Cloning and sequencing of complete *t*-crystallin cDNA from embryonic lens of *Crocodylus palustris*

RAMAN AGRAWAL, REENA CHANDRASHEKHAR, ANURAG KUMAR MISHRA, JETTY RAMADEVI, YOGENDRA SHARMA and RAMESH K AGGARWAL*

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India

**Corresponding author (Fax, +91-40-7160591; Email, rameshka@gene.ccmbindia.org)*

t-Crystallin is a taxon-specific structural protein found in eye lenses. We present here the cloning and sequencing of complete *t*-crystallin cDNA from the embryonic lens of *Crocodylus palustris* and establish it to be identical to the *a*-enolase gene from non-lenticular tissues. Quantitatively, the *t*-crystallin was found to be the least abundant crystallin of the crocodilian embryonic lenses. Crocodile *t*-crystallin cDNA was isolated by RT-PCR using primers designed from the only other reported sequence from duck and completed by 5′- and 3′-rapid amplification of cDNA ends (RACE) using crocodile gene specific primers designed in the study. The complete *t*-crystallin cDNA of crocodile comprises 1305 bp long ORF and 92 and 409 bp long untranslated 5′- and 3′-ends respectively. Further, it was found to be identical to its putative counterpart enzyme *a*-enolase, from brain, heart and gonad, suggesting both to be the product of the same gene. The study thus provides the first report on cDNA sequence of *t*-crystallin from a reptilian species and also re-confirms it to be an example of the phenomenon of gene sharing as was demonstrated earlier in the case of peking duck. Moreover, the gene lineage reconstruction analysis helps our understanding of the evolution of crocodilians and avian species.

[Agrawal R, Chandrashekhar R, Mishra A K, Ramadevi J, Sharma Y and Aggarwal R K 2002 Cloning and sequencing of complete *t*-crystallin cDNA from embryonic lens of *Crocodylus palustris*; *J. Biosci*. **27** 251–259]

1. Introduction

The lens crystallins of vertebrates comprise a complex group of soluble and conserved structural proteins, which are classified as *a*-, *b*- and *g*-crystallins of mammalian lenses, and some species-specific crystallins (Wistow and Piatigorsky 1988; Bloemendal and deJong 1991). These crystallins not only vary between species, but are also differentially expressed during lens development (Kraft *et al* 1994). Hence, it is imperative to compare crystallins of various species from different phylogenetic levels in order to understand the mechanism underlying the process of crystallin diversification and their evolutionary significance in the differentiation of eye lenses. A number of taxon-specific crystallins have been characterized from lenses of many species, which were found to be similar to some of the housekeeping enzymes in other cell-types that had probably been recruited to perform the specialized structural role in the lens (Wistow and Piatigorsky 1987). *t*-Crystallin is one such taxon-restricted crystallin, which was first identified in the lenses of lamprey and turtles (Staple and deJong 1983; Williams *et al* 1985) and subsequently in duck. Despite its identification in many species, *t*-crystallin is one of the least

Keywords. *a*-enolase; crocodile; gene sharing; lens; *t*-crystallin

Abbreviations used: GAM, Genital ridge-adrenal-metanephrosis complex; ORF, open reading frame; RACE, rapid amplification of cDNA ends; UTR, untranslated region.

GenBank accession numbers of the nucleotide sequences reported in the paper are: AF428105, AF428106, AF428107, AF428108.

studied crystallin with the only detailed characterization having been reported in the duck lens (Wistow and Piatigorsky 1987; Kim *et al* 1991). The latter study suggests it to be similar to *α*-enolase, the metabolic enzyme of the glycolytic pathway, but with a greatly reduced activity indicating its recruitment for a dual role; as an enzyme in other tissues, and a sensitive structural protein in lens. Incidentally, the only report on cloning and sequencing of the cDNA of *τ*-crystallin is also from the duck lens and none from the reptilian species wherein it is seen to be of wider occurrence based on protein studies.

Here, we provide the first report on cloning and sequencing of the complete *τ*-crystallin cDNA from embryonic eye lenses of a reptilian species, the Indian crocodile *Crocodylus palustris*. The study also shows that the *τ*-crystallin cDNA is identical to the *α*-enolase cDNA cloned from various other non-lenticular tissues of crocodilian embryo and thus both should be the product of the same gene. These results, in turn, demonstrate that the phenomenon of tissue-dependent recruitment of the same gene product for different functional roles is of ancient origin, as crocodilians are believed to be the longest living vertebrates.

2. Materials and methods

2.1 *Crocodilian lenses and tissues*

Fertilized eggs of Indian mugger (*Crocodylus palustris*) were collected from the Nehru Zoological Park, Hyderabad, on day zero of egg laying and were incubated in the laboratory at 30–32°C. The proper guidelines to handle the animals were strictly followed. Crocodilian lenses, heart, brain and genital ridge-adrenal-metanephrosis (GAM) tissues were taken from the developing embryos, sacrificed between the developmental stages 21–25 (Ferguson 1987) and stored frozen at – 70 °C until use.

