Molecular diversity and genomic organisation of the α, β and γ eye lens crystallins from the Antarctic toothfish *Dissostichus mawsoni*

Andor J. Kiss *, C.-H. Christina Cheng

Department of Animal Biology, 515 Morrill Hall, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Received 4 December 2007; received in revised form 3 February 2008; accepted 7 February 2008

Available online 19 February 2008

**Abstract**

The eye lens of the Antarctic toothfish living in the −2 °C Southern Ocean is cold-stable. To investigate the molecular basis of this cold stability, we isolated, cloned and sequenced 22 full length crystallin cDNAs. We found two α crystallins (αA, αB), six β crystallins (βA1, βA2, βA4, βB1, βB2, βB3) and 14 γ crystallins (γ N, γS1, γS2, γ M1, γ M3, γ M4, γ M5, γ M7, γ M8a, γ M8b, γ M8c, γ M8d, γ M8e, and γ M9). Alignments of α, β and γ with other known crystallin sequences indicate that toothfish α and β crystallins are relatively conserved orthologues of their vertebrate counterparts, but the toothfish and other fish γM crystallins form a distinct group that are not orthologous to mammalian γ crystallins. A preliminary Fingerprinted Contig analysis of clones containing crystallin genes screened from a toothfish BAC library indicated α crystallin genes occurred in a single genomic region of ~266 kbp, β crystallin genes in ~273 kbp, while the γ crystallin gene family occurred in two separate regions of ~180 and ~296 kbp. In phylogenetic analysis, the γM isoforms of the ectothermic toothfish displayed a diversity not seen with endothermic mammalian γ crystallins. Similar to other fishes, several toothfish γ crystallins are methionine-rich (γM isoforms) which may have predisposed the toothfish lens to biochemically attenuate γ crystallin hydrophobicity allowing for cold adaptation. In addition to high methionine content, conservation of αβ crystallins both in sequence and abundance suggests greater functional constraints relative to γ crystallins. Conversely, reduced constraints upon γ crystallins could have allowed for greater evolutionary plasticity resulting in increased polydispersity of γ crystallins contributing to the cold-stability of the Antarctic toothfish lens.

© 2008 Elsevier Inc. All rights reserved.

**Keywords:** Lens; Crystallins; Antarctic toothfish; Alpha; Beta; Gamma; cDNA; Genome organisation; Bacterial artificial chromosome; *Tetraodon*

1. Introduction

The Antarctic toothfish *Dissostichus mawsoni* is a large predatory fish that is endemic to the waters of the Southern Ocean. It is a member of the teleost suborder Notothenioidei which includes about 45% of the fish species and accounts for over 95% of the fish biomass in the sub-zero (~2 °C) continental shelf waters of Antarctica. The Antarctic notothenioid fishes represent an important and diverse taxonomic group of cold adapted ectothermic vertebrates (Eastman, 1993; Parker et al., 2002; di Prisco et al., 2007). Many of the fishes which inhabit the perennally sub-zero seawater have undergone a myriad of physiological and biochemical adaptations (Detrich, 1991; Behan-Martin et al., 1993; Eastman, 1993; Cheng and Chen, 1999), most notable of which was the evolution of a blood borne antifreeze glycoprotein which prevents these fish from freezing (DeVries, 1988; Cheng and Chen, 1999). Adaptations in the visual system of these fishes have been recently described, specifically within the retinal organisation of several Antarctic species (Pointer et al., 2005) and in the eye lens of the Antarctic toothfish *D. mawsoni* (Kiss et al., 2004).

The eye lenses of endothermic vertebrates, as well as ectothermic tropical fishes display a phenomenon known as cold cataract (Delaye et al., 1982; Banh and Sivak, 2004; Kiss et al., 2004) which does not occur within the lenses of the Antarctic toothfish at their normal body temperature of ~2 °C (Kiss et al., 2004). The phenomenon of cold-cataracts has been used to
model cataracts of non-thermal aetiology as well as other related protein aggregation diseases (sickle cell anaemia, Alzheimer’s dementia) in humans (Benedek, 1997). The common denominator in these seemingly unrelated pathologies is a protein instability caused by changes in the micro-environment of the particular proteins, which in the case of cataracts are the lens crystallins. Vertebrate eye lenses are composed of fibre cells particular proteins, which in the case of cataracts are the lens stability caused by changes in the micro-environment of these seemingly unrelated pathologies is a protein in-

Previous work has suggested that a reduced proportion of hydrophobic residues in many proteins could be an adaptation to prevent cold denaturation at temperatures close to or below 0 °C (Privalov, 1990). Although this proposal is very attractive and has been convincing in limited in vitro biophysical studies, it remains to be conclusively demonstrated as a general strategy for cold adaptation of structural (non-enzymatic) proteins in cold stenothermal organisms such as the Antarctic toothfish.

Cold-cataracts in endothermic mammals were well-documented with current opinion that cold-
cation about the cold stability of the toothfish lens. If tropical fishes and endothermic mammals were the principal factors contributing to the cold stability of the toothfish lens. If biochemical differences of the toothfish γ crystallins are in fact responsible for the increased stability, these differences might be encoded at the level of the primary structure. Moreover, as the eye lens is incredibly dense in crystallins (close to 1000 mg mL⁻¹ in fishes (Kroger et al., 1994)) contributions to cold stability undoubtedly arise from interactions between all three (α|β|γ) crystallin molecules. To address the hypothesis that biochemical properties encoded by the toothfish crystallins impart low temperature stability, we isolated, cloned and sequenced the toothfish lens crystallin cDNAs. To frame the evolution of the γ toothfish crystallins, we have analysed them in conjunction with the available sequences from the tropical zebrafish and a number of previously unidentified sequences from the temperate Tetraodon nigroviridis (spotted green pufferfish).

