

Print ISSN: 2656-0097 | Online ISSN: 0975-1491

Vol 12, Issue 7, 2020

Original Article

MODELING AND STRUCTURAL ANALYSIS OF ACETYLCHOLINESTERASE ENZYME OF FISHES

PRESSY P. PRAKASIA

Department of Botany and Centre for Research, St. Teresa's College (Autonomous), Ernakulam Email: dearpressy@gmail.com

Received: 05 Apr 2020, Revised and Accepted: 21 May 2020

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 (pl) 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: Acetlycholinesterase enzyme of *Esox lucius* was found to be the best compared to the other species, which possess a high number of active sites with Ile, Set and Glu rich active sites.

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

© 2020 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/) DOI: http://dx.doi.org/10.22159/ijpps.2020v12i7.37762. Journal homepage: https://innovareacademics.in/journals/index.php/ijpps

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 specific activity, especially for a serine hydrolase functioning at a rate approaching that of a diffusion-controlled reaction [4].

The distinctive biochemical properties and physiological significance of AChE make it an interesting target for detailed structure-function analysis. Coding sequences of AChE have been cloned from a range of diverse evolutionary vertebrate and invertebrate species that include insects, nematodes, fish, reptiles, birds and several mammals, the most important among them man. The first crystal model for AChE was determined from *Torpedo californica*, one of the main sources of AChE for research. Following, crystal structures from mouse, *Drosophila* and man were obtained and found to be fundamentally similar.

The crystal structure of AChE shows that the catalytic triad, which is formed by serine, histidine and glutamate, is located at the bottom of a narrow 20-Å-deep gorge that penetrates halfway into the enzyme and widens close to its base [5, 6]. The ligand-binding cavity is lined with aromatic residues that account for approximately 40 percent of the cavity surface. The entrance of the gorge is termed the peripheral anionic site (PAS) as it was initially believed to contain several negatively charged amino acids due to its preference of binding cationic ligands [7]. However, the crystal structure indicates that insufficient acidic amino acids are located close to the ligand-binding cavity to support this hypothesis. Instead, it has been shown that aromatic residues interact with cationic ligands [5, 8]. A similar interaction pattern can be seen at the catalytic site (CAS), and it is believed that AChE initially binds to the PAS and then rapidly diffuses down to the catalytic site [9, 5].

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 physiochemical 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 \mbox{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).



Fig. 1: Neighbour joining tree showing evolutionary relationship between various AChE sequences

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 aliphatic 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.	Name of	Accession	No. of	Molecular	Pi	Aliphatic	Gravy	Instability	Stability
No.	organism	no.	amino	weight		index		index	analysis
			acids						
1	Salveinus alpinus	XP_023826535.1	139	15991.85	4.58	76.62	-0.491	43.55	unstable
2	Oreochromis	XP_003458443.2	136	15541.31	4.50	78.31	-0.466	44.88	unstable
	niloticus								
3	Amphiprion	XP_023119368.1	139	15991.85	4.58	76.62	-0.491	43.55	unstable
	ocellaris								
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	15991.85	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

Amino	Accession number					
acid	XP_023826535.1	XP_003458443.2	XP_023119368.1	XP_022603603.1	XP_020450658.1	XP_010898117.1
residue	Organism					
	S. alpinus	0. niloticus	A. ocellaris	S. dumerili	M. albus	E. lucius
Ala	6.5%	6.6%	6.5%	6.8%	6.5%	6.2%
Arg	5.0%	4.4%	5.0%	5.3%	5.0%	4.9%
Asn	4.3%	4.4%	4.3%	6.5%	4.3%	3.7%
Asp	4.3%	4.4%	4.3%	3.0%	4.3%	4.9%
Cys	2.2%	2.2%	2.2%	5.2%	2.2%	0.0%
Gln	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	8.6%	8.8%	8.6%	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%

Table 3: Amino acid residue frequency of acetylcholinesterase (Highlighted are the highest amino acid frequencies)

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 helixes. Oreochromis niloticus has the lowest number of alpha helixes. 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 coils. 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. Salveinus alpinus

