Interpretable Knowledge Acquisition for Predicting Bioluminescent Proteins Using an Evolutionary Fuzzy Classifier Method

Hui-Ling Huang[1,2,3], Hua-Chin Lee[1,2,3], Phasit Charoenkwan[1,2], Wen-Lin Huang[4], Li-Sun Shu[5] and Shinn-Ying Ho[1,2,3]*

[1]Institute of Bioinformatics and Systems Biology, National Chiao Tung University (NCTU), Taiwan
[2]Department of Biological Science and Technology, NCTU, Taiwan
[3]Center for Bioinformatics Research, NCTU, Taiwan
[4] Department of Management Information System, Asia Pacific Institute of Creativity, Taiwan
[5]Department of Multimedia and Game Design, Overseas Chinese University, Taichung, Taiwan
*Corresponding Email: hlhuang@mail.nctu.edu.tw, syho@mail.nctu.edu.tw

***Abstract*—New applications of using bioluminescent proteins (BLPs) are constantly increasing in a variety of research fields such as protein engineering of using single-cell bioluminescent organisms to determine how animals move through water. In this study, we propose a knowledge acquisition method for characterizing BLPs and understanding their functions using a compact set of fuzzy rules. The rule set was obtained by designing an if-then fuzzy-rule-based bioluminescent protein classifier (named iFBPC) with physicochemical properties as input features. In designing iFBPC, feature selection, membership function design, and fuzzy rule base generation are all simultaneously optimized using an intelligent genetic algorithm (IGA). We used the same benchmark dataset for comparisons used in existing SVM-based prediction methods BLProt and PBLP using 100 and 15 features of physicochemical properties, respectively. The classifier iFBPC has two fuzzy rules (one for BLP and the other for non-BLP) and four physicochemical properties with test accuracy of 74.82% where BLProt and PBLP have accuracies of 80.06% and 81.79%, respectively. The four physicochemical properties are structures, protein linkers, nucleation, and membrane proteins in the AAindex database. The analysis of characterizing BLPs was conducted based on knowledge of the fuzzy rule base.**

Keywords-bioluminescent proteins; feature selection; fuzzy rules; genetic algorithm; knowledge acquisition; physicochemical properties.

I. INTRODUCTION

Bioluminescence of organisms occurs in diverse forms of morphology, with various mechanisms of light emission. Then the cited state of the emitter will emit light with a very short lifetime. After releasing the energy in the form of a photon, the reaction time just keep a few nanoseconds. At quite another situation, fluorophore is another substance, which can generate light. It can acquire its excitation energy in bypassing excitation of primary emitter. For example, the green flurocense protein (GFP) usually use the covalently bond of fluorophore. In Aequorea GFP [1], the post-translational reaction of cyclization, dehydration and oxidation of Ser65-Tyr66-Gly67 [2] because the emitting light and hence it can be easily expressed in eukaryotic and prokaryotic orgasms without losing its emitting function. No coelenterate-specific enzymes are needed to join the reaction. Understanding physicochemical properties of the bioluminescent proteins (BLPs) may help improve the applications of BLPs.

The experimental methods [3][4] to identify the BLPs could be often time-consuming expensive and have very limited scopes due to some restrictions for many enzymatic reactions. Recently, researchers have interests in computational methods, which have been developed to predict BLPs.

Kandaswamy et al. [5] first proposed a predictive method, as known as BLProt, based on support vector machine (SVM) and physicochemical properties to predict BLPs. The three different filter approaches, ReliefF, infogain, and mRMR were utilized to identify the most informative features. In 2011, Huang et al. [6] proposed a novel method using the physicochemical properties (PBLP). In that work PBLP, an efficient algorithm inheritable bi-objective genetic algorithm (IBCGA) [7] was used to select significant features, which could discriminate the two classes of BLPs. Recently, Zhao et al. [8] developed a new computational method to predict BLPs using a model based on position specific scoring matrix and auto covariance (PSSM-AC). Their results showed that accuracy of PSSM-AC model was higher than BLProt and PBLP. The existing methods [5][6][8] can predict BLPs but suffer from obtaining human-interpretable knowledge from sequences.

In our previous work, PBLP [6] investigates the optimal design of predictors for predicting from amino acid sequence using both informative features and an appropriate classifier. Furthermore, we obtained a set of relevant physicochemical properties can advance prediction performance. The proposed PBLP identified m=15 features of properties for predicting BLPs with an independent test accuracy of 81.79%. Since the set of 15 physicochemical properties performs well, we would apply it to acquire the rule-based knowledge for predicting and analyzing BLPs.

