ELSEVIER

afpCOOL: A tool for antifreeze protein prediction

Morteza Eslami [a], **Ramin Shirali Hossein Zade** [b], **Zeinab Takalloo** [c],
Ghasem Mahdevar [d], **Abbasali Emamjomeh** [e], **Reza H. Sajedi** [c], **Javad Zahiri** [f,g,*]

[a] *Department of Computer Engineering, Arak University, Arak, Iran*

[b] *Department of Computer Engineering, Sharif University of Technology, Tehran, Iran*

[c] *Department of Biochemistry, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

[d] *Department of Mathematics, Faculty of Sciences, University of Isfahan, Isfahan, Iran*

[e] *Laboratory of Computational Biotechnology and Bioinformatics (CBB), Department of Plant Breeding and Biotechnology (PBB), Faculty of Agriculture, University of Zabol, Zabol, Iran*

[f] *Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran*

[g] *School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), P.O. Box 19395-5746, Tehran, Iran*

* Corresponding author.
E-mail address: zahiri@modares.ac.ir (J. Zahiri).

Abstract

Various cold-adapted organisms produce antifreeze proteins (AFPs), which prevent the freezing of cell fluids by inhibiting the growth of ice crystals. AFPs are currently being recognized in various organisms, living in extremely low temperatures. AFPs have several important applications in increasing freeze tolerance of plants, maintaining the tissue in frozen conditions and producing cold-hardy plants by applying transgenic technology. Substantial differences in the sequence and structure of the AFPs, pose a challenge for researchers to identify these proteins. In this paper, we proposed a novel method to identify AFPs, using supportive vector machine (SVM) by incorporating 4 types of features. Results of the two used benchmark datasets, revealed the strength of the proposed method in AFP prediction. According to the results of an independent test setup, our method outperformed the current state-of-the-art methods. In addition, the comparison results of the discrimination power of different feature types revealed that

physicochemical descriptors are the most contributing features in AFP detection. This method has been implemented as a stand-alone tool, named afpCOOL, for various operating systems to predict AFPs with a user friendly graphical interface.

Keywords: Computer science, Bioinformatics, Computational biology, Mathematical biosciences, Biochemistry

1. Introduction

Organisms, which are exposed to the freezing conditions, produce special proteins, called antifreeze proteins (AFPs) [1, 2]. AFPs bind to small ice crystals to forestall additional crystallization and depress the freezing point of the solution below the melting point [3, 4, 5, 6]. Additional terms have also been proposed in the literature to name the AFPs: ice structuring proteins and thermal hysteresis proteins [4, 7].

For the first time, AFPs were found in the species of fish and insect that had adapted to extremely low temperatures [3, 8], and the structures of 5 structurally distinct AFPs were identified by Davies and Hew in 1990 [9]. AFPs have also been found in fungi, bacterial species and overwintering plants [10, 11, 12]. AFPs have potential applications in preservation, gene transformation, cryosurgery of tumors, and also in agriculture to produce economically efficient fishes and plants, resisting against extremely low temperatures [13, 14].

Recently, two computational approaches have been proposed to discriminate AFPs from non-AFPs [15, 16], but these methods do not have a satisfactory performance. By considering the substantial differences in the sequence and structure of AFPs [17, 18], it is needed to use suitable machine learning methods to predict AFPs. An appropriate machine learning algorithm offers a cost-effective approach to construct predictive models to identify AFPs by exploiting experimentally validated training data. In this article, we proposed a computational method to identify AFPs, which achieved an accuracy of 95% and 91% on two benchmark datasets and the accuracy of 96% on an independent test dataset.

