

# 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\*<sup>1</sup>

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 monooxygenases. 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 *Lymnaea stagnalis* (Teunissen et al. 1992), CYP30 from the clam, *Mercenaria mercenaria* (Brown et al. 1998) and CYP4 fragments from *Mytilus galloprovincialis* 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<sup>TM</sup> (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)

<sup>1</sup>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 CTCGAGTCGACATCGAT<sub>20</sub>), 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 MgCl<sub>2</sub>, 0.2 mM dNTPs and 0.5 µM of adaptor primer (AP: TCGATGTGCGACTCGAGTCA, complementary to oligo dTAP) and 0.5 µM 19F (a specific internal primer: GCTGCGGTCACTCCATCGTC). The reaction was heated to 95°C and 1U 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: AGGGTCAAGG TAAAATCTGGG, (ii) 19R2: CTGCCATTACAAAA GTCCC, and (iii) BgCYP5: ATGGACTTTTTGTGTT CGTTG. Controls with no RT or with an oligo dT primer were run simultaneously and checked using actin primers as above. cDNA was cleaned (Qiaquick PCR purification kit, Qiagen) and tailed with dATPs using terminal deoxynucleotidyl transferase (TdT). This 20 µl reaction contained 14 µl cleaned cDNA, denatured for 3 min at 94°C first, 1 × TdT reaction buffer (Promega), 0.8 mM dATP and 20 U TdT (Promega) and was incubated on ice for 60 min, then 10 min at 37°C and 10 min at 70°C. 1 µl tailed cDNA was amplified with the oligo dTAP, AP and specific nested primers: (i) 19R2: CTGCCATTACAAAAGTCCC, (ii) Bgcyp5: ATGGACTTTTTGTGTTTCGTTG, and (iii) BgCYP3R: GTTCACCTTCTTCCCCTCC and a second nested primer, BgCYP2: AGCGTGTTTCAGGTAGAGGAG. 25 µl reaction contained cDNA, 1× PCR buffer, 2.5 µM MgCl<sub>2</sub>, 0.5 µM each of AP and the specific primer, 0.08 µM oligo dTAP and 0.2 mM dNTP. The reaction was heated to 94°C for 5 min and cooled to 80°C. 1 U *Taq* polymerase was added through the mineral oil. Reaction conditions were 1 min at 94°C, 2 min at 48°C and 40 min at 72°C, then 30 cycles of 1 min at 94°C, 1 min at 56-60°C and 3 min at 72°C followed by 15 min at 72°C. The annealing temperature for the cycling was varied between 56-60°C depending on the specific nested primer. Simultaneous controls used untailed cDNA.

**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* 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: AGGGTCAA GGTAATCTGGG and CYP5: TATTGCGAGATCCA CTGTCC. 10 µl PCR contained 1 µl cDNA, 1× PCR buffer (Bioline), 2.5 mM MgCl<sub>2</sub>, 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. This fragment was cleaned using Qiaquick columns (Qiagen), labelled with [<sup>32</sup>P]-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 dT<sub>12</sub>VN primer. RT-PCR in duplicate, was performed using specific primers, CYP5: TATTGCGAGATCCACTG TCC and Bgcyp5: ATGGACTTTTTGTG TTCGTTG. 10 µl PCRs contained 1 µl cDNA, 1× PCR buffer (Bioline), 2.5 mM MgCl<sub>2</sub>, 0.2 mM dNTPs and 0.5 µM each primer. Reaction conditions were 35 cycles of 30 s at 95°C, 30 s at 58°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.

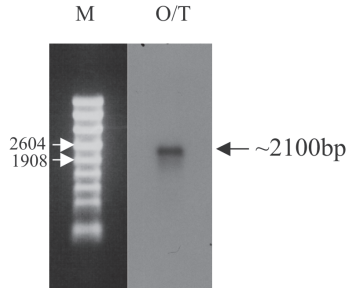


Fig. 1: Northern blot with poly(A) RNA extracted from the ovotestis (O/T) of 70% resistant strain of *Biomphalaria glabrata*, probed with BgCYP. M: RNA marker (Promega).

