ABSTRACT

Objective: Computational studies on fish brain acetylcholinesterase were conducted, expanding our views and for a deeper understanding of the activity of the fish acetylcholinesterase enzyme.

Methods: Physico-chemical properties of the fish acetylcholinesterase enzyme were studied. Homology model of the acetylcholinesterase enzyme was predicted, validated its quality and active sites were predicted. The amino acid frequency in the active sites was also compared. Similarly, the secondary structure of the sequences was predicted and compared. Phylogenetic analysis was performed by the neighbour joining tree method.

Results: Among the selected fish species stability of acetylcholinesterase was found in fish species namely *Esox lucius*. The negative GRAVY score value of enzyme in all the fish species ensured better interaction and activity in the aqueous phase. It was found that the molecular weight of the acetylcholinesterase enzyme ranged between 9113 and 15991 Da. Iso-electric (pI) of acetylcholinesterase was found to be acidic in nature. GOR IV was used to predict the secondary structure of acetylcholinesterase, which showed that random coil was dominated. Neighbor joining tree of the enzyme showed that fish species named *Amphiprion ocellaris* as the most divergent species, while the species *Oreochromis niloticus* is the most primitive one.

Conclusion: Acetylcholinesterase enzyme of *Esox lucius* was found to be the best compared to the other species, which possess a high number of active sites with 1se, Set and Ghu rich active sites.

Keywords: Acetylcholinesterase, *Esox lucius*, Phylogenetic analysis, CASTp, ProSA Web

INTRODUCTION

Acetylcholine-mediated neurotransmission is necessary for a nervous system function. Its sudden and unexpected blockade causes progressive deterioration of cognitive, autonomic and neuromuscular functions as in Alzheimer’s disease, multiple system atrophy and other conditions. The principal biological role of acetyl cholinesterase (AChE), acetylcholine hydrolase, EC 3.1.1.7) is the termination of impulse transmission at cholinergic synapses by rapid hydrolysis of the neurotransmitter acetylcholine (ACh) to yield acetic acid and choline [1-3]. In keeping with its function, AChE possesses a remarkably high activity of the fish acetylcholinesterase enzyme.

As per the data obtained by biochemists, geneticists and neurologists during the past few years, it is clear that the research in AChE continues due to the vast complexity of the enzyme’s nature. Hence the enzyme AChE enzyme needs to be characterized in all aspects to understand the structural and functional relations. Since the 3D structures of proteins can yield essential information about the functional mechanism, an initiative was taken to in silico characterization and analyses of these enzymes using molecular modeling techniques in different fish species. Protein sequences of acetylcholinesterase retrieved from NCBI were subjected to ProtParam to analyze physicochemical parameter, secondary structure prediction using GOR IV, homology modeling, phylogenetic analysis and active site prediction.

MATERIALS AND METHODS

Sequence retrieval and alignment

Protein sequence of acetylcholinesterase of 6 different fish species was retrieved from the National Centre for Biotechnology (NCBI) and made as the query sequence for the structure, properties prediction and molecular modeling.

Phylogenetic tree construction

A phylogenetic tree was constructed to find the evolutionary relationship between the different species. The relationship was analysed based on the neighbor-joining method on MEGA X software [10].

Physicochemical characterization

The sequences obtained were analyzed using various software’s available in the ExpASy server [11]. The physicochemical characterization such as molecular weight, pH, aliphatic index, Grand average of hydropathicity (GRAVY) and instability index was analyzed using ProtParam software. Percentage of amino acid occurrence was compared in these sequences.

Secondary structure prediction

The secondary structures of the sequences were predicted using the GOR IV prediction method [12]. It was performed to understand the
presence of helices, beta turns and coils in the protein chain. To obtain the graphical representation of the secondary structure of the protein data obtained from the GOR IV prediction method, PSIPRED software was used.

Table 1: The organism and accession number of fishes understudy

| Organism          | Accession number |
|-------------------|------------------|
| Salveinus alpinus | XP_023826535.1   |
| Oreochromis niloticus | XP_003458443.2 |
| Amphiprion ocellaris | XP_023119368.1 |
| Serioli dumerili  | XP_022603603.1   |
| Monopterus albus  | XP_020450658.1   |
| Esox lucius       | XP_010898117.1   |

Homology modeling of acetyl cholinesterase

Homology modeling was done using SWISS-MODEL homology modeling server [13].