2. Materials and methods

2.1. cDNA library construction

Lenses of the Antarctic toothfish, D. mawsoni (Norman) were collected from live specimens caught in McMurdo Sound, Antarctica, and stored at −80 °C. Total lens RNA was isolated using Ultraspec RNA isolation reagent (Biotecx, TX), and the poly(A)⁺ fraction was isolated from the total RNA using oligo (dT)-cellulose (Collaborative Research, MA) following standard protocols (Sambrook and Russell, 2001). A cDNA library was constructed from poly(A)⁺ RNA using the Universal Riboclon cDNA Synthesis System (Promega, WI) per manufacturer instructions except for the following modifications. In place of the EcoRI adaptors, BamHI/XmnI non-palindromic adaptor formed with complimentary oligonucleotides 5’-d(GATCCGAAGG- GTTCCG)-3’ and 5’-d(pCGAACCCCCCTTCG)-3’ (New England Biolabs, MA) was ligated to the blunt-ended doubled stranded cDNA, and then partially filled-in with dGTP and dATP. The cDNAs were then ligated to the phagemid pBK-CMV (Stratagene, CA) previously digested with Xhol and partially filled-in with dTTP and dCTP, transformed into XL1-BLUE Supercompetent cells as per manufacturer’s protocol (Stratagene, CA).

2.2. Library screening for crystallin cDNA clones

2.2.1. αβ and αβ crystallins

Recombinant clones from the lens cDNA library were screened to identify α, β, and γ crystallin cDNA clones by a combination of PCR-amplification with crystallin primers and Southern analyses of restriction digested plasmid DNA. For α crystallin, clones of the αA isoform were inadvertently identified by PCR-screening of 96 clones using a primer pair that targeted the full length αB coding sequence - alpha3 (5’-ATGGAATTCTATC-CAGATTCCTGG-3’) and alpha3 (5’-TCCACAGATGATAGGGATG-GAGTGCTGGG-3’) based on reported zebrafish (Danio rerio) αB crystallin cDNA sequence (Posner et al., 1999). AlphaB crystallin was isolated using primers DRAB1 (5’-CCTGGAATGATAGGGATG-GAGTGCTGGG-3’) and DRAB1-2 (5’-CAACACGCCTGACAGGATA-3’) specific to sequence sites within the conserved α crystallin domain of zebrafish αB cDNA. An expected amplicon of ~250 bp was obtained from clones and subsequently sequenced to contain full length αB crystallin cDNA clone.

2.2.2. β and γ crystallins

Since β and γ crystallins are members of the same superfamily and they share sequence homology in their core structural regions (Slingsby and Clout, 1999), we were able to identify putative β|γ crystallin cDNA clones by low stringency hybridisation of a Southern blot of Srl & Xhol digested clones from the library using a 32P-labeled 407 bp partial γ crystallin cDNA from D. mawsoni. This 407 nt partial γ crystallin cDNA was obtained by RT-PCR amplification using previously reported primers CPG5 (5’-GAGGACAGAATTTCTATCCAGATTCCTGG-3’) and CPG3 (5’-GCTGCGCCGCGCTGCGTGTTT-3’) designed to the Asian (or common) carp Cyprinus carpio γM2 crystallin (Chang et al., 1988); this partial cDNA was later found to correspond to nucleotide 46–453 of the toothfish γM8e. Initially, 50 randomly selected clones were restriction digested and screened in this low stringency manner. Of the 50 clones, 36 were identified by positive hybridisation which when sequenced gave five full length (3γ and 2β) and 11 partial unique β|γ crystallin cDNAs. Tentative identification of the sequences as β|γ crystallin isoforms came from nucleotide comparisons at GenBank/NClB (BLASTN algorithm). These 11 partial β|γ cDNAs were used as templates to design isoform specific oligonucleotide primers (Tₘ ≥ 65 °C) to obtain full length cDNA sequences by 5’ and 3’ RACE.
Most of the known vertebrate β crystallin cDNA isoform sequences are conserved homologues, thus partial cDNA sequences that were not initially identified in the low stringency Southern blot were obtained by designing isoform specific primers based on an alignment of known mammalian and chicken β isoform sequences. The γ crystallin gene family was screened with full length 32P-labelled cDNAs for αA and αB, and the β crystallin gene family with full length βA2 and βB2. Gamma crystallin gene regions were screened with the same 32P-labelled γM6e probe used in the initial identification (at higher stringency) of the 14 γ crystallin isoforms. Putative positive clones were then re-streaked from archived glycerol stocks to single colonies. Recombinant BAC plasmids were prepared from liquid cultures of single colonies, digested with Ncol and electrophoresed on a pulsed field system (CHEF Mapper XA System, BioRad). Low range PFG (pulsed field gel) markers (NEB) were used to estimate the size of the BAC clone inserts. To confirm the clones were positive for crystallin genes, the Ncol digested BAC DNA was vacuum blotted from the gel onto Hybond–N nylon membrane, and the membrane was hybridised with the same αA, β or γ probes (as appropriate) used in the initial BAC library macro-array screening. The verified positive clones were then digested with HindIII, and separated on a 1% agarose gel (40 cm) along with DNA Marker II for Genomic DNA Analysis (Fermentas, MD), stained with SYBR Green I (Molecular Probes/Invitrogen) and photographed using Kodak 1D Image Analysis System (Eastman Kodak Company, New Haven, CT, USA). The BAC clone fingerprint (HindIII restriction fragment mobilities) images were edited with IMAGE v3.10 (www.sanger.ac.uk/Software/Image/), and imported into FPCv8.2 (Fingerprint Contig; www.agcol.arizona.edu/software/fpc/) for construction of clone order and overlap (Nelson and Soderlund, 2005).