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 (Cys<sub>529</sub>) 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 cyp450 to be identified 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

Species	SwissProt	cyp name	Sequence
<i>Biomphalaria</i>	BGCYP	320A1	FRPQRHLDLLGRALTPQSFRPYGIGARRCVGESIASMQMFIYVANILHNF
Rabbit	P08682	2E1	EKPEHFLNEEGKFKYSDYFKPFSAGKRVCVGEGGLARMEFLFLLSAILLQHF
Rat	P11715	17A1	FMPERFLDPTGHLITPQSYLPFGAGPRSCIGEALARQELFVFTALLLQRF
Fruitfly	Q95078	18a1	FRPSRFIDTEGKVRKPEYFIPFGVGRMCLGDVLRARMEFLFLFFASFMHCF
Channel catfish	O73853	17A1	FNPERFLNEEGNSLPSLSYLPFGAGVRVCLGEALAKLELFLFLSWILQRF
Chicken	P12394	17A1	FNPGFRFLDEQQQHIPSYSYLPFGAGIRVCLGEVLAKMEFLFLFLAWVLQRF
Mouse	O54750	2J6	FNPEHFLE-NGQFKKRESFLPFSMGKRACLGEQLARSELFIFFTSMLQKF
Spiny dogfish	Q92113	17A1	FDPGRFLDEDGKYVSSSESYLPFGAGTRVCLGEMLARMEFLFLFTSWILQRF
Human	P51589	2J2	FNPDHFLE-NGQFKKREAFMPFSLGKRACLGEQLARTELFIFFTSLMQKF
Mouse	P27786	17A1	FMPERFLDPTGHLITPPSYLPFGAGPRSCIGEALARQELFIIFMALLLQRF
Rat	P51590	2J3	FNPEHFLE-NGQFKKRESFLPFSMGKRACLGEQLARSELFIIFITSLIQKF
Golden Hamster	P51581	2E1	EKPEHFLNENGGKFKYSDHFKAFSAGKRVCVGEGGLARMEFLFLLLTAILQHF
Rat	P33272	2B12	FNPEHFLDANGALKKSEAFLPFSTGKRICLGEGLARNELFIFFTAILQNF
Mouse	O55071	2B19	FNPEHFLDANGALKTNEAFMPFSTGKRICLGEGLARNELFIFFTTILQNF
Spiny lobster	Q27712	2L1	FYPEHFLDDQGFVKREHLVNFVGRVRCVGESLARMEFLFVFLSAILQNF
Mouse	O54749	2J5	FNPEHFLE-NGQFKKRESFLPFSMGKRACLGEQLAKSELFIFFSALMQKF
Mouse	P11714	2D9	FHPEHFLDAQGHFVKPEAFMPFSAAGRRSCLGEALARMEFLFLFTCLLQRF
Rabbit	P00182	2C3	FDPGHFLDESNGFKKSDYFMPFSAAGKRVGEGGLARMEFLFLLLTTILQHF
Medeka fish	P70085	17A1	FNPGFRFLSADGSSLTSSSYLPFGAGLRVCLGEALAKMEFLFLFLSWILQRF
Bovine	P05185	17A1	FMPERFLDPTGTQLISPSYLPFGAGPRSCVGEMLARQELFLFMSRLLQRF

Fig. 2: comparison with the heme-binding region of other cyp450 sequences in clan 2. Amino acids shared between the *Biomphalaria glabrata* cyp450 (CYP320A1) and other cyp450s are shaded. The cysteine residue associated with heme-binding is shaded darker grey.