Structural quality analysis

The quality of protein structure obtained from SWISS model was analyzed using ProSA web [14].

Active site prediction

Active sites of the predicted models were analyzed using CASTp [15]. Amino acid compositions of all the cavities and frequency of amino acid occurrence in the cavities of each model were analyzed.

RESULTS

Phylogenetic tree construction

The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale with branch lengths in the same unit as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved six acetylcholinesterase protein sequences. All ambiguous positions were removed for each sequence pair. The organism with accession number XP 023119368.1 is the most divergent species, while the organism with accession number XP 003458443.2 is the most primitive one (fig. 1).

Physico-chemical parameters of selected proteins

The number of amino acid in the selected fish species showed variation (table 2). It ranges from 81 (Esox lucius) to 959 (Serioli dumerili). The molecular weight of acetylcholinesterase showed a range from 9113.31 dalton to 117687.15 dalton. Serioli dumerili was found to be the organism with the highest molecular weight, while Esox lucius had the lowest molecular weight for acetylcholinesterase.

The data revealed that all the fish species under study belong to acidic nature. The alphatic index reveals that only Esox lucius is thermostable. Serioli dumerili was found to be the most hydrophilic in nature, analyzing the GRAVY index. It was also found to show high catalytic activity. Based on the instability index parameter, the stability of the enzyme was predicted. Only Esox lucius was a stable enzyme. According to the data, Ile (Isoleucine) was the dominant amino acid residue with a percentage of 14.7 (table 3).

Table 2: Protparam analysis of acetylcholinesterase sequences

| S. No | Name of organism | Accession no. | No. of amino acids | Molecular weight | Pi | Aliphatic index | Gravy | Instability index | Stability analysis |
|-------|------------------|---------------|--------------------|------------------|----|----------------|-------|-------------------|-------------------|
| 1     | Salveinus alpinus| XP_023826535.1| 139                | 1599.85          | 4.58| 76.62         | -0.491| 43.55             | unstable          |
| 2     | Oreochromis niloticus | XP_003458443.2 | 136              | 1554.13          | 4.50| 78.31         | -0.466| 44.88             | unstable          |
| 3     | Amphiprion ocellaris | XP_023119368.1 | 139              | 1599.85          | 4.58| 76.62         | -0.491| 43.55             | unstable          |
| 4     | Serioli dumerili  | XP_022603603.1| 959               | 117687.15        | 4.57| 67.85         | -0.279| 53.38             | unstable          |
| 5     | Monopterus albus  | XP_020450658.1| 139               | 1599.85          | 4.58| 76.62         | -0.491| 43.55             | unstable          |
| 6     | Esox lucius       | XP_010898117.1| 81                | 9113.31          | 4.63| 88.02         | -0.621| 38.26             | stable            |

Fig. 1: Neighbour joining tree showing evolutionary relationship between various AChE sequences
Table 3: Amino acid residue frequency of acetylcholinesterase (Highlighted are the highest amino acid frequencies)