In this paper, we design an interpretable fuzzy rule classifier based on the 15 physicochemical properties as features [6]. The proposed classifier with an accurate and compact fuzzy rule base using a scatter partition of feature space for BLPs is named iFBPC. Because BLPs from database [6] have the property of natural clustering, fuzzy

 ISBN: 978-1-61208-354-4

classifiers using a scatter partition of feature spaces often have a smaller number of rules than those using grid partitions. The design of iFBPC has three objectives to be simultaneously optimized: maximal classification accuracy, minimal number of rules, and minimal number of used physicochemical properties. In designing iFBPC, the flexible membership function, fuzzy rule, and physicochemical properties selection are simultaneously optimized. Huang et al. [9] applied an intelligent genetic algorithm (IGA) [10] to efficiently solve the design problem with a large number of tuning parameters.

The iFBPC built with 2 rules and 4 physicochemical properties have fine training accuracy of 73.67% and test accuracy of 74.82%. These results are suggested that iFBPC provides the interpretable and confidant rules that can will identify the BLPs. The results show that the membrane protein properties are most important to BLPs and the amino acids prone to locate at terminal of the alpha-helix are not preferred in BLPs. This is might be caused from the working environment of BLPs and these results would also give the biologists the considerations about the protein engineering examinations for altering the BLP stability.

The rest of this paper is organized as follows. Section II describes the materials and methods used. Section III describes the results and performance, and Section IV addresses the conclusions of this paper. Finally, the acknowledgement closes the article.

II. MATERIALS AND METHODS

We propose a fuzzy rule-based knowledge acquisition system based on interpretable if-than fuzzy classifiers (iFBPC). The design of iFBPC is provided with an accurate and compact fuzzy rule base using a scatter partition of feature space for bioluminescent protein data analysis. The framework is presented in Fig. 1.

A. Dataset

The bioluminescent proteins (BLPs) were extracted from Kandaswamy et al. [5]. More details about this data set can be found by [5][6][8]. After all, a total 441 BLPs are kept as positive dataset. The statistic of the training and test sets is shown in Table I. 300 BLPs are random selected from the 441 positive dataset and served as training dataset. The others are served as testing dataset. 300 non-BLPs are also randomly picked from seed proteins of Pfam protein families. These proteins, served as negative dataset, are unrelated to BLPs. The negative testing dataset is composed of the 141 non-BLPs Pfam protein families and are different from training non-BLPs. Finally, the testing dataset is composed of 141 BLPs and 141 non-BLPs.

TABLE I - THE STATISTIC OF THE TRAINING/TEST SETS.

Dataset	Number of BLPs	Number of non-BLPs
Training	300	300
Test	141	141

B. Feature set

Considering the BLPs data set, the set of m=15 informative properties (PCPs) identified by PBLP performs best where the best solution with accuracy of 84.11% is used [6]. The PBLP is a systematic approach to automatically identify a set of physicochemical and biochemical properties in the AAindex database to design SVM-based classifiers for predicting and analyzing BLPs. The set of m=15 PCPs is identified by PBLP, we would apply it to acquire the rule-based knowledge for predicting and analyzing BLPs data set. The set of 15 PCPs is described in Table II.

