Structural modeling of cytochrome P450 51 from a deep-sea fish points to a novel structural feature in other CYP51s

Jared V. Goldstone [1#], David C. Lamb [2#], Steven L. Kelly [2], Galina I. Lepesheva [3], and John J. Stegeman [1*]

[1] Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[2] Faculty of Medicine, Health and Life Science, Swansea University, Swansea, SA2 8PP, UK.

[3] Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, 37232, USA

[#] These authors contributed equally to the work

Keywords:

Cytochrome P450; CYP51; deep sea; pressure; adaptation; fish

[*] Corresponding author: John J. Stegeman, Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

E-mail: jstegeman@whoi.edu

Abstract

Cytochromes P450 (CYP), enzymes involved in the metabolism of endogenous and xenobiotic substrates, provide an excellent model system to study how membrane proteins with unique functions have catalytically adapted through evolution. Molecular adaptation of deep-sea proteins to high hydrostatic pressure remains poorly understood. Herein, we have characterized recombinant cytochrome P450 sterol 14α-demethylase (CYP51), an essential enzyme of cholesterol biosynthesis, from an abyssal fish species, *Coryphaenoides armatus*. *C. armatus* CYP51 was heterologously expressed in *Escherichia coli* following N-terminal truncation and purified to homogeneity. Recombinant *C. armatus* CYP51 bound its sterol substrate lanosterol giving a Type I binding spectra (K_D 15 μM) and catalyzed lanosterol 14α-demethylation turnover at 5.8 nmol/min/nmol P450. *C. armatus* CYP51 also bound the azole antifungals ketoconazole (K_D 0.12 μM) and propiconazole (K_D 0.54 μM) as determined by Type II absorbance spectra. Comparison of *C. armatus* CYP51 primary sequence and modeled structures with other CYP51s identified amino acid substitutions that may confer an ability to function under pressures of the deep sea and revealed heretofore undescribed internal cavities in human and other non-deep sea CYP51s. The functional significance of these cavities is not known.

Prologue

This paper is dedicated in memory of Michael Waterman and Tsuneo Omura, who as good friends and colleagues enriched our lives. They continue to inspire us.

1. Introduction

Structural and functional analysis of cytochrome P450 (CYP or P450) genes and proteins in an environmental context has provided critical information on the roles, regulation, and evolution of the P450 superfamily [1]. For example, comparisons of expression and induction of CYP1 genes in fish inhabiting polluted and clean sites have pointed to a pollution-driven evolutionary adaptation in the aryl hydrocarbon (AH) receptor pathway [2]. P450 adaptations to environmental conditions include gene duplication (e.g. in pesticide resistance [3]), as well as single nucleotide substitutions, as found in the sexual selection of CYP2J19 to produce red carotenoids in birds [4]. Other environmental selection pressures may lead to sequence changes that result in functional adaptation to the stressor. This paper concerns the structural adaptation in P450s to an extreme environment, the deep sea, and the possible impact of environmental chemicals known to occur in the deep ocean.

Molecular adaptation to life in the deep sea is essentially an adaptation to maintain functional control at high hydrostatic pressures [5]. Pressures in the ocean increase by approximately 0.1 atmospheres (atm) (10^5 pascals, Pa) per m of depth. At 4,000 m, the average depth of the North Atlantic Ocean, the pressure is roughly 400 atm (40 MPa). The pressures in the deep ocean are high enough to necessitate evolution of P450 protein structural modifications to account for pressure influence(s) on volumes associated with ordered water molecules in channels and around amino acid side-chains, as well as volume changes during catalysis, to continue enzyme function in a controlled way [6], [7]. Increased intracellular concentrations of osmoregulatory compounds such as trimethylamine oxide compensate for some water effects, but

typically such molecules cannot penetrate into protein water channels [8]. We previously described positionally unique amino acid residues in the CYP1A protein from a deep-sea fish, *Coryphanoides armatus* [9], which occurs in the ocean worldwide, at depths between 2,000 and 5,000 m [10]. *C. armatus*, also known as the abyssal grenadier, typically grows to 40 cm, may live to 30 years or more, and maintains low metabolic rates during foraging [11]. The amino acid changes observed in CYP1A were ones that were on the surface of the protein rather than in access channels or the active site, and which we predict reduces the overall ordering of water molecules around the protein [9].

In the present study, we have now examined structural models of an essential P450, CYP51, also from *C. armatus*. In many animals CYP51 catalyzes a critical step during cholesterol biosynthesis, the stepwise oxidative demethylation of lanosterol [12] **(Figure 1).** The objective for this study was two-fold. First, we hypothesize that, like CYP1A, CYP51 will show positionally unique amino acid residues on the surface of the protein. Second, we are interested in CYP51 as a possible target for pollutant chemicals that occur throughout the world and which penetrate even the abyssal and hadal depths of the oceans [13]–[16]. The apparent induction of CYP1A protein in deep sea fish, including *C. armatus*, is associated with polychlorinated biphenyl (PCB) pollutants in the deep ocean and in tissues of the fish [15], [17]. Given the essential role of CYP51 in sterol biosynthesis, we are interested in this protein as a possible target for azoles or other inhibitors of CYP51 that can occur in the ocean. Plastic particles are now ubiquitous in the oceans including in the abyssal realms [18]–[20], and are being found in deep sea fish [21]). UV-protectants contained in plastic particles include compounds

such as hydroxybenzophenones and benzotriazoles, that potentially can bind to CYP51 proteins. Indeed, 4,4' dihydroxybenzophenone was reported to bind to CYP51 from *Mycobacterium tuberculosis* [22] **(Figure 1).** Such chemicals were examined for binding using recombinant *C. armatus* CYP51. In addition to changes in amino acid residues on the surface of the protein, computed protein structures of the deep sea CYP51 and other CYP51s were examined for architectural differences, which identified here-to-fore unknown supplemental cavities in human and other CYP51 proteins.

