An Insight into the Structural Analysis of α-crystallin of Habitat-specific fish – A Computational Approach

Aparajita Chakraborty  
St Xavier's College

Sayak Ganguli  
St Xavier's College

Priyanka De  
St Xavier's College

Sudipa Saha (sudipa@sxccal.edu)  
St Xavier's College

Research Article

Keywords: α-crystallin, habitat-specific fish, computational analysis, structure, sequence, phylogenetic relationship, homology modelling

Posted Date: April 3rd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2751121/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

α-crystallin is a major eye lens protein, comprising up to 40% of total lens protein. It is composed of two subunits, αA and αB and share a common central domain of about 90 residues with variable N- and C-terminal extensions. For the establishment of an evolutionary inter-relationship, an elucidation of the structure and alignment of protein sequences is crucial. In the present study, a bioinformatics approach was adopted to explore the possible structure, sequence and phylogenetic diversity of α-crystallin (both subunits αA and αB) from ten habitat-specific fish species, (freshwater and saltwater) and compared with a standard sequence of Bos taurus species. The sequence of Bos taurus was predicted to be a close homologue of the fish species. Homology modelling has been performed in order to generate a possible ‘3D’ structure of the crystallin proteins using SWISS-MODEL. Our analysis shows that the secondary structures of bovine α-crystallin revealed no considerable differences as compared to that of the crystallins of the habitat-specific fish and that the presence of β-sheets was predominant in all structures. Though no significant differences in the αA subunits were revealed yet some structural variations were observed for αB subunits which had been confirmed by MSA analysis. The 3D structure of the protein hasn’t been elucidated yet so a computational analysis estimated no major differences in structures of crystallin for either bovine or the fish species except that saltwater fish proteins possess more favourable states and higher reliabilities. In agreement with previous literature, α-crystallin has a molecular weight of 20kDa approx. and a theoretical pI of 5.75; 55.1% of its sequence is composed of hydrophilic amino acids and it is a cytosolic protein. Considerable improvement of the currently available tools is being done for a detailed understanding of the structure/function relationships of α-crystallin proteins.

I. Introduction

α-crystallin, a major protein of the vertebrate eye lens of most species is a member of the small heat-shock protein family and possesses chaperone-like function. They are composed of two subunits, αA and αB; αA is expressed mostly in the eye lens whereas αB is expressed in other tissues, including heart, skeletal muscle, kidney, lung and brain (Horwitz, 2003, Augustyn, 2004, Horwitz, 2009).

Crystallin proteins comprise over 90% of the dry weight of human eye lens and exhibit a higher refractive index than average proteins. They exist at very high concentration in the eye lens, up to 400 mg/ml in humans or even higher in some aquatic species (over 1000 mg/ml) (Slingsby et al., 2013, Sprague-Mercy et al., 2021). The vertebrate lens cell organisation is particularly important for lens transparency and focusing but most of the refractive power of the lens is conferred by high concentrations of proteins, the highly abundant protein being the crystallins. Nonchordates, on the other hand use quite different proteins inspite of possessing superficially different lenses. The most widespread crystallins belong to three classes, namely α, β and γ (Basha et al., 2012). The β- and γ-crystallins are not related to α-crystallins but are the members of another protein superfamily which has restricted phylogenetic and tissue distribution.

In humans the genes (CRYAA and CRYAB) encode proteins αA- and αB- crystallin (HSPB4 and HSPB5) which are 53% identical, which coassemble in the lens to form α-crystallin of molecular weight around 800 kDa. CRYAB is widely expressed in other tissues whereas CRYAA is more highly expressed in the lens than CRYAB, but not responsive to stress. Experiments performed on bovine, monkey or human lens provide additional evidence about the abundance of these crystallin subunits in lens fiber cells but a low concentration in lens epithelium and their chaperone function for maintenance of the lens transparency (Horwitz et al., 1998, Wang et al., 2004, Timsina and Mainali, 2021).

Eye lens may vary in case of fish species such as for instance, in case of Antarctic nototheneid fishes which inhabit the perennially freezing Antarctic seawater are transparent at -2°C, whereas the cold-sensitive and tropical fish lenses exhibit cold-induced cataract at 20°C and 7°C respectively. As cold temperatures depress the rates of biochemical function along with the stability of protein structure, thereby these fish species display a suite of biochemical and physiological adaptations to low temperature which includes cold-effective protein translocation or membrane phospholipid unsaturation (Eastman et al., 2000, Kiss et al., 2004, Kiss and Cheng, 2008). Various studies have indicated that the structure of eye or retina are the same like those of sub-tropical fish but whether the lens and their constituent proteins differ from temperate and tropical species is yet to be revealed. The model system for studies on vertebrate lens has been that of the cow/bovine lens. The bovine lens consists of a small insoluble albuminoid fraction and a large soluble crystallin fraction, where α-crystallin is predominant (Delaye and Benedek, 1982, Bloemendal and de Jong, 1991, Narberhaus, 2002). Comparative studies on α-crystallin in vertebrates have been performed previously to determine whether the properties of the respective protein have changed during the course of evolution.