### Table 1

| Dissostichus mawsoni eye lens crystallin cDNA sequences and their GenBank accession numbers |
|---------------------------------|------------------|------------------|
| **Sequence name**              | **cDNA length**  | **Amino acid length** |
| **(bp)**                       |                  | **(predicted)**    |
| αA                             | 1012             | 176               | DQ143965 |
| αB                             | 1452             | 164               | DQ147910 |
| βA1                            | 1301             | 196               | DQ147540 |
| βA2                            | 850              | 196               | DQ143966 |
| βA4                            | 712              | 196               | DQ143967 |
| βB1                            | 777              | 230               | DQ143968 |
| βB2                            | 864              | 208               | DQ143669 |
| βB3                            | 1015             | 253               | DQ143670 |
| γN                             | 752              | 183               | EU016230 |
| γS1                            | 737              | 176               | DQ143671 |
| γS2                            | 891              | 175               | DQ143672 |
| γM1                            | 620              | 178               | DQ143673 |
| γM3                            | 603              | 176               | DQ143675 |
| γM4                            | 613              | 180               | DQ143676 |
| γM5                            | 765              | 176               | DQ143678 |
| γM7                            | 645              | 175               | DQ143679 |
| γM8a                           | 656              | 182               | DQ268581 |
| γM8b                           | 623              | 184               | DQ143681 |
| γM8c                           | 677              | 183               | DQ143682 |
| γM8d                           | 619              | 176               | DQ143683 |
| γM8e                           | 628              | 177               | DQ143674 |
| γM9                            | 672              | 178               | DQ143677 |

**2.3. Genomic Southern Blots of D. mawsoni DNA**

Approximately 20 μg of genomic DNA from *D. mawsoni* liver was digested overnight with 50 units of EcoRI. Digests were resolve on a 1% agarose gel, vacuum blotted to Hybond-N nylon membrane (Amersham Biosciences), UV-crosslinked and hybridised with 32P-labeled probes (see figure legends) in PerfectHyb (Sigma–Aldrich, St Louis, MO, USA). The blot was washed to 55°C in 0.1XSSC/0.5%SDS, and autoradiographed on Kodak Biomax X-ray film.

**2.4. Screening and analyses of BAC clones containing crystallin genes**

A large-insert DNA bacterial artificial chromosome (BAC) library was constructed for *D. mawsoni* using red blood cells based on previously published methods (Miyake and Amemiya, 2004). A total of 67,584 recombinant clones, equivalent to about 6X genome (2C) coverage were robotically picked and archived in 384-well master culture plates, as well as printed on nylon hybridisation membranes as a macro-array for screening target genes. The α crystallin gene family was screened with full length 32P-labelled cDNAs for αA and αB, and the β crystallin gene family with full length βA2 and βB2. Gamma crystallin gene regions were screened with the same 32P-labelled γM6e probe used in the initial identification (at higher stringency) of the 14 γ crystallin isoforms. Putative positive clones were then re-streaked from archived glycerol stocks to single colonies. Recombinant BAC plasmids were prepared from liquid cultures of single colonies, digested with Ncol and electrophoresed on a pulsed field system (CHEF Mapper XA System, BioRad). Low range PFG (pulsed field gel) markers (NEB) were used to estimate the size of the BAC clone inserts. To confirm the clones were positive for crystallin genes, the Ncol digested BAC DNA was vacuum blotted from the gel onto Hybond–N nylon membrane, and the membrane was hybridised with the same αA, β or γ probes (as appropriate) used in the initial BAC library macro-array screening. The verified positive clones were then digested with HindIII, and separated on a 1% agarose gel (40 cm) along with DNA Marker II for Genomic DNA Analysis (Fermentas, MD), stained with SYBR Green I (Molecular Probes/Invitrogen) and photographed using Kodak 1D Image Analysis System (Eastman Kodak Company, New Haven, CT, USA). The BAC clone fingerprint (HindIII restriction fragment mobilities) images were edited with IMAGE v3.10 (www.sanger.ac.uk/Software/Image/), and imported into FPCv8.2 (Fingerprint Contig; www.agcol.arizona.edu/software/fpc/) for construction of clone order and overlap (Nelson and Soderlund, 2005).
2.5. Nomenclature

Nomenclature of the α and β crystallins from *D. mawsoni* was based on homology with vertebrate crystallins. Gamma crystallin designations for toothfish were based on the previous nomenclature “γM” of zebrafish and carp as much as possible (Chang et al., 1988; Wistow et al., 2005). Further designations were based on BLASTN (GenBank/NCBI), our phylogenetic reconstruction, identity matrices and manual inspection of aligned amino acid sequences.

2.6. Bioinformatics

DNA sequences were edited for quality and to remove vector and other non-cDNA portions. Amino acid translations of the crystallin cDNAs were aligned using ClustalW in BioEdit v7.0.5.2 (Hall, 1999), and manually checked for alignment quality. Amino acid identity matrices (IDENTIFY, BLOSUM62, GONNET) were generated using BioEdit v7.0.5.2.

For phylogenetic analyses, γ crystallins from *D. mawsoni* and other ectothermic species (see Table 1 Suppl. Tables 1 and 2 for GenBank accession nos.) were first aligned by the translated amino acid sequences. Nucleotide sequences were aligned based on this amino acid alignment using CodonAlign2.0 (Hall, 2004). All phylogenetic reconstructions were performed using DNA sequence alignments unless indicated otherwise. Tree construction was done by MrBayes3.1.1 using the evolutionary model GTR+I+Γ (Waddell and Stell, 1997) as selected by Mr. Modeltest v2.0 (Nylander, 2004). Four runs using four chains and 4,000,000 generations were performed. All other parameters in MrBayes 3.1.1 were left at default. Stationarity was assessed using the “sump burnin=2500” command and examining the.

---

Fig. 1. Alignment of the *in silico* translation of cDNAs of the Antarctic toothfish (*Dm*) αA (A) and αB (B) crystallins with zebrafish (Dr), Asian catfish (Cf, Cb), cow (Bt) and human (Hs). (A) The identity between the αA homologues is very high (black background) in the typically more divergent N-terminal domain, as well as the conserved α crystallin domain. (B) Although the α crystallin domain of the αB sequences is also highly conserved, the more divergent N-terminal domain and the C-terminal extension are more variable than in the αA crystallin. The domain region designations are based on previous reports (Caspers et al., 1995; Narberhaus 2002). Grey regions indicate ≥ 50% similarity between sequences.
plot of the generation versus the log likelihood values. A 50% majority rule consensus tree was generated using the command
“sumpt burnin=2500”.