Conf:		110011111	1 1111 2011120		Conf:			122 111 2222		
Pred:	-0				Pred:		-0)		
Pred:	ссссннннннннн	нннсссссо	CCCCCEEECC	CEEEE	Pred:	ссссснннннннссо	ссннннннн	нессссссс	22222	
AA:	MHTSIVFLLPFFLLSI	LASSSSSQS	SDSPELIVSMR	AGRVR	AA:	PSGGNDTELVDCLRN	KSPQELIDQEN	QVMPWSAIF	RESEV	
	10	20	30	40		290	300	310	320	
Conf:	}=====================================			JEEC	Conf:	3		100000 110 00	J===f	
Pred:	$\rightarrow \rightarrow $	<u> </u>			Pred:	f)		<u> </u>	$ \rightarrow $	
Pred:	EEEEECCCCCEEEEEE	0000000000	222222222222	22222	Pred:	сссссссссснини	HCCCCCCCCE	EEEECCCCCC	CEEEE	
AA:	GVRLPTPDRSHVTAFI	GIPFAEPPI	JGKKRFRKAEP	KKPWS	AA:	PVIDGVVLPDNPEAM	LSSGNFKDTQI	LLGVNQDEGS	SFFLL	
	50	60	70	80		330	340	350	360	
Conf:]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]			l l l l l l l l l l l l l l l l l l l	Conf:				l l l l l l	
Pred:					Pred:					
Pred:	CCEECCCCCCCCCCCC	0000000000	222222222222	CCCCC	Pred:	cccccccccccccc	ннннннн	СССССНННН	нннн	
AA:	DLFEASSYPNACYQF	DTSFPGFQC	TEMWNPNREM	SEDCL	AA:	YGAPGFSKDNESLIS	REDFLESVKLS	SVPHANEIGQE	EAVVL	
	90	100	110	120		370	380	390	400	
Conf:	}		1222 1 22222)cccc	Conf:	} 			l l l l l l l l	
Pred:		\rightarrow			Pred:			->-0		
Pred:	CEEEEECCCCCCCCC	EEEEEECCO	222222222222	22222	Pred:	ннссссссссннннн	нннннсссе	EECCHHHHH	нннн	
AA:	ILNIWVPISPRPHNLI	140	150	IDGRI	AA:	QYTDWFDENNGVKNRI	DALDDVVGDHN	WICPLQHFAN	RSYAQ	
	150	140	150	100		410	420	430	440	
Conf:	}==== =1111111111111				Conf:					
Pred:	\longrightarrow		Ð		Pred:			_	<u>~</u>	Conf: 1000000000000000000000000000000000000
Pred:	CCCCCCEEEEEECCCC	CCCCCCCCC	CCCCCCCHHH	HHHHH	Pred:	несссссссссссссс		CCEEEEEEE	EECCC	Pred:
AA:	170	180	1GSSEAFGNVG	200	AA:	FHSNQGANNGQQAGN	AGAMTKDGNS	QGGVYLYLFD	HRASN	<pre>Pred: CCCCCCCCCCCHHHHHHHHHHHHHHHHHHCCCHHHHHHH</pre>
	170	100	150	200		450	460	470	480	570 580 590 600
Conf:]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]			BBBBBBBBBBBBB	Conf:					
Pred:		\rightarrow	Ð		Pred:			0	_	Conf: 1
Pred:	HHHHHHHHHHHHHCCCC	CCCCCEEECO	СНННННННН	HHCCC	Pred:	000000000000000000000000000000000000000		СССССННННН	нннн	Pred: HUBBCCCBBBBBBBBBBCCCCCCCCC
AA:	210	220	230	240	AA:	LAWPEWMGVIHGYEI	EFVFGMPLEKE	RLNYTAEEEKI	LSRRM	AA: KVEFHRWSSYMMHWKSQFDHYSKQERCTDL
	220	220	200	2.0		490	500	510	520	610 620 630
Conf:]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]			11111	Conf				la a la antes de la composición de la composicinde la composición de la composición de la composición	Legend:
Pred:		Ð			Pred:					= helix Conf:],] [= confidence of prediction
Pred:	CHHHHHHHHHHHCCCCC	CCCCCCCCHE	HHHHHHHHHHHH	HHCCC	Pred:	ннннннннссссссо			CEEEE	= + = strand Pred: predicted secondary structure
AA:	250	260	270	280	AA:	MRYWANFARTGNPNI	FDGSVDNRR	WPQFTNSEQ	KHVGL	
	200	200	270	200		530	540	550	560	- COII AN: Carger Sequence