2.2 *Isolation of lens crystallins*

The 10–15 lenses collected from crocodile embryos were homogenized in 0·05 M Tris, pH 7·5, 0·05 M NaCl, 0·02% sodium azide, 1 mM EDTA, centrifuged and the supernatant collected. The supernatant (2 ml) was fractionated on a Bio-Gel A-1·5 m column properly calibrated by standard molecular weight markers to calculate the native molecular mass. The fractions were analysed by SDS-PAGE to identify different types of crystallin and estimate their relative abundance in the lens proteins.

2.3 *RNA isolation and RT-PCR*

Total RNA was extracted from embryonic lens, brain, heart and GAM using the TRIzol reagent kit and cDNA synthesized using Superscript II RNaseH Reverse Transcriptase kit (Life Technologies, USA). Primers for *τ*-crystallin were designed from the duck *τ*-crystallin gene sequence (table 1). The cDNA synthesized from the total RNA

Table 1. Primers used for cloning of the full-length *τ*-crystallin cDNA from crocodilian lens and *α*-enolase from other tissues.

Primer code	Primer sequence (5′ → 3′)	Nucleotide positions	Reference
Duck specific primers for isolation of crocodile homologue of *τ* cDNA			
d-*τ*-1 (forward)	GGTGTTCAAGATGTCCATTCTC	18–39 (Acc. No. M20749)	Wistow *et al* 1988
d-*τ*-2 (reverse)	GCAGCTTAGTTGATACGG	1321–1338 (Acc. No. M20749)	Wistow *et al* 1988
For 3′-RACE			
Q_T	CCAGTGAGCAGAGTGACGAGGACTC-GAGCTCAAGCTTTTTTTTTTTTTTTT		Frohman *et al* 1988
Q_o	CCAGTGAGCAGAGTGACG		Frohman *et al* 1988
Q_I	GAGGACTCGAGCTCAAGC		Frohman *et al* 1988
C-GSP$_1$	GTGGTGCCTCAACTGGAATC	202–221 (Acc. No. AF428105)	Present study
C-GSP$_2$	GTTCACTGCCTGTGTGGACA	1010–1029 (Acc. No. AF428105)	Present study
For 5′-RACE			
C-GSP$_3$	AGCAGCACCAGCTTTGCACA	356–376 (Acc. No. AF428105)	Present study
C-GSP$_4$	CAGACGCTCAGATCTACAAG	1285–1304 (Acc. No. AF428105)	Present study
For *α*-enolase from non-lenticular embryonic tissues			
C-GSP$_5$	GGTATTCAAAATATGTCAGTTCTC	83–104 (Acc. No. AF428105)	Present study
C-GSP$_6$	TTATCTCGAGAACAGCAGCTGCC	1411–1431 (Acc. No. AF428105)	Present study

preparation from embryonic crocodile lenses as described above, was used for PCR reaction on a MJ Research Gradient Thermal Cycler under standard PCR conditions. The annealing temperature gradient of 45°–60°C was employed. The product was checked on a 1·5% agarose gel, cut out, eluted and subjected to DNA sequencing using the same set of primers. The sequence (1320 bp) was analysed by BLAST software program of NCBI (Altschul *et al* 1990) and was used to design gene-specific primers (GSPs, table 1) to obtain the full-length sequence of *t*-crystallin.

2.4 *3′-end and 5′-end RACE*

The 3′-UTR was obtained by RACE PCR using GSP_1 and Q_o primers (table 1) and cDNA template synthesized using QT primer (Frohman *et al* 1988); cloned into pMOS Blue vector (Pharmacia-Amersham) and then sequenced using GSP_2 primer. Similarly, the 5′-UTR was cloned and sequenced by 5′-RACE using the kit (version 2) available from Gibco BRL and GSPs (table 1). Briefly, the cDNA was synthesized using GSP_4 and after tailing reaction, the first and nested PCR was done using abridged and tailed primers. The product obtained was directly sequenced using GSP_3 primer on automated DNA sequencer.

2.5 *a-Enolase cDNA from non-lenticular tissues*

In order to address the question of gene sharing between *t*-crystallin and *a*-enolase, the GSPs were designed from the completed crocodile *t*-crystallin cDNA sequence (table 1) and were used for amplification and sequencing of related transcripts, if any, from the non-lenticular embryonic tissues namely, brain, heart and GAM tissues. In each case, cDNA made by RT-PCR from total RNA isolated from the respective tissue was used as template.

2.6 *Sequence analysis and primer design*

The completed cDNA sequences were analysed for open reading frame using NCBI-ORF finder and homology searches were done using BLAST sequence analysis software. The alignment of proteins as well as nucleotide sequence was done using the program Multialign version 5.4.1 (Corpet 1988). Gene and species specific primers were designed using the software Primer-3 (Rozen and Skaletsky 1998).