3. Results

3.1. \(\alpha\) Crystallins

The \emph{in silico} translation of the \(\alpha\)A and \(\alpha\)B lens crystallin cDNA sequences from the Antarctic toothfish \textit{D. mawsoni} are shown in Fig. 1 (GenBank Accession nos. in Table 1). Toothfish sequences were aligned with zebrasfish \textit{D. rerio} (Dr) (Posner et al., 1999; Runkle et al., 2002), catfish \textit{Clarias fuscus} (Cf) and catfish \textit{Clarias batrachus} (Cb) (Yu et al., 2004), with cow \textit{Bos taurus} (Bt) and human \textit{Homo sapiens} (Hs) sequences (GenBank accession nos: NP_776714, NP_000385, NP_776715, AAP36581). Designations of the N-terminal region/domain, \(\alpha\) crystallin domain and C-terminal extensions are based on previously published reports (de Jong et al., 1993; Caspers et al., 1995; Narberhaus, 2002).

The \(\alpha\)A and \(\alpha\)B alignments (Fig. 1) highlight the conservation of the \(\alpha\) crystallin domain between these five fish and mammalian taxa. The amino acid identities between the \(\alpha\)A sequences range from a low of 68.7% between toothfish and human, to a high of 94.2% between cow and human (Table 2). Among the fish species analysed, \(\alpha\)B sequences are more divergent from each other than are the \(\alpha\)A sequences (Table 2). The greatest variation is in the N-terminal region of the \(\alpha\) crystallin (Fig. 1B). In contrast, the N-terminal region of the \(\alpha\)A crystallin is highly conserved between fish and mammals, thus accounting for the greater sequence similarity of \(\alpha\)A between these phylogenetically distant species (Fig. 1A).

Southern blot of \textit{EcoRI} digested \textit{D. mawsoni} genomic DNA probed with \(\beta\)-P labeled full length \textit{D. mawsoni} cDNAs for \(\alpha\)A and \(\alpha\)B resulted in hybridisation to three distinct bands (3, 1, 2) in decreasing intensity (Fig. 2). The presence of three bands is consistent with the presence of three \(\alpha\) isoforms in zebrasfish and catfish (Yu et al., 2004; Smith et al., 2006). It is tempting to suggest that the intensity of the three bands in the Southern blot is representative of the gene copy number, isoform type and possibly expression levels. However, attempts to isolate a complete second \(\alpha\)B cDNA sequence were unsuccessful. A partial second \(\alpha\)B sequence was isolated giving an \emph{in silico} translation tentatively designated \(\alpha\)B, “TKDGVEITGKHEDRKDEHG-FVSRSTFRKYTLPSNTDVKEVNSL”, corresponding to position 86 to 130 of the full length \textit{D. mawsoni} \(\alpha\)B sequence (Fig. 1) which spans the conserved ‘\(\alpha\) crystallin’ domain.

3.2. \(\beta\) Crystallins

The \(\beta\) crystallin protein sequences from the toothfish are presented in Fig. 3 and aligned with the zebrasfish, cow and human sequences. The alignment highlights the highly conserved four domains (coloured underscores) which form the Greek-key fold motifs. Within the \(\beta\) crystallins, the \(\beta\)A3 and \(\beta\)A1 differ by their initiation codons; the translational start of the \(\beta\)A3 transcript is upstream of the \(\beta\)A1 (McDermott et al., 1997; Bloemendal et al., 2004). The six \(\beta\) crystallin isoforms isolated from toothfish are clear homologues of the \(\beta\) crystallin family. They are unambiguously identified when contrasted with each other (Table 3) and compared to mammalian orthologues (Fig. 3).

---

**Table 2**

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Dm (\alpha)A</th>
<th>Dr (\alpha)A</th>
<th>Cf (\alpha)A</th>
<th>Bt (\alpha)A</th>
<th>Hs (\alpha)A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dm (\alpha)A</td>
<td>100</td>
<td>77.2</td>
<td>72.7</td>
<td>71.0</td>
<td>68.7</td>
</tr>
<tr>
<td>Dr (\alpha)A</td>
<td>100</td>
<td>81.5</td>
<td>73.2</td>
<td>72.1</td>
<td></td>
</tr>
<tr>
<td>Cf (\alpha)A</td>
<td>100</td>
<td>71.0</td>
<td>70.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bt (\alpha)A</td>
<td>100</td>
<td>94.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hs (\alpha)A</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Dm (\alpha)B</th>
<th>Dr (\alpha)B</th>
<th>Cf (\alpha)B</th>
<th>Bt (\alpha)B</th>
<th>Hs (\alpha)B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dm (\alpha)B</td>
<td>100</td>
<td>47.0</td>
<td>51.4</td>
<td>48.2</td>
<td>47.7</td>
</tr>
<tr>
<td>Dr (\alpha)B</td>
<td>100</td>
<td>56.5</td>
<td>60.5</td>
<td>59.4</td>
<td></td>
</tr>
<tr>
<td>Cf (\alpha)B</td>
<td>100</td>
<td>62.5</td>
<td>61.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bt (\alpha)B</td>
<td>100</td>
<td>97.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hs (\alpha)B</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The toothfish \( \beta A1, \beta A2, \beta A4 \) sequences show a high degree of sequence identity (100% identity – black background) or sequence similarity (\( \geq 50\% \) similarity – grey background) to both the zebrafish and mammalian (cow and human) \( \beta \) crystallins (Fig. 3). The \( \beta B1, \beta B2, \) and \( \beta B3 \) toothfish sequences again show conservation within their domains (coloured underscores) to zebrafish, cow and human (Fig. 3). The proline–arginine rich repeat, designated a ‘PAPA-arm’ (Bloemendal et al., 1984; Hejtmancik et al., 1986; Coop et al., 1998) that is found N-terminal to the first domain of the \( \beta B1 \) mammalian crystallin, is similar to the PAPA domain found in the C-terminal domain of the toothfish \( \beta B3 \) crystallin. An analogous PNPN domain is also found in the C-terminal domain of zebrafish \( \beta B3 \) (Fig. 3).

