGT GAGCCGAGGC GTTGAGGCGC GGCGCCCACG
CCACTGTCCC GAGAGGACGC AGGTGGAGCG GGCGCGGCTT CGCGGAACCC GGCGCCGGCC GCCGCAGTGG TCCCAGCCTA CACCGGGTTC CGGGGACCCG
GCCGCCAGTG CCCGGGGAGT AGCCGCCGCC GTCGGCTGGG CACC ATG AAC AGC AGC AGC GCC AAC ATC ACC TAC GCC AGT CGC AAG CGG CGG
AAG CCG GTG CAG AAA AC gtgagtgtcc cgagcgcgtc ctcatcgcgg gggctgggcg ctcaggcacg cgggtgcggg aggcagcccc accccgcccc
caaatccctg cgatcctggg attaggtcca ttcccggcac tgcccgtgga atcgaggttt ggaggccggc ----Intron A total size approx 11.5
Kb---- ggaagatttt agaaagactt acgtaaactt taagtagact ttaaaagttt gttgtgttag agaaatattt gaggagatgt tataatgcaa
tagaaatttt tgctttatat tttttaaagg attttttatg gtattttgtt tgtttttcag A GTA AAG CCA ATC CCA GCT GAA GGA ATC AAG TCA
AAT CCT TCC AAG CGG CAT AGA GAC CGA CTT AAT ACA GAG TTG GAC CGT TTG GCT AGC CTG CTG CCT TTC CCA CAA GAT GTT ATT
AAT AAG TTG GAC AAA CTT TCA GTT CTT AGG CTC AGC GTC AGT TAC CTG AGA GCC AAG AGC TTC TTT GAT G gtaagacaga
agggtttaat ttgtctacaa taacgtataa aaaatacttg tactagatat agctgtttct gtgttaataa ctacaaaaat tagccgtatt tggaaaattt
attgctgtat agtaaaattt cagtggcaaa gccagattta ----Intron B total size approx 15 Kb----- tttattattt accttgtttt
agccaaatac taaaaatacc agtggatgcc agtatttcaa attgtctttg tttggtgttc agaagttttc tattatagct ctttactctt gcttactttt
aaaatcattg tttttccttt ttttccatag TT GCA TTA AAA TCC TCC CCT ACT GAA AGA AAC GGA GGC CAG GAT AAC TGT AGA GCA GCA
AAT TTC AGA GAA GGC CTG AAC TTA CAA GAA GGA GAA TTC TTA TTA CAG gtaaatttta gtaaatatag tttcttacac taaggacagt
tgtaaatgga aaatgaatta ataagtcttt tagtaattcc ctgtttactt aggatttgct caatgttttt tgccactttc atctgaactg cgatagtgaa
tgatgcttaa ----Intron C total size 4.7 Kb---- gtccaggagt gtatgttttg gctgtgtttg tgaaatgtga caatttttaac tattttgaag
agaagaattt tcagagataa aagtaataac ctttatctga tggtcaatat taagtcatat tactaatttt agaacttcct ttccttgtag GCT CTG AAT
GGC TTT GTA TTA GTT GTC ACT ACA GAT GCT TTG GTC TTT TAT GCT TCT TCT ACT ATA CAA GAT TAT CTA GGG TTT CAG CAG
gtaagtatat attatttata tcatttatat tattatatca tttacttttt atatatattt aggacacagt tggccatttg tatgtataaa gtatcaatat
atgctgtaga atacagtatg tatgtgtagc taataattnn ctctaatatt ----Intron D total size 2Kb---- tcttaggtga ctagggaatt
ttaggaatca ttcaattcgt attcatcacc actagcaagc acccactaat ctaaataggc tttaaaatta atttagccat atttttttaat cagtcctttt
gttgtattgc cttgtatctt ttttctttag TCT GAT GTC ATA CAT CAG AGT GTA TAT GAA CTT ATC CAT ACC GAA GAC CGA GCT GAA TTT
CAG CGT CAG CTA CAC TGG GCA TTA AAT CCT TCT CAG TGT ACA GAG TCT GGA CAA GGA ATT GAA G gtaagaattg atggtacaaa
aaatagtgtt ggcagttttt aaatatgagt ctgtgaaagg aggctgggaa cctgtagggt catagaactc catgataggg aagtagtgga aagatgtaaa
gtctgcaatc ataggccaca gctaggtcac ----Intron E total size 0.68 Kb---- tagtttcttg tcattagctc ttttgaaaat gatttttttg
tattcagaac acagactcca gtttagaaac taatacaaat tttacctatt caagtgctta attttacagc aaaatggaaa gtaaattttg ttttgccttt
atttctacag AA GCC ACT GGT CTC CCC CAG ACA GTA GTC TGT TAT AAC CCA GAC CAG ATT CCT CCA GAA AAC TCT CCT TTA ATG GAG
AGG TGC TTC ATA TGT CGT CTA AGG TGT CTG CTG GAT AAT TCA TCT GGT TTT CTG gtaaggtaca aaatttttatg atactggctt
ttactattgt tacaataaaa gcttgaggca aatttaattt agcaaaatat aattcagcag agaactattc ccaaatcagg caaccctcag aaccagagag
cttcagagag ctcacttaaa ----Intron F total size 3031 bp---- aggagtgaag gaaactaaca gttctcagct tcaggaatat ttatggaatt
atcaagaacc ctagaggcag ccataatgga gccttttaaa aaggcactat tttatattgt taattttaat gaacttttt gttgttgttg ctttttttaag GCA
ATG AAT TTC CAA GGG AAG TTA AAG TAT CTT CAT GGA CAG AAA AAG AAA GGG AAA GAT GGA TCA ATA CTT CCA CCT CAG TTG GCT
TTG TTT GCG ATA GCT ACT CCA CTT CAG CCA CCA TCC ATA CTT GAA ATC CGG ACC AAA AAT TTT ATC TTT AGA ACC AAA CAC AAA
CTA GAC TTC ACA CCT ATT GGT TGT GAT GCC AA gtaagtgaga ctttttcact tttattttat tggatgtaca ttatgtttca gtaagtctct
cttagcatgt aaaacataca gtgtatgtaa tattgtttat tattagattg gctattatcg tacattcttc cagtgtngcc ttttacagta ----Intron G
total size 772 bp---- tgtttaattt cacatctact tatgtgaaat tctaaatgtg aactaaaaca tattgcagaa actagcgtaa aaccaatgaa
tttatcttgg ttatttcatt tatgttaaat cttaatccat tcttatttta ccttttttta ttttaaacag A GGA AGA ATT GTT TTA GGA TAT ACT