| Amino acid residue | Accession number | S. alpinus | O. niloticus | A. ocellaris | S. dumerili | M. albus | E. lucius |
|--------------------|------------------|------------|-------------|-------------|-------------|----------|----------|
| Ala                | XP_023826535.1   | 6.5%       | 6.6%        | 6.5%        | 6.8%        | 6.5%     | 6.2%     |
| Arg                | XP_003458443.2   | 5.0%       | 4.4%        | 5.0%        | 5.3%        | 5.0%     | 4.9%     |
| Asn                | XP_023119368.1   | 4.3%       | 4.4%        | 4.3%        | 6.5%        | 4.3%     | 3.7%     |
| Asp                | XP_022603603.1   | 4.3%       | 4.4%        | 4.3%        | 3.0%        | 4.3%     | 4.9%     |
| Cys                | XP_020450658.1   | 2.2%       | 2.2%        | 2.2%        | 5.2%        | 2.2%     | 0.0%     |
| Gin                | XP_010898117.1   | 0.0%       | 0.0%        | 0.0%        | 0.3%        | 0.0%     | 0.0%     |
| Glu                |                 | 11.5%      | 11.8%       | 11.5%       | 11.3%       | 11.5%    | 13.6%    |
| Gly                |                 | 4.3%       | 3.7%        | 4.3%        | 2.2%        | 4.3%     | 4.9%     |
| His                |                 | 5.0%       | 5.1%        | 5.0%        | 3.4%        | 5.0%     | 7.4%     |
| Ile                |                 | 14.4%      | 14.7%       | 14.4%       | 10.4%       | 14.4%    | 16.0%    |
| Leu                |                 | 3.6%       | 3.7%        | 3.6%        | 4.4%        | 3.6%     | 4.9%     |
| Lys                |                 | 0.0%       | 0.0%        | 0.0%        | 0.1%        | 0.0%     | 0.0%     |
| Met                |                 | 1.4%       | 1.5%        | 1.4%        | 3.2%        | 1.4%     | 0.0%     |
| Phe                |                 | 0.0%       | 0.0%        | 0.0%        | 2.5%        | 0.0%     | 0.0%     |
| Pro                |                 | 1.4%       | 1.5%        | 1.4%        | 2.4%        | 1.4%     | 2.5%     |
| Ser                |                 | 12.9%      | 13.2%       | 12.9%       | 7.7%        | 12.9%    | 8.6%     |
| Thr                |                 | 9.6%       | 9.8%        | 9.6%        | 8.6%        | 8.6%     | 8.6%     |
| Trp                |                 | 0.0%       | 0.0%        | 0.0%        | 0.4%        | 0.0%     | 0.0%     |
| Tyr                |                 | 0.0%       | 0.0%        | 0.0%        | 2.2%        | 0.0%     | 0.0%     |
| Val                |                 | 0.0%       | 0.0%        | 0.0%        | 1.1%        | 0.0%     | 0.0%     |
| Pyl                |                 | 4.3%       | 3.7%        | 4.3%        | 6.9%        | 4.3%     | 1.2%     |
| Sec                |                 | 2.2%       | 2.2%        | 2.2%        | 3.3%        | 2.2%     | 1.2%     |

Percentage of amino acid sequence forming secondary structure in GOR IV and PSIPRED

The secondary structure of the protein was predicted using GOR IV. The data reveal the percentage of alpha helix, beta-helix, extended strands and random coils of the AChE enzyme under study (table 4). Based on the random coils, the dominant ones are selected. *Monopterus albus* and *Esox lucius* has the highest number of alpha helices. *Oreochromis niloticus* has the lowest number of alpha helices. *Salveins alpinus* have the lowest number of extended strand and *Serioli dumerili* has the highest number of extended strands. *Serioli dumerili* has the lowest number of random coils and *Salveins alpinus* has the highest number of random coil. The graphical representation of the data was predicted using PSIPRED. In PSIPRED, the data of the GOR IV are represented using symbols for helix, strands and coils.

1. Salveins alpinus
2. Oreochromis niloticus

3. Amphiprion ocellaris

4. Serioli dumerilli
5. *Monopterus albus*

![Graphical representation of the data obtained from GOR IV visualized using Psipred](image)

Fig. 2: Graphical representation of the data obtained from GOR IV visualized using Psipred

| S. No. | Name of organism     | Accession no. | G or IV  |
|--------|----------------------|---------------|----------|
|        |                      |               | Alpha helix % | Extended strand % | Random coils % |
| 1      | Salvelinus alpinus   | >XP_023826535.1 | 23.33% | 20.79% | 55.87% |
| 2      | Oreochromis niloticus| >XP_003458443.2 | 22.66% | 22.19% | 55.16% |
| 3      | Amphiprion ocellaris| >XP_023119368.1 | 24.08% | 21.88% | 52.72% |
| 4      | Seriola dumerili     | >XP_022603603.1 | 23.56% | 24.36% | 52.08% |
| 5      | Monopterus albus     | >XP_020450658.1 | 25.40% | 21.88% | 52.72% |
| 6      | Esox lucius          | >XP_010898117.1 | 25.40% | 21.88% | 52.72% |

Table 4: Percentage of amino acid sequence forming secondary structure in GOR IV (Highlighted in yellow show the highest percentage in each category)

Analysis of the structural quality using ProSA Web

The quality of the predicted structure was analyzed using ProSA Web by NMR and X-ray method. The results are shown in fig. 3.

