Isolation and characterization of the full-length cDNA encoding a member of a novel cytochrome p450 family (CYP320A1) from the tropical freshwater snail, Biomphalaria glabrata, intermediate host for Schistosoma mansoni

Anne E Lockyer, Leslie R Noble*, David Rollinson, Catherine S Jones*/+*

Wolfson Wellcome Biomedical Laboratory, The Natural History Museum, London, UK *School of Biological Sciences, Aberdeen University, Tillydrone Avenue, Aberdeen. AB24 2TZ, Scotland

Cytochrome p450s (cyp450s) are a family of structurally related proteins, with diverse functions, including steroid synthesis and breakdown of toxins. This paper reports the full-length sequence of a novel cyp450 gene, the first to be isolated from the tropical freshwater snail Biomphalaria glabrata, an important intermediate host of Schistosoma mansoni. The nucleotide sequence is 2291 bp with a predicted amino acid sequence of 584aa. The sequence demonstrates conserved cyp450 structural motifs, but is sufficiently different from previously reported cyp450 sequences to be given a new classification, CYP320A1. Initially identified as down-regulated in partially resistant snails in response to S. mansoni infection, amplification of this gene using RT-PCR in both totally resistant or susceptible snail lines when exposed to infection, and all tissues examined, suggests ubiquitous expression. Characterization of the first cyp450 from B. glabrata is significant in understanding the evolution of these metabolically important proteins.

Key words: cytochrome p450 - Biomphalaria glabrata - Schistosoma mansoni - RACE - cyp450 - CYP320A1

The cytochrome p450 (cyp450) superfamily of proteins contains a large number of heme-containing monoxygenases. These proteins oxidize a variety of structurally unrelated compounds including steroids, fatty acids and xenobiotics (for reviews see Nelson et al. 1996, Chang & Kam 1999, Omura 1999). The cyp450s have been classified according to sequence similarity into families (Nelson et al. 1996) and a higher order of classification of 'clans' (Nelson 1998). A large number of these proteins have been identified in individual species, particularly those for which the genome is totally or nearly sequenced. For example the 'superfamily' website genome assignments [http://supfam.mrc-lmb.cam.ac.uk: (Gough et al. 2001, Madera et al. 2004)] has identified 266 cytochrome p450 proteins in Arabidopsis thaliana, 99 in Drosophila melanogaster and 130 in Anopheles gambiae, while in humans, mouse and rat, 104, 164, and 127 cyp450s respectively, have so far been identified. Despite the large numbers found in these organisms, for molluscs only 7 cyp450 genes or sequence fragments are currently submitted to GenBANK, including CYP10 from Lymnea stagnalis (Teunissen et al. 1992), CYP30 from the clam, Mercenaria mercenaria (Brown et al. 1998) and CYP4 fragments from Mytilus gallo-provincialis and Haliotis rufescens (Snyder 1998).

A sequence fragment with homology to the cyp450 superfamily was isolated from the freshwater tropical snail Biomphalaria glabrata (Lockyer et al. 2000), which is an intermediate host to Schistosoma mansoni, the causative agent of human intestinal schistosomiasis. The identified 241 bp sequence fragment, designated BgCYP, did not provide enough information to allow the classification of this sequence to a family of cytochrome p450 protein sequences (Lockyer et al. 2000) but it was demonstrated that it was expressed in ovotestis, mantle and nephridium tissue from snails that were 70% resistant to schistosome infection. This paper reports the full-length sequence of the first cytochrome p450 to be identified from B. glabrata, and expression analysis of this gene in snails differing in their susceptibility to schistosome infection.

MATERIALS AND METHODS

Snail material - Three strains of B. glabrata with different susceptibilities to schistosome infection were investigated. A partially resistant strain (NHM accession number 1778), from which the BgCYP gene fragment was initially identified, was used to isolate the full-length sequence. Other strains used were a 100% resistant strain, (NHM accession number 1981) and a susceptible strain (NHM accession number 1742). Adult snails were swiftly killed by cutting off the head/foot through the shell and preserved in RNAlater™ (Ambion).

RNA isolation - Total RNA was extracted from ovotestis, mantle and brain tissue, using SV total isolation system (Promega) according to the manufacturers instructions. This kit includes a DNase treatment step to remove contaminating genomic DNA. Poly (A) RNA was extracted from total RNA with RNeasy extraction kit (Qiagen) according to the protocol supplied with the kit.