Figure 3 Molecular structure of the human AhR gene. The nucleotide in bold, marked (+1), indicates the furthest known upstream transcription initiation site. The nucleotide in bold, marked (*), indicates an additional and probably more common transcription initiation site. Single underlined nucleotides in the 5' untranslated region indicate regions of ambiguity between published sequences. Start and stop codons are in bold and single underlined. Exons are shown in upper case text; translated regions are shown in frame. Introns are shown in lower case text; canonical intron termini are in bold. The putative polyadenylation site is in bold and double underlined.

case for reactions containing AhRF1/AhRR9 and AhRF5/AhRR13. This suggested that the distance between primers AhRF5 and AhRR9 was simply too great to allow amplification directly from a complex template such as genomic DNA.

Screening of the human P1 library identified six potentially positive clones, stab cultures of which were obtained from the ICRF. DNA prepared from these clones was investigated by PCR using primer pairs AhRF1/AhRR4 and AhRF9/AhRR8, which should generate am-

GAA GCA GAG CTG TGC ACG AGA GGC TCA GGT TAT CAG TTT ATT CAT GCA GCT GAT ATG CTT TAT TGT GCC GAG TCC CAT ATC CGA

A gtaagttgta gttccttatg aacatgtcag aagaaaacgg catatactgt tgtacatgtt tcaaattctt acgtaatgta aagtgtttaa gtaaagtata

tggataaact tctacttagt aacatatcat tatgattaac cagggttaca ----Intron H total size 649 bp---- atgcctttta ttatttttg

tttaaaattt attgtctttg gggataaagg aaatacatcc agaactatgt cacaagagct ttgttttagg aataatcttt actatattga tttgggggtt

tgataattta attttttaat tttattttag TG ATT AAG ACT GGA GAA AGT GGC ATG ATA GTT TTC CGG CTT CTT ACA AAA AAC AAC CGA