2.7 *Evolutionary status*

To understand the evolutionary history of the gene lineage, related reference sequences were retrieved from NCBI viz. *t*-crystallin of duck (Acc. No. M20749), *a*-enolase sequence of alligator (Acc. No. AF072586), trachemys (Acc. No. AF072588), chicken (Acc. No. D37900), sceloporus (Acc. No. AF072587), python (Acc. No. AF072589), rat (Acc. No. NM_012554), mouse (Acc. No. AK002336), bovine (Acc. No. AF149256) and human (Acc. No. NM_001428). The sequences were used to construct a phylogenetic tree depicting gene lineage, using the clustering program DNAML. The reliability of the constructed phenogram was tested by the Bootstrap analysis. Both the DNAML and Bootstrap analysis were done using the software package PHYLIP version 3·6 (Felsenstein 1994).

3. Results and discussion

3.1 *Characterization of lens crystallins*

The Bio-Gel A-1·5 m column chromatogram of soluble lens homogenate of crocodile embryonic lens revealed five distinct and well-resolved peaks (figure 1a). The latter when individually separated on the SDS-PAGE (figure 1b) showed the presence of: (i) *a*H, *a*L, and *d*-crystallin, as the major crystallins of the crocodilian lenses in 1st to 3rd peaks respectively, (ii) *e*-crystallin of 37 kDa (Hendriks *et al* 1988; Chiou *et al* 1989) in the descending portion of the 3rd peak, (iii) *t*-crystallin identified in the fractions from the first half of the 4th peak based on its hydrodynamic properties i.e. molecular weight both under native as well as denaturing conditions, and (iv) *b*L and *i*-crystallins in the second half of 4th and 5th peaks respectively. Crocodilian *t*-crystallin appeared as a monomeric protein having a molecular weight of about 48 kDa as seen on SDS-PAGE (figure 1b, lane 4). This crystallin is present only to the extent of 3–4% of the total lens proteins and is the least abundant of all the crystallins found in the crocodilian lens. This is in contrast to the observation in turtles where it is among the major crystallins (Williams *et al* 1985).

3.2 *Cloning and sequencing of t-crystallin gene*

Since the sequence of *t*-crystallin gene from related species such as turtle or alligator is not known, primers were designed from the duck *t*-crystallin sequence (table 1). Figure 2 (lane 5) shows the RT-PCR product analysed on agarose gel, which correlated well with the expected size of 1320 bp as per the duck *t*-crystallin sequence. This product was gel eluted, purified and sequenced. Sequence analysis of the cloned product revealed a continuous ORF confirming the candidate clone to be coding for *t*-crystallin, but was incomplete as it did not have start and stop codons. Full-length sequence of *t*-crystallin gene

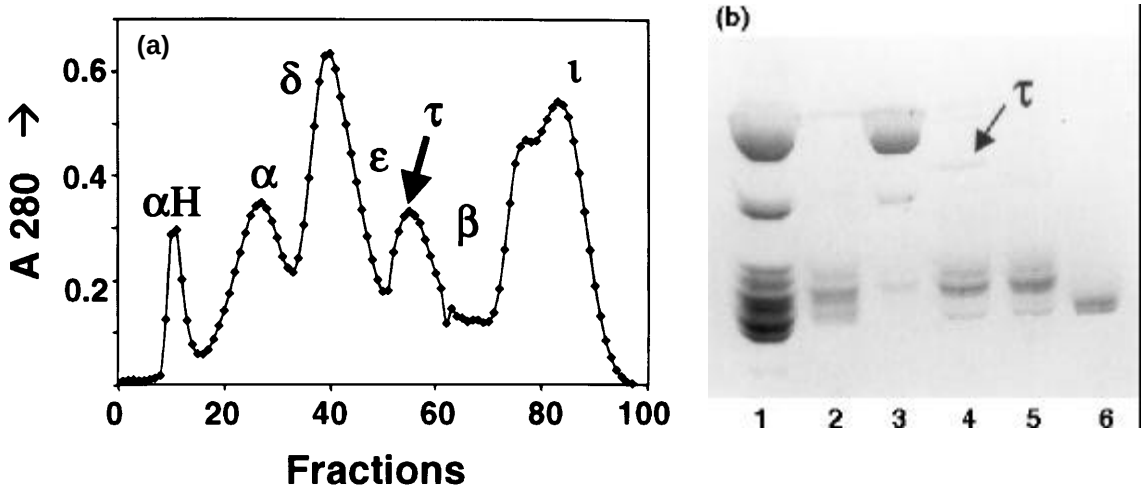

Figure 1. Protein profiles of the crocodile lens. (**a**) Bio-Gel A-1·5 m profile of the total lens homogenate of crocodilian embryos, in the 50 mM Tris buffer, 100 mM NaCl, 1 mM EDTA, 0·02% sodium azide. The various crystallins eluted in the peaks are marked. (**b**) SDS-PAGE profiles of all crystallins. Lanes: (1) total lens homogenate; (2) *a*H + *a*-crystallin (peaks 1, 2); (3) *d*-crystallin and *e*-crystallin (peak 3); (4) *t*-crystallin of about 47 kDa, marked with an arrow, and *b*L-crystallin (peak 4); (5) descending fractions of peak 4 containing *b*L-crystallin; (6) *i*-crystallin (peak 5).

could be obtained by performing 5′- and 3′-end RACE with GSP primers designed from the partial crocodile *t*-crystallin cDNA sequence. The complete 1806 bp sequence of the crocodile *t*-crystallin cDNA is shown in figure 3, which also includes both 5′- and 3′-UTR regions (Acc. No. AF428105). The open reading frame translates into a putative polypeptide of 434 amino acid residues comparable to the putative *a*-enolase and *t*-crystallin proteins of alligator (Mannen and Li 1999) and duck respectively, having a calculated molecular mass of 47510·18 Dalton, a net charge of – 3 and theoretical isoelectric point of 6·24.