2. Methods

2.1. Dataset

In this study, we had used two independent datasets (Fig. 1) to evaluate the strength of the predictor and also compare the prediction performance of the proposed method with the current state-of-the-art methods.
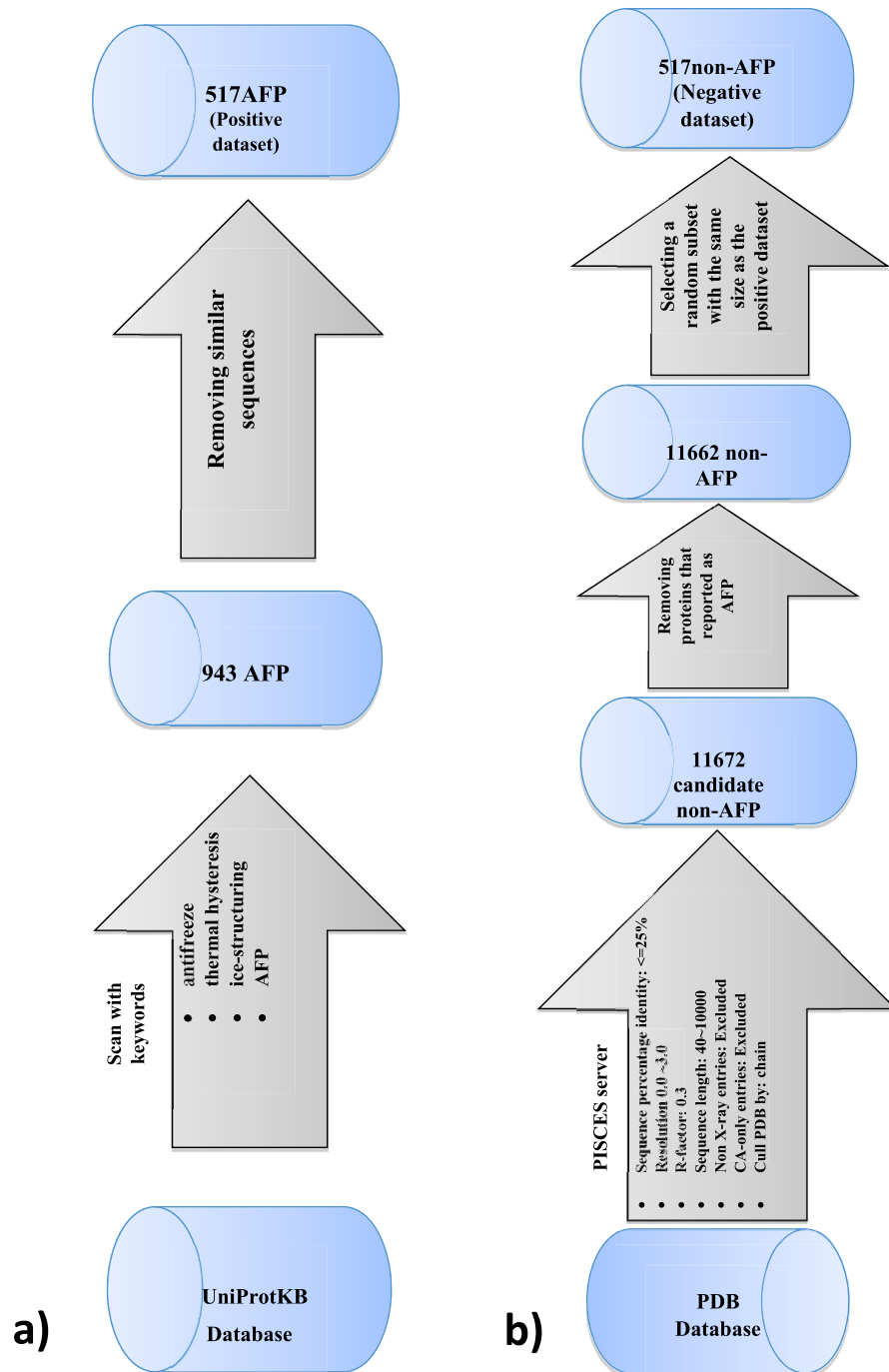
Fig. 1. Schematic representation of AFP517 dataset construction. a) Positive dataset contains 843 AFPs. b) Negative dataset contains 843 non-AFPs.

2.1.1. AFP481 dataset

The first benchmark dataset (AFP481) was extracted from Kandaswamy et al. [15]. This dataset contains 481 AFPs and 9493 non-AFPs, which were used for the

positive and the negative datasets, respectively. In this dataset, the proteins with $\geq 40\%$ sequence similarity were omitted using CD-HIT [19]. Train dataset consists of 300 out of all the 481 AFPs and 300 out of all the 9493 non-AFPs, which had been selected randomly for each positive and negative datasets, respectively. Also, the remaining 181 AFPs and 9193 non-AFPs were used as an independent test dataset.

2.1.2. AFP517 dataset

For a better assessment of the strength of the proposed method, a more comprehensive benchmark dataset, named AFP517, was assembled as following (Fig. 1b). Antifreeze protein sequences were extracted from UniProtKB [20]. For this goal, UniProtKB was scanned with a list of keywords, implying antifreeze proteins. The "antifreeze", "thermal hysteresis", "ice-structuring", and "AFP" were used as keywords. We retrieved 943 proteins in total, where the number of proteins corresponding to "antifreeze", "thermal hysteresis", "ice-structuring" and "AFP" keywords were 734, 22, 52, and 135, respectively. Finally, the proteins with $\geq 90\%$ sequence similarity were removed using CD-HIT, which resulted in 517 AFP proteins, used as positive instances.

To select negative examples (non-AFP proteins), we took advantage of PISCES [21] as a public server to cull sets of protein sequences from the Protein Data Bank (PDB) [22] by determining sequence identity and structural quality criteria (Fig. 1). We used the same number of AFP and non-AFP proteins to construct a balanced dataset.

2.2. Features

We trained our model to detect AFPs by exploiting four types of descriptors including hydropathy (3 descriptors), physicochemical properties (218 descriptors), amino acid composition (20 descriptors), and evolutionary profile (400 descriptors).

2.2.1. Hydropathy descriptors

According to the hydropathy, 20 amino acids were categorized into 3 feature groups as following: strongly hydrophilic (RDENQKH), strongly hydrophobic (LIVAMF), and weakly hydrophilic or weakly hydrophobic (STYW). For each feature of these three groups in a protein sequence, the number of occurrences of each group was computed and divided by the length of the sequence.

