International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:4, No:5, 2010

Prediction of Protein Subchloroplast Locations using Random Forests

Chun-Wei Tung, Chyn Liaw, Shinn-Jang Ho, Shinn-Ying Ho

***Abstract*—**Protein subchloroplast locations are correlated with its functions. In contrast to the large amount of available protein sequences, the information of their locations and functions is less known. The experiment works for identification of protein locations and functions are costly and time consuming. The accurate prediction of protein subchloroplast locations can accelerate the study of functions of proteins in chloroplast. This study proposes a Random Forest based method, ChloroRF, to predict protein subchloroplast locations using interpretable physicochemical properties. In addition to high prediction accuracy, the ChloroRF is able to select important physicochemical properties. The important physicochemical properties are also analyzed to provide insights into the underlying mechanism.

***Keywords*—**Chloroplast**,** Physicochemical properties, Protein locations, Random Forests.

I. Introduction

CHLOROPLASTS are typical organelles in plant cells and are developed and differentiated from proplastids. Chloroplasts play important roles in cellular metabolism and several biological processes, including amino acid biosynthesis and photosynthesis. Chloroplasts are originated from cyanobacteria. But, most of their genes are transferred to the nucleus of the cell and their autonomy is lost during evolution [1]. The initiations of chloroplast proteome projects [2], [3], [4], [5] point out the importance of identification and characterization of chloroplast proteins.

Previous computational studies mainly focus on prediction and identification of chloroplast proteins. For example, TargetP [6] and ChloroP [7] were developed to predict proteins in plastid and chloroplast by recognizing transit peptides. Also, some studies applied these tools to identify candidate chloroplast proteins in a genome-wide manner [8], [9], [10]. However, the information of subchloroplast locations is still not available for a large number of chloroplast proteins. Due to the relation between protein subchloroplast locations and functions, it is desirable to develop computational methods for predicting and analyzing protein subchloroplast locations

Recently, a tool of SubChlo was developed for predicting subchloroplast locations that is based on an evidence-theoretic k-nearest neighbor classifier. It can predict subchloroplast locations with an accuracy of 67.18% on S60 dataset consisting of proteins with less than 60% sequence similarity. However, the utilized feature of pseudo-amino acid composition is hard to provide interpretable information of the underlying mechanism of protein localizations [11].

Physicochemical properties, one of the most intuitive and interpretable features, were applied to predict subchloroplast locations of proteins. Physicochemical properties of proteins such as hydrophobicity and charge play vital roles in molecular recognitions and protein localizations and are extensively used in bioinformatics for prediction and analysis of various problems. Examples include the prediction and analysis of peptide immunogenicity [12], protein ubiquitylation sites [13] and HIV coreceptor usage [14]. Most importantly, apart from prediction accuracies, physicochemical properties are able to provide human interpretable knowledge concerning protein sorting mechanisms [15], [16].

In this study, a method named ChloroRF is proposed to predict subchloroplast locations of proteins. ChloroRF based on Random Forests (RF) classifiers [17] and 531 physicochemical properties obtained from AAindex database [18] can predict subchloroplast locations with an accuracy of 67.43% that is comparable with SubChlo. The advantages of the RF classifier include less overfitting problems [19], [20] and its native method for estimating feature importance. The property of avoidance of overfitting problems is especially important when analyzing a small dataset in this study.

In addition to an accurate prediction method, the feature importance can provide insights into the underlying mechanism of protein sorting in chloroplast. Two criteria of mean decreases of accuracy and Gini index are applied to separately select the corresponding top 30 physicochemical properties. Among the 60 properties, four hydrophobicity-related properties are important for determining protein locations and three properties are directly associated with membrane locations. Finally, a total of 12 physicochemical properties found common in the two property sets are identified for further analysis.

Chun-Wei Tung is with the Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, 300 Taiwan (e-mail: cwtung@livemail.tw).

Chyn Liaw is with the Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, 300 Taiwan (e-mail: chynliaw@gmail.com).

Shinn-Jang Ho is with the Department of Automation Engineering, National Formosa University, Yunlin 632, Taiwan. (e-mail: sjho@sunws.nhit.edu.tw)

Shinn-Ying Ho is with the Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, 300 Taiwan. (corresponding author; phone: 303-555-5555; fax: 303-555-5555; e-mail: syho@mail.nctu.edu.tw).