The three-dimensional structure of α-crystallin is yet to be deciphered and thus there are limited studies regarding the sequence or structure of this protein. Keeping this in mind and basing on the importance of crystallin protein, a bioinformatics approach has been utilised to focus on its sequence, structure, phylogenetic analysis and homology modelling for structural studies. 20 protein sequences of both subunits (αA and αB) of α-crystallin from 10 different habitat-specific fish have been analysed.

Freshwater species Danio rerio, Bagarius yarrelli, Xiphophorus hellerii, Morone saxatilis, Salmo salar and Saltwater species Cyprinodon variegates, Thunnus macoyii, Stegastes partitus, Paralichthys olivaceus, Mastacembelus armatus have been retrieved from NCBI protein database. The bovine protein (Bos taurus) has been used as a reference here. The present study was thus undertaken to investigate the sequence, structure prediction and the phylogenetic relationship among ten habitat-specific fish, (five freshwater and five saltwater), and bovine using standard bioinformatic tools and techniques.
II. Materials And Methods

A. Sequence and Structural Analysis of α-crystallin proteins

The FASTA sequences of α-crystallin protein (both subunits) from ten habitat-specific fish (five freshwater species, five saltwater species) under study were obtained from NCBI database. The sequence of Bos taurus was retrieved and used as a control for comparison. (National Center for Biotechnology Information (nih.gov))

The secondary structure prediction tools namely GOR4 (Prediction of the Secondary Structure by GOR (ocha.ac.jp), SOPMA (NPS@: SOPMA secondary structure prediction (ibcp.fr), PSIPred (Daniel and David, 2019) were used to predict the 2-D structure of α-crystallin subunits of the ten fish and Bos taurus under study. Prediction of the 3D (3-dimensional structure) of the proteins under study was done by SWISS-MODEL (SWISS-MODEL (expasy.org)) and the structures with a greater Q-mean Z score and maximum sequence identity with the target sequence were selected as the models. Furthermore, the 3-D structures were visualised in RasMol (RasMol and OpenRasMol) (Guex and Peitsch, 1997, Pembroke, 2000) and finally confirmed in Ramachandran Plot (Kopp and Schwede, 2004, Guex and Schwede, 2009).

B. Physiochemical Characterisation

The physiochemical analyses of α-crystallin were performed using two tools, Expasy Protparam (ExPASy - ProtParam tool) (Lusiana and Irfannuddin, 2022) and Expasy ProtScale (ExPASy - ProtScale) (Du et al., 2022). ProtParam allows the computation of certain parameters such as calculation of molecular weight, theoretical isoelectric point, amino acid composition, atomic composition, estimated half life, hydrophaticity index, among others whereas ProtScale provides an analysis of hydrophaticity of the protein, i.e. a value < 0 indicates the presence of a greater number of hydrophilic (water soluble) amino acids and a value > 0 indicates greater presence of hydrophobic (water insoluble) amino acids, (water-insoluble protein).

Using the help of TMHMM server, it was used to determine the presence of transmembrane helices in α-crystallin.

C. Multiple Sequence Alignment of α-Crystallin Proteins

ClustalW (Multiple Sequence Alignment - CLUSTALW (genome.jp)) and Clustal Omega (Dahlman et al., 2005, Sievers et al., 2011) (Clustal Omega < Multiple Sequence Alignment < EMBL-EBI) tools were used for multiple sequence alignment (MSA) in order to predict the presence of conserved amino acid residues in α-crystallin of fish and bovine. The amino acid FASTA sequences retrieved from NCBI were loaded in Clustal tool for MSA. Moreover, crystallin protein sequences of the fish taken for the study, were compared with the standard sequences like Bos taurus, retrieved from SWISS-MODEL as well and predicted to be the closest homologue with the fish.

D. Phylogenetic Analysis

In order to search for evolutionary relationships among the protein sequences of fish and Bos taurus, cladograms were constructed using Clustal omega tool (Sievers et al., 2011) (Clustal Omega < Multiple Sequence Alignment < EMBL-EBI).