6. *Esox lucius*

![Graphical representation of the data obtained from GOR IV visualized using Psipred](image)

Fig. 2: Graphical representation of the data obtained from GOR IV visualized using Psipred

Prediction of an active site using CASTp

The active sites of the enzyme were analyzed using CASTp tool to find out the sites for ligand attachment. The blue highlighted amino acids are the active sites for binding the ligands (fig. 4).
1. *Amphiprion ocellaris*

2. *Esox lucius*

3. *Monopterus albus*

4. *Oreochromis niloticus*
5. *Serioli dumerili*

Fig. 3: NMR and X ray data images of the sequences visualized using Pro SA web

1. *Amphiprion ocellaris*

2. *Esox lucius*

3. *Monopterus albus*
4. *Oreochromis niloticus*

![Image](image1.png)

Fig. 4: The predicted active sites of each species enzyme in 3D and sequence form

**DISCUSSION**

AChE has long been of academic and industrial interest and studied extensively at the biochemical, biophysical, and molecular because this enzyme is a target of palliative Alzheimer drugs, nerve agents, and insecticides [16]. The evolutionary history of the tree constructed from nucleotide sequences of the AChE gene of six fish species was inferred using the Neighbor-Joining method in MEGA X software. A phylogenetic tree or evolutionary tree is a branching diagram or tree showing the evolutionary relationships among various biological species. The phylogeny is based upon the similarities and differences in their physical or genetic characteristics. Phylogenetics enriches the understanding of the evolution of genes, genomes and species [17]. In the present study, the evolutionary analyses reveal that the fish species with accession number XP 023119368.1 (*Amphiprion ocellaris*) is the most divergent species, while fish species with accession number XP 003458443.2 (*Oreochromis niloticus*) is the most primitive one.

ProtParam is a Bioinformatics tool which allows the computation of various physical and chemical parameters for a given protein. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Instability index provides an estimate of the stability of the protein. Based on the instability index, the stability of the AChE enzyme retrieved from six different fish species was predicted. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. And it was found that the AChE enzyme of *Esox lucius* was stable enzyme and other AChE enzyme fish species are unstable. The Aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of the thermostability of globular proteins [18]. GRAVY (Grand Average of Hydropathy) value for a peptide or protein is calculated as the sum of hydropathy value of all the amino acids divided by the number of residues in the sequence [19]. In the present study, it was found that the GRAVY value was negative in all species studied. Theoretical pI of proteins is defined as the pH at which a protein has no net charge. When the pH>pI, a protein has a net negative charge and when the pH<pI, a protein has a net positive charge. In the present study, pI value showed that the enzyme is acidic in nature. Generally, it was observed that towards acidic pI values, the GRAVY tends to be low.

The secondary structure of the protein was predicted using GOR IV and PSIPRED. The method is based on information theory and an assumption that the information function of a protein chain can be approximated by a sum of information from single residues and pairs of residues. The analysis of the frequencies of occurrence of secondary structure for singlets and doublets of residues in a protein database enables prediction of secondary structure for new amino acid sequences. It analyses sequences to predict alpha helix, beta-sheet, turn, or random coil secondary structure at each position based on 17-amino-acid sequence windows. AChE enzyme from *Monopterus albus* and *Esox lucicus* had the highest number of alpha-helices while *Oreochromis niloticus* had the lowest. AChE enzyme of *Serioli dumerili* had the highest number of extended strands and *Salvinus alpinus* had the lowest. AChE enzyme of *Salvinus alpinus* had the highest number of random coils.

Homology modeling, also known as comparative modeling of a protein refers to constructing an atomic-resolution model of the target protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein.
Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structures [20]. Among the six species, Esox lucius had the highest Z score. The result indicates the quality of the enzyme from Esox lucius to be best among the rest of the species.