2. Experimental procedures

2.1. Animals and Materials

Individuals of *C. armatus* were retrieved from 3000 m depth in Monterey Bay Canyon during 2009. Within 1 h of trawl recovery, specimens were sorted on ice. Livers were excised and frozen in liquid nitrogen and kept at − 80 °C until further processing, as discussed elsewhere [9]. All chemicals, unless otherwise stated, were obtained from Sigma Chemical Company (Poole, United Kingdom). Growth media, sodium ampicillin, Isopropyl β-d-1-thiogalactopyranoside (IPTG) and 5-aminolevulenic acid were obtained from Foremedium Ltd. (Hunstanton, United Kingdom). Ni^{2+}-nitrilotriacetic acid (NTA) agarose affinity chromatography matrix was obtained from Qiagen (Crawley, United Kingdom).

2.2. Total RNA extraction and CYP51 gene sequence determination

Total RNA was isolated from livers using the Total Fatty and Fibrous Tissue kit from Bio-Rad (Hercules, USA), and submitted for RNA-seq analysis at the Tufts University Genomics Core. 100bp paired end reads were generated on an Illumina Hi-Seq 2000

instrument, while 250bp paired end reads were generated from the same samples on a MiSeq. Details of the gene expression patterns will be published elsewhere. Following demultiplexing, FastQC, and read trimming (trimmomatic v0.3), reads were assembled using Trinity [23] and annotated using blastx. The CYP51 gene sequence used here was obtained from this data. Alignment to all available full length teleost CYP51 sequences was performed using Clustal Omega [24] (See Supplemental Information.)

2.3 Construction of the *C. armatus* CYP51 expression vector.

The *C. armatus* CYP51 sequence, determined from liver RNAseq, was missing sequence encoding the first 35-37 amino acids (~ 111 bp), which will encode the N-terminal membrane anchor. To engineer the *C. armatus* CYP51 for heterologous expression, a primary sequence alignment with *Aspergillus fumigatus* CYP51B (up to the conserved CYP51 proline Pro^{50} in *A. fumigatus* CYP51B) allowed for replacement of the existing 15 N-terminal amino acids with the 5-amino acid sequence fragment MAKKT- [25]. In addition, a six-histidine extension (CATCACCATCACCATCAC) was engineered immediately before the stop codon to facilitate protein purification by Ni^{2+}-NTA agarose affinity chromatography. Thus, this *C. armatus CYP51* cDNA was synthesized by Eurofins MWG Operon (Ebersberg, Germany), incorporating an *Nde*I restriction site at the 5′ end and a *Hin*dIII restriction site at the 3′ end of the sequence cloned into the pUC57 plasmid. The *C. armatus CYP51* cDNA was excised by *Nde*I/*Hin*dIII restriction digestion followed by cloning into the pCWori^{+} *Escherichia coli* expression vector using Roche T4 DNA ligase. The integrity of the synthesized cDNA was re-confirmed by sequencing.

2.4 Heterologous expression and purification of recombinant CarmCYP51

protein.

E. coli DH5α competent cells were transformed with 100 ng of the *C. armatus* CYP51 expression plasmid (see above). A single colony of bacteria was used to inoculate 5 ml of LB medium containing 0.1 mg/ml ampicillin. An overnight culture was incubated at 37 °C and 250 rpm for 16 h and then diluted into 6 × 500 ml of Terrific Broth medium (2.4% (w/v) yeast extract, 1.2% (w/v) tryptone, 0.1 M (w/v) potassium phosphate, pH 7.5, 0.1% (v/v) glycerol) containing 0.1 mg/ml ampicillin. Cells were grown for 7 h at 37 °C, 200 rpm. Expression was induced by the addition of IPTG (1 mM final concentration) and 5-aminolevulenic acid (1 mM final concentration), followed by incubation at 27 °C and 160 rpm for 20 h. Protein isolation was according to the methods of Arase *et al*. [26]. The solubilized *C. armatus* CYP51 protein was purified by Ni^{2+}-NTA agarose affinity chromatography as previously described for bovine CYP21 [26], followed by dialysis against 5 liters of 20 mM potassium phosphate (pH 7.5), 1% (v/v) sodium cholate and 10% (v/v) glycerol. Protein purity was assessed by SDS-polyacrylamide gel electrophoresis, followed by staining with Coomassie brilliant blue R-250. Ni^{2+}-NTA agarose-purified *C. armatus* CYP51 was used for all subsequent spectral and activity determinations.

2.5 CYP51 reconstitution assay system.

Sterol 14α-demethylase reconstitution assays (500-µl final reaction volume) were performed as previously described [27], containing 0.6 µM *C. armatus* CYP51, 1.8 µM *Aspergillus fumigatus* cytochrome P450 reductase (AfPOR; UniProtKB accession number Q4WM67), 50 µM lanosterol, 50 µM dilaurylphosphatidylcholine, 4% (v/v) (2-hydroxypropyl)-β-cyclodextrin (HPCD), 0.4 mg/ml isocitrate dehydrogenase, 25 mM

trisodium isocitrate, 50 mM NaCl, 5 mM $MgCl_2$, and 40 mM morpholinepropanesulfonic acid (MOPS; pH ~7.2). Reactions were incubated for 10 min at 30 °C prior to assay initiation by the addition of 4 mM β-NADPH-Na_4, final concentration. Samples were shaken for 30 min at 30°C. Sterol metabolites were recovered by extraction with ethyl acetate, followed by derivatization with *N*,*O*-bis(trimethylsilyl)trifluoroacetamide and tetramethylsilane in pyridine prior to analysis by gas chromatography mass spectrometry [28]. CYP51 levels were measured by reduced CO difference spectra using established methods [29]. For CYP51-lanosterol binding experiments, Type I spectral difference was recorded after the incremental addition of lanosterol, and the K_D of the CYP51/lanosterol complex was determined as described previously for recombinant zebrafish CYP51 [30].