2.2.2. Physicochemical descriptors

To compute physicochemical descriptors, 544 different physicochemical indices were extracted from AAINDEX database [23], which is a database of numerical indices representing various physicochemical and biochemical properties of

amino acids. For reducing the biases, which can be a result of having many correlated indices, we considered each pair of indices with a correlation coefficient greater than 0.8 and less than −0.8 as redundant indices. Finally, a subset of 218 non-redundant indices were selected to encode proteins. More precisely, the 20-dimensional amino acid composition vector was multiplied by the AAINDEX matrix (dimension: 218×20), resulted into a 218-dimensional feature vector.

2.2.3. Amino acid composition

Each protein was encoded by a 20-dimensional feature vector, indicating the amino acid composition of the protein. Each element of this vector indicates the frequency of an amino acid of the protein sequence.

2.2.4. Evolutionary information

It has been shown that evolutionary information is effective in detecting diverse properties of proteins [16, 24, 25]. We used evolutionary information of the protein in the form of Position-Specific Scoring Matrix (PSSM). Position Specific Iterated BLAST (PSI-BLAST) was used against the NCBI non-redundant dataset with three iterations and the e-value of 0.0001, to generate PSSMs for all the proteins. Regarding substitution scores in the PSSMs, each protein was encoded as a 400-dimensional feature vector. Each element of this vector is the sum of all positive substitution scores of an amino acid to one of the twenty standard amino acids.

2.2.5. Normalization

In all of the four above-mentioned descriptor types, every element was normalized according to the length of the sequence to avoid the bias, which can be the result of different sequence lengths.

2.3. Support vector machines

In recent years, support vector machine (SVM) has been widely used in various prediction problems in bioinformatics [26]. SVM classifies the input samples, represented in the form of n-dimensional feature vectors, into two classes using an optimal hyper-plane in the feature space. In this study, the input of the SVM classifier is a 641-dimensional vector that encodes the features of a given protein, and the output is a binary label, indicating whether the given protein is an AFP or not. We used the SVM implementation by means of WEKA package [27] with Pearson VII function-based universal kernel (PUK).

2.4. Evaluation parameters for the prediction performance

A 10 fold cross-validation (10-fold CV) approach was used to evaluate the performance of the proposed prediction model. Positive and negative instances were distributed randomly into 10 folds. In each of the 10 iterative steps, 9 of the 10 folds were used to train the classifier, and then the classifier was evaluated by using the remaining data (test data). The predictions made for the test instances in all the 10 iterations were combined and used to compute the following performance measures:

$$\text{Sensitivity (or Recall)} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{MCC} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$\text{F} - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where, TP and TN are the correctly predicted AFP and non-AFP instances, respectively. Similarly, FP and FN are the number of proteins, wrongly predicted as AFP and non-AFP. In addition to the mentioned measures, we used the receiver operating characteristic (ROC) curve, which is an important graphical tool for assessing the classification performance. ROC plots sensitivity (true positive rate) against false positive rate and shows the trade-off between sensitivity and specificity. We also used the area under the ROC curve (AUC), as a reliable performance measurement. For further assessment of the afpCOOL, in addition to the mentioned 10-fold cross validation, we also obtained the prediction results by applying leave-one-out-cross-validation procedure (LOOCV). LOOCV is similar to the 10-fold CV but the number of folds is equal to the number of instances.

3. Results and discussion

We compared afpCOOL with two recently published methods, to show the strength of the proposed method regarding the current state-of-the-art methods: AFP-Pred and AFP-PSSM [16]. The prediction results of afpCOOL in the two benchmark

datasets, AFP481 and AFP517, and also an independent test set are described in the following.

3.1. Results on the AFP481 dataset

Table 1 shows the most important prediction performance measures for afpCOOL on the AFP481 dataset. This dataset contains 300 AFPs and 300 non-AFPs. According to the 10-fold cross validation results, our method performed well in AFP detection: the method achieved an accuracy rate of 93% with an f-measure of 93%.

3.2. Results on the AFP517 dataset

This dataset contains 1668 proteins with equal number of positive (AFPs) and negative (non-AFPs) instances. The various performance measures in a 10-fold cross validation procedure revealed the strength of afpCOOL (Table 1). Our method achieved an accuracy of 91.9%, precision and sensitivity of 92% with an MCC of 84%. Fig. 2 shows the ROC curve of the method on the two mentioned benchmark datasets. The proposed model performed better on the larger dataset (AFP517 dataset), as these curves disclose. It was clear that the results of LOOCV show the robustness of the method.

3.3. Comparison with the current state-of-the-art methods

For further evaluation of our method, an independent test set was used to compare the method with the current state-of-the-art methods. We used a test dataset, which has been recently exploited by Zhao et al [16] as the benchmark for comparison purposes; this dataset contains 181 AFPs and 9193 non-AFPs. We have trained our model with the same data that has been used by the two competitor methods (AFP481), to have a fair comparison.

As Table 2 shows, AFP-Pred performed better in sensitivity measurement, but on the other hand it's the worse method according to the specificity and accuracy. The

Table 1. Prediction performance of the afpCOOL on two benchmark datasets in a 10-fold cross validation (10-fold CV) and leave-one-out (LOOCV) procedures. The AFP481 dataset contains 300 AFPs and 300 non-AFPs; and, the AFP517 dataset contains 517AFPs and 517non-AFPs.