3.3 *Cloning and sequencing of crocodile a-enolase gene*

Attempts to amplify the related cDNAs of *a*-enolase from three non-lenticular tissues i.e. heart, brain and GAM of crocodile embryo using GSPs designed from the completed crocodile *t*-crystallin cDNA (table 1) produced the expected size PCR product (figure 2, lanes 2–4). The sequencing of the amplified cDNAs (Acc. Nos. AF428106-108) after cloning in pMOS vector, revealed these to be identical to the *t*-crystallin sequence obtained from the lens tissue (Acc. No. AF428105). These results thus suggest that the gene coding for *t*-crystallin in eye lens and *a*-enolase in non-lenticular tissues must be one and the same as would be expected if the phenomenon of gene sharing exists in crocodiles similar to what was shown earlier in detailed studies in the peking duck by Wistow *et al* (1988).

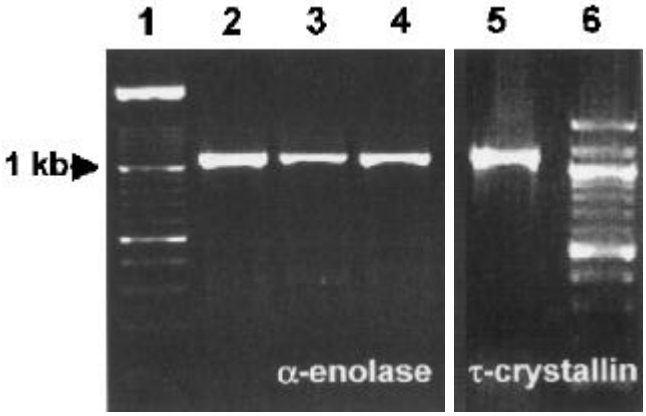

Figure 2. RT-PCR products for *t*-crystallin from lens, and *a*-enolase from other tissues. Lanes 1 and 6, Molecular weight standards; lanes 2, 3 and 4, *a*-enolase from brain, heart and GAM, respectively; lane 5, *t*-crystallin from lens. Arrow shows the expected 1320 bp PCR product in all the lanes.

3.4 *Taxon- and species-specific variations in t-crystallin*

The homology analysis of the *t*-crystallin/*a*-enolase sequence of crocodile lens with the existing sequences in gene databank revealed the alligator *a*-enolase (Mannen and Li 1999) as the closest relative having a very high level of sequence identity both at the nucleotide (only 2·9% mismatches in the coding region, table 2) and protein level (2·53% mismatches corresponding to 11 out of 434 deduced amino acid residues, figure 4, table 2). In comparison, the crocodile *t*-crystallin sequence differed

much more significantly from that of duck *t*-crystallin both at the level of nucleotide sequence and protein residues. Overall, the two showed sequence dissimilarity of 21% for the whole available sequence and 12·2% for the coding region, differences being even more in the 5′-/3′-UTR regions (table 2). Similarly, the variation at the protein level was 5·52% which was almost double that of alligator *a*-enolase (table 2).

The 3′-UTR of crocodile *t*-crystallin that is 409 bp long was found to be quite similar (92·2%) to the corresponding region of *a*-enolase of the alligator (table 2). There is, however, one prominent variation. The alligator