E. Homology Modelling

The theoretical models of α-crystallin were generated using SWISS-MODEL. Homology modelling was performed using SWISS-MODEL to get the complete structures of α-crystallin of the ten fish and bovine. The best models were generated and selected based on the one possessing the highest sequence identity and complete coverage (Guex and Peitsch, 1997, Pembroke, 2000) (SWISS-MODEL Interactive Workspace (expasy.org) along with the best QMEAN profiles.

III. Results Obtained

A. Sequence and structural analysis of α-crystallin proteins

The 20 amino acid sequences of α-crystallin proteins of 10 habitat-specific fish had been retrieved from NCBI database for our study. For comparison, α-crystallin from Bos taurus was predicted and found to be the closest homologue for the crystallins of the ten fish.

The secondary structure of all the 20 amino acid sequences of α-crystallin protein were predicted using tools like GORIV, SOPMA,PSIPRED listed in ExPASy proteomics server. The 2D structure tools have shown α-helices, β-pleated sheets and coils in all 20 sequences at respective places (Fig. 1, supplementary table 1). The 2D structures of α-crystallin of the sequences have been presented in the form of PSIPRED (Fig. 2I-VI).

Table 1 The number of 2D structures predicted by secondary structure prediction tools.
(H:Alpha Helix, S: Beta-pleated Sheets, C: Coils)
The 3D structures of α-crystallin protein sequences of the five saltwater fish species (Fig. 3–5) revealed an overall structural similarity with higher Z scores, indicating their favourable states. The GMQE values of αB-crystallins of saltwater fish species are generally higher than that of freshwater fish species, thereby indicating that saltwater fish species have favourable states and higher reliabilities, compared to the freshwater species. An exception is sheepshead minnow αB crystallin (2ygd.1.A) which possesses a lower GMQE as compared to zebrafish αB crystallin (2ygd). The αA crystallins of some species possess similar GMQE values for instance, αA crystallins of zebrafish (2ygd) and sheepshead minnow, Green Swordtail with Damselsh (6t1r), Atlantic Salmon with Zigzag Eel. The Ramachandran Plots for the species (Fig. 3–5) showed that all the residues were mostly present in the favourable/allowed regions.

II. Physiochemical Characterisation

The various physiochemical parameters of the α-crystallin proteins of 10 fish species are shown in Fig 6 (supplementary table 2). Some proteins, such as αB-crystallin of two freshwater fish Striped Bass, Atlantic Salmon possess instability indices of 36.22 and 37.53 respectively, indicating their stabilities. On the other hand, αA-crystallin of saltwater fish Zigzag eel is found to be a stable protein with an instability index of 39.89.