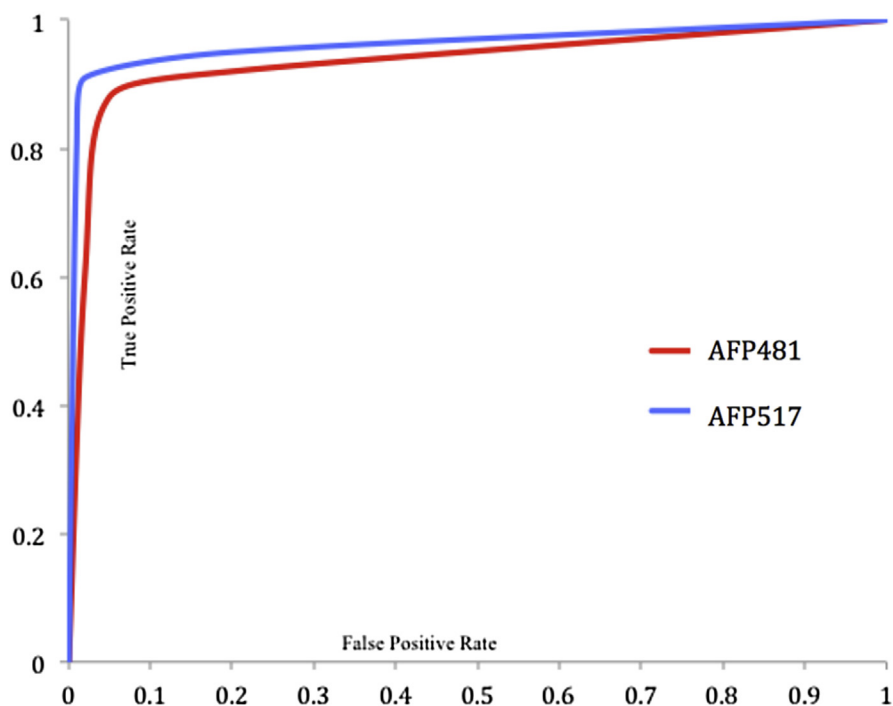| Dataset		Performance measures					
		Accuracy	**Precision**	**Sensitivity**	**F-Measure**	**MCC**	**AUC**
10-Fold CV	AFP481	0.93	0.93	0.93	0.93	0.87	0.93
	AFP517	0.91	0.92	0.92	0.92	0.84	0.92
LOOCV	AFP481	0.89	0.90	0.88	0.86	0.78	0.91
	AFP517	0.91	0.90	0.89	0.90	0.81	0.90

Fig. 2. The receiver operating characteristic (ROC) curves of our method on the two benchmark dataset calculated from the ten-fold cross validation. The AFP481 dataset contains 300 AFPs and 300 non-AFPs; and, the AFP517 dataset contains 517AFPs and 517non-AFPs.

Table 2. Performance comparison of the proposed method (afpCOOL) with the two current state-of-the-art methods in AFP prediction. All methods are trained on the AFP481 dataset in a 10-fold cross validation procedure and tested on an independent test dataset with 181 AFP and 9193 non-AFPs.

Method	Performance Measure		
	Sensitivity	Specificity	Accuracy
afpCOOL	0.72	0.98	0.96
AFP-Pred (Griffith and Yaish 2004)	0.85	0.82	0.83
AFP-PSSM (Zhao et al. 2012)	0.76	0.93	0.93

accuracy of afpCOOL (96%) is higher than the accuracy rate of AFP-Pred (83%) and AFP-PSSM (93%). It was clear that our method outperformed the other methods in terms of specificity.

3.4. afpCOOL tool

We have developed the afpCOOL as a tool, enabling fast *in-silico* AFP detection. This tool has been implemented as a stand-alone java application for various operating systems with a user-friendly graphical interface (Fig. 3). Users can use the
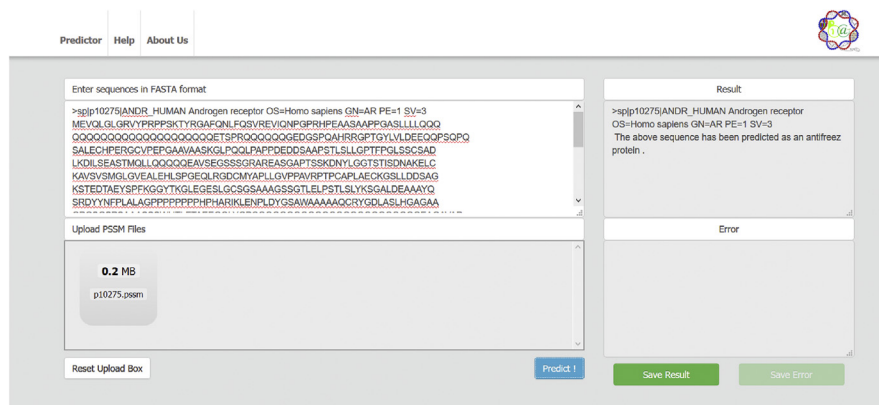
Fig. 3. Graphical user interface of afpCOOL.

afpCOOL simply by providing sequences (in fasta format) and PSSMs of the interested proteins; afpCOOL extracts features from the provided inputs and then uses the trained SVM-based model to assign AFP or non-AFP labels to the queries. The afpCOOL is freely downloadable for non-commercial use at http://biocool.ir.