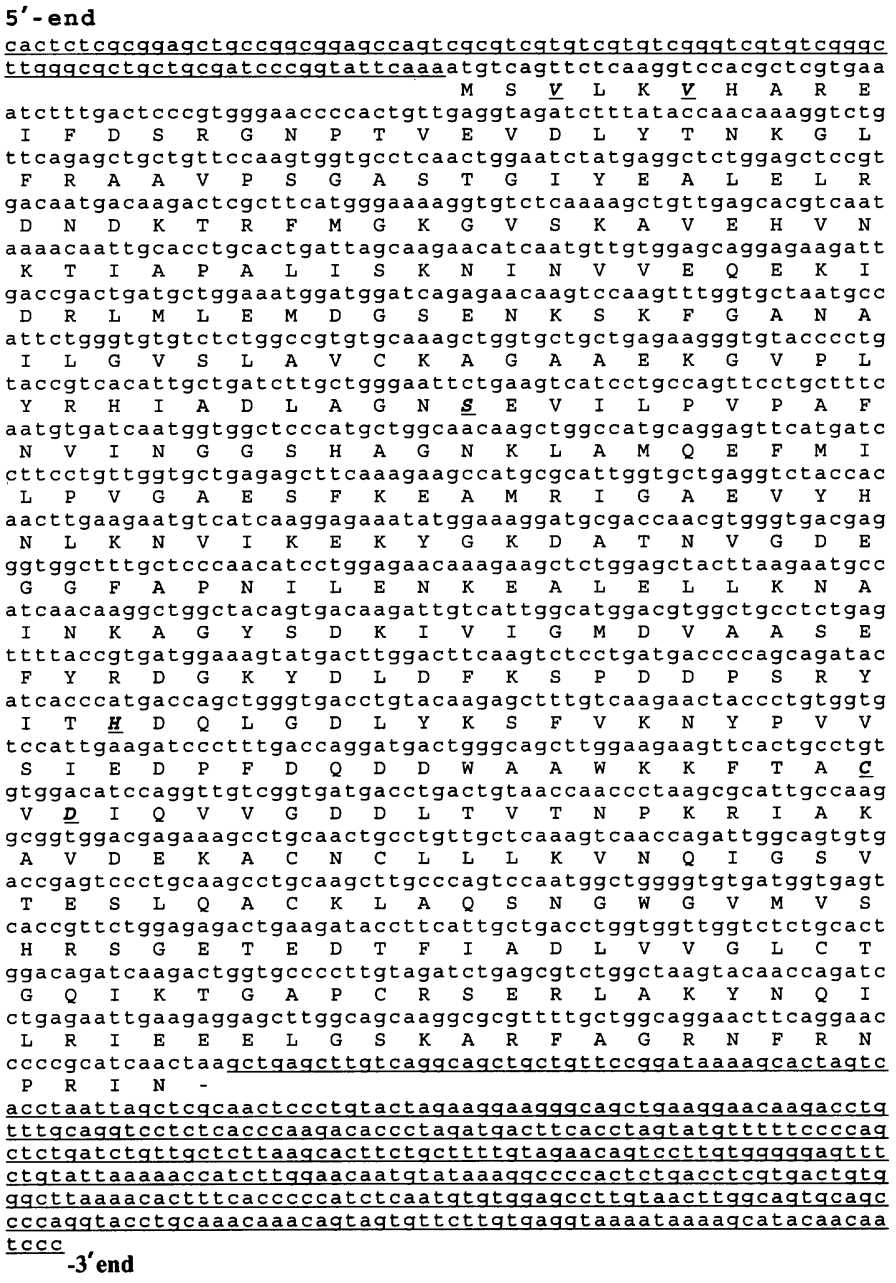

Figure 3. 1806 bp long complete sequence of *t*-crystallin cDNA of crocodile. The nucleotides in the 5′-UTR and 3′-UTR regions are underlined. The deduced protein sequence is indicated below the codons of the ORF. The sequence codes for a putative protein of 434 amino acid residues with expected molecular mass of 47·5 kDa. The amino acid residues unique to crocodilian lens protein compared to alligator and duck are shown in bold, italicized underlined format.

a-enolase has three copies of the pentameric repeat motif 'GTTTT' while in case of crocodile, there is only one such repeat motif at 153 bp downstream of the termination codon (data not shown). It was interesting to note that the duck, an outgroup species to both crocodile and alligator, also has only one such pentameric repeat motif in the 3′-UTR of its *t*-crystallin cDNA, which otherwise differs at the nucleotide level from that of crocodile. These results would imply that the ancestral form of the *t*-crystallin/*a*-enolase gene had only one such repeat motif and that the two additional repeat motifs seen in alligator *a*-enolase gene has probably been selectively acquired sometime later during the course of evolution of the gene after its separation from the common ancestral form.

The 92 bp long 5′-UTR of crocodile *t*-crystallin, unlike the 3′-UTR region, does not seem to be well conserved except for the first 20 bp stretch immediately upstream of the initiation codon ATG with that of *a*-enolase from alligator and *t*-crystallin of duck. This apparent absence of conservation is probably because the available 5′-UTR sequences of alligator and duck are only 48 and 27 bp long and are not complete.

A comparison of the 434 amino acid residues long putative proteins of *t*-crystallin/*a*-enolase from crocodile, duck and alligator, show that they are fairly conserved (figure 4) and have nearly identical molecular weights. On the whole only 29 amino acid residues were found to be variable in the protein sequences of crocodile, duck and alligator. Of the variable residues, seven are unique to crocodile *t*-crystallin, four to alligator *a*-enolase and eighteen to duck *t*-crystallin (figure 4). Many of these replacements are associated with the loss of hydrophobic residues, P141S, P273H, A277G in crocodile, E67A in alligator and L171P, V173C, A302G, A309G in duck. It was interesting to note that out of the total 29 variable amino acid residue positions, there was not a single position at which all the three proteins differed from one another, indicating that the observed variations were not random in crocodile/alligator/duck. This apparent specificity would suggest specific adaptive changes as a result of selective selection pressure during the course of evolution and are thus expected to be phylogenetically informative.