2.6 Bioinformatics, modeling and docking:

Homology modeling of CYP51 was performed using Modeller (v9.16) [31]. 100 independent models were produced based on human CYP51 structures (PDB codes 3LD6, 3JUV, 3JUS). The model with the highest DOPE-HR (high resolution discrete optimized potential energy) score was further refined by solvating with water and relaxing the structure using molecular dynamics (NAMD v2.13). Protein sequences for cod (*G. morhua*) and hadal snailfish (*P. swirei*) were predicted from either reference cDNA sequences (cod) or the sequenced genome (snailfish) and modeled using the same homology reference sequences. Protein structure and model alignment was performed using POSA [32]. Alphafold models were constructed with Alphafold v2.3.1 using the Google Colaboratory. Volume differences were visualized using Pymol (v2.5.2, Schrodinger LLC), and analyzed using Hollow [33]; grid spacing 0.5, interior

probe size 1.4 A, surface probe size 8.0 A). The volume occupied by these probes was analyzed using the Voss Volume Voxelator (3V), [34] with a probe radius of 1.4 A. Docking was performed using Smina [35] Sequence analysis was performed in BioEdit (v7.1.3.0) [36].

3.0 Results

3.1 Sequence analysis

CYP51 is extremely well conserved. CYP51 genes are present in plants, fungi, protists, and bacteria [37], and most animals (with the exception of some invertebrate groups [38]–[42]). The isolated *C. armatus* CYP51 cDNA, determined from RNAseq data as previously described [9], consisted of 1545 bp encoding a protein of 515 amino acids. The upstream sequence of 111 bp (5'), corresponding to the first 37 amino acids was missing in our RNAseq data. We anticipate that genomic sequence data, when available, will encode the transmembrane anchor; liver microsomes from C. armatus show a typical P450 spectrum [43], indicating full length P450s. Other metazoan CYP51s including fish CYP51 we have previously cloned, sequenced, and expressed all have a N-terminal membrane anchor and are not soluble P450s [30]. BLASTP analysis shows that the abyssal fish CYP51 shares ~ 80-90 % sequence identity with CYP51 orthologs of surface fish, with the highest identity being to CYP51 of the shallow water cod *Gadus morhua* (89.62 %; see **Supplemental Information**).

Sequence comparison of the *C. armatus* CYP51 with 180 other fish CYP51 sequences available in GenBank (**see Supplemental Table 1**; average ungapped amino acid identity of 81%) identified 13 amino acid positions in the *C. armatus* CYP51

protein that are positionally unique **(Table 1).** (See also **Supplemental Figure S1**). The majority of these amino acid substitutions are located at solvent exposed surfaces, although there is not a consistent pattern of increasing hydrophobicity, assuming the simplifying assumption of a linear substitution scale for the Kyte-Doolittle hydrophobicity scale [44].

3.2 Expression, purification, and spectral characterization of *C. armatus* CYP51.

The yield of expressed *C. armatus* CYP51 was ~50-200 nmol/liter of *E. coli* culture as determined by reduced CO difference spectra [29] after purification by Ni^{2+}-NTA agarose chromatography. SDS-polyacrylamide gel electrophoresis confirmed the purity of the Ni^{2+}-NTA agarose-eluted *C. armatus* CYP51 protein to be greater than 90% when assessed by Coomassie brilliant blue R-250 staining intensity, with an apparent molecular weight of about 50 KDa, which was close to the predicted value of 51,932 Da and included the six-histidine C-terminal tag extension. Similar to other CYP51 orthologs [45]–[47], *C. armatus* CYP51 was obtained in the ligand-free ferric low spin form after purification from *E. coli*. The heme iron was readily reduced by sodium dithionite, and the complex with CO was formed very rapidly, with an absorbance maximum at 451 nm. Additionally, the presence of (inactive) cytochrome P420 was detected (**Figure 2A)**. A presence of P420 is not uncommon in substrate-free P450s.

3.3 Ligand Binding and Activity

C. armatus CYP51 bound its sterol substrate, lanosterol, to produce a Type I difference binding spectrum, with a spectral maximum at ~406 nm and spectral minimum at ~426 nm (**Figure 2B**). The production of a Type I spectral change is indicative of putative substrate molecules for P450s (low- to high- spin transition of the heme iron due to

displacement of a water molecule, which in the ligand-free CYP51s occupies the sixth Fe^{3+} coordination sphere). A spectral dissociation constant (K_D) value for lanosterol of 15 μM was determined. Type II difference binding spectra were observed between *C. armatus* CYP51 and the azole antifungal compounds ketoconazole and propiconazole, yielding a spectral maximum at ~430 nm and a spectral minimum at ~410 nm (**Figures 2C and 2D**). Type II spectra arise from the coordination of the imidazole N3 of ketoconazole and the triazole N4 of propiconazole as a sixth ligand with the *C. armatus* CYP51 heme Fe [48]. The estimated K_D values were 0.12 μM and 0.54 μM for ketoconazole and propiconazole, respectively, indicating that ketoconazole has a higher affinity for *C. armatus* CYP51 than propiconazole.

Sterol 14α-demethylase activity was reconstituted between *C. armatus* CYP51 and *A. fumigatus* cytochrome P450 reductase (AfPOR). AfPOR has been previously shown to drive recombinant CYP51 catalytic activities (e.g., [49]). *C. armatus* CYP51 catalyzed the demethylation of lanosterol to 4, 4-dimethyl-5α-cholesta-8,14,24-triene-3ß-ol with an approximate turnover value of ~5.8 nmol/min/nmol P450. Negative control reactions, from which NADPH or AfPOR were omitted, had no demethylase activity. Our present activity and ligand binding results indicate that the *C. armatus* CYP51 is functionally similar to other CYP51s. The only other fish for which CYP51 activity and binding characteristics has been determined in vitro is zebrafish (*Danio rerio*) [43]. At the same reaction conditions, zebrafish CYP51 displayed a turnover rate of 3.2 nmol/min/nmol P450, about 40% lower than the *C. armatus* rates determined here.