Prakasia

2. Oreochromis niloticus

Conf:					Conf:		111111111111		lannef.						
Pred:		0			Pred:	-	-0								
Pred:	СНННННННННННСС	ссссснин	ннннсссснн	нннн	Pred:	снинининнессссо	ссссссснин	нннынны	ннннс						
AA:	SPAEARRRATQLAKEVG	CNGGNDTE	TIDCLRSKSPC	QLID	AA:	MKESRRLLQIHRPYSI	ITKMOPTVVP	LFSSIFLLSF	IIPVÇ						
	290	300	310	320		10	20	30	40						
Conf:					Conf:	3000000000000000000000									
Pred:					Pred:										
Pred:	ннннссссссссссссссссссссссссссссссссссс		ссснннннсс	CCCC	Pred:	CCCCCCCCEEECCCE	EEEEEEEC	CCCEEEEEEC	ccccc						
AA:	HEWQVLPWSALFRFSFV	PVVDGEFL	PDTPEAMLNSG	NFKD	AA:	FSQNDGDLVVQTRTG	RVRGIRQLVP	DQSYITAFLG	IPFAE						
	330	340	350	360		50	60	70	80						
Conf:			000000000000000000000000000000000000000		Conf	10000000000000000000000000000000000000	1111111111								
Pred:			0	_	Pred:										
Pred:	CCEEEEEECCCCCEEEE	000000000	ссссссснини	нннн	Pred:	000000000000000000000000000000000000000	22222222222	ccccccccc	ccccc						
AA:	TQILLGVNQNEGSYFLI	YGAPGFSK	DNDSLISREDF	LESY	AA:	PPLGNRRFRPAEPKRI	WTGVLEAKE	YPNACYQYVD	TAFPG						
	370	380	390	400		90	100	110	120						
Conf					Conf.										
Pred:	-				Drod.										
Pred:	нниссссснынныны	HCCCCCCC	ССНННННННН	нннс	Pred.		CCCCCEECO	CCCCCCCCEE	FFFFF						
AA:	RLSVPHANDIGLEAVVI	QYTDWMDE	NNEIKNRDAMD	DIVG	AA:	FQGSEMWNPNREMNEI	CLYLNVWVP	SSPRPHNLTV	MVWIY						
	410	420	430	440		130	140	150	160	Conf:	300000000000	1220222			l l l l l l l l l l l l l l l l l l l
Conf.					a.e.t.	III.			lana al	Pred:	$ \rightarrow $	-	0		
Pred.					Conr:					Pred:	CCCCCCEEEE	cccccc	сссссенни	нннннннн	ннннс
Pred.	ССЕЕВССНИННИНИНИ	HCCCCCCC	coccoccccc	CCCC	Pred.		CCCCCCCCE	FFFFFCCCCC	00000	AA:	TVSEQKHVGL	NTESLKI	HKGLRNQM	CAFWNRFLPR	LLNIŢ
AA:	DHNVICPLAHFARSYAC	HNALKANV	GGMNFGGVNSG	NSQG	AA:	GGGFYSGSSSLDVYD	GRYLAHSETV	IVVSMNYRIG.	AFGFL		5	70	580	590	600
	450	460	470	480		170	180	190	200						
a.e.e.		In all the							In-n	Conf:					
Conf:					Conf:					Pred:	-0		Ð		
Dred.			~~~~~~~~~	0000	Pred:					Pred:	ССННННННН	НННННСС	сниннини	ннннннссс	ccccc
AA:	GVYLYLFDHRASNLAWF	EWMGVIHG	YEIEFVFGLPL	VKRL	AA:	ALHGSSEAPGNVGLLI	ORMALOWVO	DNIHFFGGNP	KQVTI	AA:	DNIDEAERQW	KVEFHR	VSSYMMHWK	SQFDHYSKQE	RCTDL
	490	500	510	520		210	220	230	240		6	10	620	630	640
										Lege	nd:				
Conf:				de a la	Conf:				Hand t	$\overline{\mathbf{O}}$	= helix	Conf:	3	confidence	of prediction
Pred:	-0				Pred:		-0		-		~		- +		
Pred: AA:	CCCHHHHHHHHHHHHHHH NYTRDEEKLSRRMMKYW	HHHHCCCC ANFARTGN	CCCCCCCCCCC	CCCC WPVF	Pred:	ECCCHHHHHHHHHHH	CCCCHHHHHH	HHHHCCCCCC ATLOSGVPNS	CCCCC		<pre>> = strand</pre>	Pred:	predicted	secondary st	ructure
	530	540	550	560		250	260	270	280		= coil	AA: ta	arget seque	nce	
						200	2.50	210	200					1007020	