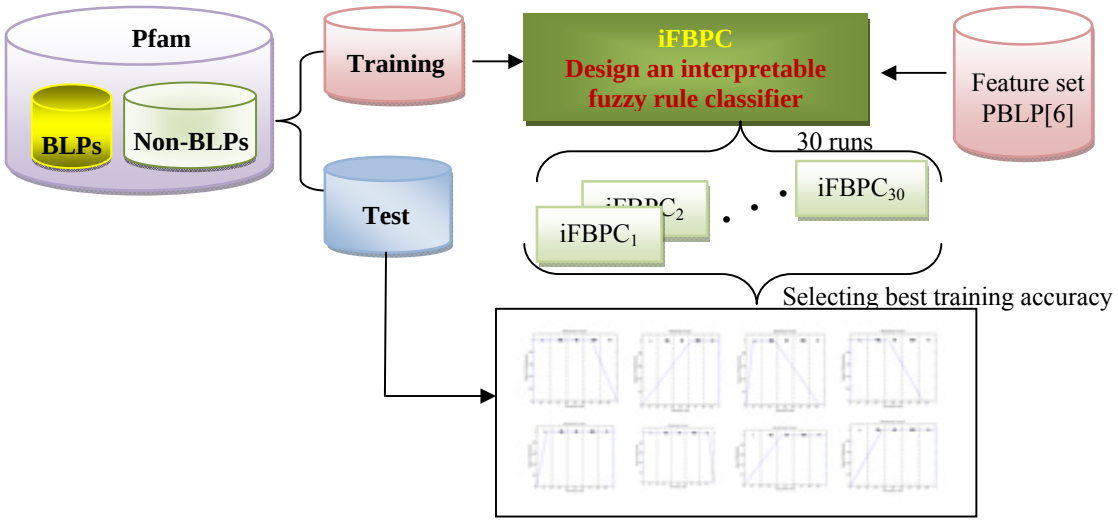

Figure 1. The framework of if-than fuzzy rule-based classifier for bioluminescent proteins (iFBPC).

TABLE II - THE PBLP INDENTED A SET OF M=15 PHYSICOCHEMICAL PROPERTIES ON BLPs.

Feature ID	AAindex ID	Description
8	BHAR880101	Positional flexibilities of amino acid residues in globular proteins
13	BROC820102	The isolation of peptides by high-performance liquid chromatography using predicted elution positions
18	BUNA790103	1H-nmr parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH
95	FINA910104	Physical reasons for secondary structure stability: alpha-helices in short peptides
107	GEIM800111	Amino acid preferences for secondary structure vary with protein class
202	NAKH920101	The amino acid composition is different between the cytoplasmic and extracellular sides in membrane proteins
223	PALJ810101	Protein secondary structure
310	RACS820111	Differential geometry and polymer conformation. 4. Conformational and nucleation properties of individual amino acids
380	VENT840101	Hydrophobicity parameters and the bitter taste of L-amino acids
439	PARS000102	Protein thermal stability: insights from atomic displacement parameters (B values)
473	MITS020101	Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces
475	TSAJ990102	The packing density in proteins: standard radii and volumes
489	PUNT030101	A knowledge-based scale for amino acid membrane propensity
491	GEOR030101	An analysis of protein domain linkers: their classification and role in protein folding

An FGPMF $\mu(x)$ with a single fuzzy set is defined as

C. Acquisition of the rule-based knowledge

The performance of iFBPC mainly arises from two aspects. One is to simultaneously optimize all parameters in the design of iFBPC where all the elements of the fuzzy classifier design have been transformed into parameters of a large parameter optimization problem. The other is to use an efficient optimization algorithm IGA, which is a specific variant of the intelligent evolutionary algorithm [10]. The intelligent evolutionary algorithm uses a divide-and-conquer strategy to effectively solve large parameter optimization problems. IGA is shown to be effective in the design of accurate classifiers with a concise fuzzy rule base using an evolutionary scatter partition of feature space [11].

The proposed iFBPC design involves: 1) flexible generic parameterized membership functions (FGPMFs) and a hyperbox-type fuzzy partition of feature space, 2) determining a fuzzy reasoning method and fuzzy if-then rules corresponding to fuzzy regions, and 3) determining a fitness function and a chromosome representation for using IGA to optimize the system's tuning parameters.

$$\mu(x) = \begin{cases} 0 & \text{if } x \le a \text{ or } x \ge d \\ \dfrac{x-a}{b-a} & \text{if } a < x < b \\ \dfrac{d-x}{d-c} & \text{if } c < x < d \\ 1 & \text{if } b \le x \le c \end{cases} \quad (1)$$

where $x \in [0, 1]$ and $a \le b \le c \le d$. Some illuminations of FGPMF are shown in Fig. 2. The variables a, b, c and d determining the shape of a trapezoidal fuzzy set are the parameters to be optimized. This transformative scheme of training patterns and the encoded parameters of the IGA's chromosomes have been described more detail in previous research [9].

D. Fuzzy if–then rule and Fuzzy reasoning method

The following fuzzy if–then rule base for n-dimensional classification problems are used in the design of iFBPC:

R_j : If x_1 is A_{j1} and . . . and x_n is A_{jn} then class CL_j with CF_j, $j = 1, \ldots, N$.

where R_j is a rule label, x_i denotes a variable of physicochemical property, A_{ji} is an antecedent fuzzy set, C is a number of classes, $CL_j \in \{1, . . ., C\}$ denotes a consequent class label, CF_j is a certainty grade of this rule in the unit interval [0, 1], and N is a number of initial fuzzy rules in the training phase. In this study, C=2 (two classes for BLPs and non-BLPs), n=15 (initial number in the feature set to be selected), and N=3C (initial number in the rule set to be selected).