252 **3.4 Molecular modeling**

253 Molecular modeling of the *C. armatus* CYP51 protein based on human and other
254 crystalized CYP51s produced models that are very close to known structures. The
255 overall RMSD of equivalent residues in CYP51s ranged from 0.38-1.91 Å, similar to the
256 range observed between CYP51 crystal structures [50]. As noted above, most of the
257 amino acid substitutions are on the surface of the protein (**Figure 3**). Interestingly, four
258 of the unique amino acids fall in helix D, three on sequential turns of the solvent-facing
259 side of the helix (P174, Q178, E181), and one on the interior side (I184). These
260 substitutions are neutral or increase the hydrophobicity of that region of the protein
261 (**Table 1**). Two of the substitutions (T440, G460) are on K''' and L helices, immediately
262 adjacent to sites predicted to be key in POR-P450 binding (K423 and E438 in CYP2C9
263 [51]–[53]. These substitutions either increase (A to T) or decrease (N to G)
264 hydrophobicity, with uncertain results on POR-CYP51 coupling. *C. armatus* POR has
265 been cloned and expressed [9] but not yet used in protein-protein binding studies. The
266 substitutions at internal sites were in helices D, E, and I **(Figure 3).**

267 Notably, the three positionally unique amino acids on the interior surface of the
268 protein where in locations close to previously undescribed voids (supplemental cavities),
269 a set of connected side pockets seen in the comparator, human CYP51. Searching
270 other CYP51 structures revealed that this cavity (or cavities) is present in both ligand-
271 free and ligand-bound vertebrate, fungal, and protist CYP51 structures, and is formed in
272 part by the beta-3 sheet (**Figure 4**). While this cavity is nominally not solvent exposed,
273 nor are crystallographically-defined waters present in this location in the deposited
274 structures, sufficient volume is present to contain numerous water molecules (ranging

from a calculated 202 to 578 Å^3, **Table 2**; a single water molecule occupies approximately 30 Å^3). Based on our experience, ligand-free CYP51s are more flexible than substrate- or inhibitor-bound structures. The slightly larger supplemental cavity volumes in the ligand-free structures could reflect a possibility that the more relaxed protein state allows expansion of the supplemental cavities. These supplemental cavities do not have any volume (do not exist) in the computed *C. armatus* structure, due to the specific amino acid substitutions. An Alphafold model of C. armatus CYP51 also did not show these cavities. Large supplemental cavities are present in CYP51 models from the closest related species with a sequenced gene, Atlantic cod (*Gadus morhua*), which inhabits shallow depths. Examining the predicted structure of CYP51 gleaned from the genome of snailfish (*Pseudoliparis swirei*), which lives in hadal depths down to 8,000 m [54], showed a very small supplemental cavity (30 Å$^{3)}$. (See Supplemental Figures). We also note that the liparid CYP51 had position unique substitutions on the surface of the protein.

3.5 Molecular docking

Molecular docking of a variety of potential substrates or inhibitors resulted in a wide range of computed affinities. Lanosterol binds in the optimal orientation, with the 14α-methyl group nearest to the heme, and a very high computed affinity (see **Table 3**). This result is similar to what we observed previously with models of zebrafish (*Danio rerio*) and sergeant major (*Abudefduf saxatilis*) CYP51 [30]. The K_D values determined spectrally are often called “apparent K_Ds”, because they are obtained at certain, not necessarily “optimal” conditions. Lanosterol is a very lipophilic molecule, and its apparent binding affinity to the enzyme in aqueous solutions is likely to be much lower.

We also performed docking simulations for a number of different plastic UV light stabilizers that may be transported unmodified to the deep ocean via the sinking of plastic particles. Dihydroxybenzophenone compounds appear to computationally bind very strongly in the active site, but not directly with the heme Fe possibly due to the lack of N atoms in their structures. Spectroscopic titrations of *C. armatus* CYP51 with these compounds did not reveal either Type I or Type II binding spectra for any of the benzophenones, despite the computed affinity. At this time these results stand in contrast to the crystallization of the soluble *M. tuberculosis* CYP51, which was accomplished with 4,4' dihydroxybenzophenone bound to the protein [22].

4.0 Discussion

This study of CYP51 in an abyssal fish species is part of our ongoing efforts to understand the gene-environment interactions involving P450s. The deep sea is one of the most expansive and species-rich habitats on Earth, now threatened [55]. About 80% of the volume occupied by life is found at depths below 1000 m. The hydrostatic pressure (HP) produced by the overlying water can reach 1100 bar (110 MPa) in the deepest trenches, yet microbes, invertebrates and vertebrates (fishes) have evolved the ability to live at these extreme pressures [5]. As an evolutionary force, HP possibly leads to changes in the number of genes present in a given species (gains and losses), modulate their levels of expression, and lead to subtle changes of translated protein sequence to allow optimal functionality at high HP.

The linear increase of HP with depth in the oceans affects the vertical distribution

of animals, in accordance with the barosensitivity of their enzymatic arsenal [56]. Studies with deep-sea fish have shown HP tolerance of the few proteins that have been studied when compared to surface fish. In lactate dehydrogenase a surprisingly small number of amino acid substitutions (21) were involved (e.g. [6]). More recent genome sequencing of several deep sea snailfish has uncovered substitutions in a number of different inferred protein sequences that may be related to barotolerance, but no protein work has been performed [54], [57]. Structural features of HP tolerance thus remain largely unknown in deep-sea animals.

The number of position specific amino acid substitutions seen in *C. armatus* CYP51 (13) is somewhat greater than the number of substitutions (5) seen in CYP1A from the same species [9]. The majority of the substitutions were, however, distributed on the surface of the protein, as was also the case in CYP1A. While the substitutions in CYP1A represent both polar and non-polar substitutions, overall the substitutions in both cases were also ones that would tend to reduce the ordering of water molecules around the protein, which would tend to reduce the overall volume change in catalysis and contribute to pressure tolerance.

One might also expect to find amino acid substitutions in the active site, or in access channels or indeed in any part of the internal structure where water molecules may participate in function. In contrast to CYP1A, we did observe internal substitutions in *C. armatus* CYP51. The location of these, in comparison with other CYP51s, revealed a set of small cavities in human and other CYP51s that are completely absent in the *C. armatus* CYP51 protein models. These cavities, discovered by comparison of the structures of the different CYP51s, have not previously been identified to our

knowledge, perhaps due to the lack of crystallographically-defined waters or due to crystallization conditions themselves. The fact that such a cavity does not appear to exist in the *C. armatus* CYP51 protein, and while present but extremely small in the hadal snailfish, suggests their absence or near absence might contribute to overall pressure tolerance in this protein. An equally interesting question is whether this previously unknown set of small cavities may have some functional significance. Further studies, including crystallization of the *C. armatus* CYP51 protein and site-directed mutagenesis of human, mycobacterium or other CYP51s to convert them *to C. armatus* sequences might eliminate the voids and affect function.