TGG ACT TGG GTC CAG TCT AAT GCA CGC CTG CTT TAT AAA AAT GGA AGA CCA GAT TAT ATC ATT GTA ACT CAG AGA CCA CTA AC

gtaagcacaa ataatgtttc ctgttttaac agttttgttt tcataagtcc tcttatgtga aagcataaaa ataattcaag cagactttag tctgtaaata

aaaattgaaa agtttaattc atctagaaag aagagcacag gtgagagaca ----Intron I total size 2.9 Kb---- tcatgagaat aattgaggtg

aaaataaaat atgtcccttt ctgaattcaa ttacaatgta ttttgcttta tgtttttctt ttttaaatta tttttatttt aaaatgtttg atagaatttt

tttctaagac ttttttgtac acaattttag A GAT GAG GAA GGA ACA GAG CAT TTA CGA AAA CGA AAT ACG AAG TTG CCT TTT ATG TTT

ACC ACT GGA GAA GCT GTG TTG TAT GAG GCA ACC AAC CCT TTT CCT GCC ATA ATG GAT CCC TTA CCA CTA AGG ACT AAA AAT GGC

----------Nucleotide sequence as contained in Genbank file L19872, positions 1678-2685 inclusively.----------

ATG TAT CAG TGC CAG CCA GAA CCT CAG CAC ACC CAC GTG GGT CAG ATG CAG TAC AAT CCA GTA CTG CCA GGC CAA CAG GCA TTT

TTA AAC AAG gtaagggtgt tatcaaactg aattaaatct ttcagtgatt ctttttacct tatagacatg ttacacattt tttatgtcag ctgattttaa

tcggttatct acagcattca tggagacagc attttttatt atatctgtga ctacctttt ----Intron J total size 2.6 Kb---- tgaaagattt

aaaatttagc aacagtaaag ggacttgaag tttacaactc tcaggggtaa gattttaaaa atacatgtta atgttattta ctggcttaag atacttggaa

gatctattcc aataagttgc atcaccattt ttgttttcag TTT CAG AAT GGA GTT TTA AAT GAA ACA TAT CCA GCT GAA TTA AAT AAC ATA

AAT AAC ACT CAG ACT ACC ACA CAT CTT CAG CCA CTT CAT CAT CCG TCA GAA GCC AGA CCT TTT CCT GAT TTG ACA TCC AGT GGA

TTC CTG **TAA** TTCCAAGCCC AATTTTGACC CTGGTTTTTG GATTAAATTA GTTTGTGAAG GATTATGGAA AAATAAAACT GTCACTGTTG GACGTCAGCA

AGTTCACATG GAGGCATTGA TGCATGCTAT TCACAATTAT TCCAAACCAA ATTTTAATTT TTGCTTTTAG AAAAGGGAGT TTAAAAATGG TATCAAAATT

---------Nucleotide sequence as contained in Genbank file L19872, positions 3113-5012 inclusively.----------

AACCTTTTAT TATAAGTCTT ACATAAACCA TTTTTGTTAC TCTCTTCCAC ATGTTACTGG ATAAATTGTT TAGTGGAAAA TAGGCTTTTT AATCATGAAT

ATGATGACAA TCAGTTATAC AGTTATAAAA TTAAAAGTTT GAAAAGCAAT ATTGTATATT TTTATCTATA TAAAATAACT AAAATGTATC TAAGAAT**AAT**

AAAATCACGT

Figure 3 Continued.

plicons at the 5′ and 3′ ends of the AhR gene, respectively. Three clones, ICRFP700A2368, ICRFP700P24121 and ICRFP700N15113, generated PCR products with AhRF9/AhRR8, but only the last one also generated PCR product with AhRF1/AhRR4. This indicated that clone ICRFP700N15113 contained the entire AhR gene.

Using primers AhRF5/AhRR9 and the higher copy number/lower complexity DNA prepared from the P1 clone (ICRFP700N15113) as a PCR template, it was possible to produce a single high yield amplicon of about 15 kilobases in size (fig 1).

All PCR amplicons described above were successfully cloned into pMOSBlue T-vector and used as templates for automated sequencing. In every case it was possible to read the appropriate cDNA sequence, a canonical intron/exon boundary, and variable amounts of intronic sequence. All but one clone contained contiguous cDNA sequence, demonstrating that they contained only a single intron. The one exception contained the fragment generated with primers AhRF4 and AhRR1. Subsequent sequencing of this clone indicated that all of the coding sequence unaccounted for was present within a single exon approximately 0·7 kilobases downstream of primer AhRF4.

The data presented above, and in fig 1, permitted the construction of structural maps for the entire transcribed region of the human AhR gene. These show that the gene is composed of 11 exons and spans about 50 kilobases of genomic DNA (figs 2 and 3).