To enhance interpretability of fuzzy rules, linguistic variables in fuzzy rules can be used. Each variable xi has a linguistic set U = {S (small), SM (small medium), M (medium), ML (medium large), L (large)}. Each linguistic value of xi equally represents 1/5 of the domain [0, 1]. Examples of linguistic antecedent fuzzy sets are shown in Fig. 3.

 ISBN: 978-1-61208-354-4

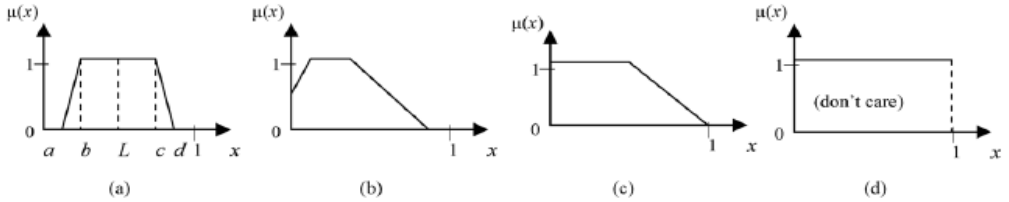

Figure 2. Illuminations of FGPMF: (a) $a>0$ and $d< 1$; (b) $a<0<b$, (c) $b \leqq 0$; (d) $b \leqq 0$ and $c \geqq 1$.

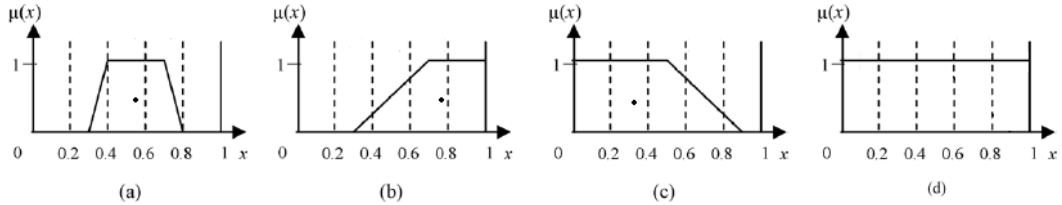

Figure 3. Examples of an antecedent fuzzy set A_{ji} with linguistic values (L: low, ML: medium low, M: medium, MH: medium high, H: high): (a) A_{ji} represents {ML, M, MH}; (b) A_{ji} represents {ML, M, MH, H}, i.e., not Low; (c) A_{ji} represents {L, ML, M, MH, H} or ALL.

In the training phase, all the variables CL_j and CF_j are treated as parametric genes encoded in a chromosome and their values are obtained using IGA. The following fuzzy reasoning method is adopted to determine the class of an input pattern $x_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$ based on voting using multiple fuzzy if–then rules:

Step 1: Calculate score $S_{\text{Classv}}(v = 1, \ldots, C)$ for each class as follows:

$$S_{\text{Classv}} = \sum_{\substack{R_j \in FC \\ CL_j = Class\ v}} \mu_j(x_p) CF_j,\ \mu_j(x_p) = \prod_{i=1}^{n} \mu_{ji}(x_{pi}), \quad (2)$$

where FC denotes the fuzzy classifier, and $\mu_{ji}(\cdot)$ represents the membership function of the antecedent fuzzy set A_{ji}.

Step 2: Classify x_p as the class with a maximal value of S_{Classv}.

Notably, x_p is classified into the BLP or non-BLP class for one iFBPC. The final classification of x_p is determined using the proposed classifier iFBPC in the study.

E. IGA to optimize the system's tuning parameters

A GA-chromosome consists of control GA-genes for selecting useful features and significant fuzzy rules, and parametric GA-genes for encoding the membership functions and fuzzy rules. The control GA-gene comprises two types of parameters. One is parameter r_j, j=1,…N, represented by one bit for eliminating unnecessary fuzzy rules. The other is parameter f_i, i=1,…N, represented by one bit or eliminating useless features. The parametric genes determine variables of three types: $V_{ji}^{t} \in [0, 1]$, t=1, …, 5, for determining the antecedent fuzzy set A_{ji}, CL_j for determining the consequent class label of rule R_j, and $CF_j \in [0, 1]$ for determining the certainty grade of rule R_j , where j=1, ..., N and i=1, ..., n. A rule base with N fuzzy rules is represented as an individual. The detailed explanation of the chromosome representation and implementation can be referred to [9]. The design of an efficient fuzzy classifier is formulated as a large parameter optimization problem. Once the solution of IGA is obtained, an accurate classifier with a concise fuzzy rule base can be obtained.