Financial support: The Wellcome Trust, The Biological and Biotechnology Research Council (Advanced fellowship to CSJ, 1/AF09056)
*Corresponding author. E-mail: c.s.jones@abdn.ac.uk
Received 18 November 2004
Accepted 28 March 2005
**3′RACE PCR** - 1 µg total RNA from *B. glabrata* strain 1778, was reverse transcribed in a 20 µl reaction containing 1 × RT buffer (Promega), 5 µM oligo dTAP (TGA CTCGAGTGCAGCATCC GT), 1.25 mM dNTPs (Bioline), 40 U RNAsin (Promega) and 200 U MMLV-RT (moloney murine leukemia virus reverse transcriptase, Promega). The RNA and primer mix was denatured at 70°C and cooled on ice, prior to adding the rest of the reaction components. This was incubated at 22°C for 10 min, 42°C for 60 min, then 70°C for 10 min to inactivate the RT. Controls without reverse transcriptase were performed and tested with *B. glabrata* specific actin primers that span an intron to check for contaminating genomic DNA (Lockyer et al. 2000). The synthesized cDNA was used for a 25 µl PCR containing 1 µl cDNA, 1× PCR buffer (Bioline), 2.5 mM MgCl₂, 0.2 mM dNTPs and 0.5 µM of adaptor primer (AP: TCGTGTGTCAGCATCC GTCA, complementary to oligo dTAP) and 0.5 µM 19F (a specific internal primer: GCTGGCGTCAGCATCC GTCTCC). The reaction was heated to 95°C and 1 U diluted Taq polymerase (Bioline) added through the mineral oil. Reaction conditions were 35 cycles of 30 sec at 95°C, 1 min at 58°C and 3 min at 72°C, followed by 5 min at 72°C.

**5′RACE** - 5 µg poly (A) RNA from *B. glabrata* strain 1778 was reverse transcribed (as above) using specific primers (0.1 µM). Three rounds of 5′ RACE were required to recover the full-length cDNA. Primers used for first strand cDNA synthesis were (i) 19R: AGGGTCAAGGGGTAATCCGTGGG, (ii) 19R: TGTCGAGTGCAGCATCC GTCA, (iii) BgcypR7: ATGGACTTTGGTTGTTGTTGTTGTTG, Controls with no RT or with an oligo dT primer were run simultaneously and checked using specific primers, CYPF: TATTGCGAGATCCACTGTGGC, and (iii) BgcypR5: ATGGACCTTTGGTTGTTGTTGTTGTTGTTG. Controls with no RT or with an oligo dT primer were used simultaneously and checked using specific primers as above. cDNA was cleaned (Qiagen) and tailed with dATPs using terminal deoxynucleotidyl transferase (TdT). Reaction conditions were 30 cycles (30 s at 95°C, 30 s at 60°C and 30 s at 72°C), followed by 5 min at 72°C.

Cloning and sequencing - RACE PCR products were analysed on electrophoresis gels and excised fragments purified using QIAEX II Agarose gel extraction kit (Qiagen). These were reamplified if necessary, ligated into pGEM-T easy vector (Promega) and used to transform JM109 cells (Promega). Positive clones were sequenced on both strands from the purified plasmid (Wizard Miniprep, Promega) using the SequiTHERM EXCEL™ II DNA sequencing kit-LC (Cambio, UK) according to the protocol supplied with the kit and run on a LICOR LongReadIR automatic sequencer (MWG Biotech, UK). Sequences were analysed using e-seq™ (version 1.0) and aligned using Align IR (version 1.2).

**Sequence analysis** - The full-length sequence was compared to GenBank using BlastN and BlastP (Altschul et al. 1990). It was translated using proteomics tools at the EXPASY website (http://ca.expasy.org/sprot/). Protein alignments with the closest Blast matches were made by eye.

**Northern blotting** - 5 µg poly (A) RNA from *B. glabrata* strain 1778 strain ovotestis and 5 µg of RNA marker (Promega) were electrophoresed on a denaturing agarose gel (Sambrook et al. 1989) and transferred onto Hybond N membrane (Amersham) according to the manufacturer’s instructions. A 853 bp PCR product was amplified for use as a probe using specific primers 19R: AGGGTCAAGGGGTAATCCGTGGG and CYPF: TATTGCGAGATCCACTGTGGC. The reaction was heated to 95°C and 1 U diluted Taq polymerase (Bioline) added through the mineral oil. Reaction conditions were 30 cycles of 30 s at 95°C, 30 s at 60°C and 30 s at 72°C, followed by 5 min at 72°C. This fragment was cleaned using Qiagous columns (Qiagen), labelled with [32P]-dATP (Prime-a-gene kit, Promega) and hybridized overnight to the membrane. The hybridized labelled probe was visualized with autoradiography for 16–48 h at –80°C with an intensifying screen.