Our present enzyme activity and ligand binding results indicate that the *C. armatus* CYP51 is functionally similar to other CYP51s from fish that do not inhabit depths as does *C. armatus*. The only other fish for which CYP51 activity and binding characteristics has been determined *in vitro* is zebrafish (*Danio rerio*) [30]. Zebrafish CYP51 displayed a turnover rate of 3.2 nmol/min/nmol P450, about 40% lower than the *C. armatus* rates determined here. Binding of ketoconazole (K_D = 0.26 μM) and propioconazole (K_D = 0.64 μM) to zebrafish CYP51 was also slightly weaker (**Table 2**). These values have been determined with assays at 1 bar (1 atm). Thus far we have not been able to accomplish experimental studies of enzyme function under high pressure. We predict that the function of CYP51 from *C. armatus* would be little influenced by pressures at least to 400 atm, while human and other CYP51s would be affected.

Part of our selection of CYP51 as a focus is the possibility that environmental chemicals may bind to and possibly inhibit this essential enzyme in deep sea animals [21], [58]. We speculate that structural adaptation to pressure might even render the

deep-sea protein more susceptible to such compounds. The binding studies with azole compounds did not reveal much difference among the different proteins at this time (see **Table 2**). We also examined compounds that may be carried with plastics, which are now present in both abyssal and hadal environments [59]. Plastic particles have been identified in deep sea fish [21], [58]. This suggests that the particles or associated chemicals, including plasticizers could act on fishes (or other species in deep sea ecosystems). The UV stabilizer 4,4'- hydroxybenzophenone did not bind to produce a spectral change, nor did other plasticizers that produce good docking scores, in contrast to binding with soluble *M. tuberculosis* CYP51 [22]. We note that in a manner analogous to human CYP51, the detergents used in purifying *C. armatus* CYP51 may have deleteriously affected the binding of dihydroxybenzophenones to the protein. For example, *M. tuberculosis* CYP51 can bind estriol, while detergent purified human CYP51 cannot, ostensibly due to carryover of detergent.

5. Conclusion

In summary, in this study we identified positionally unique amino acid substitutions in CYP51 from an abyssal fish which we hypothesize will confer pressure tolerance, allowing controlled catalytic function of this essential protein at habitat pressures. In molecular modeling locations of the amino acid substitutions and comparing them to human CYP51, we found a supplemental cavity present in human CYP51 but not *C. armatus* CYP51. Examining CYP51 proteins from other species reveals cavities are commonly present but the functional significance of this void pocket is unknown. Having now found the presence of supplemental cavities in CYP51s from human or near surface fish, the fact that these cavities are either absent or an order of

magnitude smaller in CYP51s from an abyssal and a hadal fish suggests that this could be an evolutionary adaptation to the deep-sea environment. Analysis of additional CYP51s from the deep sea are warranted. X-ray structural analysis of *C. armatus* CYP51 is currently underway in our laboratories, which we hope will shed new insights on our experimental and modelled findings described herein.

Acknowledgements

Support for this research was provided by the Woods Hole Center for Oceans and Human Health (the U.S. Oceans and Human Health Program (NIH grant P01ES028938 and National Science Foundation grant OCE-1840381) (J.J.S.), the Ocean Vision 2030 Fund and its generous donors through the Woods Hole Oceanographic Institution (J.V.G.), the Vertebrate Genome Nomenclature Committee NIH grant U24HG003345 (J.V.G.). Funding at Swansea University was supported by the European Regional Development Fund/ Welsh European Funding Office via the BEACON project (S.L.K.) and a UK-US Fulbright Scholarship (D.C.L.). Funding at Vanderbilt University was supported by NIH grant R01 GM067871 (G.I.L.). We thank anonymous reviewers for their very helpful comments.

Author contributions

JJS, JVG and DCL - Conceptualization; JVG and DCL - formal analysis and methodology; All - data interpretation. JJS, DCL and JVG - original drafting; All - review and editing.

References

[1] J. J. Stegeman and J. J. Lech, “Cytochrome P-450 monooxygenase systems in aquatic species: carcinogen metabolism and biomarkers for carcinogen and pollutant exposure.,” *Environ. Health Perspect.*, vol. 90, pp. 101–109, Jan. 1991, doi: 10.1289/ehp.90-1519513.

[2] I. Wirgin, N. K. Roy, M. Loftus, R. C. Chambers, D. G. Franks, and M. E. Hahn, “Mechanistic basis of resistance to PCBs in Atlantic tomcod from the Hudson River.,” *Science*, vol. 331, no. 6022, pp. 1322–1325, Mar. 2011, doi: 10.1126/science.1197296.

[3] C. T. Zimmer *et al.*, “Neofunctionalization of Duplicated P450 Genes Drives the Evolution of Insecticide Resistance in the Brown Planthopper.,” *Curr. Biol. CB*, vol. 28, no. 2, pp. 268-274.e5, Jan. 2018, doi: 10.1016/j.cub.2017.11.060.

[4] H. Twyman, S. Andersson, and N. I. Mundy, “Evolution of CYP2J19, a gene involved in colour vision and red coloration in birds: positive selection in the face of conservation and pleiotropy.,” *BMC Evol. Biol.*, vol. 18, no. 1, p. 22, Feb. 2018, doi: 10.1186/s12862-018-1136-y.

[5] P. M. Oger and M. Jebbar, “The many ways of coping with pressure.,” *Res. Microbiol.*, vol. 161, no. 10, pp. 799–809, Dec. 2010, doi: 10.1016/j.resmic.2010.09.017.

[6] A. A. Brindley, R. W. Pickersgill, J. C. Partridge, D. J. Dunstan, D. M. Hunt, and M. J. Warren, “Enzyme sequence and its relationship to hyperbaric stability of artificial and natural fish lactate dehydrogenases.,” *PloS One*, vol. 3, no. 4, p. e2042, Apr. 2008, doi: 10.1371/journal.pone.0002042.