3.5. Need for a reliable computational tool for AFP prediction: BLAST cannot effectively detect the AFPs

We ran BLAST by using each of the AFPs as the query against the last update (August 12, 2015) of the UniProt Archive (UniParc) database with e-value ≤1e-3, to show the strength of the afpCOOL. It should be mentioned that 801 proteins have been mapped to UniParc identifiers, and as a result, 801 BLAST searches were used for this analysis. More details were presented in Fig. 4.a. One can see that the BLAST does not have a satisfactory specificity in AFP detection; more than 62% (539 out of 801) and more than 87% of the BLAST searches had specificity
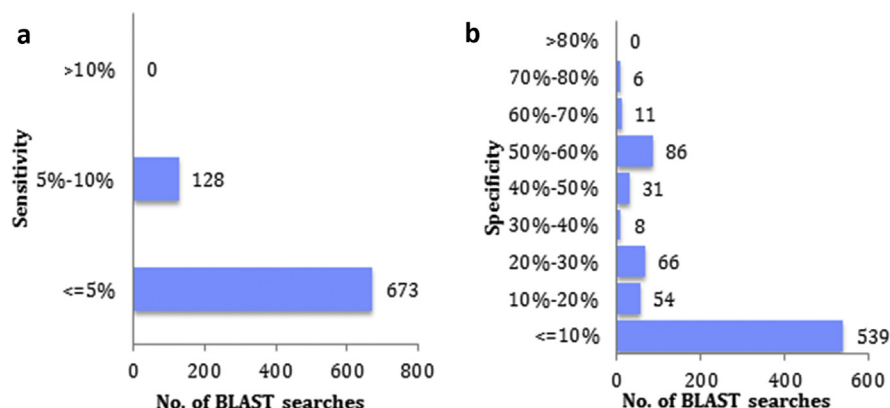

Fig. 4. Sensitivity (a) and specificity(b) of the BLAST when using each of the 801 AFPs as the query against the last update.

rates of less than 10% and ≤50%, respectively. Considering the large dataset, used for BLAST searches, it may be expected to have a low specificity and very good sensitivity. Nevertheless, by considering the sensitivity of the BLAST searches, Fig. 4.b, the condition is worse than the specificity: there is no BLAST search

**Table 3.** Anti-freeze proteins with no any hit in the BLAST search (35 proteins).

Proteins (UniProt ID)	Organism
Q9DF23; P04367; P04368	Myoxocephalus scorpius (Shorthorn sculpin) (Cottusscorpius).
Q1AMR6; Q1AMR7	Parachaenichthyscharcoti (Charcot's dragonfish) (Chaenichthyscharcoti).
Q1AMR9; Q9DFU2	Gymnodraco acuticeps (Antarctic dragonfish).
P02733; P02734	Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
P20421; P20617	Myoxocephalus aenaeus (Grubby sculpin) (Cottusaenaeus).
P11920; P11921	Eleginus gracilis (Saffron cod) (Gadusgracilis).
Q9S8C6; Q9S8C5	Secale cereale (Rye).
Q90402	Dissostichus mawsoni (Antarctic cod).
Q8JHE3	Notothenia microlepidota.
Q1AMR4	Chaenocephalus aceratus (Blackfinicefish) (Chaenichthysaceratus).
Q6JIC7	Liparis atlanticus (Atlantic seasnail).
F8UWP2	Tautogolabrus adspersus (Cunner).
Q1AMS2	Pogonophryne cerebropogon.
Q9DF18	Myoxocephalus octodecemspinosus (Longhorn sculpin) (Cottus octodecemspinosus).
Q6JIC6	Liparis gibbus (variegated snailfish).
B3EWE8	Trapa natans (Water chestnut).
P84794	Solanum tuberosum (Potato).
P02732	Pagothenia borchgrevinki (Bald rockcod) (Trematomusborchgrevinki).
P86268	Antarctomycespsychrotrophicus.
Q3HYD3	Tenebrio molitor (Yellow mealworm beetle).
D7PBP2	Hypogastrura harveyi.
Q9S9D9	Nicotiana tabacum (Common tobacco).
H0SH19	Bradyrhizobium sp. ORS 375.
P85102	Cullen corylifolia (Malaysian scurfpea) (Psoralea corylifolia).
K0P291	Cardinium endosymbiont cEper1 of Encarsia pergandiella.
B4RIE8	Phenylobacterium zucineum (strain HLK1).
Q091Z2	Stigmatella aurantiaca (strain DW4/3-1).

Table 4. Anti-freeze proteins with 0% sensitivity in the BLAST search (11 proteins).

Proteins (UniProt ID)	Organism
J7HY61; H2KMI2; H2KMI4; J7I1U4; J7I1K6	Boreogadus saida (Polar cod).
A8P5U5	Coprinopsis cinerea (strain Okayama-7/130/ ATCC MYA-4618/FGSC 9003)
C6KF34	Ammopiptanthusnanus.
F2XFX1	Dissostichus mawsoni (Antarctic cod).
Q1AMR5	Chaenocephalus aceratus (Blackfinicefish) (Chaenichthys aceratus).
H2KMI6	Gadus ogac (Greenland cod).
Q1AMR8	Gymnodraco acuticeps (Antarctic dragonfish).

with sensitivity higher than 10%; and also, 84% of the BLAST searches have sensitivity $\leq 5\%$ (673 search out of 801 BLAST searches).