Table 2 Physiochemical parameters of α-crystallin of Bos taurus and the ten habitat-specific fish, computed from Expasy ProtParam.
<table>
<thead>
<tr>
<th>Species</th>
<th>Number of amino acids</th>
<th>Molecular Weight</th>
<th>Theoretical p.I.</th>
<th>Total number of negatively charged residues (Asp + Glu)</th>
<th>Total number of positively charged residues (Arg + Lys)</th>
<th>Extinction Coefficient</th>
<th>Instability Index</th>
<th>Aliphatic Index</th>
<th>Grand average of hydropathicity (GRAVY)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bos taurus(aA)</strong></td>
<td>173</td>
<td>19790.14</td>
<td>5.78</td>
<td>25</td>
<td>20</td>
<td>14440</td>
<td>57.45</td>
<td>72.08</td>
<td>-0.494</td>
</tr>
<tr>
<td><strong>Bos taurus(aB)</strong></td>
<td>175</td>
<td>20036.79</td>
<td>6.76</td>
<td>25</td>
<td>24</td>
<td>13980</td>
<td>50.92</td>
<td>77.43</td>
<td>-0.507</td>
</tr>
<tr>
<td><strong>Freshwater Species</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danio rerio(aA)</td>
<td>173</td>
<td>19713.18</td>
<td>5.84</td>
<td>25</td>
<td>21</td>
<td>17545</td>
<td>41.16</td>
<td>66.42</td>
<td>-0.643</td>
</tr>
<tr>
<td>Danio rerio(aB)</td>
<td>168</td>
<td>19977.59</td>
<td>5.77</td>
<td>25</td>
<td>20</td>
<td>33920</td>
<td>48.40</td>
<td>71.18</td>
<td>-0.591</td>
</tr>
<tr>
<td>Bagarius yarrelli(aA)</td>
<td>275</td>
<td>30033.23</td>
<td>5.21</td>
<td>41</td>
<td>29</td>
<td>12170</td>
<td>46.27</td>
<td>67.02</td>
<td>-0.543</td>
</tr>
<tr>
<td>Bagarius yarrelli(aB)</td>
<td>197</td>
<td>22017.40</td>
<td>4.71</td>
<td>34</td>
<td>25</td>
<td>7115</td>
<td>64.65</td>
<td>70.71</td>
<td>-0.561</td>
</tr>
<tr>
<td>Xiphophorus helleri(aA)</td>
<td>176</td>
<td>19890.08</td>
<td>5.84</td>
<td>25</td>
<td>21</td>
<td>17420</td>
<td>40.37</td>
<td>69.26</td>
<td>-0.585</td>
</tr>
<tr>
<td>Xiphophorus helleri(aB)</td>
<td>164</td>
<td>19200.75</td>
<td>5.49</td>
<td>23</td>
<td>18</td>
<td>44585</td>
<td>42.39</td>
<td>66.52</td>
<td>-0.484</td>
</tr>
<tr>
<td>Morone saxatilis(aA)</td>
<td>176</td>
<td>19912.96</td>
<td>5.70</td>
<td>25</td>
<td>21</td>
<td>17420</td>
<td>49.51</td>
<td>61.48</td>
<td>-0.634</td>
</tr>
<tr>
<td>Morone saxatilis(aB)</td>
<td>164</td>
<td>19255.78</td>
<td>6.06</td>
<td>24</td>
<td>21</td>
<td>45950</td>
<td>36.22</td>
<td>75.43</td>
<td>-0.552</td>
</tr>
<tr>
<td>Salmo salar(aA)</td>
<td>177</td>
<td>20045.28</td>
<td>5.69</td>
<td>25</td>
<td>21</td>
<td>17545</td>
<td>48.34</td>
<td>58.93</td>
<td>-0.605</td>
</tr>
<tr>
<td>Salmo salar(aB)</td>
<td>149</td>
<td>16300.58</td>
<td>5.62</td>
<td>18</td>
<td>13</td>
<td>8605</td>
<td>37.53</td>
<td>80.40</td>
<td>-0.219</td>
</tr>
<tr>
<td><strong>Saltwater Species</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprinodon variegates(aA)</td>
<td>175</td>
<td>19715.82</td>
<td>5.84</td>
<td>25</td>
<td>21</td>
<td>17420</td>
<td>41.15</td>
<td>70.74</td>
<td>-0.567</td>
</tr>
<tr>
<td>Cyprinodon variegates(aB)</td>
<td>164</td>
<td>19071.72</td>
<td>5.63</td>
<td>22</td>
<td>19</td>
<td>33460</td>
<td>46.19</td>
<td>71.34</td>
<td>-0.395</td>
</tr>
<tr>
<td>Thunnus maccocyii(aA)</td>
<td>176</td>
<td>19978.12</td>
<td>5.70</td>
<td>25</td>
<td>21</td>
<td>15930</td>
<td>49.27</td>
<td>65.90</td>
<td>-0.569</td>
</tr>
<tr>
<td>Thunnus maccocyii(aB)</td>
<td>143</td>
<td>16588.72</td>
<td>5.50</td>
<td>21</td>
<td>16</td>
<td>37470</td>
<td>47.92</td>
<td>66.08</td>
<td>-0.548</td>
</tr>
<tr>
<td>Stegastus partitus(aA)</td>
<td>176</td>
<td>19890.08</td>
<td>5.84</td>
<td>25</td>
<td>21</td>
<td>17420</td>
<td>40.37</td>
<td>69.26</td>
<td>-0.585</td>
</tr>
<tr>
<td>Stegatus partitus(aB)</td>
<td>162</td>
<td>19208.67</td>
<td>5.85</td>
<td>23</td>
<td>19</td>
<td>47440</td>
<td>45.56</td>
<td>59.51</td>
<td>-0.501</td>
</tr>
<tr>
<td>Paralichthys olivaceus(aA)</td>
<td>176</td>
<td>20071.47</td>
<td>5.71</td>
<td>25</td>
<td>21</td>
<td>18910</td>
<td>47.25</td>
<td>65.28</td>
<td>-0.634</td>
</tr>
<tr>
<td>Paralichthys olivaceus(aB)</td>
<td>164</td>
<td>19264.39</td>
<td>5.77</td>
<td>23</td>
<td>18</td>
<td>46075</td>
<td>48.78</td>
<td>58.84</td>
<td>-0.691</td>
</tr>
<tr>
<td>Mastacembelus armatus(aA)</td>
<td>176</td>
<td>19959.09</td>
<td>5.76</td>
<td>25</td>
<td>20</td>
<td>17545</td>
<td>39.89</td>
<td>64.77</td>
<td>-0.559</td>
</tr>
<tr>
<td>Mastacembelus armatus(aB)</td>
<td>164</td>
<td>19225.76</td>
<td>6.22</td>
<td>23</td>
<td>20</td>
<td>42970</td>
<td>47.44</td>
<td>72.50</td>
<td>-0.601</td>
</tr>
</tbody>
</table>
The Kyle-Doolittle Plots of α-crystallin of the 10 habitat-specific fish revealed no considerable differences i.e. the number of hydrophilic amino acids was found to be a little greater than the number of hydrophilic amino acids for both α-crystallin subunits as reflected from the respective positions 85-90, 105-110, 205-220, 400-480, 550-599, 601-699, 750-800, 905-915 for αA- subunit of the freshwater and saltwater fish species and 100-105, 250-290, 701-720, 801-810 for αB- subunits of the fish species displaying hydrophobic sites (Fig 7-8). They bear resemblances to the hydropathy plot of bovine α-crystallin which too exhibited a similar trend, with positions 35-40, 45-59, 65-79, 123-141 in case of bovine αA- and positions 38-42, 86-98, 122-142 in bovine αB- subunit showing the presence of hydrophobic amino acids.