3. Amphiprion ocellaris

				-					and and the state	
Conf:					Conf:					
Pred: AA:	CCCCHHHHHHHHCCCC NGGNDTELVDCLRNK	CHHHHHHHHH TPQELIDQEW	HCCCCCCCCC QVLPWSALFF	CCCCCC	Pred: AA:	CCCHHHHH MQPSVLHI	HHHHHHH LSSIFLL	HHHHCCCCCCC	CCCCCEEECC	CEEEE
	290	300	310	320			10	20	30	40
Conf:				ling	Conf:			1=11 11111		
Pred:		-			Pred:		_	>		
Pred: AA:	CCCCCCCCCCHHHHH VVDGEVLPDTPEAMLM	HCCCCCCCCE NSGNFKDTQI	EEEEECCCCC	EEEEC	Pred: AA:	EEEEECCO GIRLPVPI	CCEEEEE	ECCCCCCCCC MGIPFAEPPV	CCCCCCCCCC GRRRFRPAEP	CCCCC KHPWT
	330	340	350	360			50	60	70	80
Conf:					Conf:			00 00000 000	0000 00000	
Pred:			0		Pred:					
Pred: AA:	CCCCCCCCCCCCCH GAPGFSKDNESLISR	HHHHHHHHH EDFLEGVKLS	CCCCCHHHHH	HHHHH AVVLQ	Pred: AA:	CCCCCCCC GVYDANAY	CCCCCCCC	CCCCCCCCCC VDTSFPGFQG	CCCCCCCCCC SEMWNPNREM	CCCCC SEDCL
	370	380	390	400			90	100	110	12
Conf:	3	IIIIIIIIIIIII			Conf:	300000				la a a a a a a a a a a a a a a a a a a
Pred:			<u>}_1</u>	_	Pred:					
Pred: AA:	CCCCCCCCCHHHHHHH YTDWMDENNGLKNRD	HHHHHHCCCE AMDDIVGDHN	EECCHHHHHH VICPLAHFAF	HHHHC	Pred: AA:	CCCCCCCC	SPRPHNL	EEEEEEECCC TVMVWIYGGG	CCCCCCCCCC	CCCCC
	410	420	430	440			130	140	150	16
Conf:	3222200000000000000		1111111111		Conf:					
Pred:			<u> </u>		Pred:		<u> </u>		0	-
Pred: AA:	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCE	EEEEEEECCC YLYLFDHRAS	NLAWP	Pred: AA:	CCCCCCER LAHSETVI	EEEEECC	CCCCCCCCCC	CCCCCCCCHH GSSEAPGNVG	HHHHH LLDQR
	450	460	470	480			170	180	190	20
Conf:	100100000000000000000000000000000000000				Conf:					
Pred:			4)	_	Pred:	-			0	_
Pred: AA:	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCC GLPLEKRLNY	CHHHHHHHHH ILEEEKLSRF	MMRYW	Pred: AA:	HHHHHHH MALQWVQI	HHHHHCCC	CCCCCCEEECC NPKQVTIFGE	CCHHHHHHHH SAGGASVGFH	HCCC
	490	500	510	520			210	220	230	24
Conf:	100000100000000			linet I	Conf:					
Pred:			\rightarrow		Pred:	Ð		0		_
Pred: AA:	HHHHHCCCCCCCCCC	CCCCCCCCCC RGKRWPVFTV	CCCCEEEECC	CCCCCC	Pred: AA:	CHHHHHHH	HHHCCCC	CCCCCCCCCHH NCPWASVSPA	HHHHHHHHH EARRATLLG	HCCC
	530	540	550	560			250	260	270	20