[7] A. Gibbs and G. N. Somero, “Pressure adaptation of Na+/K+-ATPase in gills of marine teleosts.,” *J. Exp. Biol.*, vol. 143, pp. 475–492, May 1989, doi: 10.1242/jeb.143.1.475.

[8] P. H. Yancey and J. F. Siebenaller, “Co-evolution of proteins and solutions: protein adaptation versus cytoprotective micromolecules and their roles in marine organisms.,” *J. Exp. Biol.*, vol. 218, no. Pt 12, pp. 1880–1896, Jun. 2015, doi: 10.1242/jeb.114355.

[9] B. Lemaire *et al.*, “Molecular adaptation to high pressure in cytochrome P450 1A and aryl hydrocarbon receptor systems of the deep-sea fish Coryphaenoides armatus.,” *Biochim. Biophys. Acta Proteins Proteomics*, vol. 1866, no. 1, pp. 155–165, Jan. 2018, doi: 10.1016/j.bbapap.2017.06.026.

[10] A. J. Jamieson, T. Fujii, M. Solan, A. K. Matsumoto, P. M. Bagley, and I. G. Priede, “Liparid and macrourid fishes of the hadal zone: in situ observations of activity and feeding behaviour.,” *Proc. Biol. Sci.*, vol. 276, no. 1659, pp. 1037–1045, Mar. 2009, doi: 10.1098/rspb.2008.1670.

[11] O. A. Bergstad, R. H. Hunter, N. J. Cousins, D. M. Bailey, and T. Jørgensen, “Notes on age determination, size and age structure, longevity and growth of co-occurring macrourid fishes.,” *J. Fish Biol.*, vol. 99, no. 3, pp. 1032–1043, Sep. 2021, doi: 10.1111/jfb.14801.

[12] G. I. Lepesheva and M. R. Waterman, “Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms.,” *Biochim. Biophys. Acta*, vol. 1770, no. 3, pp. 467–477, Mar. 2007, doi: 10.1016/j.bbagen.2006.07.018.

[13] J. Cui *et al.*, “Occurrence of Halogenated Organic Pollutants in Hadal Trenches of the Western Pacific Ocean.,” *Environ. Sci. Technol.*, vol. 54, no. 24, pp. 15821–15828, Dec. 2020, doi: 10.1021/acs.est.0c04995.

[14] O. Froescheis, R. Looser, G. M. Cailliet, W. M. Jarman, and K. Ballschmiter, “The deep-sea as a final global sink of semivolatile persistent organic pollutants? Part I: PCBs in surface and deep-sea dwelling fish of the north and south Atlantic and the Monterey Bay Canyon (California).,” *Chemosphere*, vol. 40, no. 6, pp. 651–660, Mar. 2000, doi: 10.1016/s0045-6535(99)00461-0.

[15] J. J. Stegeman, P. J. Kloepper-Sams, and J. W. Farrington, “Monooxygenase Induction and Chlorobiphenyls in the Deep-Sea Fish Coryphaenoides armatus.,” *Science*, vol. 231, no. 4743, pp. 1287–1289, Mar. 1986, doi: 10.1126/science.231.4743.1287.

[16] S. Takahashi *et al.*, “Organohalogen compounds in deep-sea fishes from the western North Pacific, off-Tohoku, Japan: Contamination status and bioaccumulation profiles.,” *Mar. Pollut. Bull.*, vol. 60, no. 2, pp. 187–196, Feb. 2010, doi: 10.1016/j.marpolbul.2009.09.027.

[17] S. Koenig, P. Fernández, and M. Solé, “Differences in cytochrome P450 enzyme activities between fish and crustacea: relationship with the bioaccumulation patterns of polychlorobiphenyls (PCBs).,” *Aquat. Toxicol. Amst. Neth.*, vol. 108, pp. 11–17, Feb. 2012, doi: 10.1016/j.aquatox.2011.10.016.

[18] R. Nakajima *et al.*, “Massive occurrence of benthic plastic debris at the abyssal seafloor beneath the Kuroshio Extension, the North West Pacific.,” *Mar. Pollut. Bull.*, vol. 166, p. 112188, May 2021, doi: 10.1016/j.marpolbul.2021.112188.

[19] Nurlatifah *et al.*, “Plastic additives in deep-sea debris collected from the western North Pacific and estimation for their environmental loads.,” *Sci. Total Environ.*, vol. 768, p. 144537, May 2021, doi: 10.1016/j.scitotenv.2020.144537.

[20] S. Zhao, E. R. Zettler, R. P. Bos, P. Lin, L. A. Amaral-Zettler, and T. J. Mincer, “Large quantities of small microplastics permeate the surface ocean to abyssal depths in the South Atlantic Gyre.,” *Glob. Change Biol.*, vol. 28, no. 9, pp. 2991–3006, May 2022, doi: 10.1111/gcb.16089.

[21] L. Soliño, L. Vidal-Liñán, P. Pérez, S. García-Barcelona, F. Baldó, and J. Gago, “Microplastic occurrence in deep-sea fish species Alepocephalus bairdii and Coryphaenoides rupestris from the Porcupine Bank (North Atlantic).,” *Sci. Total Environ.*, vol. 834, p. 155150, Aug. 2022, doi: 10.1016/j.scitotenv.2022.155150.

[22] A. N. Eddine, J. P. von Kries, M. V. Podust, T. Warrier, S. H. E. Kaufmann, and L. M. Podust, “X-ray structure of 4,4’-dihydroxybenzophenone mimicking sterol substrate in the active site of sterol 14alpha-demethylase (CYP51).,” *J. Biol. Chem.*, vol. 283, no. 22, pp. 15152–15159, May 2008, doi: 10.1074/jbc.M801145200.

[23] M. G. Grabherr *et al.*, “Full-length transcriptome assembly from RNA-Seq data without a reference genome.,” *Nat. Biotechnol.*, vol. 29, no. 7, pp. 644–652, May 2011, doi: 10.1038/nbt.1883.

