Computers in Biology and Medicine 43 (2013) 817–821

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

An insight into the molecular basis for convergent evolution in fish antifreeze Proteins

Abhigyan Nath [a,1], Radha Chaube [a,*], Karthikeyan Subbiah [b,**]

[a] Zoology/Bioinformatics section, Mahila Mahavidyalaya, Banaras Hindu University, Varanasi-221005, India
[b] Department of Computer Science, Banaras Hindu University, Varanasi-221005, India

ARTICLE INFO

Article history:
Received 10 July 2012
Accepted 23 April 2013

Keywords:
Convergent evolution of fish AFPs
t-test
Amino acid preference/avoidance
Amino acid physicochemical property groups

ABSTRACT

Antifreeze proteins (AFPs) prevent the growth of ice-crystals in order to enable certain organisms to survive under sub-zero temperature surroundings. These AFPs have evolved from different types of proteins without having any significant structural and sequence similarities among them. However, all the AFPs perform the same function of anti-freeze activity and are a classical example of convergent evolution. We have analyzed fish AFPs at the sequence level, the residue level and the physicochemical property group composition to discover molecular basis for this convergent evolution. Our study on amino acid distribution does not reveal any distinctive feature among AFPs, but comparative study of the AFPs with their close non-AFP homologs based on the physicochemical property group residues revealed some useful information. In particular (a) there is a similar pattern of avoidance and preference of amino acids in Fish AFP subtypes II, III and IV—Aromatic residues are avoided whereas small residues are preferred, (b) like other psychrophilic proteins, AFPs have a similar pattern of preference/avoidance for most of the residues except for Ile, Leu and Arg, and (c) most of the computed amino acids in preferred list are the key functional residues as obtained in previous predicted model of Doxey et al. For the first time this study revealed common patterns of avoidance/preference in fish AFP subtypes II, III and IV. These avoidance/preference lists can further facilitate the identification of key functional residues and can shed more light into the mechanism of antifreeze function.

1. Introduction

The environment plays a key role in the structural and functional changes in an organism for their survival in it. It was believed that the extreme cold condition is uninhabitable for any living species because the ice formation in the intra-cellular fluid is lethal to the living cells. However it has been observed that certain fishes, insects, bacteria, fungi, etc. were able to survive in the weather conditions where the temperature dips below the freezing point of their intra-cellular fluids. This leads to the conclusion that these living species have some kind of anti-freeze mechanism within it. The detailed study on anti-freeze mechanism had revealed that the anti-freeze proteins are responsible for the anti-freeze effect. These proteins were first discovered by De Vries in the blood plasma of marine teleosts [1,2]. Antifreeze proteins are a diverse group of proteins which inhibit the growth of ice crystals [3]. These proteins have evolved as an adaptation to cold temperatures and are found in different organisms—fish, insect, bacteria, fungi, plants and diatoms [4,5]. Antifreeze proteins results in non-colligative (i.e. lowering of freezing point is not in proportion to its concentration), non-equilibrium lowering of the freezing point of the extracellular fluids to safe level [6]. It is interesting to observe the diversity of AFPs in fishes. The fish AFPs have been classified into five distinct types namely AFGP (Antifreeze Glycoprotein), Type I, Type II, Type III and Type IV based on their origins and properties [7]. These are unrelated and possess distinct characteristics both in terms of structure as well as sequence composition though all of them perform the same antifreeze function.

It has been speculated that different fish AFPs have each evolved independently as an adaptation to progressively cooling environmental conditions [8,9]. This is supported by their wide diversity and distribution across the fishes. Different proteins must have evolved to perform the antifreeze function in order to survive and resist the adverse effects of freezing.

Research studies on AFPs have shown that AFP types are radically different in their primary sequences and tertiary structures and yet they all bind to ice and depress the freezing point below the melting point which is also known as thermal hysteresis.

In this study, we have attempted to infer the common similarities at the residue level between convergently evolved fish AFPs. To achieve this, AFP protein sequences were analyzed with their

* Corresponding author. Tel.: +91 9336847252.
** Corresponding author. Tel.: +91 9473967721.
E-mail addresses: abhigyannath01@gmail.com (A. Nath), radhachaube72@gmail.com (R. Chaube), karthinikita@gmail.com (K. Subbiah).
[1] Tel.: +91 9956015187.

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.04.013

homologous non-antifreeze proteins. With such study one can get insight into the properties of the selective adaptation process which has lead to the molecular evolutionary process.

2. Materials and Methods

2.1. Dataset

Two different datasets were created for our study. First dataset was created to analyze AFPs at the residues level to discover any common features present among them and comprises of all the five subtypes of fish AFPs. The second dataset was created to discover specific changes that occurred in the evolutionary process and comprises of all the fish AFP subtypes together with their respective close non-AFP homologs.

2.1.1. Data set for residue level analysis

For the In Silico characterization of different fish AFP subtypes, we have downloaded fish AFP sequences from the UniProt [10] database and used CD-HIT [11] to cluster the sequences at 80% sequence identity in order to remove very similar sequences that could bias our result.