---

**Fig. 3.** Alignment of the *in silico* translation from cDNAs of the Antarctic toothfish (Dm) \( \beta \) crystallins against zebrafish (Dr), cow (Bt) and human (Hs). Immediately apparent is the extent of the identity of residues across the phylogenetically diverse taxa. Within the four conserved domains regions (alternating blue and green bars) there is high levels of identity across all of the \( \beta \) crystallins (black background) as well as substantial similarity (\( \geq 50\% \) grey background). It is likely that only the \( \beta A1 \) is present for the toothfish (Dm, \( \beta A1 \)) as an upstream start codon (making an \( \beta A3 \) isoform) was not found in the cloned cDNA. The N-terminal regions of the toothfish \( \beta A1, \beta A2, \beta A4, \beta B2 \) and \( \beta B3 \) crystallins are highly conserved across all four taxa. Toothfish \( \beta B1 \) crystallin shows a high degree of conservation with zebrafish of the N-terminal linker element. Note the presence of a large poly-(PA) element in the C-terminal region of the toothfish \( \beta B3 \) crystallin, similar to the poly-(PN) element in the zebrafish \( \beta B3 \) crystallin.
3.3. γ Crystallins

A total of 14 distinct γ crystallin cDNAs from the lens of the Antarctic toothfish (Dm) were isolated, cloned and sequenced (Table 1). In silico translations of the γ cDNAs and their alignment with three γ crystallins from the Asian (common) carp C. carpio (Cc) (Chiou et al., 1986) and 16 γ crystallins from zebrafish D. rerio (Dr) (Wistow et al., 2005), 14 ‘unamed protein products’ (which we propose are in fact γ; see below and Fig. 4) crystallins from spotted green pufferfish T. nigroviridis, as well as several other γ crystallins from lipshark, blind Mexican cavefish, African clawed frog, Japanese firebelly newt, iguana and sea squirt are shown (Fig. 4). There is a high degree of sequence identity among the various toothfish isoforms (40.7% to 92.3%; Table 4) as well as high conservation between ectothermic species (black background in Fig. 4) and similarities (≥50% grey background Fig. 4). Of the fourteen Antarctic γ crystallin sequences, two (γS1 and γS2), are likely paralogues of the γS isoform and one was clearly an orthologue of a γN isoform (Pan et al., 1997; Bloemendal et al., 2004; Wistow et al., 2005). Based on alignment (Fig. 4), percent identities (Table 4) and phylogenetic analysis (Fig. 5), the remaining eleven Antarctic γ crystallin cDNA sequences were designated (γM) originally proposed for methionine rich γ crystallins from carp (Chang et al., 1988), but now adopted for other fishes as well as other aquatic animals such as some frogs (Lu et al., 1996). In naming the toothfish γ crystallins we used the orthologous name as assigned previously for carp and zebrafish (γM1, γM3, γM4, γM5, γM7) where possible. In cases where there were not orthologues of toothfish γ crystallins, we continued to use the same γM crystallin nomenclature, but used the next available “γM#”.

Five related γ crystallins from the toothfish, four of which all share a conserved nine amino acid N-terminal extension “MS(N/T)T(G/D)M(N/S)M(R/-)” and did not have clear homologues in zebrafish or carp, were assigned as γM8a-e (Figs. 4 and 5). These four γM8 toothfish crystallins (γM8a-d) consistently formed a distinct clade in phylogenetic trees, regardless of the method of tree construction (Fig. 5). In addition to the toothfish specific γM8 isoforms there are clearly identified γM1 crystallins with the three residue conserved “MMF” insertion between positions 102 and 104, which is also present in γM1 of zebrafish, carp, pufferfish and lipshark (Ccx_γ1) (Fig. 4). No γMX sequence was identified from our toothfish lens cDNA library. However, given the resolution of our phylogram, we place the previously reported zebrafish γMX isoform (Wistow et al., 2005) very close to the γS group indicating that this might be a γS ‘caught in the act’ of evolving into a more γM-like isoform (Fig. 5).

A phylogram of ectothermic and largely obligate aquatic γ crystallins is presented (Fig. 5) as a Bayesian 50% majority rule consensus tree. The evolutionary model used was the Generalised Time Reversible plus Invariant plus Gamma (GTR + I + Γ) model of substitution as chosen by MrModelTest (Nylander 2004). We used the coding regions of the nucleotide sequences

Table 3
Amino acid identities toothfish β crystallins (percent identical) based upon pairwise alignment against each other

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>βA1</th>
<th>βA2</th>
<th>βA4</th>
<th>βB1</th>
<th>βB2</th>
<th>βB3</th>
</tr>
</thead>
<tbody>
<tr>
<td>βA1</td>
<td>100.0</td>
<td>59.1</td>
<td>67.8</td>
<td>32.0</td>
<td>36.2</td>
<td>26.6</td>
</tr>
<tr>
<td>βA2</td>
<td>100.0</td>
<td>54.0</td>
<td>35.8</td>
<td>38.5</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>βA4</td>
<td>100.0</td>
<td>32.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>βB1</td>
<td>100.0</td>
<td>48.2</td>
<td>40.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>βB2</td>
<td>100.0</td>
<td>42.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>βB3</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of the cDNAs (sans 5'-UTRs, 3'-UTRs and stop codons) aligning the nucleotide sequences based on the amino acid alignments (Suppl. Fig. 1). The outgroup choice for this tree was urochordate sea squirt *Ciona intestinalis* because it represents an βγ sequence from a basal chordate and an aquatic marine ectotherm. (Franck et al., 2004; Shimeld et al., 2005; Riyahi and Shimeld, 2007).

In the construction of the phylogram we included the toothfish cDNAs cloned and sequences from this study as well as the recently reported zebrafish γ crystallins (Wistow et al., 2005).