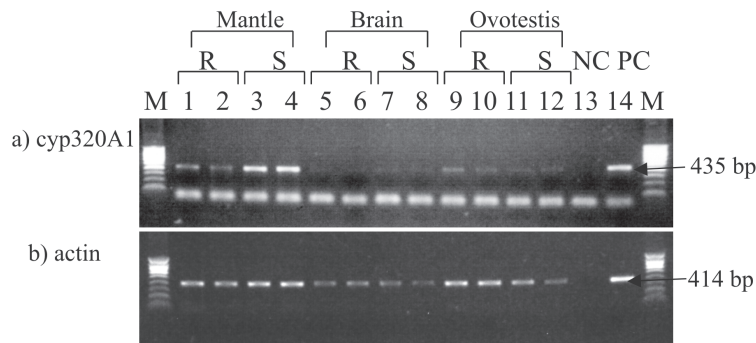


Fig. 3: RT-PCR performed on two snail lines (100% resistant strain and a susceptible strain) shows CYP320A1 is expressed in both snail lines and in all tissues examined (ovotestis, mantle and brain). RT-PCR using a) *Biomphalaria glabrata* CYP320A1 specific primers (435 bp) and b) *B. glabrata* actin primers (414 bp). cDNA samples were from lanes - 1-2, 100% resistant mantle; 3-4, susceptible mantle; 5-6, 100% resistant brain; 7-8, susceptible brain; 9-10, 100% resistant ovotestis; 11-12, susceptible ovotestis; 13, No DNA control, NC and 14, 70% resistant ovotestis (positive control, PC). A fragment of approximately 50 bp present in all lanes (1-14) in Fig. 3a) is primer-dimer. M. DNA size standard (Hyperladder IV Bioline).

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

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Brown DJ, Clark GC, Van Beneden RJ 1998. A new cytochrome P450 (CYP30) family identified in the clam, *Mercenaria mercenaria*. *Comp Biochem Physiol C-Pharmacol Toxicol Endocrinol* 121: 351-360.
- Chang GWM, Kam PCA 1999. The physiological and pharmacological roles of cytochrome P450 isoenzymes. *Anaesthesia* 54: 42-50.
- Chaty S, Rodius F, Vasseur P 2004. A comparative study of the expression of CYP1A and CYP4 genes in aquatic invertebrates (freshwater mussel, *Unio tumidus*) and vertebrate (rainbow trout, *Oncorhynchus mykiss*). *Aquatic Toxicol* 69: 81-93.
- Gough J, Karplus K, Hughey R, Chothia C 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313: 903-919.
- Lockyer AE, Jones CS, Noble LR, Rollinson D 2000. Use of differential display to detect changes in gene expression in the intermediate snail host *Biomphalaria glabrata* upon infection with *Schistosoma mansoni*. *Parasitology* 120: 399-407.
- Madera M, Vogel C, Kummerfeld SK, Chothia C, Gough J 2004. The SUPERFAMILY database in 2004: additions and improvements. *Nuc Acid Res* 32: D235-D239.
- Nelson DR 1998. Metazoan cytochrome P450 evolution. *Comp Biochem Physiol C-Pharmacol Toxicol Endocrinol* 121: 15-22.
- Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW 1996. P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 6: 1-42.
- Omura T 1999. Forty years of cytochrome P450. *Biochem Biophys Res Com* 266: 690-698.
- Sambrook J, Fritsch EF, Maniatis T 1989. *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Snyder MJ 1998. Cytochrome P450 enzymes belonging to the CYP4 family from marine invertebrates. *Biochem Biophys Res Com* 249: 187-190.
- Snyder MJ 2000. Cytochrome p450 enzymes in aquatic invertebrates: recent advances and future directions. *Aquatic Toxicol* 48: 529-547.
- Teunissen Y, Geraerts WPM, Vanheerikhuizen H, Planta RJ, Joosse J 1992. Molecular cloning of a cDNA encoding a member of a novel cytochrome P450 family in the mollusc *Lymnaea stagnalis*. *J Biochem* 112: 249-252.