We define the fitness function of IGA for designing iFBPC as follows:

$$\max Fit(FC) = ACC - W_r N_r - W_f N_f \quad (3)$$

where W_r and W_f are positive weights. In this study, the fitness function is used to optimize the three objectives:

1) to maximize the classification accuracy ACC,
2) to minimize the number N_r of fuzzy rules, and
3) to minimize the number N_f of selected features.

The trade-off between prediction accuracy and conciseness of the rule base can be determined by tuning the weights W_r and W_f. For obtaining an easily-interpretable and compact knowledge rule base with concise iFBPC, the small values of N_r and N_f are preferred. Therefore, we used large values of penalty weights $W_r = 0.6$ and $W_f = 0.1$. If the high accuracy of an individual iFBPC is the most important objective, small values of penalty weights are preferred. The simulation results show that the weights W_r and W_f are not very sensitive to the accuracy of the obtained solutions using IGA. To further advance the prediction accuracy of predicting BLP is utilized.

III. RESULTS

The parameter settings of IGA [10] are $N_{pop} = 20$, $P_c = 0.7$, $P_s = 1 - P_c$, $P_m = 0.01$ and $\alpha = 15$. Because the search space of the optimal design of iFBPC is proportional to the number N_p of parameters to be optimized, the stopping condition is suggested to use a fixed number $100N_p$ of fitness evaluations.

A. Prediction performance evaluation

The training samples with 15 properties in the dataset BLPs are represented as 15-dimensional feature vectors. This set of 15 physicochemical properties is identified by PBLPs [12]. Due to the non-deterministic characteristic of genetic algorithms, the average performance of 30 independent

 ISBN: 978-1-61208-354-4

iFBPC is given in Table III. The top six of high selected frequency PCPs in the 30 runs are shown in Table IV.

TABLE III. THE AVERAGE VALUES OF 30 INDEPENDENT RUNS OF THE PROPOSED iFBPC.

Training			Test
Accuracy. (%)	Feature no.	Rule no.	Accuracy (%)
73.67	3.67	2	74.82%

TABLE IV. THE TOP SIX OF HIGH SELECTED FREQUENCY PCPs IN THE 30 RUNS.

Freq.	Feature No.	AAindex No.	Category
20	489	PUNT030101	Membrane Protein
17	202	NAKH920101	Membrane Protein
12	491	GEOR030101	Protein linker
10	223	PALJ810101	Structure
10	475	TSAJ990102	Structure
10	502	ZHOH040103	Hydrophobicity

B. *Rule-based knowledge*

We selected one $iFBPC_1$ with best training accuracy in the independent 30 runs, to illustrate the rules for bioluminescent proteins mechanism. The $iFBPC_1$ has training accuracies of 73.67%, the test accuracies of 74.82%, the feature numbers N_f of 4, and the rule numbers N_r of 2, respectively. The selected physicochemical properties are BHAR880101 (Protein linker), GEIM800111 (Structure), PUNT030101 (Membrane Protein) and FINA910104 (Nucleation), shown in Fig 4. The fuzzy rules are linguistically interpretable as follows:

Fuzzy Classifier $iFBPC_1$:

R1: if BHAR880101 is ALL, GEIM800111 is {medium, large}, PUNT030101 is {small, medium} and FINA910104 is {small, medium}, then BLPs with CF=0.714.

R2: if BHAR880101 is ALL, GEIM800111 is ALL, PUNT030101 is {medium, large} and FINA910104 is {medium, large}, then non- BLPs with CF= 0.267.