[24] F. Sievers and D. G. Higgins, “Clustal Omega for making accurate alignments of many protein sequences.,” *Protein Sci. Publ. Protein Soc.*, vol. 27, no. 1, pp. 135–145, Jan. 2018, doi: 10.1002/pro.3290.

[25] G. I. Lepesheva *et al.*, “Structural insights into inhibition of sterol 14alpha-demethylase in the human pathogen Trypanosoma cruzi.,” *J. Biol. Chem.*, vol. 285, no. 33, pp. 25582–25590, Aug. 2010, doi: 10.1074/jbc.M110.133215.

[26] M. Arase, M. R. Waterman, and N. Kagawa, “Purification and characterization of bovine steroid 21-hydroxylase (P450c21) efficiently expressed in Escherichia coli.,” *Biochem. Biophys. Res. Commun.*, vol. 344, no. 1, pp. 400–405, May 2006, doi: 10.1016/j.bbrc.2006.03.067.

[27] G. I. Lepesheva *et al.*, “Sterol 14alpha-demethylase as a potential target for antitrypanosomal therapy: enzyme inhibition and parasite cell growth.,” *Chem. Biol.*, vol. 14, no. 11, pp. 1283–1293, Nov. 2007, doi: 10.1016/j.chembiol.2007.10.011.

[28] J. E. Parker *et al.*, “Prothioconazole and prothioconazole-desthio activities against Candida albicans sterol 14-α-demethylase.,” *Appl. Environ. Microbiol.*, vol. 79, no. 5, pp. 1639–1645, Mar. 2013, doi: 10.1128/AEM.03246-12.

[29] T. OMURA and R. SATO, “THE CARBON MONOXIDE-BINDING PIGMENT OF LIVER MICROSOMES. II. SOLUBILIZATION, PURIFICATION, AND PROPERTIES.,” *J. Biol. Chem.*, vol. 239, pp. 2379–2385, Jul. 1964.

[30] A. M. S. Morrison, J. V. Goldstone, D. C. Lamb, A. Kubota, B. Lemaire, and J. J. Stegeman, “Identification, modeling and ligand affinity of early deuterostome CYP51s, and functional characterization of recombinant zebrafish sterol 14α-demethylase.,” *Biochim. Biophys. Acta*, vol. 1840, no. 6, pp. 1825–1836, Jun. 2014, doi: 10.1016/j.bbagen.2013.12.009.

[31] N. Eswar *et al.*, “Comparative protein structure modeling using Modeller.,” *Curr. Protoc. Bioinforma.*, vol. Chapter 5, p. Unit-5.6, Oct. 2006, doi: 10.1002/0471250953.bi0506s15.

[32] Z. Li, P. Natarajan, Y. Ye, T. Hrabe, and A. Godzik, “POSA: a user-driven, interactive multiple protein structure alignment server.,” *Nucleic Acids Res.*, vol. 42, no. Web Server issue, pp. W240-245, Jul. 2014, doi: 10.1093/nar/gku394.

[33] B. K. Ho and F. Gruswitz, “HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures.,” *BMC Struct. Biol.*, vol. 8, p. 49, Nov. 2008, doi: 10.1186/1472-6807-8-49.

[34] N. R. Voss and M. Gerstein, “3V: cavity, channel and cleft volume calculator and extractor.,” *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W555-562, Jul. 2010, doi: 10.1093/nar/gkq395.

[35] D. R. Koes, M. P. Baumgartner, and C. J. Camacho, “Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise.,” *J. Chem. Inf. Model.*, vol. 53, no. 8, pp. 1893–1904, Aug. 2013, doi: 10.1021/ci300604z.

[36] Hall, T., “BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.,” *Nucl Acids Symp Ser*, vol. 41, pp. 95–98, 1999.

[37] D. C. Lamb *et al.*, “Concerning P450 Evolution: Structural Analyses Support Bacterial Origin of Sterol 14α-Demethylases.,” *Mol. Biol. Evol.*, vol. 38, no. 3, pp. 952–967, Mar. 2021, doi: 10.1093/molbev/msaa260.

[38] W. Dermauw, T. Van Leeuwen, and R. Feyereisen, “Diversity and evolution of the P450 family in arthropods.,” *Insect Biochem. Mol. Biol.*, vol. 127, p. 103490, Dec. 2020, doi: 10.1016/j.ibmb.2020.103490.

[39] J. V. Goldstone *et al.*, “Cytochrome P450 1 genes in early deuterostomes (tunicates and sea urchins) and vertebrates (chicken and frog): origin and diversification of the CYP1 gene family.,” *Mol. Biol. Evol.*, vol. 24, no. 12, pp. 2619–2631, Dec. 2007, doi: 10.1093/molbev/msm200.

[40] J. L. Humble *et al.*, “Genome-wide survey of cytochrome P450 genes in the salmon louse Lepeophtheirus salmonis (Krøyer, 1837).,” *Parasit. Vectors*, vol. 12, no. 1, p. 563, Nov. 2019, doi: 10.1186/s13071-019-3808-x.

[41] D. R. Nelson, J. V. Goldstone, and J. J. Stegeman, “The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s.,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 368, no. 1612, p. 20120474, Feb. 2013, doi: 10.1098/rstb.2012.0474.

[42] K. V. Pankov, A. G. McArthur, D. A. Gold, D. R. Nelson, J. V. Goldstone, and J. Y. Wilson, “The cytochrome P450 (CYP) superfamily in cnidarians.,” *Sci. Rep.*, vol. 11, no. 1, p. 9834, May 2021, doi: 10.1038/s41598-021-88700-y.

[43] John Stegeman, “Hepatic microsomal monooxygenase activity and the biotransformation of hydrocarbons in deep benthic fish from the western North Atlantic.,” *Can. J. Fish. Aquat. Sci.*, vol. 40: (Suppl. 2), pp. 78–85.

[44] J. Kyte and R. F. Doolittle, “A simple method for displaying the hydropathic character of a protein.,” *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, May 1982, doi: 10.1016/0022-2836(82)90515-0.