**Fig. 4.** Amino acid alignment of the 14 distinct γ crystallins from the Antarctic toothfish (Dm) with 51 other γ crystallin sequences from ectothermic animals. The four major Greek–Key fold motifs are indicated in alternating blue/green bars at the bottom of the figure. Boundaries between the domains are in red (≠). The numbering (+1) of the crystallins begins where the first glycine (G) residue appears in the majority of sequences. The toothfish appears to have unique class of γM8 crystallins which are absent in the other species. Identity (100%) between each species’ γ crystallins is indicated by a black background, whereas similarity (≥ 50%) is indicated by grey background. Species are designated by a two (or three in the case of lipshark, i.e. *Chiloscyllium colax* = Ccx_γ1) letter abbreviation composed of *Genus* species name and one of the five different isoform classes (γ+/+, M, S, N, MX).
In addition, we included a number of other fishes and ectotherms, specifically; *C. fuscus*, whitespotted clarias; *Chiloscyllium colax*, lipshark; *Astyanax mexicanus*, blind Mexican cavefish; *C. carpio*, Asian (common) carp, *Xenopus laevis*, African clawed frog; *Cynops pyrrhogaster*, Japanese firebelly newt and *Iguana iguana*, iguana (Suppl. Table 2). Furthermore, there are 15 previously ‘unidentified proteins’ from *T. nigroviridis* (spotted green pufferfish) used in this analysis which were deposited into the GenBank/NCBI database by Genoscope (France) and the Whitehead Institute (Massachusetts Institute of Technology) during the course of the *Tetraodon* genome project (www.genoscope.cns.fr/externe/English/Projets/Projet_C/organisme_C.html). We have identified these 15 sequences as γ crystallins and thus propose isoforms names
The γ crystallins can be divided into three major isoform groups (γN, γS, and γM which includes amphibian γ) largely based on the phylogram (Fig. 5). The γN isoforms form a distinct clade (green, Fig. 5) indicating their ancestral position and high conservation across species. Most of the γS isoforms (blue, Fig. 5) are basal to the γM isoforms (purple, Fig. 5) with exception of the γS1 and γS2 from C. fuscus whitespotted claris. The γ crystallins included from the amphibians (X. laevis and C. pyrrhogaster) form a clade nested within the γM crystallins of the fishes (red, Fig. 5). From the Bayesian tree...
(Fig. 5) the phylogenetic distance (expected changes per site) between the γ crystallin cDNA coding sequences can be inferred from the branch lengths in the tree (Fig. 5), thus it appears that the amphibian γ crystallins are more closely related to the γS isoforms than to some of the more derived γM fish crystallins. The remaining γM crystallins show clustering based on their relatedness to the previously named carp and zebrafish sequences. A Southern blot of EcoRI digested toothfish genomic DNA probed with a conserved centre section (407 nt of γ toothfish crystallin) revealed 17 bands (Fig. 6). The number of distinct bands in the Southern blot is within the magnitude of number of γ cDNAs isolated and sequenced.

### 3.4. Genomic organisation

A BAC library macro-array was screened with 32P-labeled toothfish α (αA and αB), β (βA2 and βB2) and γ (γM8e) crystallin probes. Seventeen BAC α clones were initially identified from the macro-array and further re-analysed by HindIII digest Southern blot (Suppl. Fig. 2). Thirteen of these clones (a1, a3 ~ a11, a13 ~ a15, a17) were found to be overlapping and form one contig 46 consensus band CB units long (Fig. 7A). Each CB unit was estimated to be ~5800 bp long based on the mean restriction fragment sizes in the HindIII DNA fingerprint analysis (Suppl. Figs. 2, 3 and 4), therefore 46 CB units would cover a region of 266.8 kbp (Castellarin et al., 2006). Clones a12, a16 did not overlap with the single contig. Clone a2 was also not in the contig and was not analysed further as it had only two bands at ~8000 and 5700 bp in the HindIII DNA fingerprint and very weak Southern Blot hybridisation (Suppl. Fig. 2). Southern blot analysis of the other 16 α clones indicated strong hybridisation for all the clones except weak hybridisation for clone a16. Hybridisation patterns were similar between the clones, except for clones a5 and a12 which also had different HindIII DNA fingerprint patterns (Suppl. Fig. 2). Only clone a12 showed both a different fingerprint/hybridisation pattern than that of the other α BAC clones.

Six β crystallin clones identified in the initial macro-array screen had extremely similar HindIII fingerprints and almost identical hybridisation patterns by Southern blot analysis (Suppl. Fig. 3). These data were confirmed in FPC analysis which placed all the β BAC clones within the same 47 CB (272.6 kbp) contig (Fig. 7B).

There were twenty seven positive clones found for γ crystallins in the BAC macro-array screen. Based on our FPC analysis the γ crystallin BAC clones divide up into three groups. The first group forms a contig of 31 CB units (179.8 kbp) and includes clones g1, g3–5, and g12 (Fig. 7C and Suppl. Fig. 4) and have similar Southern Blot hybridisation (Suppl. Fig. 4). The second group forms a separate contig of 51 CB units (295.8 kbp) and includes clones g6–11, g13–20 and g22 (Fig. 7D and Suppl. Fig. 4). All of these clones within this second contig group also have a similar Southern Blot hybridisation, which differs from the first γ contig group. Two clones that did not contig with either of the two previous groups (g2 and g21, Suppl. Fig. 4) also shared similar Southern Blot hybridisation patterns to each other. These two clones failed to contig even after re-analysis with adjusted Tolerance and Cut-Off parameters in the FPC analysis program (Soderlund et al., 2000).

### 4. Discussion

The Antarctic toothfish D. mawsoni lives in the perennially sub-zero seawater (~2°C) of the Southern Ocean at the lower thermal limit of marine vertebrate ectotherms. At this subzero temperature, the toothfish has a completely transparent lens that is similar to other mammalian vertebrates (Kiss et al., 2004). To investigate the possible molecular basis of the cold stable toothfish lens, we have obtained the crystallin sequences by cDNA cloning and sequencing.

There were two complete α crystallin (αA and αB) cDNAs isolated from toothfish lens. Toothfish αA crystallin show high sequence identity to αA from both fishes and mammals in contrast to αB sequences which are more divergent (Table 2). The homology of the toothfish αA amino acid sequence with other vertebrates suggests that αA crystallin is under greater functional constraint than the αB isoform, therefore maybe acting as the primary sHSP within the eye lens as has been previously hypothesized (Bova et al., 1997; Rajaraman et al., 2004).
hybridised to the toothfish DNA. There were 17 positive bands.

implementing a GTR+I+ evolutionary model. Each major isoform class in the tree is colour coded (red for βB crystallins). While it is clear that within recombinant experiments this is the case, it remains unclear what are the native function(s) of N-terminal and C-terminal extensions in an intact lens.