An analysis by TMHMM server revealed no transmembrane segments in the respective proteins of either species, indicating that α-crystallin is a cytosolic protein (Fig 9).

III. Multiple Sequence Alignment of α-Crystallin Proteins

The results of multiple sequence alignment using Clustal 2.1 and Clustal Omega programme revealed considerable amount of conservation at certain sites such as methionine residues at position 60, methionine, aspartic acid, glutamine, isoleucine, histidine, alanine, arginine, phenylalanine, glycine residues at position 16, 68, 177 for α-crystallin of freshwater fish species [Fig 10 (A), (B)]; aspartic acid, asparagine, glycine, histidine, serine, glutamic acid, methionine, valine, glutamine, arginine, lysine residues at position 114 for crystallin of saltwater fish species [Fig 10 (C) (D)]. An analysis of MSA results of crystallins of both freshwater and saltwater fish species revealed conservation at respective sites i.e. arginine, leucine, aspartic acid, asparagine, tyrosine, serine, isoleucine, lysine, alanine, proline at positions 50, 97, 151 respectively. The considerable amount of conservation might probably be due to the high sequence similarity among the α-crystallin of ten habitat-specific fish.

The MSA analysis of the crystallins of the fish species, along with Bos taurus revealed the presence of conserved serine residues at positions 173 for bovine as well as the fish species; at the same time presence of conserved residues proline and lysine at position 175 of bovine protein was detected.

IV. Phylogenetic Analysis

The cladograms constructed with the help of tree viewer tool (Clustal Omega). From the phylogram presented in Fig 11, it is evident that Bos taurus and the other fish species are more or less located on a single branch, except for αB crystallin of catfish, a freshwater species and αB crystallin of Sheepshead Minnow, a saltwater species which are located on a single branch, whereas α-crystallins of freshwater fish species Striped Bass, Damselfish and Zigzag eel, both saltwater fish species are located on a single branch as well.

The above results emphasises the close phylogenetic relationship among these organisms with respect to the amino acid sequence of α-crystallin protein.

V. Homology Modelling

Models of α-crystallin were generated using the homology modelling approach, by SWISS-PDB Viewer tool. The best selected templates were done based on a greater sequence identity and maximum coverage, for instance freshwater species zebrafish α-crystallin (2ygd.1.X) was selected with an identity of 53.89 and 100% coverage which bear similarities with bovine (Fig 12.i.).

Similarly, on the other hand the template for catfish α-crystallin (3N3E) was selected with a sequence identity of 77.67 and 60% coverage (Fig 12.ii). On the other hand saltwater species, Zigzag eel (ID: 2YGD) α-crystallin was selected with a sequence identity of 71.10 and 99.99% coverage (Fig 12.iii.) The structures for the twenty amino acid fish sequences were overall similar, with greater sequence identities and maximum coverages; Bos taurus α-crystallin (PDB ID:2YGD) was considered as a standard template bearing closest resemblance with the crystallin protein of the ten habitat-specific fish under study (Fig 12.iv.). Such a homology might probably be due to the higher sequence similarity and sequence coverage with twenty protein sequences of fish.