		10	20	30	40
nf:			0-00 011110]]]]]
ed:		11000		Server Scotles	
ed: AA:	EEEEECO	CCCCEEEEE DRSYVTAF	ECCCCCCCCC MGIPFAEPPV	CCCCCCCCCC GRRRFRPAEP	CCCCC KHPWT
		50	60	70	80
f:			00 0000000 00	00000000000]]]]]]]]]
d:					
d: A:	CCCCCCC GVYDANA	CCCCCCCCC	CCCCCCCCCC VDTSFPGFQG	CCCCCCCCCC SEMWNPNREM	CCCCC SEDCL
		90	100	110	12
f:					Jaco C C
d:					
d: A:	CCCCCCC	SSPRPHNL	EEEEEEECCC TVMVWIYGGG	CCCCCCCCCC FYSGSSSLDV	CCCCC
		130	140	150	16
f:					
d:		×		0	
d: A:	CCCCCCE	EEEEEECC	CCCCCCCCCC	CCCCCCCCHH GSSEAPGNVG	HHHHH LLDQR
		170	180	190	20
f:					
d:				0	
d: A:	HHHHHHH MALQWVQ	HHHHHHCCC DNIHFFGG	CCCCCCEEECC NPKQVTIFGE	CCHHHHHHHH SAGGASVGFH	HHCCC
		210	220	230	24
f:					10- 11
d:	Ð		0		
d: A:	CHHHHHH	HHHHHCCCC	CCCCCCCCCHH NCPWASVSPA	HHHHHHHHH EARRATLLG	HHCCC
		250	260	270	28

	Ð			Pred:	YDGRY	YSGSSSLDV	TVMVWIYGGGF
HCCCC	CHHHHHHHHHH NIDEAERQWKV	HHHHHHHCC	CCCCCCHHHHHHHHH HKGLRNQMCAFWNRFI	Pred: AA:			
600	590	580	570			0	
		CCCCCCCC KQERCTDL	HHHHHHHHHHHHHHCO SSYMMHWKSQFDHYSP	Conf: Pred: Pred: AA:	200	190 2200000000	180 180 180 180
		620	610		HHHCCC	CHHHHHHHH AGGASVGFH	CCCCCCEEECCC
			i:	Leger	240	230	220
of predicti	= confidence	: 3	= helix Conf:	0	100-000		
ructure	secondary st	: predicted	<pre>> = strand Pred:</pre>		HHCCC	нняннян	сссссссния
	lence	target sequ	= coil AA: 1	-	GKLVGC	ARRRATLLG	NCPWASVSPAE