3.5 *Evolutionary status*

Reconstruction of gene lineage using crocodile *t*-crystallin sequence and *t*-crystallin/*a*-enolase sequences of many reference vertebrate species confirms it to be a form of *a*-enolase that is ancient in origin mainly restricted to the archosaurial lineage (figure 5). The sequence information and gene lineage reconstruction presented here also help our understanding of the evolution of crocodilians and avian species. The data on significant taxon-specific differences in the crocodile *t*-crystallin and alligator *a*-enolase cDNA sequences, viz., 8 transversions in the coding region, 8 non-synonymous replacements of amino acid residues and presence of 2 additional pentameric repeats in the 3′-UTR region, suggest early separation and independent evolution of the *Crocodylus* and *Alligator* species for a very long time.

Table 2. Summary of the variations in the *t*-crystallin/*a*-enolase cDNA of crocodile compared to *t*-crystallin cDNA of peking duck and *a*-enolase of alligator.

Parameter	Crocodile (*t*-cDNA/ *a*-enolase cDNA)	Peking duck (*t*-cDNA Acc. No. M20749)	Alligator (*a*-enolase cDNA Acc. No. AF072586)
cDNA			
Total size (bp)	1806	1715	1772
5′-untranslated region	Complete 92 bp	Incomplete, 27 bp	Incomplete, 48 bp
bp changes (% dissimilarity)*		77 (83·5)	54 (58·7)
3′-untranslated region	409	383	419
bp changes (% dissimilarity)*		144 (34·7)	33 (7·8)
ORF	1305	1305	1305
bp changes (% dissimilarity)*		154 (12·2)	38 (2·9)
Transitions (bp)		109	30
Transversions (bp)		45	8
Putative protein			
Size (a.a.)	434	434	434
a.a. changes (% dissimilarity)*		25 (5·76)	11 (2·53)
Synonymous changes		14	4
Non-synonymous changes		11	7

*Compared to the corresponding *t*-cDNA sequence domain of the crocodile.

This observation conforms to our earlier findings on evolution of crocodilians that clearly suggested crocodile and alligators representing two different lineages separated very early during their evolution (Aggarwal *et al* 1994). Similarly, in the gene lineage tree there is a bifurcation point just after Sceloporus distinctly separating all mammalian species from the rest, suggesting that the birds are the closest relatives of crocodiles (figure 5). This observation also lends support to the debated phylogenetic status of avian to be archosaurial (Janke and Arnason 1997; Tracy and Hedges 2000) rather then them being closer to mammals as proposed by haemothermia hypothesis (Lovtrup 1985).

The crocodile *t*-crystallin cDNA codes for a putative 47·5 kDa protein, which exists as monomer under normal gel filtration conditions. On the other hand, the enzymatically active form of the protein *a*-enolase has been shown to be dimeric (Wistow *et al* 1988), suggesting that the tissue-specific role would have compelled it to exist as monomer with concomitant loss of enzymatic activity

```
          1 C  C                                A                              50
Crocodile MSVLKVHARE IFDSRGNPTV EVDLYTNKGL FRAAVPSGAS TGIYEALELR
Alligator MSILKIHARE IFDSRGNPTV EVDLYTSKGL FRAAVPSGAS TGIYEALELR
     Duck MSILKIHARE IFDSRGNPTV EVDLYTNKGL FRAAVPSGAS TGIYEALELR
Consensus MS!LK!HARE IFDSRGNPTV EVDLYTnKGL FRAAVPSGAS TGIYEALELR

          51     D           A D                  DA     D    D   D 100
Crocodile DNDKTRFMGK GVSKAVEHVN KTIAPALISK NINVVEQEKI DRLMLEMDGS
Alligator DNDKTRFMGK GVSKAVAHVN KTIAPALISK NISVVEQEKI DRLMLEMDGS
     Duck DNDKTRYMGK GVSKAVEHIN KTIAPALISK NVNVVEQDKI DKLMLDMDGS
Consensus DNDKTR%MGK GVSKAVeH!N KTIAPALISK N!nVVEQ#KI DrLML#MDGS

          101                                                  C      150
Crocodile ENKSKFGANA ILGVSLAVCK AGAAEKGVPL YRHIADLAGN SEVILPVPAF
Alligator ENKSKFGANA ILGVSLAVCK AGAAEKGVPL YRHIADLAGN PEVILPVPAF
     Duck ENKSKFGANA ILGVSLAVCK AGAAEKGVPL YRHIADLAGN PEVILPVPAF
Consensus ENKSKFGANA ILGVSLAVCK AGAAEKGVPL YRHIADLAGN pEVILPVPAF

          151                    D D  D                          200
Crocodile NVINGGSHAG NKLAMQEFMI LPVGAESFKE AMRIGAEVYH NLKNVIKEKY
Alligator NVINGGSHAG NKLAMQEFMI LPVGAESFKE AMRIGAEVYH NLKNVIKEKY
     Duck NVINGGSHAG NKLAMQEFMI PPCGADSFKE AMRIGAEVYH NLKNVIKEKY
Consensus NVINGGSHAG NKLAMQEFMI lPVGA#SFKE AMRIGAEVYH NLKNVIKEKY

          201                              D    D        D          250
Crocodile GKDATNVGDE GGFAPNILEN KEALELLKNA INKAGYSDKI VIGMDVAASE
Alligator GKDATNVGDE GGFAPNILEN KEALELLKNA INKAGYSDKI VIGMDVAASE
     Duck GKDATNVGDE GGFAPNILEN KEALELLKTA IGKAGYSDKV VIGMDVAASE
Consensus GKDATNVGDE GGFAPNILEN KEALELLKnA InKAGYSDK! VIGMDVAASE

          251                           DC    C       D               300
Crocodile FYRDGKYDLD FKSPDDPSRY ITHDQLGDLY KSFVKNYPVV SIEDPFDQDD
Alligator FYRDGKYDLD FKSPDDPSRY ITPDQLADLY KSFVKNYPVV SIEDPFDQDD
     Duck FYRDGKYDLD FKSPDDPSRY ISPDQLADLY KGFVKNYPVV SIEDPFDQDD
Consensus FYRDGKYDLD FKSPDDPSRY ItpDQLaDLY KsFVKNYPVV SIEDPFDQDD

           D       DC  C                          DA               350
Crocodile WAAWKKFTAC VDIQVVGDDL TVTNPKRIAK AVDEKACNCL LLKVNQIGSV
Alligator WAAWKKFTAS VGIQVVGDDL TVTNPKRIAK AVDDKACNCL LLKVNQIGSV
     Duck WGAWKKFTGS VGIQVVGDDL TVTNPKRIAK AVEEKACNCL LLKVNQIGSV
Consensus WaAWKKFTas VGIQVVGDDL TVTNPKRIAK AV##KACNCL LLKVNQIGSV

          351                                                        400
Crocodile TESLQACKLA QSNGWGVMVS HRSGETEDTF IADLVVGLCT GQIKTGAPCR
Alligator TESLQACKLA QSNGWGVMVS HRSGETEDTF IADLVVGLCT GQIKTGAPCR
     Duck TESLQACKLA QSNGWGVMVS HRSGETEDTF IADLVVGLCT GQIKTGAPCR
Consensus TESLQACKLA QSNGWGVMVS HRSGETEDTF IADLVVGLCT GQIKTGAPCR

          401       D                          434
Crocodile SERLAKYNQI LRIEEELGSK ARFAGRNFRN PRIN
Alligator SERLAKYNQI LRIEEELGSK ARFAGRNFRN PRIN
     Duck SERLAKYNQL LRIEEELGSK ARFAGRNFRN PRIN
Consensus SERLAKYNQi LRIEEELGSK ARFAGRNFRN PRIN
```