The 14 distinct γ crystallin isoforms found in the Antarctic toothfish lens fall into three isoform groups. We found a single γN isoform, which based on gene structure is believed to be an evolutionary bridge between the β and γ isoforms (Wistow et al., 2005). Two more cDNAs belonged to the γM isoform (Bloemendal et al., 2004). The N-terminal PAPA-arm of βB1 (Fig. 3) of cow and human lenses has been shown to be strongly associated with the membrane (insoluble) component of the toothfish lens during isolation of the lens proteins (Kiss, 2005). Although the N-terminal extension in the toothfish βB1 does not have a PAPA-arm, there is a long PAPA-arm in the C-terminal of the toothfish βB3 crystallin. Interestingly, this toothfish C-terminal PAPA-arm appears to have a counterpart C-terminal ‘PNPN-arm’ in zebrafish βB3 crystallin (Fig. 3).

Although the hydrophobic PAPA-arm of bovine βB1 is thought to insert itself into the lipid membrane, the substitution of asparagine (N) for alanine (A) in the zebrafish PNPN-arm would make the PNPN-arm very polar and thus its function could be quite different than what is hypothesized for the mammalian and possibly toothfish PAPA-arms.

Recent structural studies of recombinant mammalian β crystallins (Bateman et al., 2001; Bateman et al., 2003; Van Montfort et al., 2003) have suggested that linker regions as well as the extensions (N- and C-terminal) are involved in oligomerisation of the β crystallins. While it is clear that within recombinant experiments this is the case, it remains unclear what are the native function(s) of N-terminal and C-terminal extensions in an intact lens.

The 14 distinct γ crystallin isoforms found in the Antarctic toothfish lens fall into three isoform groups. We found a single γN isoform, which based on gene structure is believed to be an evolutionary bridge between the β and γ isoforms (Wistow et al., 2005). Two more cDNAs belonged to the γS (formerly βS (Björk, 1961)) group and the remaining eleven were γM isoforms (Chang et al., 1988; Wistow et al., 2005) so named originally for their high methionine content. The very high methionine content in the γM crystallins of the toothfish and other fishes is a curious adaptation (Chang et al., 1988), and so far remains unexplained. Unlike other non-polar residues, methionine has special properties which may in fact contribute significantly to the stability of fish lenses, both in terms of their.

Beta crystallin protein sequences from toothfish lens are aligned with zebrafish, cow and human (Fig. 3). They display a high degree of conservation both for each β isoform and between each species analysed. We did not isolate a βA3 isoform, which in mammals has a longer N-terminal extension that normally ‘incorporates’ the βA1 isoform by means of an earlier, upstream translational start codon (Bloemendal et al., 2004).

There are three basic β crystallin isoforms (βB1, βB2, βB3) in toothfish, as in mammals. The ‘A or ‘B’ designation after the β- referring to either an acidic [(βA#) or basic (βB#) class of the β crystallin isoform (Bloemendal et al., 2004). The N-terminal PAPA-arm of βB1 (Fig. 3) of cow and human lenses has been shown to be strongly associated with the membrane (aquous insoluble) portion of the lens (Bloemendal et al., 1984; Hejtmancik et al., 1986; Coop et al., 1998; Bateman et al., 2001). We did observe a significant proportion of β crystallins associate with the membrane (insoluble) component of the toothfish lens during isolation of the lens proteins (Kiss, 2005).

The converse, that αB likely has a diversity of non-lenticular roles (Bennardini et al., 1992; Platigorsky 1998; Dahlman et al., 2005) is also supported by our toothfish αB crystallin sequence data. Interestingly, we had some difficulty obtaining αB DNA from the toothfish lens, which may reflect the low expression levels of the αB message (mRNA). Proteomic analysis of αA and αB crystallins from toothfish as well as detailed chaperone-like assays of recombinant toothfish αA and αB crystallins is in progress and will be published elsewhere.

2001; Dahlman et al., 2005). The converse, that αB likely has a diversity of non-lenticular roles (Bennardini et al., 1992; Platigorsky 1998; Dahlman et al., 2005) is also supported by our toothfish αB crystallin sequence data. Interestingly, we had some difficulty obtaining αB DNA from the toothfish lens, which may reflect the low expression levels of the αB message (mRNA). Proteomic analysis of αA and αB crystallins from toothfish as well as detailed chaperone-like assays of recombinant toothfish αA and αB crystallins is in progress and will be published elsewhere.

Fig. 5. Phylogram of ectothermic γ crystallins from the Antarctic toothfish, zebrafish, pufferfish and select other fishes, amphibians and iguana. Tree was constructed from the aligned nucleotides of the coding region from the cDNAs (Suppl. Fig. 1). The 50% majority-rule consensus tree was generated by the program Mrbayes, implementing a GTR+I+ evolutionary model. Each major isoform class in the tree is colour coded (γN= green, γS= blue, amphibian γ = red, γM= purple). Only the γN isoform is monophyletic. The cross-species conserved γS isoforms are polyphyletic in this analysis. The amphibian γ crystallins (red) form a clade nested basally within the γM isoform group, close to the γS isoform class indicating that they may be intermediate between the ancestral γS and the fish γM isoforms.