Discussion

α-crystallin is an important protein of the eye lens whose structural conformation plays a vital role in maintaining the lens transparency. In the present study we explored a computational prediction to determine the structure of α-crystallin for ten fish species of two different habitats, freshwater and saltwater using the sequence of bovine crystallin as a template. The results obtained from using different secondary structure analytical tools showed the β-sheet content of α-crystallin of the saltwater fish species to be 22%, a bit more as compared to that of freshwater species (19%), while the rest consists of coils while the α-helix of freshwater species is 19%, which is greater as compared to saltwater fish (12%) (Supplementary table 1) which makes the protein stable and good for 3D modelling. The bovine α-crystallin is shown to be a good template as it possesses about 14% β-sheet indicating its stability. Previous Studies performed on zebrafish α-crystallin had revealed that there were no major differences in crystallin structure of the respective fish species, in comparison to mammalian α-crystallin (human) (Posner et al., 2003, Dahlman et al., 2005). The predominance of β-sheet in human α-crystallin (50-60%) in comparison to α-helix (3%) was revealed which reflected somewhat in our obtained results (Augustyn et al., 2004, Biswas et al., 2011). Another study which was performed on α-crystallin of six ectothermic species stated the presence of similar abundance of β-sheets in each species, as demonstrated by far-UV CD spectroscopy (Ghahghaei et al., 2009, Posner et al., 2012). A similar trend existed in our results except for the fact that the amount of β-sheet was slightly more in saltwater fish species crystallins while α-helices predominated in freshwater.
species. Wet lab experiments performed on human $\alpha$-crystallin had shown an abundance of $\beta$-sheet structures which reflected the relevance of using bovine $\alpha$-crystallin (mammal) as a template in our study as the results were more or less similar (Posner et al., 2003, Sprague-Piercy et al., 2021).

We are well aware that the prediction of the three dimensional protein structures through wet lab or experimental methods is really expensive and time consuming. 3D prediction tools are important of linking the protein to a specific function as well-known proteins with similar structures and well defined functions are used as templates for constructing the particular model (Silva et al., 2020). Moreover it is difficult to obtain the 3D structure of $\alpha$-crystallin owing to the high heterogeneity of its complex and its inaccessibility in the crystalline form, thus the structure is still unknown (Shamsi et al., 2022). Therefore a computational-based approach to determine its structure was carried out using SWISS MODEL. No previous work to elucidate the 3D structure of $\alpha$-crystallins of habitat-specific fish using computational methods had been performed, so this is the first time we have tried to decipher the respective structures using SWISS MODEL tool to validate our results with the wet lab experiments (we aim to perform experiments in wet lab in the upcoming days). SWISS MODEL provides an approach for successful 3D protein model generation and modelling. The Z- score indicates the overall quality of the model and even measures the total energy of the structure with respect to an energy distribution which had been derived from random conformations (de Jong et al., 1988, Persson, 2000, Shamsi et al., 2022). The structures of crystallins from the saltwater fish species showed higher Z values (>2) as compared to the freshwater fish species which indicated that they have more favourable states and higher reliabilities. From the figures, it can be seen that the Z-score values are almost similar to that of the template which recommends that the obtained models are reliable and close enough to the experimentally determined structure. The Ramachandran plot analyses showed that the main-chain conformations for 91.6% residues are within the favoured or allowed region, 2.16% lies within disallowed regions. It is known that a score close to 100% implies a good stereochemical quality of the model (Acharya et al., 2014) which shows that the models selected possess regular secondary and stable structures.

The physicochemical parameters such as number of amino acids, molecular weight, and total number of negatively and positively charged residues, grand average of hydropathicity (GRAVY), theoretical pl and the individual composition of each amino acid have been determined for each of the proteins, which show that $\alpha$B crystallins of two freshwater fish Striped Bass, Atlantic Salmon are stable. On the other hand, the $\alpha$A-crystallin of saltwater fish Zigzag eel is found to be a stable protein as well. The instability index of $\alpha$B-subunit of all species was 46.33, slightly greater as compared to $\alpha$A subunit (which is 45.22) which highlights the fact that $\alpha$A-subunit possesses more stability than $\alpha$B, as analysed by previous studies (Silva et al., 2020). The importance of computing physicochemical parameters lies in the transition from in silico to in vitro studies. For instance, the isoelectric point is useful for purification studies by isoelectric focusing (Silva et al., 2020, Valanciute et al., 2023). While knowing the theoretical pl of $\alpha$-crystallin of bovine and fish species to be 5.73, in vitro experiments requiring pH control to be carried out can be performed using that particular point as a start and control for in vitro experiments (Acharya et al., 2014). The average GRAVY score of $\alpha$-crystallin for the corresponding freshwater/saltwater fish species is -0.565 and -0.575 respectively with respect to bovine protein having a score -0.5005, stating that the protein is predominantly composed of hydrophilic amino acids.