4. Serioli dumerili

Conf:					
Pred:	- 0			-	
Pred: AA:	CCCHHHH	HHHHHHHH LLSSIFLL:	HHHHCCCCCC SLLIPTCLSQ	CCCCCEEECC	CEEEE
		10	20	30	40
Conf:	30000000		0×00 000000000000000000000000000000000		l III af
Pred:	->				
Pred: AA:	EEEEECC GIRLPVP	CCCEEEEEI	ECCCCCCCCC MGIPFAEPPV	GRRRFRPAEPI	CCCCC KHPWT
		50	60	70	80
Conf:]======			0000 000000	l annt
Pred:					
Pred: AA:	CCCCCCC GVYDANA	CCCCCCCC	CCCCCCCCCC	CCCCCCCCCCCC SEMWNPNREM:	CCCCC SEDCL
		90	100	110	120
Conf:)]]
Pred:			<u> </u>	and the second	
Pred: AA:	CCCCCCCC	CCCCCCCCCI SSPRPHNL	EEEEEECCC TVMVWIYGGG	CCCCCCCCCC	CCCCC
		130	140	150	160
Conf:	30000		000000000000000000000000000000000000000		
Pred:				0	_
Pred: AA:	CCCCCCE	EEEEEECCO	CCCCCCCCCC	CCCCCCCCHHI GSSEAPGNVG	HHHHH
		170	180	190	200
Conf:					
Pred:				0	
Pred: AA:	HHHHHHH MALQWVQ	HHHHHCCCO DNIHFFGGI	CCCCCEEECC	CCHHHHHHHH SAGGASVGFH	HCCC
		210	220	230	240
Confr		112110			
Drod.					
Prod.	CUUUUUU	ниниссос	COCCCCCHH		HICCC
AA:	SRPTFTR	AILQSGVPI	NCPWASVSPA	EARRRATLLG	KLVGC
		250	260	270	280





Prakasia

						5. <i>M</i> C	onopter	rus albus	S	
e : } I n		00			Conf:) Doool	
·			\rightarrow	_	Pred:		0			
CCC MQI	CHHHHHHHHHHHH MSVLLLLATVLLL	HHHCCCCCCCC	CCCCCEEECCO SEADFIVQTR	CEEEE	Pred: AA:	CCCCHHHHHHHHHCCC NGGNDTELVDCLRNK	CHHHHHHHHH	HCCCCCCCCCC VQVLPWSALFRE	SEVP	
	10	20	30	40		290	300	310	320	
				BE BE	Conf:	3			Baan Baan	
_					Pred:	0				
EEI	EEEECCCCEEEEE	CCCCCCCCCC	CCCCCCCCCCC	CCCCC	Pred:	CCCCCCCCCCCHHHHH	HCCCCCCCCE	EEEEECCCCCE	EEEC (FLLY	
	50	60	70	80		330	340	350	360	
]==:				l l l l l l l l l l l l l l l l l l l	Conf:]]]]]	
_		1.174 - 1.144 - 4.14424			Pred:) - 0		
CCC GV	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCCCCCCC DTSFPGFQG	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	CCCCC SEDCL	Pred: AA:	CCCCCCCCCCCCCCH GAPGFSKDNESLISR	HHHHHHHHHH EDFLEGVKLS	CCCCCHHHHHF SVPHANDIGLE	HHHHH AVVLQ	
	90	100	110	120		370	380	390	400	
				la a a f	Conf:	3			Jan Bart	
_	~ ~				Pred:	0		10		
CCO	CEECCCCCCCCCCC	EEEEEECCC	CCCCCCCCCCC	CCCCC	Pred: AA:	CCCCCCCCCHHHHHH	HHHHHHCCCH	EECCHHHHHH	HHHC	
	130	140	150	160		410	420	430	440	
					Conf:				E	Conf. 19999139913910101010101099110911099
			0		Pred:			~		Pred:
CCO	CCCCEEEEEECCC	CCCCCCCCCC	CCCCCCCCHHI	HHHH	Pred: AA:	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC		EEEEEEECCCC	CCCCC	Pred: CCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
	170	180	190	200		450	460	470	480	570 580 590 600
				1 1 1 1	Conf:					Conf: BUDDEDDDDDDDDDDDDDDDDDDDDDDD
_			-0	<u> </u>	Pred:			Ð		Pred:
HHI MAI	HHHHHHHHHHCCCO LQWVQDNIQFFGGI	CCCCCEEECC	CCHHHHHHHH	ICCCC LLSPG	Pred: AA:	CCCCCCCCCCCCCCC EWMGVIHGYEIEFVF	GLPLEKRLNY	CHHHHHHHHHH	HHHHH MRYW	Pred: CCCHHHHHHHHHHHHHHHHHCCCCCCCCC AA: HRWSSYMMHWKSQFDHYSKQERCTDL
	210	220	230	240		490	500	510	520	610 620
					Conf:				llooof	Legend:
Ð		0			Pred:			\rightarrow	_	- +
CHI	HHHHHHHHHHCCCCC PTFTRAILQSGVP1	CCCCCCCCHH	HHHHHHHHHH	HCCC	Pred: AA:	HHHHHCCCCCCCCC ANFARTGNPNEGLAD	RGKRWPVFT	CCCCCEEEECCC	CCCCC	<pre>= strand Pred: predicted secondary structure</pre>
	250	260	270	280		530	540	550	560	= coil AA: target sequence