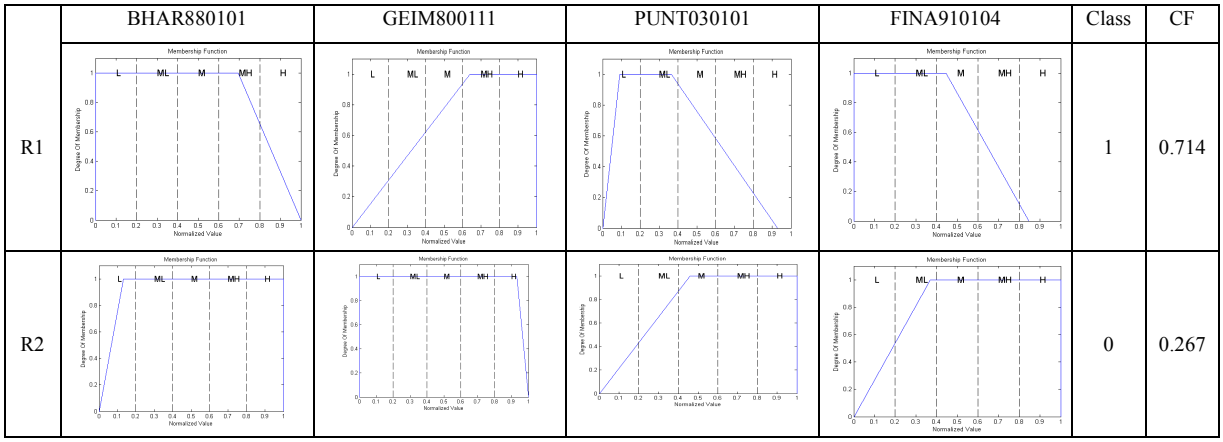

Figure 4. Fuzzy rules of the selected 4 PCPs, the training ACC is 73.67% and testing ACC is 74.82%. Class: 1 for BLPs and 0 for non-BLPs.

C. *The physicochemical properties of BLPs*

The 15 informative PCPs are further classed into 5 categories that are structures, hydrophobicity, protein linkers, nucleation and membrane proteins. The importance of bioluminescence protein versus protein linker, hydrophobicity, structure is documented in previous work [11][12]. The BLPs works, sometimes they will meet a hydrophobic environment that caused by the luciferin, a quite hydrophobic substance [13]. The rules, which could stabilized the structures of BLPs at hydrophobic environments, of membrane protein folding could be also considered.

From the fuzzy rules, the BHAR880101, the structure of flexibility, and GEIM800111, aperiodic indices for alpha/beta-proteins, should be considered both in BLPs or non-BLPs. PUNT030101, a membrane protein properties, and FINA910104, the helix termination parameter at position, show reversed results in BLP and non-BLPs. In BLPs, the property, PUNT030101, should be small to medium. In original study [14] of this properties, the authors defined that a negative value indicated a high membrane propensity. This can be interpreted that the BLPs would have some properties that membrane proteins also have as mentioned in previous study [6].

It is interesting that the FINA910104 also shows reversed results in BLPs and non-BLPs. This property, the index about the amino acids locate at the C-terminal of alpha-helix, is driven from the nucleus structure study. From previous study, some proteins mainly composed of beta-sheets will transform to alpha-helices in partial organic solvent suggesting that this kind of solvent could keep the alpha-helical structure [15]. Forming the C-Capping will increase in alpha-helicity [16] and would cause the BLPs to form alpha-helix from native structures, which are not alpha-helix in such partial organic solvent. The BLPs would use the

 ISBN: 978-1-61208-354-4

strategy that to avoid being composed of the amino acids, which favors to locate at the alpha-helix terminal and often forms the C-Capping structure to escape alpha-helixes. This can keep the original structure of BLPs away from the structural transformation and would maintain the biological function of BLPs.

IV. CONCLUSION

The iFBPC is a high performance sequence-based classifier for identifying the BLPs based on fuzzy-rule classifier. It provides a high confident fuzzy rule that could identify the BLPs well and also provide some useful knowledge. In BLPs, the membrane properties are important because BLPs work at the partial organic solvent, which will change the folding nature of proteins and make the proteins lose their functions. BLPs would use the strategy that avoiding being composed of the amino acids favoring to locate at the terminal of alpha-helices. This strategy could provide the protein engineers a new though for protein engineering.

ACKNOWLEDGMENT

This work was funded by National Science Council of Taiwan under the contract number NSC-102-2221-E-009-169-, and "Center for Bioinformatics Research of Aiming for the Top University Program" of the National Chiao Tung University and Ministry of Education, Taiwan, R.O.C. for the project 102W962. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

[1] F. H. Johnson, O. Shimomura, Y. Saiga, L. C. Gershman, G. T. Reynolds, and J. R. Waters, “Quantum efficiency of Cypridina luminescence, with a note on that of Aequorea,” J. Cell. Comp. Physiol, vol. 60, no. 1, 1962.