II. Methods

A. Dataset

The dataset of RAW consisting of 737 protein Swissprot IDs were obtained from the website of SubChlo [11] (http://bioinfo.au.tsinghua.edu.cn/subchlo). The RAW dataset was extracted by keyword search on Swissprot database [21]. However, because three Swissprot IDs of the RAW dataset are incomplete, we retrieved a slightly different dataset of 734 protein IDs and their sequences were retrieved from Swissprot database. The four compartments of plant chloroplast associated with proteins of RAW dataset are envelope, stroma, thylakoid membrane and thylakoid lumen. Due to the preprocessing work of removing proteins annotated with more than one compartment [11], each protein of RAW dataset is associated with only one compartment of chloroplast. All three missing proteins belong to the compartment of thylakoid membrane.

To avoid overestimating the prediction performance, a tool CD-HIT [22], [23] is applied to remove highly redundant sequences of the RAW dataset. A threshold of 60% the same as previous study [11] was applied, and a final dataset S60 consisting of 261 protein sequences was used for all subsequent analyses. Please note that there is only one missing protein sequence of the constructed dataset, compared to the reported S60 dataset consisting of 262 protein sequences [11]. The numbers of proteins of S60 are 40, 49, 128 and 44 for compartments of envelope, stroma, thylakoid membrane and thylakoid lumen, respectively.

B. Physicochemical properties

Due to the importance and interpretability of physicochemical properties, they are widely used for prediction and analysis in bioinformatics studies [12], [13], [15], [14]. In this study, 544 physicochemical properties were retrieved from the amino acid indices (AAindex) database of version 9.0 [18]. The AAindex database is a collection of many published indices representing physicochemical properties of amino acids. For each physicochemical property, a set of 20 numerical values for amino acids are used to represent the property. A total of 531 physicochemical properties are used for the following studies by removing 13 physicochemical properties having the value 'NA' in their amino acid indices.

To encode a protein sequence for classification and prediction, a two-step method is applied as follows. First, given a protein sequence of length l, 531 index vectors $\mathrm{X}_p = (x_1,...,x_l)$, $p = 1,...,531$, for 531 physicochemical properties are obtained by substituting the amino acids with corresponding index values. Second, the final feature vector for representing a protein sequence is defined as $\mathrm{V} = (v_1,...,v_{531})$, where v_p is the averaged value of elements in X_p.

C. Random Forests (RF)

The Random Forests (RF) classifier based on a large ensemble of decision trees is an extensively used ensemble learning method [17]. The RF improves prediction performances of classification and regression trees (CART, [24]) by growing many weak CART trees. Every tree is built by using a fixed number of randomly selected features for tree splitting and based on a bootstrap sample of the whole training dataset. In this study, the number of selected features is set to a recommended default value 23, which is nearly equal to the square root of the total number of features square root of the total number of features (531 physicochemical properties).

The RF is useful for estimating prediction errors and evaluating feature importance. The prediction error is estimated by using out-of-bag (OOB) data. For each tree of RF, the OOB data consisting of approximately one-third of the training dataset is applied to test the decision tree that is constructed by using the remaining training dataset with no pruning procedure. Finally, the overall prediction error is then calculated by majority voting for classification and averaging for regression over all trees.

D. Feature importance

The feature importance can provide insights into the major factors determining a specific problem. There are two indices for evaluating feature importance: the means of decreased Gini index and accuracy. The feature with largest decreased values of means of Gini index or accuracy is the most important feature because it contributes most to prediction performances.

The estimation of feature importance utilizes random permutation method on a specific feature to measure the corresponding decreased performances. A three-step method is applied as follows. First, for each feature, its feature values of corresponding OOB data of constructed trees in RF classifier are randomly permuted. Second, the permuted OOB data is applied to evaluate performances of constructed trees. The performance measurement can be accuracy or Gini index. Finally, the feature importance can be obtained by calculating the difference between the performances using original and permuted OOB data. The Gini index is a measure of impurity that can be defined as $1-\sum_j p^2(j|t)$, where $p(j|t)$ denotes the estimated class probabilities for a node t in a decision tree and class $j=1,\ldots,J$. In this study, J=4 denotes the four subchloroplast locations.