Figure 4. Alignment of the deduced protein sequence of crocodile *t*-crystallin, alligator *a*-enolase and duck *t*-crystallin. Note high degree of identity between the three proteins with only 29 putative amino acid replacements that are boxed. The C, A or D letters on the boxed residue indicate it to be unique for crocodile, alligator or duck proteins, respectively.

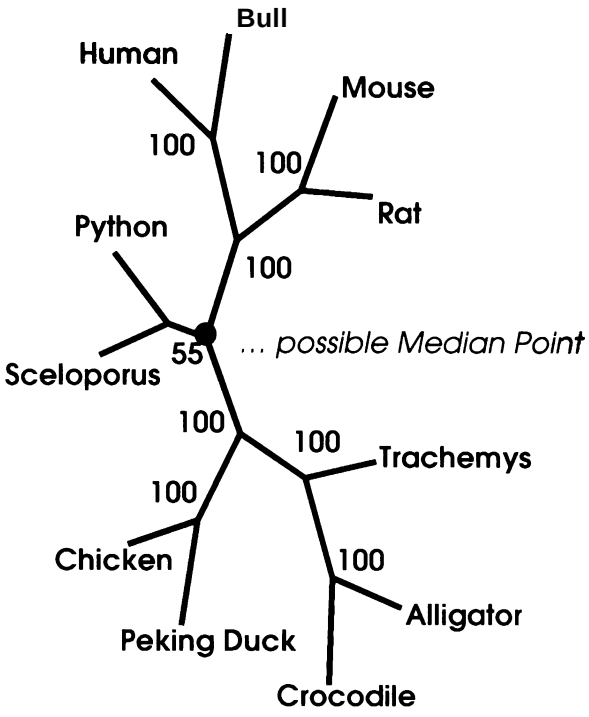

Figure 5. DNAML phenogram showing gene tree lineage for *t*-crystallin/*a*-enolase genes of crocodile and other reference vertebrates. The numbers at nodes denote the Bootstrap values, which are generally high indicating reliability of the clustering.

in lenticular tissue. It would be interesting to study the temporal and structural nature of the elusive cellular factor(s) that are responsible for this cell type specific structural/functional dimorphism of *t*-crystallin/*a*-enolase. We have now initiated studies in this direction using the recombinant *t*-crystallin protein (Mishra *et al* 2002). Such information will help us in understanding the mechanism of bi-functionality and the possible nature of the causative adaptive evolutionary pressure responsible for the phenomenon. Any effort in this direction would call for isolation and expression analysis of *t*-crystallin from different ladders of vertebrate evolution. We believe that the sequence information presented here will be useful in the further cloning of *t*-crystallins from other species.