Interestingly, there are 35 proteins without any search results in the BLAST search and 11 proteins with a sensitivity rate of 0%. These proteins are listed in Tables 3 and 4, respectively. As it can be clear 5 out of 11 (45%) proteins with sensitivity rates of 0% were extracted from *Boreogadus saida* (Polar cod). In addition, the proteins of these two categories have a significantly different amino acid composition comparing to the other AFPs. As more details were presented in Fig. 5, alanine
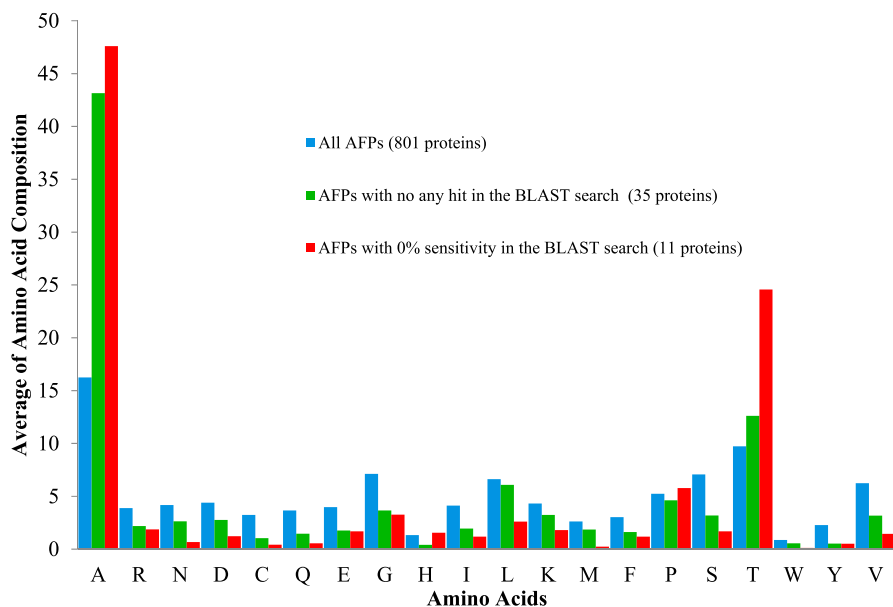
Fig. 5. Amino acid composition of AFP proteins that have been used in the BLAST searches.

was the major building blocks of these AFPs (43% in the AFPs without any hit in the BLAST search and 47% in the AFPs with 0% sensitivity in the BLAST search).

3.6. Feature importance

We have integrated the scores of three feature selection (FS) methods including, gain ration, information gain and PCA, to select the most important features. To use the scores obtained from the two first FS methods, scores have been normalized into the range $0-1$. For PCA, $ev_i \times w_i$ for each feature was used as a score, where ev_i is the Eigenvalue of the ith PC and w_i is the weight of the feature in the corresponding Eigenvector. Then, the PCA scores was also normalized into the range $0-1$. Finally, the sum of these three normalized scores, have been used as the feature importance score. Fig. 6 shows the contribution of each feature type in the 100 most important features. As it is apparent, the physicochemical descriptors are the most important features.

According to the top 10 important features, presented in Table 5, physicochemical descriptors are the most informative features to discriminate between AFPs and non-AFPs. The importance of physicochemical properties of AFPs' amino acids
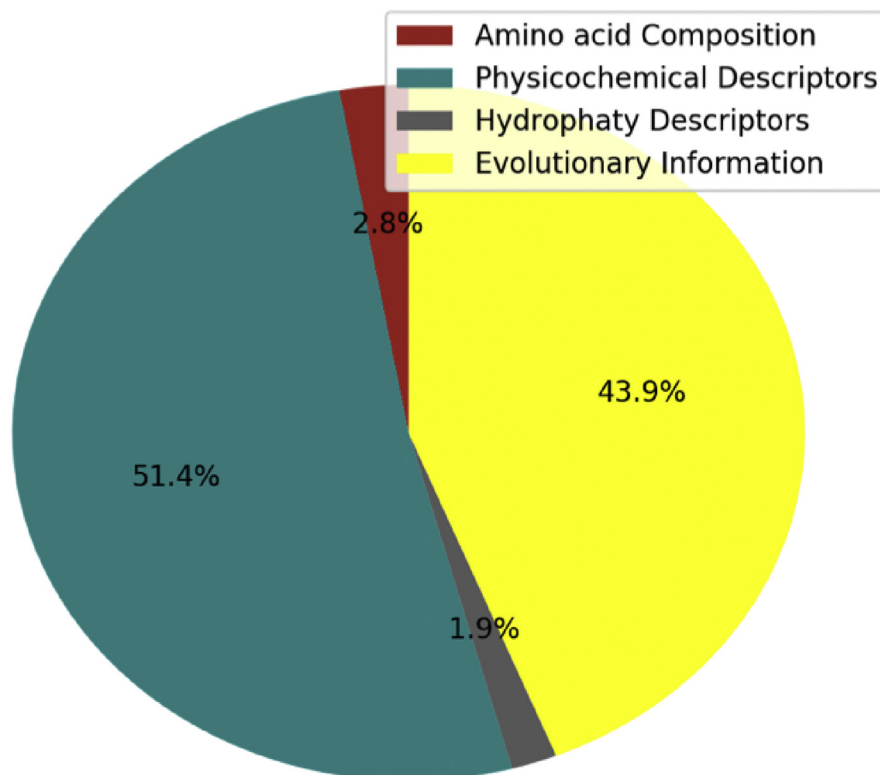

Fig. 6. The percent of each feature type score regarding the total score of the 100 most important features.