Fig. 6. Genomic Southern blot of D. mawsoni EcoRI digested DNA. DNA was probed with 407 nucleotide 32P-labeled PCR amplicon from the coding region of γM8e (D. mawsoni). Labeled arrows to the left of the blot indicate bands that hybridised to the toothfish DNA. There were 17 positive bands.
high density and in terms of their ability to readily adapt to a range of temperatures. Methionine has both a structural plasticity as well as the ability to have its hydrophobicity altered by reversible oxidation of its thioether (Gellman, 1991). In the context of the extremely dense fish lens where protein concentrations are upwards of 1000 mg mL\(^{-1}\) (Kroger et al., 1994), the multiple isoforms of γM crystallins would undoubtedly contribute to the protein stability by increasing the crystallin polydispersity, thus preventing catastrophic crystallization at such high concentrations. Furthermore, γM crystallins would impart ‘flexibility’, or sponginess to the γM crystallin surface which would further discourage crystal lattice formation. Additionally, the abundance of methionines may have biochemically predisposed the γM crystallins to cold adaptation by allowing the fish lens to reversibly oxidize their thioethers via enzymatic means (Marchetti et al., 2005; Sagher et al., 2006) thereby affording a mechanism to attenuate the hydrophobicity of these γ crystallins. Consequently, a primary adaptation of toothfish lenses by increasing the protein stability at high density by generating polydisperse γM crystallin isoforms, could have leant itself to a secondary adaptation of cold stability.

Phylogenetic analysis of the γ crystallin isoforms was done using a Bayesian approach employing a GTR+I+Γ model of molecular evolution (Waddell and Stell, 1997). This model of evolution is particularly well-suited for ancient protein coding sequences as it acknowledges that (i) there has been reversion, (ii) that there are invariant sites and (iii) that not all sites are under the same selective pressures. Furthermore, using the nucleotides that code for the amino acids aligned as codon units may facilitate greater resolution depth in the tree (Simmons et al., 2002). Phylogenetic reconstruction of the toothfish γ
crystallins in comparison with other ectothermic species shows that the γN isoforms form a monophyletic clade, whereas conserved γS crystallins are polyphyletic. The placement of the recently discovered γN class of crystallins ancestrally to the γS clade suggests that they pre-date the evolution of γS, as has been suggested from comparative analysis of their gene structure in D. rerio (Wistow et al., 2005). Based on phylogenetic analysis, there is a range of γS isoforms with a decreasing gradient of similarity from γN to γM. The two C. fuscus γS isoforms (CF-γS1 and CF-γS2, Figs. 4 and 5) could very well be misnamed as they seem to be well-situated with the γM group (specifically γM2) in both our alignment and Bayesian phylogenetic analysis. The multiplicity of the toothfish γM crystallin isoforms are emphasized in the phylogram, many of which are distinct from the zebrafish and pufferfish (Fig. 5). In our previous study (Kiss et al., 2004), we proposed that the γ crystallin component of the toothfish lens as the most likely candidate for the transparency of the toothfish lens at −2°C. However, while there are several γM crystallins isoforms unique to the toothfish, comparative hydrophobicity plots did not suggest major differences from other fish crystallins. It is possible that a few select amino acid changes, in addition to post-translational modifications of the γ crystallins could have profound implications on the cold stability of the lens. Regions between the domains of the γ crystallins termed ‘linker’ regions, and N-terminal extensions (such as the long methionine rich ones in the γM class) may also have significant influence on cold stability by affecting the solution dynamics (Wu et al., 2005). To fully address these issues, extensive biochemical analysis and proteomics experiments would likely be informative. An outline of the genomic organisation of the α, β and γ crystallin genes was obtained by screening a BAC library and analyses of the DNA fingerprints by FPC. Using this approach we found that α and β crystallin separately formed a single α or β contig each whereas the γ genes formed two contigs (Fig. 7). The two γ contigs differ both in HindIII fingerprint pattern as well as Southern blot hybridisation patterns. The second γ contig appears to have a Southern blot hybridisation pattern similar to that of the β contig Southern blot (Suppl. Fig. 3). This similarity is not unexpected as both β and γ crystallins share conserved domains and are members of the same gene superfamily (Bloemendal et al., 2004; Wistow et al., 2005). The probe used to hybridize to the γ gene regions did have variable hybridisation strength to some BAC clones, of which most are represented in the second contig. Consequently, these results suggest that some β and γ genes are in spatial proximity, or alternatively that the second contig of γ BAC clones could very well be β crystallins. Two γ clones (g2 and g21) do not form a contiguous group with either of the γ gene contigs even though their Southern blot hybridisation pattern is similar to the β and second γ gene contig groups. Within mouse, rat and humans, the γ crystallin genes are clustered as γA–F with γS (p53) a short distance apart, interspersed with highly repetitive sequences (Willard et al., 1985; den Dunnen et al., 1987; Skow et al., 1988; den Dunnen et al., 1989). Based on annotation in NEIBank, zebrafish γS isoforms are present on two different chromosomes (16 and 9) (Wistow, 2002). Without more information at the DNA sequence level, it is difficult to state with certainty how the βγ gene organisation in the Antarctic toothfish is arranged. Current and proposed BAC projects argue persuasively for the sequencing of the Antarctic toothfish which will enable questions regarding γ crystallin evolution to be addressed more fully (NRC, 2003; Clark et al., 2004; Cheng et al., 2007).

Comparison of toothfish αβγ crystallin sequences with other vertebrates indicates that α and β isoforms are well-conserved both in number and isoform type. Our current work illustrates the homology of α and β toothfish crystallin sequences to other phylogenetically distant vertebrate species (Bloemendal et al., 1984; Hejtmanick et al., 1986; Behrens et al., 1998; Chen et al., 2001; Runkle et al., 2002) reiterating the likely conservation of function of these two types of crystallins. Conversely, toothfish γ crystallins exist as multiple polydisperse isoforms, some of which appear to be paralogues and some unique. Given this nature of the toothfish γ crystallins, it is attractive to suggest that they are more evolutionarily plastic without the functional constraints of the αβ isoforms. Thus, the combination of an increased methionine content allowing for attenuation of protein hydrophobicity, along with increased polydispersity of γM isoforms in the dense eye lens may have predisposed the toothfish lens for biochemical adaptations to the extreme cold.

Acknowledgments

The authors wish to thank Prof. Mason Posner (Dept. Biology, Ashland University, Ashland, Ohio, USA) for suggestions on redesigning of the toothfish αB isoform specific primers (DRAB-1 and DRAB1-2), as well as helpful discussions during the preparation of this manuscript. In addition the authors thank the anonymous reviewers for their helpful comments and critiques. This work was supported by NSF grant OPP 02-31006 to C-H.C.C. and A.L.D.

Appendix A. Supplementary data


References