2.1.2. Data set for evolutionary change and common feature analysis

In order to study the common theme of adaptation (preference/avoidance) at the level of amino acids and physicochemical grouping of residues, we created this dataset consisting of positive examples (sequences of specific subtypes of fish AFPs) with negative examples (homologous non AFPs). Out of five different fish AFP subtypes only three of them show significant homology with proteins of other families on a similarity search on the current non-redundant database of proteins at NCBI. We created three sets of sequences each for type II, III and IV consisting of AFP sequences with its homologous non AFP sequences. The sequences of type II, III and IV were taken from the first non-redundant dataset. We used BLAST [12] for each of the sequences of fish AFP subtypes and obtained the highest scoring non-AFP homolog. If the same non-AFP homolog was obtained as highest scoring for different AFP sequences, then it is paired with the AFP with which it is having high identity based on maximum score of BLAST result.

2.2. t-test data analysis

In our study we used *t*-test to compare the mean frequencies of single amino acids and 11 different property groups at 5% significance level between the type II and its homologous non-AFP counterpart C-lectin proteins (degrees of freedom (df)=12), likewise type III with SAF superfamily proteins (degrees of freedom (df)=4) and type IV with apolipoproteins (degrees of freedom (df)=6).

The *t*-test is used to compare two small sets of quantitative data when samples are collected independently of one another and essentially a good tool for comparing the two independent groups of data. This test determines a probability that two populations are the same with respect to the variable tested.

2.3. Sequence Parameters

We have chosen few sequence parameters like percentage frequency of occurrence of all 20 amino acids and frequency percentage of 11 amino acid property groups for our study.

The property groups that we have selected for our study are Tiny amino acids group (Ala, Cys, Gly, Ser, Thr), Small amino acids group (Ala, Cys, Asp, Gly, Asn, Pro, Ser, Thr and Val), Aliphatic amino acids group (Ile, Leu and Val.), Non-polar amino acid groups (Ala, Cys, Phe, Gly, Ile, Leu, Met, Pro, Val, Trp and Tyr), Aromatic amino acid group (Phe, His, Trp and Tyr), Polar amino acid group (Asp, Glu, His, Lys, Asn, Gln, Arg, Ser, and Thr), Charged amino acid group (Asp, Glu, His, Arg, Lys), Basic amino acid group (His, Lys and Arg), Acidic amino acid group (Asp and Glu), Hydrophobic acid group (Ala, Cys, Phe, Ile, Leu, Met, Val, Trp, Tyr), Hydrophilic acid group (Asp, Glu, Lys, Asn, Gln, Arg) [13,14]. Here some of the amino acids fall in more than one property groups.

The sum of frequencies of amino acids that fall in each property groups are calculated for all the fish AFPs subtypes with their corresponding homologous non AFP protein groups for comparative analysis.

3. Results and discussion

3.1. Common feature analysis

Fig. 1 shows the distribution of amino acids in fish AFP subtypes against the background frequencies [15]. It is clear from the figure that different subtypes of fish AFPs have diverse sequence composition. The mean compositions per residue for the different subtypes of AFPs are different from background frequencies notably for Alanine, Arginine, Glutamic acid, Glycine, Threonine, and Tyrosine. It gives a panoramic view of the sequence diversity existing among the fish AFP subtypes.

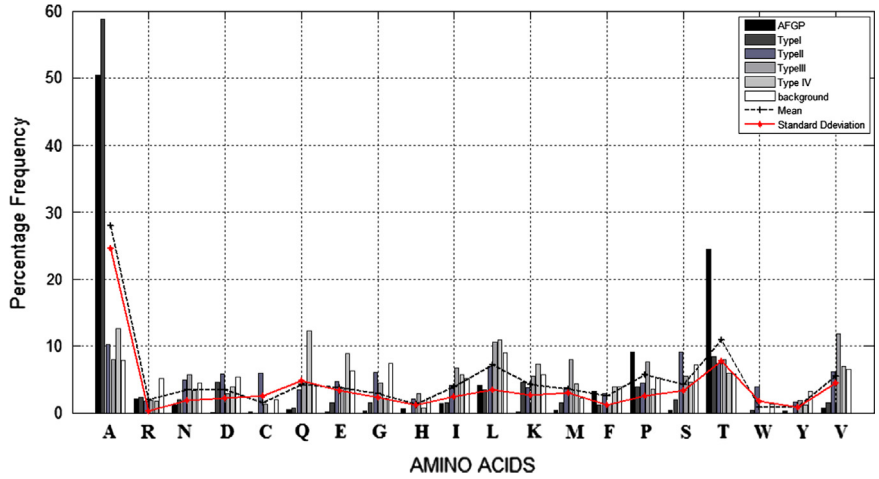

Fig. 1. Showing the percentage frequency of amino acids in fish AFPs suptypes alongwith background percentage frequency.