Table 5. The 10 most important features.

Rank	Feature Type	Feature's Detail
1	Physicochemical	The number of atoms in the side chain labeled 2 + 1 (AAIndex Code: CHAM830104)
2	Physicochemical	Normalized positional residue frequency at helix termini C (AAIndex Code: AURR980118)
3	Physicochemical	Average non-bonded energy per residue (AAIndex Code: OOBM770104)
4	Physicochemical	Free energy change of epsilon(i) to alpha(Rh) (AAIndex Code: WERD780104)
5	Physicochemical	The number of bonds in the longest chain (AAIndex Code: CHAM830106)
6	PSSM-based	H to H
7	Physicochemical	Normalized positional residue frequency at helix termini C4 (AAIndex Code: AURR980120)
8	PSSM-based	H to G
9	Physicochemical	Relative population of conformational state C (AAIndex Code: VASM830102)
10	Physicochemical	Loss of Side chain hydropathy by helix formation (AAIndex Code: ROSM880103)

has been suggested in many studies. For example, it was suggested that AFPs mainly bind to ice surfaces through hydrogen bonding or polar interaction between water molecules and hydrophilic side chains of Thr, Gln and Glu [28, 29]. However, a number of mutagenesis studies revealed that hydrogen bonds may not be necessarily essential for the ice surface binding [18, 28, 30]. Also, there were reports that the ice-binding sites of AFPs mainly consist of hydrophobic residues [31]. However, it has been shown that the hydrophilic residue of serine (Ser) may be partially responsible for its inferior antifreeze activity [18, 30].

In addition, the existence of α-helical structure is one the most important characteristics of the different AFP types [3, 18, 28, 32, 33]. In agreement with the mentioned AFPs' property, four of the most informative features (out of ten) were related to the helix structures.

4. Conclusion

We developed a novel SVM-based method to predict AFPs. In this method, each protein has been encoded by four features (evolutionary profile, amino acid composition, Hydropathy, and physicochemical properties). The results indicated that these types of features can significantly improve the prediction of AFPs. The obtained

results on two benchmark datasets revealed the strength of our method in AFP detection. Also, the results on an independent test set confirmed the better performance of the proposed method regarding the current state-of-the-art AFP prediction methods. In addition, the more analysis we did, disclosed the poor performance of BLAST in AFP detection and so indicates the critical need for an accurate tool for this purpose. Finally, evolutionary profile and amino acid composition showed the most applicable power in discriminating AFPs from non-AFPs.

Declarations

Author contribution statement

Morteza Eslami: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Ramin Shirali Hossein Zade: Performed the experiments; Analyzed and interpreted the data.

Zeinab Takalloo, Emamjomeh Abbasali: Analyzed and interpreted the data.

Ghasem Mahdevar, Reza H. Sajedi: Analyzed and interpreted the data; Wrote the paper.

Javad Zahiri: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

Competing interest statement

The authors declare no conflict of interest.

Additional information

afpCOOL is freely available at http://bioinf.modares.ac.ir:8080/AFPCOOL/page/afpcool.jsp.

References

[1] S. Jørgensen, S. Keskin, D. Kitsios, P. Commerou, B. Nilsson, Antifreeze Proteins - the Applications of Antifreeze Proteins in the Food Industries, 2008.

[2] L.A. Graham, R.S. Hobbs, G.L. Fletcher, P.L. Davies, Helical antifreeze proteins have independently evolved in fishes on four occasions, PLoS One 8 (2013), e81285.

[3] P.L. Davies, J. Baardsnes, M.J. Kuiper, V.K. Walker, D. Hall, M.A. Marahiel, M. Smallwood, D. Smith, T. Haymet, C. Knight, Structure and function of antifreeze proteins, Philos. Trans. R. Soc. B Biol. Sci. 357 (2002) 927−935.

[4] R. Drori, P.L. Davies, I. Braslavsky, Experimental correlation between thermal hysteresis activity and the distance between antifreeze proteins on an ice surface, RSC Adv. 5 (2015) 7848−7853.

[5] R. Drori, P.L. Davies, I. Braslavsky, When are antifreeze proteins in solution essential for ice growth inhibition? Langmuir 31 (2015) 5805−5811.

[6] Y. Yeh, R.E. Feeney, Antifreeze proteins: structures and mechanisms of function, Chem. Rev. 96 (1996) 601−618.

[7] S. Venketesh, C. Dayananda, Properties, potentials, and prospects of antifreeze proteins, Crit. Rev. Biotechnol. 28 (2008) 57−82.