E. Performance evaluation

Three measurements were used to evaluate ChloroRF using five-fold cross-validation (5-CV) on the dataset S60, namely percentage accuracy (ACC_i) and area under the ROC (receiver operating characteristic) curve (AUC_i) for the i^{th} compartment, i=1, …, 4, and overall accuracy (OA) for all classes:

$$\mathrm{ACC}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} \times 100\%, \quad (1)$$

$$\mathrm{OA} = \sum \frac{\mathrm{TP}_i}{N}, \quad (2)$$

where TP_i, TN_i, FP_i and FN_i are the number of true positives, true negatives, false positives and false negatives, respectively.

N (=261) is the total number of sequences. The AUC is a robust measurement for binary-class problem. For multiclass problem, a generalized method is applied as following. For each class c, a four-class problem is transformed to binary-class problem by merging the other three classes and the AUC measurement can be applied to the binary-class problem to calculate the corresponding AUC_i.

III. Results

A. Prediction of subchloroplast locations

The Random Forests (RF) classifier with interpretable features of 531 physicochemical properties is applied to construct a prediction method named ChloroRF for prediction of subchloroplast. The number of trees used in developing ChloroRF is 100. The five-fold cross-validation (5-CV) is applied to evaluate prediction performances of ChloroRF. The 5-CV procedure is applied as follows. First, dataset S60 is divided into five data subsets. Second, for each test fold $h = 1,...,5$, its prediction accuracy is calculated by applying the model constructed by using the other four data subsets to independently test data in fold h. Finally, the performances of five test folds are averaged to represent 5-CV performances of ChloroRF.

Table I shows the prediction performances using 5-CV in terms of ACC, AUC and OA. The overall performance of the proposed method ChloroRF is comparable with SubChlo with a slightly better OA=67.43% for ChloroRF than OA=67.18% for SubChlo. The prediction performances for envelop, stroma, thylakoid lumen and thylakoid membrane are 47.50%, 57.14%, 38.64% and 87.50% for ACC and 0.767, 0.839, 0.838 and 0.846 for AUC, respectively.

TABLE I
Comparison of Prediction Performances using Five-Fold Cross-Validation

Compartment	SubChlo	ChloroRF (%)	
	ACC (%)	ACC (%)	AUC
Thylakoid lumen	43.18	38.64	0.838
Stroma	67.35	57.14	0.839
Thylakoid membrane	83.72	87.50	0.846
Envelope	40.00	47.50	0.767
Overall accuracy (OA)	67.18	67.43	

B. Analysis of important physicochemical properties

One of the most useful functions of RF classifier is its ability to estimate and rank features according to their importance. The function is applied to analyze important physicochemical properties to give insights into the underlying mechanisms of protein sorting in chloroplasts.

Two measures were applied to estimate the importance of physicochemical properties, including mean decrease in accuracy and mean decrease in Gini index. The physicochemical property with a largest value of mean decrease in accuracy or Gini is with highest importance for

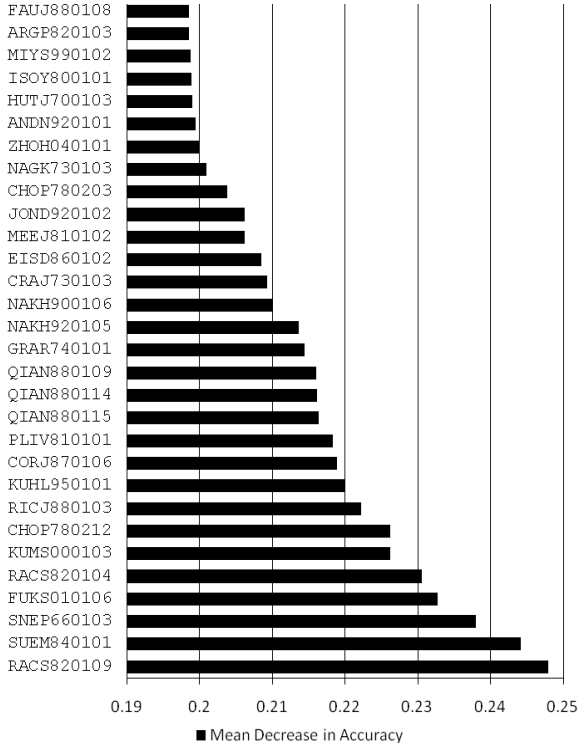

Fig. 1 Top 30 physicochemical properties ranked by mean decrease in accuracy.