About 150 base pairs (bp) of unambiguous nucleotide sequence was obtained at both ends of each intron. These data were accepted by GenBank and have been issued the accession numbers U27656, U27657, and U28060 to U28066 consecutively. The nucleotide sequence of regions flanking exons 7, 8 and 9 was confirmed as identical with that contained in GenBank file D38044, and hence was not submitted.

Discussion

Our results show that the intron/exon structure of the human AhR gene is homologous to that of the mouse gene, previously characterised by Schmidt *et al.*[18] However, if we assume that like all of the others, the size of mouse intron B (which was not determined by Schmidt *et al*) is also smaller than human intron B, then the human gene is probably some 20 kilobases larger overall.

To date, two groups have published cDNA sequences of the AhR (GenBank L19872/D16354).[14,15] The major difference between the two sequences is the position of the stop codon, which is at codon number 849 in the former and 809 in the latter. Sequencing of our cloned genomic fragment (AhRF9-AhRR8) confirmed the position of a stop codon at codon number 849. It is, however, entirely possible that this difference is due to the presence of different alleles with alternative stop codons. Differential stop codon usage has been well documented

in the mouse and now requires further investigation in humans.[19 20]

The sequences of three genomic clones have also been published.[21-23] Two of these clones (GenBank D31708/D28768) contain sequence from about 2 kilobases upstream of the transcription start site to several hundred bases inside the 5′ end of intron A. The sequence of the third clone (GenBank D38044) starts in intron F and terminates in intron I. We have considered this sequence information in our investigations and have come across several interesting anomalies, which are discussed below.

Despite considerable attempts, we have been unable to amplify successfully any fragments upstream of the translation start site. PCR reactions using forward primers positioned between the transcription and translation start sites have resulted in multiple non-specific products; reactions using forward primers upstream of the transcription start site have generated no products at all as yet. The same results (not shown) were observed when either genomic DNA or P1 clone DNA was used as the PCR template. It is unlikely that the two published sequences are incorrect as they were obtained independently and are virtually identical. The problems with primers positioned immediately upstream of the translation start site may be due to the very high GC content of this region. However, no unusual sequences are apparent further upstream; we are therefore currently unable to explain our inability to amplify fragments from this region.

With respect to the third genomic clone (GenBank D38044), we have sequenced in from the 5′ end of our clone containing intron F to overlap with this sequence, thus spanning the complete intron. However, this overlap did not occur until position 191 of the D38044 sequence. Database searches using the first 190 bp of the D38044 sequence did not reveal significant homology with any other published sequence, suggesting that they were not part of the vector. Therefore, although a cloning artefact still seems the most probable explanation for this anomaly, other reasons cannot be excluded.

In this paper we present a complete structural map of the human AhR gene and about 150 bases of consensus sequence data at both ends of each intron. These data will facilitate the identification and localisation of human polymorphisms within this gene. Indeed, Southern blot analysis has already identified a *Msp*I restriction fragment length polymorphism and demonstrated its association with a significantly increased risk of lung cancer.[24 25] Using the data we and others have reported, it should now be possible to locate this polymorphism, thereby allowing a considerably more practical PCR based genotyping assay to be developed.

Because of the considerable size of certain introns within this gene, it is likely that they will contain interesting features such as microsatellite repeat sequences and possibly additional regulatory elements. However, their size also means that full sequence determination will incur significant financial implications; for this reason the authors are prepared to distribute freely all clones to any interested groups. (This excludes P1 clones which must be obtained via the ICRF.)

We are grateful to the Faculty of Medicine, University of Birmingham and the Sir Jules Thorn Charitable Trust for their financial support.