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

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

S. No.	Name of organism	Accession no.	G or IV						
			Alpha helix %	Extended strand %	Random coils %				
1	Salveins 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%	21.99%	53.93%				
4	Serioli dumerili	>XP_022603603.1	23.56%	24.36%	52.08%				
5	Monopterus albus	>XP_020450658.1	25.40%	21.88%	52.72%				
6	Esox lucicus	>XP_010898117.1	25.40%	21.88%	52.72%				

Analysis of the structural quality using ProSA Web

Prediction of an active site using CASTp

The quality of the predicted structure was analyzed using ProSA Webby NMR and Xray method. The results are shown in fig. 3.

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).



Score

2-score

Zuscore



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

1. Amphiprion ocellaris





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 pl, 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 alphahelices while *Oreo chromis niloticus* had the lowest. AChE enzyme of *Serioli dumerili* had the highest number of extended strands and *Salveins 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 lucis* had the highest Z score. The result indicates the quality of the enzyme from *Esox lucis* 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 *Serioli dumerili* (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 Ile, Ser and Glu rich protein with mostly acidic pI value. It was classified stable only in *Esox lucius* and also it was with high aliphatic index and least GRAVY value, ensuring its best activity in a wide range of temperatures and better aqueous interactions. The quality prediction affirmed AChE of *Esox lucius* to be of high quality. It was found to be rich in ligand-protein complex sites and contains many dominant amino acid active sites. Evolutionary divergence was also exhibited by some variation in secondary structure features and differences in overall amino acid sequences.

Encompassing within the range of taxa, the present structural analysis provides an evolutionary perspective on the acetylcholinesterase enzyme. Focusing on the common core containing the catalytic site of the enzyme, this analysis is beneficial for future molecular modeling strategies and structural analysis of fish acetylcholinesterase enzyme models.

ACKNOWLEDGMENT

The author thanks 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. Pressy 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 JI. 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 Protein Struct Mol Enzymol 1982;704:52-8.
- Manjunatha KS, Manu CP, Satyanarayan, Nayak KN, Vinay MS, Vineetha, *et al.* Acetylcholinesterase inhibitory effect of 3-(1hindol-3-yl)-1, 3-diphenylpropan-1-one derivatives. Asian J Pharm Clin Res 2017;10:83-6.
- Malathi S, Vidyashree HM, Ravindran R. Restoration of memory and acetylcholinesterase activity by michelia champaca in chronically noise-stressed Wistar Albino rats. Asian J Pharm Clin Res 2016;9:210-4.
- Sussman JL, Harel M, Frolow 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.
- Raves ML, Harel M, Pang YP, Silman I, Kozikowski AP, Sussman JL. Structure of acetylcholinesterase complexed with the nootropic alkaloid (-)-huperzine A. Nat Struct Mol Biol 1997;4:57–63.
- Nolte HJ, Rosenberry TL, Neumann E. Effective charge on acetylcholinesterase active sites determined from the ionic strength dependence of association rate constants with cationic ligands. Biochemistry 1980;1916:3705–11.
- Dougherty DA, Stauffer DA. Acetylcholine binding by a synthetic receptor: implications for biological recognition. Science 1990;250:1558–60.
- 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.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406-25.
- 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.
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381-5.
- 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. Dundas J, Ouyang Z, Tseng J, Binkowski A, 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.
- Lemmon AR, Moriarty EC. The importance of proper model assumption in bayesian phylogenetics. Syst Biol 2004;53:265-77.
- Ikai AJ. Thermostability and aliphatic index of globular proteins. | Biochem 1980;88:1895-8.
- Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105-32.
- Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. EMBO | 1986;54:823-6.