[2] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier, “Primary Structure of the Aequorea-Victoria Green-Fluorescent Protein,” Gene, vol. 111, no. 2, Feb 15, 1992, pp. 229-233.

[3] O. Shimomura, F. H. Johnson, and Y. Saiga, “Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea,” J Cell Comp Physiol, vol. 59, Jun, 1962, pp. 223-39.

[4] P. A. Vidi, and V. J. Watts, “Fluorescent and bioluminescent protein-fragment complementation assays in the study of G protein-coupled receptor oligomerization and signaling,” Mol Pharmacol, vol. 75, no. 4, Apr, 2009, pp. 733-9.

[5] K. K. Kandaswamy, G. Pugalenthi, M. K. Hazrati, K. U. Kalies, and T. Martinetz, “BLProt: prediction of bioluminescent proteins based on support vector machine and relieff feature selection,” BMC Bioinformatics, vol. 12, 2011, pp. 345.

[6] H.-L. Huang, Y.-F. Liou, H.-C. Lee, W.-L. Huang, and S.-Y. Ho, “Designing predictors of bioluminescence proteins using an efficient physicochemical property mining method,” IEEE International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2012), 2012.

[7] S. Y. Ho, J. H. Chen, and M. H. Huang, “Inheritable genetic algorithm for biobjective 0/1 combinatorial optimization problems and its applications,” Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics, vol. 34, no. 1, Feb, 2004, pp. 609-620.

[8] X. W. Zhao, J. K. Li, Y. X. Huang, Z. Q. Ma, and M. H. Yin, “Prediction of Bioluminescent Proteins Using Auto Covariance Transformation of Evolutional Profiles,” International Journal of Molecular Sciences, vol. 13, no. 3, Mar, 2012, pp. 3650-3660.

[9] H. L. Huang, F. L. Chang, S. J. Ho, L. S. Shu, W. L. Huang, and S. Y. Ho, “FRKAS: Knowledge Acquisition Using a Fuzzy Rule Base Approach to Insight of DNA-Binding Domains/Proteins,” Protein and Peptide Letters, vol. 20, no. 3, Mar, 2013, pp. 299-308.

[10] S. Y. Ho, L. S. Shu, and J. H. Chen, “Intelligent evolutionary algorithms for large parameter optimization problems,” Ieee Transactions on Evolutionary Computation, vol. 8, no. 6, Dec, 2004, pp. 522-541.

[11] S. A. Moore, and M. N. G. James, “Common Structural Features of the Luxf Protein and the Subunits of Bacterial Luciferase - Evidence for a (Beta-Alpha)(8) Fold in Luciferase,” Protein Science, vol. 3, no. 11, Nov, 1994, pp. 1914-1926.

[12] A. J. Fisher, T. B. Thompson, J. B. Thoden, T. O. Baldwin, and I. Rayment, “The 1.5-A resolution crystal structure of bacterial luciferase in low salt conditions,” J Biol Chem, vol. 271, no. 36, Sep 6, 1996, pp. 21956-68.

[13] G. W. J. Moss, N. P. Franks, and W. R. Lieb, “Modulation of the General Anesthetic Sensitivity of a Protein - a Transition between 2 Forms of Firefly Luciferase,” Proceedings of the National Academy of Sciences of the United States of America, vol. 88, no. 1, Jan, 1991, pp. 134-138.

[14] M. Punta, and A. Maritan, “A knowledge-based scale for amino acid membrane propensity,” Proteins-Structure Function and Genetics, vol. 50, no. 1, Jan 1, 2003, pp. 114-121.

[15] K. Shiraki, K. Nishikawa, and Y. Goto, “Trifluoroethanol-Induced Stabilization of the Alpha-Helical Structure of Beta-Lactoglobulin - Implication for Non-Hierarchical Protein-Folding,” Journal of Molecular Biology, vol. 245, no. 2, Jan 13, 1995, pp. 180-194.

[16] J. P. Schneider, and W. F. DeGrado, “The design of efficient alpha-helical C-capping auxiliaries,” Journal of the American Chemical Society, vol. 120, no. 12, Apr 1, 1998, pp. 2764-2767.

 ISBN: 978-1-61208-354-4