The Kyle-Doolittle plots revealed that the number of hydrophilic amino acids is greater for both subunits of $\alpha$-crystallin for both bovine and the fish which may be indicative of the fact that the presence of hydrophilic amino acids in the C-terminal end of $\alpha$-crystallin maintains solubility and may bind to opposing charged residues of unfolding proteins (Derham and Harding, 1999, Acharya et al., 2014, Shamsi et al., 2022). The presence of such amino acids may signify the presence of amino acid sequences corresponding to the exon boundaries, a characteristic of domain interfaces lying at the interface of the protein (Augustyn et al., 2004). As per a previous study, the tertiary structure of $\alpha$-crystallin determined by fluorescence and UV-CD spectroscopy experiments revealed a lot of exposed hydrophobic pockets (Biswas et al., 2011). Reddy et al., 2006 stated the presence of hydrophobic amino acids in positions 50-54 and 79-99 of $\alpha$A-crystallin and in positions 75-103 of $\alpha$B-crystallin which differed in case of our obtained results. Our results showed the presence of hydrophobic amino acids at positions 35-40, 45-59, 65-79, 123-141 in case of bovine $\alpha$A- and positions 38-42, 86-98, 122-142 in bovine $\alpha$B- and in positions 85-90, 105-110, 205-220, 400-480, 550-599, 601-699, 750-800, 905-915 for $\alpha$A- subunit of the freshwater and saltwater fish species and 100-105, 250-290, 701-720, 801-810 for $\alpha$B subunits of the fish species. Literature studies revealed that the presence of hydrophobic amino acids at such definite positions mostly correlate with chaperone activity of $\alpha$-crystallin i.e. these amino acids interact with exposed hydrophobic sites of denaturing substrate proteins and prevent their aggregation (Reddy et al., 2006). The presence of hydrophobic amino acids at such defined positions may thus contribute to the chaperone activity of the protein. It is well known that $\alpha$-crystallin is a cytosolic protein thus our result could not detect any transmembrane segment from TMHMM server (D’Agostino et al., 2013, Nielsen, 2017).

The MSA sequences retrieved from Clustal tools revealed the presence of conserved methionine, aspartic acid, isoleucine, tryptophan, proline, glycine, arginine residues at positions 18, 20 and 120 of $\alpha$A subunit of $\alpha$-crystallin of bovine and freshwater fish and saltwater fish species whereas considerable differences with the exception of conserved methionine residues exist in the $\alpha$B subunits of bovine as well as for the freshwater and saltwater fish species. In previous studies performed on $\alpha$-crystallin of vertebrates, it has been observed that the diversity of $\alpha$-crystallin structure/function differed due to the differences in $\alpha$B subunit but not $\alpha$A which did not have any significant change (Posner et al., 2003, Bari and Sharma, 2020, Shamsi et al., 2022) which was evident with our results as well. This may lead us to deduce that the differences in $\alpha$-crystallin structure or their functions may develop due to the presence of different amino acid residues in $\alpha$B subunits of the corresponding fish species. The phylogenetic tree which was generated determined the close relationship existing between the freshwater and saltwater fish species especially Bluefin Tuna, Striped Bass, Catfish, Damselfish which are located in a single branch stating their interrelationship. Literature studies performed before stated that $\alpha$-crystallins are closely related to each other (Persson, 2000, Dahlman et al., 2005) in terms of structure/function especially in vertebrates of the same kingdom (Pisces). Bovine $\alpha$-crystallin is evolutionary closer to the fish species, as evident from the cladogram which states the similarities in crystallin
structure or function; if α-crystallin possessed a different function we would have expected to obtain longer branches. Our results are evident of the fact that no significant structural differences have arisen in α-crystallin in terms of evolution.

Homology modelling generally uses experimentally determined protein structures (templates) to predict the 3D structure of another protein bearing similarities to the target. The similarity between the target and the template is important for selecting the best templates i.e. the most similar sequences determine the 3D model of α-crystallin. For particular target-template pairs possessing sequence identities greater than 70%, models are generally predicted to be fairly accurate, whereas models having identity below 30% often own have different 3D structures and functions (Carver et al., 2003, Waterhouse et al., 2018) thus it was good enough to use the crystallographic structure of 2YGD (bovine) as a template to obtain a high quality alignment for structure prediction. The respective structures which were exhibited general similarities and subtle differences with the template 2YGD i.e. the structures of crystallins of zebrafish showed an identity of 100% whereas that of zigzag eel with an identity of 99.99% determined a stronger homology with the template. Furthermore, the Z-score of most of the fish species was about 2, which is almost similar to the score of the template (approx. 2.1) and thereby recommends that the obtained model is reliable.

Our work mainly aimed to reveal any structural differences of α-crystallin of ten fish species from two different habitats, using bovine protein as a template. From our estimated results no significant differences could be revealed in the structures of the proteins irrespective of the habitat and the minute details which may differentiate them is probably due to the presence of different amino acid residues in αB subunit of α-crystallin; whether the same holds in wet lab studies is yet to be elucidated in the upcoming days.