1 Le Beau MM, Carver LA, Espinosa R, Schmidt JV, Bradfield CA. Chromosomal localization of the human AHR locus encoding the structural gene for the Ah receptor to 7p21-p15. *Cytogenet Cell Genet* 1994;**66**:172–6.
2 Burbach KM, Poland A, Bradfield CA. Cloning of the Ah receptor cDNA reveals a distinctive ligand activated transcription factor. *Proc Natl Acad Sci USA* 1992;**89**:8185–9.
3 Perdew GH. Association of the Ah receptor with the 90-kDa heat shock protein. *J Biol Chem* 1988;**263**:13802–5.
4 Okey AB, Bondy GP, Mason ME, Kahl GF, Eisen HJ, Guenthner TM, *et al.* Regulatory product of the Ah locus. *J Biol Chem* 1979;**254**:11636–48.
5 Pollenz RS, Sattler CA, Polland A. The aryl-hydrocarbon receptor and aryl-hydrocarbon receptor nuclear translocator protein show distinct subcellular localizations in Hepa1C1C7 cells by immunofluorescence microscopy. *Mol Pharmacol* 1994;**45**:428–38.
6 Matsushita N, Sogawa K, Ema M, Yoshida A, Fujii-Kuriyama Y. A factor binding to the xenobiotic responsive element (XRE) of P-4501A1 gene consists of at least two helix loop helix proteins, Ah receptor and Arnt. *J Biol Chem* 1993;**268**:21002–6.
7 McGuire J, Whitelaw ML, Pongratz I, Gustafsson JA, Poellinger L. A cellular factor stimulates ligand-dependent release of hsp90 from the basic helix-loop-helix dioxin receptor. *Mol Cell Biol* 1994;**14**:2438–46.
8 Bock KW. Aryl hydrocarbon or dioxin receptor: biologic and toxic responses. *Rev Physiol Biochem Pharmacol* 1994;**125**:1–42.
9 Liu RM, Vasiliou V, Zhu H, Duh JL, Tabor MW, Puga A, *et al.* Regulation of the [Ah] gene battery of enzymes and glutathione levels by 5,19-dihydroindeno[1,2-b]indole in mouse hepatoma cell lines. *Carcinogenesis* 1994;**15**:2347–52.
10 Okey AB, Riddick DS, Harper PA. Molecular biology of the aromatic hydrocarbon (dioxin) receptor. *Trends Pharmacol Sci* 1994;**15**:226–32.
11 Nebert DW. Proposed role of drug-metabolizing enzymes: Regulation of steady state levels of the ligands that effect growth, homeostasis, differentiation, and neuroendocrine functions. *Mol Endocrinol* 1991;**5**:1203–14.
12 Nebert DW, Puga A, Vasiliou V. Role of Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal transduction. *Ann NY Acad Sci* 1993;**685**:624–40.
13 Minn A, Ghersi-Egea JF, Perrin R, Leninger B, Siest G. Drug metabolizing enzymes in the brain and cerebral microvessels. *Brain Res* 1991;**16**:65–82.
14 Dolwick KM, Schmidt JV, Carver LA, Swanson HI, Bradfield CA. Cloning and expression of a human Ah receptor cDNA. *Mol Pharmacol* 1993;**44**:911–17.
15 Itoh S, Kamataki T. Human Ah receptor cDNA: analysis for highly conserved sequences. *Nucleic Acids Res* 1993;**21**:3578.
16 Cheng S, Fockler C, Barnes WM, Higuchi R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc Natl Acad Sci USA* 1994;**91**:5695–9.
17 Sykes B. DNA in heritable disease. *Lancet* 1983;**ii**:787–8.
18 Schmidt JV, Carver LA, Bradfield CA. Molecular characterization of the murine Ahr gene. *J Biol Chem* 1993;**268**:22203–9.
19 Chang C, Smith DR, Prasad VS, Sidman CL, Nebert DW, Puga A. Ten nucleotide differences, five of which cause amino acid changes, are associated with the Ah receptor locus polymorphism of C57bl/6 and DBA/2 mice. *Pharmacogenetics* 1993;**3**:312–21.
20 Poland A, Palen D, Glover E. Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol Pharmacol* 1994;**46**:915–21.
21 Eguchi H, Hayashi S, Watanabe J, Gotoh O, Kawajiri K. Molecular cloning of the human Ah receptor gene promoter. *Biochem Biophys Res Commun* 1994;**203**:615–22.
22 Takahashi Y, Itoh S, Shimojima T, Kamataki T. Characterization of Ah receptor promoter in human liver cell line, HepG2. *Pharmacogenetics* 1994;**4**:219–22.
23 Hayashi S, Watanabe J, Nakachi K, Eguchi H, Gotoh O, Kawajiri K. Interindividual difference in expression of human Ah receptor and related P450 genes. *Carcinogenesis* 1994;**15**:801–6.
24 Jones JE, Huckaby CS, Stafford MD, Linnoila RI. An MspI RFLP of the human AHR gene. *Hum Mol Genet* 1994;**3**:2083.
25 Jones JE, Hannigan JL, Shaw GL, Linnoila RI. A novel polymorphism of the human aromatic hydrocarbon (Ah) receptor: association with increased lung cancer risk. *FASEB J* 1995;**9**:134.