[8] G.L. Fletcher, S.V. Goddard, Y.L. Wu, Antifreeze proteins and their genes: from basic research to business opportunity, Chemtech 29 (1999) 17−28.

[9] P.L. Davies, C.L. Hew, Biochemistry of fish antifreeze proteins, FASEB J. 4 (1990) 2460−2468.

[10] C.H.C. Cheng, Evolution of the diverse antifreeze proteins, Curr. Opin. Genet. Dev. 8 (1998) 715−720.

[11] K.V. Ewart, Q. Lin, C.L.C.N.-C. Hew, Structure, function and evolution of antifreeze proteins, Cell Mol. Life Sci. 55 (1999) 271−283.

[12] J.M. Logsdon, W.F. Doolittle, Origin of antifreeze protein genes - a cool tale in molecular evolution, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 3485−3487.

[13] F.S.A.N. Zealand, Ice structuring protein as a processing aid in ice cream and edible ices: a safety assessment / Food Standards Australia New Zealand, Final Assess. Rep. (2005) 1−92.

[14] R. Wang, P. Zhang, Z. Gong, C.L. Hew, Expression of the antifreeze protein gene in transgenic goldfish (Carassius auratus) and its implication in cold adaptation, Mol. Mar. Biol. Biotechnol. 4 (1995) 20−26. http://www.ncbi.nlm.nih.gov/pubmed/7749462.

[15] K.K. Kandaswamy, K.C. Chou, T. Martinetz, S. Möller, P.N. Suganthan, S. Sridharan, G. Pugalenthi, AFP-Pred: a random forest approach for

predicting antifreeze proteins from sequence-derived properties, J. Theor. Biol. 270 (2011) 56–62.

[16] X. Zhao, Z. Ma, M. Yin, Using support vector machine and evolutionary pro-files to predict antifreeze protein sequences, Int. J. Mol. Sci. 13 (2012) 2196–2207.

[17] M. Griffith, W.F.M. Yaish, Antifreeze proteins in overwintering \nplants: a tale of two activities, Trends Plant Sci. 9 (2004) 399–405.

[18] Z. Jia, P.L. Davies, Antifreeze proteins: an unusual receptor-ligand interaction, Trends Biochem. Sci. 27 (2002) 101–106.

[19] W. Li, L. Jaroszewski, A. Godzik, Clustering of high homologuous sequences to reduce the size of large protein database, Bioinformatics 17 (2001) 282–283.

[20] A. Bairoch, R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, L.S.L. Yeh, The universal protein resource (Uni-Prot), Nucleic Acids Res. 33 (2005) D154–D159.

[21] G. Wang, R.L. Dunbrack Jr., PISCES : a protein sequence culling server, Bio-informatics 19 (2003) 1589–1591.

[22] H.M. Berman, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[23] S. Kawashima, H. Ogata, M. Kanehisa, AAindex: amino acid index database, Nucleic Acids Res. 27 (1999) 368–369.

[24] J. Zahiri, M. Mohammad-noori, R. Ebrahimpour, S. Saadat, J.H. Bozorgmehr, T. Goldberg, A. Masoudi-nejad, LocFuse: human protein − protein interaction prediction via classifier fusion using protein localization information, Geno-mics 104 (2014) 4–11.

[25] J. Zahiri, O. Yaghoubi, M. Mohammad-noori, R. Ebrahimpour, A. Masoudi-nejad, Genomics PPIevo: protein − protein interaction prediction from PSSM based evolutionary information, Genomics 102 (2013) 237–242.

[26] J. Zahiri, J. Bozorgmehr, A. Masoudi-Nejad, Computational prediction of protein–protein interaction networks: algorithms and resources, Curr. Genom. 14 (2013) 397–414.

[27] H.W.I. Hall Mark, Eibe Frank, Geoffrey Holmes, Pfahringer Bernhard, Peter Reutemann, The WEKA data mining software: an update, SIGKDD Ex-plor. 11 (2009) 10–18.

[28] F.P. Cid, J.I. Rilling, S.P. Graether, L.A. Bravo, M. De La Luz Mora, M.A. Jorquera, Properties and biotechnological applications of ice-binding proteins in bacteria, FEMS Microbiol. Lett. 363 (2016) 1−12.

[29] R.K. Kar, A. Bhunia, Biophysical and biochemical aspects of antifreeze proteins: using computational tools to extract atomistic information, Prog. Biophys. Mol. Biol. 119 (2015) 194−204.

[30] J.G. Duman, Animal ice-binding (antifreeze) proteins and glycolipids: an overview with emphasis on physiological function, J. Exp. Biol. 218 (2015) 1846−1855.

[31] C. Wang, S. Pakhomova, M.E. Newcomer, B.C. Christner, B.H. Luo, Structural basis of antifreeze activity of a bacterial multi-domain antifreeze protein, PLoS One 12 (2017), e0187169.

[32] G. Deng, D.W. Andrews, R.A. Laursen, Amino acid sequence of a new type of antifreeze protein, from the longhorn sculpin Myoxocephalus octodecimspinosis, FEBS Lett. 402 (1997) 17−20.

[33] H. Nada, Y. Furukawa, Antifreeze proteins: computer simulation studies on the mechanism of ice growth inhibition, Polym. J. 44 (2012) 690−698.