Declarations

Funding

This research did not receive any particular funding from any funding agencies in the public, commercial, or not-for-profit sectors.

Authors Contributions

The study was designed by AC, SG, SS, PD. AC performed the dry lab analyses using various online bioinformatic tools, preparation of figures and preparing the manuscript. SG, SS, PD edited and revised the entire manuscript. All the authors have read and approved the final manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data availability statement.

All information collated in this study was gotten from the Google search engine. Data were sourced via relevant references from different journals.

References


**Supplementary Tables 1 And 2**

Supplementary Tables 1 and 2 are not available with this version

**Figures**
Figure 1

A comparison of 2D structures of Freshwater Fish Species with Saltwater Fish Species, with relation to *Bos taurus* a. GOR b. SOPMA
Figure 2

I. 3D structure of α-crystallin protein of *Bos taurus* predicted by PSIPRED. a. αA of *Bos taurus*. b. αB- crystallin of *Bos taurus*. ii. PSIPRED analysis of α-crystallin a.i. Zebrash (freshwater) αA- crystallin

II. Zebrash αB- crystallin. b. i. Sheepshead Minnow (saltwater) αA- ii. Sheepshead Minnow αB.

III. PSIPRED analysis of αA- crystallin a. i. Catfish (freshwater fish) ii. Bluefin Tuna (saltwater fish) αB- crystallin b. i. Catfish (freshwater fish) ii. Bluefin Tuna (saltwater fish)

IV. PSIPRED analysis of αA- crystallin a. i. Green Swordtail (freshwater fish) ii. Damselfish (saltwater fish) αB- crystallin b. i. Green Swordtail (freshwater fish) ii. Damselfish (saltwater fish)

V. PSIPRED analysis of αA- crystallin a. i. Striped Bass (freshwater fish) ii. Flounder fish (saltwater fish) αB- crystallin b. i. Striped Bass (freshwater fish) ii. Flounder Fish (saltwater fish)

VI. PSIPRED analysis of αA- crystallin a. i. Atlantic Salmon (freshwater fish) ii. Zigzag Eel (saltwater fish) αB- crystallin of b. i. Atlantic Salmon (freshwater fish) ii. Zigzag Eel (saltwater fish)
Figure 3

i. 3D structure of α-crystallin of Freshwater Fish Zebrafish in RasMol viewer

ii. Q-mean Z-score

iii. Ramachandran Plot

Figure 4

i. 3D structure of α-crystallin of Freshwater Fish Catfish in RasMol viewer

ii. Q-mean Z-score

iii. Ramachandran Plot
Figure 5

i. 3D structure of α- crystallin of Saltwater Fish Damselfish in RasMol viewer

ii. Q-mean Z-score

iii. Ramachandran Plot

Figure 6

A comparison of the various physicochemical parameters computed by EXPASY ProtParam tool of freshwater fish species with saltwater fish species, in relation to *Bos taurus*.
Figure 7

Kyle-Doolittle Plots of the α-crystallin of *Bos taurus* and freshwater Fish Species computed from Expasy ProScale. The number of hydrophilic amino acids is slightly greater than the number of hydrophilic amino acids, for both α-crystallin subunits. a. αA of *Bos taurus* b. αB of *Bos taurus* c. αA of Freshwater Fish Species d. αB of Freshwater Fish Species
Figure 8

Kyle-Doolittle Plots of $\alpha$-crystallin of saltwater fish species, computed by ExPasy ProScale. The number of hydrophilic amino acids is slightly greater than the no. of hydrophobic amino acids, for both subunits $\alpha_A$, $\alpha_B$.

Figure 9

The prediction of transmembrane segments in $\alpha$-crystallin by TMHMM server revealed no transmembrane segments, stating that crystallin is a cytosolic protein.
Figure 10

Clustal Omega results for α-crystallin amino acid sequences for the 10 habitat-specific fish and *Bos taurus* a. αA crystallins of *Bos taurus* with Freshwater Fish Species and Saltwater Fish species b. αA crystallins of *Bos taurus* with Freshwater Fish Species and Saltwater Fish species c. αB crystallins of *Bos taurus* with Freshwater Fish Species and Saltwater Fish species d. αB crystallins of *Bos taurus* with Freshwater Fish Species and Saltwater Fish species.
Figure 11

Phylogenetic tree relates the relatedness of habitat-specific fish, and bovine with respect to α-crystallin.

Figure 12

Models of α-crystallin generated by Homology modelling in SWISS MODEL i. Zebrafish(freshwater)

ii. Catfish iii. Zigzag Eel (Saltwater Fish iv. Bos taurus(bovine))