It is intriguing to note that *t*-crystallin has been observed more in water-dwelling reptiles (turtles, alligators, crocodiles) and birds (duck), and not in other animals of the same taxa, which have a terrestrial habitat (such as in chicken). This interesting conjecture makes a strong ground to look at its possible role, if any, in water adaptability of the lenticular tissue. *t*-Crystallin is not expressed in mammalian lenses, indicating that *t*- and other taxon-specific crystallins were lost in mammals not because of the lack of their adaptability in higher lenses but due to the lack of any strong selective pressure to retain them. It would be interesting to see if water-dwelling mammals have *t*-crystallin expressed in their lenses.

Acknowledgements

We thank the Central Zoo Authority of India for permission and K N Banerji, Nehru Zoological Park, Hyderabad for providing the crocodile eggs. We also thank Dr Lalji Singh for all the support.

References

Aggarwal R K, Majumdar K, Lang J W and Singh L 1994 Generic affinities among crocodilians: as revealed by DNA fingerprinting using a Bkm-derived probe; *Proc. Natl. Acad. Sci. USA* **91** 10601–10605

Altschul S F, Gish W, Miller W, Myers E W and Lipman D J 1990 Basic local alignment search tool; *J. Mol. Biol.* **215** 403–410

Bloemendal H and deJong W W 1991 Lens proteins and their genes; *Prog. Nucleic Acid Res. Mol. Biol.* **41** 259–280

Chiou S H, Chang W P and Chen C C 1989 Characterization and comparison of epsilon-crystallin and lactate dehydrogenases in the lenses of vertebrates and invertebrates; *Biochem. Int.* **18** 1093–1100

Corpet F 1988 Multiple sequence alignment with hierarchical clustering; *Nucleic Acids Res.* **16** 10881–10890

Felsenstein J 1994 *PHYLIP* (*Phylogeny Inference Package*). Version 3.6. Department of Genetics, University of Washington, Seattle, USA

Ferguson M W J 1987 Post-laying stages of embryonic development for crocodilians; in *Wildlife management: Crocodiles and alligators 1987* (eds) G J W Webb, S C Manolis and P J Whitehead (Chipping Norton; Surrey Beatty) pp 427–444

Frohman M A, Dush M K and Martin G R 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer; *Proc. Natl. Acad. Sci. USA* **85** 8998–9002

Hendriks W J, Mulders W M, Bibby M A, Slingsby C, Bloemendal H and deJong W W 1988 Duck lens *e*-crystallin and lactate dehydrogenase B4 are identical: a single gene product with two distinct functions; *Proc. Natl. Acad. Sci. USA* **85** 7114–7118

Janke A and Arnason U 1997 The complete genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles); *Mol. Biol. Evol.* **14** 1266–1272

Kim R Y, Lietman T, Piatigorsky J and Wistow G J 1991 Structure and expression of the duck alpha-enolase/tau-crystallin-encoding gene; *Gene* **103** 193–200

Kraft H J, Voorter C E, Wintjes L, Lubsen N H and Schoenmakers J G 1994 The developmental expression of taxon-specific crystallins in the duck lens; *Exp. Eye Res.* **58** 389–395

Lovtrup S 1985 On the classification of the taxon tetrapoda; *Syst. Zool.* **34** 463–470

Mannen H and Li S S 1999 Molecular evidence for a clade of turtles; *Mol. Phylogenet. Evol.* **13** 144–148

Mishra A K, Chandrashekhar R, Aggarwal R K and Sharma Y

2002 Crocodilian *t*-crystallin: overexpression, purification and characterization; *Protein Expres. Purif.* (in press)

Rozen S and Skaletsky H J 1998 Primer3. Code available at *http://www.genome.wi.mit.edu/genome_software/other/primer3.html*

Staple S O and deJong W W 1983 Lamprey 48 kDa lens protein represents a novel class of crystallins; *FEBS Lett.* **162** 305–309

Tracy M R and Hedges S B 2000 Evolutionary history of the enolase gene family; *Gene* **259** 129–138

Williams L A, Ding L, Horwitz J and Piatigorsky J 1985 *t*-Crystallin from the turtle lens: purification and partial characterization; *Exp. Eye Res.* **40** 741–749

Wistow G J and Piatigorsky J 1987 Recruitment of enzymes as lens structural proteins; *Science* **236** 1554–1556

Wistow G J and Piatigorsky J 1988 Lens crystallins: evolution and expression of proteins for a highly specialized tissue; *Annu. Rev. Biochem.* **57** 479–504

Wistow G J, Lietman T, Williams L A, Stapel S O, deJong W W, Horwitz J and Piatigorsky J 1988 *t*-crystallin/*a*-enolase: one gene encodes both an enzyme and a lens structural protein; *J. Cell Biol.* **107** 2729–2736

MS received 22 March 2002; accepted 5 April 2002

Corresponding editor: DURGADAS P KASBEKAR