**Semi-quantitative RT-PCR** - 1 µg total RNA, extracted from mantle, brain and ovotestis dissected from resistant 1981 and susceptible 1742 snail lines, was reverse transcribed as described above, but using an anchored oligo dT12 VN primer. RT-PCR in duplicate, was performed using specific primers, CYPF: TATTGCGAGATCCACTGTGGC and BgcypR5: ATGGACTTTTTGTGTTGCGTTG. 10 µl PCRs contained 1 µl cDNA, 1× PCR buffer (Bioline), 2.5 mM MgCl₂, 0.2 mM dNTPs and 0.5 µM each primer. Reaction conditions were 35 cycles of 30 s at 95°C, 30 s at 60°C and 30 s at 72°C, followed by 5 min at 72°C. The number of cycles during the PCR was limited to the linear phase where the amount PCR product is proportional to the amount of starting material. Semi-quantitative RT-PCR using actin primers under the same conditions, but limiting the number of cycles to 25 was carried out for comparison to the specific products.

**RESULTS**

The full-length sequence of the *B. glabrata* cyp450 (BgCYP) gene was obtained using 3′ and 5′ RACE. Primers were designed from the original sequence fragment. Four overlapping fragments, amplified using 3′ RACE and three rounds of 5′ RACE, were sequenced and aligned. The length of the full sequence was 2291 bp. A northern blot (Fig. 1) using poly (A) RNA extracted from ovotestis of the partially resistant strain, the tissue from which the sequence fragment was originally isolated, showed only one size of transcript identified by the BgCYP sequence, which was estimated to be approximately 2100 bp.
The full length sequence has a long open reading frame (ORF) of 1752 bp coding for a predicted protein of 584 amino acids with 195 bp 5' untranslated region (UTR) and 344 bp 3' UTR (GenBANK accession number AY922309). The 3' UTR contains a polyadenylation signal. The predicted start Met residue is encoded by the first ATG in frame with the long ORF. A BLASTP search using the full length BgCYP predicted peptide sequence identified cyp450 domains from both CYP2 and CYP17 families as well as many other CLAN2 cyp450s. Comparison with these cyp450s from Clan2 shows that the predicted protein for BgCYP contains the cyp450 family signature including a highly conserved cysteine (Cy529) residue, which binds a heme molecule (Fig. 2). The predicted peptide sequence was submitted to the cytochrome p450 nomenclature committee and named CYP320A1, with a sufficiently distinct amino acid sequence to be the first protein in a novel group of cytochrome p450 proteins, which fall into the CLAN 2 (D. Nelson, personal communication), confirming our previous conclusion (Lockyer et al. 2000).

The protein sequence of the B. glabrata cytochrome p450 demonstrates that it shows less than 40% identity to any known cyp450 family and therefore it has been classified as a novel family. As the first member of the CYP320 family, the function of this protein remains unknown and cannot be hypothesized on the basis of sequence homology to other, more characterized, cyp450s.

RT-PCR was performed on two snail lines, a 100% resistant strain and a susceptible strain and showed that the gene is expressed in both these snail lines, in the ovotestis, mantle and brain tissue (Fig. 3). Semi-quantitative RT-PCR using actin primers mirrors that of CYP320A1, suggesting no large difference in the level of expression between the tissues or strains investigated.

**DISCUSSION**

The full-length sequence of BgCYP has enabled us to determine that it codes for a novel cytochrome p450 and the RT-PCR result shows that it is expressed in all tissues examined. This sequence represents the first cytochrome p450 to be identified from B. glabrata and only the 8th sequence from a mollusc. However, cytochrome p450 activity has been examined as a biomarker in response to xenobiotic exposures such as pollution both by biochemical assay in several marine molluscs (reviewed by Snyder 2000) and by RT-PCR in, for example, the freshwater mussel Unio tumidus (Chaty et al. 2004).

The gene was initially identified as down-regulated in partially resistant snails in response to schistosome infection (Lockyer et al. 2000). However this has not been observed in either totally resistant or susceptible snails lines exposed to infection (results not shown). Therefore it appears that BgCYP is not intimately connected with the passage of schistosome infection through the snail intermediate host. However the characterization of the first cyp450 from B. glabrata is significant in understanding
the evolution of these metabolically important proteins. It seems likely that more cyp450 genes will be identified in *B. glabrata*, given that many have been identified in other organisms, and elucidating the role that they play in snail physiology may enable a better understanding of the evolution and diversity of function of these proteins.

**ACKNOWLEDGEMENTS**

To Sarah Martin and Mike Anderson at NHM for snail culture and Dr Fred Lewis, Biomedical Research Institute, Rockville, MD, US, for providing the NHM with the resistant snail line.

**REFERENCES**