[45] T. Y. Hargrove, Z. Wawrzak, J. Liu, W. D. Nes, M. R. Waterman, and G. I. Lepesheva, “Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14alpha-demethylase (CYP51) from Leishmania infantum.,” *J. Biol. Chem.*, vol. 286, no. 30, pp. 26838–26848, Jul. 2011, doi: 10.1074/jbc.M111.237099.

[46] G. I. Lepesheva *et al.*, “CYP51 from Trypanosoma cruzi: a phyla-specific residue in the B’ helix defines substrate preferences of sterol 14alpha-demethylase.,” *J. Biol. Chem.*, vol. 281, no. 6, pp. 3577–3585, Feb. 2006, doi: 10.1074/jbc.M510317200.

[47] G. I. Lepesheva, W. D. Nes, W. Zhou, G. C. Hill, and M. R. Waterman, “CYP51 from Trypanosoma brucei is obtusifoliol-specific.,” *Biochemistry*, vol. 43, no. 33, pp. 10789–10799, Aug. 2004, doi: 10.1021/bi048967t.

[48] C. R. Jefcoate, J. L. Gaylor, and R. L. Calabrese, “Ligand interactions with cytochrome P-450. I. Binding of primary amines.,” *Biochemistry*, vol. 8, no. 8, pp. 3455–3463, Aug. 1969, doi: 10.1021/bi00836a049.

[49] A. G. S. Warrilow *et al.*, “Isavuconazole and voriconazole inhibition of sterol 14α-demethylases (CYP51) from Aspergillus fumigatus and Homo sapiens.,” *Int. J. Antimicrob. Agents*, vol. 54, no. 4, pp. 449–455, Oct. 2019, doi: 10.1016/j.ijantimicag.2019.07.011.

[50] T. Y. Hargrove *et al.*, “Relaxed Substrate Requirements of Sterol 14α-Demethylase from Naegleria fowleri Are Accompanied by Resistance to Inhibition.,” *J. Med. Chem.*, vol. 64, no. 23, pp. 17511–17522, Dec. 2021, doi: 10.1021/acs.jmedchem.1c01710.

[51] M. Sellner, A. Fischer, C. G. Don, and M. Smieško, “Conformational Landscape of Cytochrome P450 Reductase Interactions.,” *Int. J. Mol. Sci.*, vol. 22, no. 3, Jan. 2021, doi: 10.3390/ijms22031023.

[52] E. Hazai, Z. Bikádi, M. Simonyi, and D. Kupfer, “Association of cytochrome P450 enzymes is a determining factor in their catalytic activity.,” *J. Comput. Aided Mol. Des.*, vol. 19, no. 4, pp. 271–285, Apr. 2005, doi: 10.1007/s10822-005-4995-4.

[53] C. D. Bostick, K. M. Hickey, L. A. Wollenberg, D. R. Flora, T. S. Tracy, and P. M. Gannett, “Immobilized Cytochrome P450 for Monitoring of P450-P450 Interactions and Metabolism.,” *Drug Metab. Dispos. Biol. Fate Chem.*, vol. 44, no. 5, pp. 741–749, May 2016, doi: 10.1124/dmd.115.067637.

[54] K. Wang *et al.*, “Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation.,” *Nat. Ecol. Evol.*, vol. 3, no. 5, pp. 823–833, May 2019, doi: 10.1038/s41559-019-0864-8.

[55] R. Danovaro *et al.*, “Ecological variables for developing a global deep-ocean monitoring and conservation strategy.,” *Nat. Ecol. Evol.*, vol. 4, no. 2, pp. 181–192, Feb. 2020, doi: 10.1038/s41559-019-1091-z.

[56] G. N. Somero, “Biochemical ecology of deep-sea animals.,” *Experientia*, vol. 48, no. 6, pp. 537–543, Jun. 1992, doi: 10.1007/BF01920236.

[57] Y. Mu *et al.*, “Whole genome sequencing of a snailfish from the Yap Trench (~7,000 m) clarifies the molecular mechanisms underlying adaptation to the deep sea.,” *PLoS Genet.*, vol. 17, no. 5, p. e1009530, May 2021, doi: 10.1371/journal.pgen.1009530.

[58] J. M. Pereira, Y. Rodríguez, S. Blasco-Monleon, A. Porter, C. Lewis, and C. K. Pham, “Microplastic in the stomachs of open-ocean and deep-sea fishes of the North-East Atlantic.,” *Environ. Pollut. Barking Essex 1987*, vol. 265, no. Pt A, p. 115060, Oct. 2020, doi: 10.1016/j.envpol.2020.115060.

[59] G. Peng, R. Bellerby, F. Zhang, X. Sun, and D. Li, “The ocean’s ultimate trashcan: Hadal trenches as major depositories for plastic pollution.,” *Water Res.*, vol. 168, p. 115121, Jan. 2020, doi: 10.1016/j.watres.2019.115121.

621 **Figure legends**

622

623 **(Figures 1 and 2 are one column; Figures 3 and 4 are two column.)**

624 **Figure 1. CYP51 reaction and ligand structures (one column)**

625 **Figure 2. Characterization of Recombinant *C. armatus* CYP51**. (**A**) Dithionite-
626 reduced, CO difference spectrum of the purified *C. armatus* CYP51. (**B**) Induced Type I
627 binding spectrum of *C. armatus* CYP51 by lanosterol. (**C**) Induced Type II binding
628 spectra by the medical azole antifungal ketoconazole and (**D**) Induced Type II binding
629 spectra by the agricultural azole antifungal propiconazole.

630

631 **Figure 3. Modeling Location of Positionally Unique Amino Acids**
632 Model of *C. armatus* CYP51 showing positionally unique amino acids (yellow). Note that
633 these are mostly, but not exclusively, on the surface of the protein (**see Table 1**).
634 Locations of selected amino acids are shown. **A** and **B** are rotated views of the same
635 model.

636

637 **Figure 4. Novel Cavities in CYP51. (A)** Crystal structure of human CYP51 (PDB:
638 3JUV). **(B)** Crystal structure of the *Aspergillus* CYP51 (PDB: 6CR2). The figures show
639 void regions (black) in these structures that are absent from the *C. armatus* CYP51.

640