The active site of the enzyme was predicted using CASTp. The active site is the region of an enzyme where substrate molecules bind and undergo a chemical reaction. The active site consists of residues that form temporary bonds with the substrate (binding site) and residues that catalyse a reaction of that substrate (catalytic site). In the active site, amino acids of the enzyme protein will bind to the substrate. The active sites and the dominant amino acids were studied from the structure and sequence. In the present study, the number of active sites was calculated to be the highest in the Seriolidae (XP_022603603.1) acetylcholinesterase enzyme compared to the rest of the species.

CONCLUSION

The study provides a better insight of the enzyme acetylcholinesterase in fishes. The acetylcholinesterase enzyme was found to be 1k, Ser and Glu rich protein with mostly acidic pI value. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structures [20]. Among the six species, Esox lucius had the highest Z score. The result indicates the quality of the enzyme from Esox lucius to be best among the rest of the species.

ACKNOWLEDGMENT

The authors thank Dr. Vinod Kumar Nathan, Assistant Professor, School of Chemical and Biotechnology, Sastra University for the advice and guidance during the conduct of this study.

FUNDING

Nil

AUTHORS CONTRIBUTIONS

Dr. Prassy P Prakasia conducted the experiment and wrote the manuscript.

CONFLICT OF INTERESTS

Declared none

REFERENCES

1. Barnard EA. Neuromuscular transmission-enzymatic destruction of acetylcholine. In: Hubbard JL. editors. The Peripheral Nervous System. New York: Plenum publishers; 1974. p. 201–24.
2. Hasinoff BB. Kinetics of acetylthiocholine binding to electric eel acetylcholinesterase in glycerol/water solvents of increased viscosity. Evidence for a diffusion-controlled reaction. Biochem Biophys Acta 1982;704:52-8.
3. Manjunatha KS, Manu CP, Satyanarayana, Nayak KN, Vinay MS, Vineetha, et al. Acetylcholinesterase inhibitory effect of 3-(1h-indol-3-yl)-1, 3-diphenylpropan-1-one derivatives. Asian J Pharm Clin Res 2017;10:83-6.
4. Malathi S, Vidyaashree HM, Ravindran R. Restoration of memory and acetylcholinesterase activity by michelina champaca in chronically noise-stressed Wistar Albino rats. Asian J Pharm Clin Res 2016;9:210-4.
5. Sussman JL, Harel M, Frolov F, Oefner C, Goldman A, Toker L, et al. Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science 1991;253:872–9.
6. Raves ML, Harel M, Pang YP, Silman L, Kozikowski AP, Sussman JL. Structure of acetylcholinesterase complexed with the nootropic alkaloid (−)-huperzine A. Nat Struct Mol Biol 1997;4:57–63.
7. Nozke HJ, Rosenberry TL, Neumann E. Effective charge on acetylcholinesterase active sites determined from the ion strength dependence of association rate constants with cationic ligands. Biochemistry 1980;19:3705–11.
8. Dougherty DA, Staufer DA. Acetylcholine binding by a synthetic receptor: implications for biological recognition. Science 1990;245:1643-9.
9. Taylor BK, Holloway D, Printz MP. A unique central cholinergic deficit in the spontaneously hypertensive rat: physostigmine reveals bradycardia associated with sensory stimulation. J Pharmacol Exp Ther 1994;268:1081–90.
10. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406-25.
11. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPAsy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 2003;31:3784–8.
12. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. Method Enzymol 1996;266:540-53.
13. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modelling server. Nucleic Acids Res 2003;31:339-45.
14. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007:35:407-10.
15. Wierenga K, Kabasakal S, Leung K, Ouyang Z, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with the structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 2006;34:116-8.
16. Taylor P, Radic Z. The cholinesterases: from genes to proteins. Annu Rev Pharmacol Toxicol 1994;34:281–320.
17. Lemmon AR, Moriarty EC. The importance of proper model assumption in bayesian phylogenetics. Syst Biol 2004;53:265-77.
18. Iki A. Thermostability and aliphatic index of globular proteins. J Biochem 1980;88:1895-8.
19. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105-32.
20. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;5:223-6.