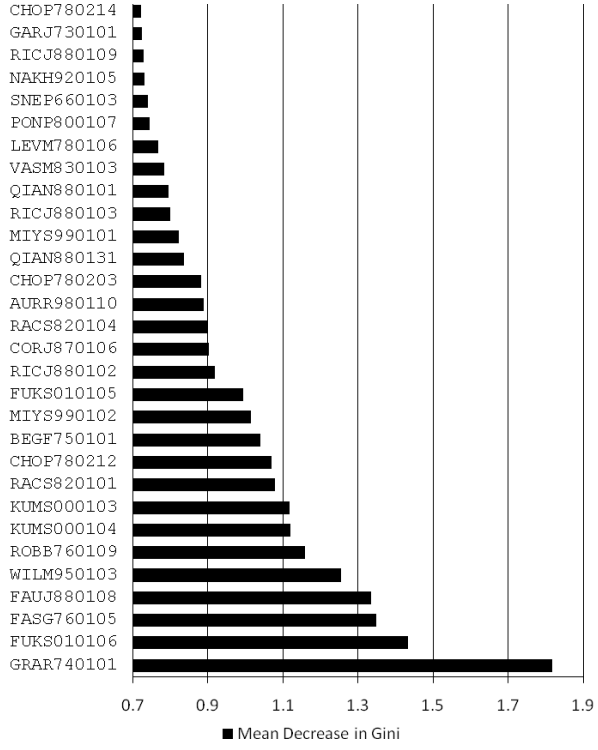

Fig. 2 Top 30 physicochemical properties ranked by mean decrease in Gini index.

TABLE II
COMMON PHYSICOCHEMICAL PROPERTIES SELECTED BY USING BOTH MEAN DECREASES IN ACCURACY AND GINI

AAindex ID	Description	Rank by mean decrease in accuracy	Rank by mean decrease in Gini
SNEP660103	Principal component III (Sneath, 1966)	3	26
FUKS010106	Interior composition of amino acids in intracellular proteins of mesophiles (percent) (Fukuchi-Nishikawa, 2001)	4	2
RACS820104	Average relative fractional occurrence in EL(i) (Rackovsky-Scheraga, 1982)	5	16
KUMS000103	Distribution of amino acid residues in the alpha-helices in thermophilic proteins (Kumar et al., 2000)	6	8
CHOP780212	Frequency of the 1st residue in turn (Chou-Fasman, 1978b)	7	10
RICJ880103	Relative preference value at N-cap (Richardson-Richardson, 1988)	8	21
CORJ870106	ALTLS index (Cornette et al., 1987)	10	15
GRAR740101	Composition (Grantham, 1974)	15	1
NAKH920105	AA composition of MEM of single-spanning proteins (Nakashima-Nishikawa, 1992)	16	27
CHOP780203	Normalized frequency of beta-turn (Chou-Fasman, 1978b)	22	18
MIYS990102	Optimized relative partition energies - method A (Miyazawa-Jernigan, 1999)	28	12
FAUJ880108	Localized electrical effect (Fauchere et al., 1988)	30	4

protein localization. Fig. 1 and Fig. 2 show the top 30 properties ranked by using mean decrease in accuracy and Gini, respectively.

The most important properties with AAindex IDs of RACS820109 [25] and GRAR740101 [26] represent an average relative fractional occurrence in AL (i-1) and a composition for sets of accuracy and Gini, respectively. Four properties with AAindex IDs of WILM950103 [27], KUHL950101 [28], PONP800107 [29] and EISD860102 [30] are associated with hydrophobicity that is important for determining protein locations. Three properties with AAindex IDs of ARGP820103 [31], NAKH920105 [32] and CORJ870106 [33] are directly correlated with membrane localizations.

By comparing the property sets for accuracy and Gini, a total of 21 properties are selected in both sets (shown in Table II). Interestingly, two out of three properties ranked as top 10 in both sets associated with propensities of mesophile and thermophile (AAindex IDs of FUKS010106[34] and KUMS000103[35], respectively) mean that the mesophilicity and thermophilicity might play roles in protein localization. The other property with AAindex ID of CORJ870106 [33] represents an index for detecting amphipathic proteins that is associated with membrane proteins. The property with AAindex ID of NAKH920105 [32] representing amino acid composition of single-spanning proteins is directly related to protein subchloroplast locations.

IV. CONCLUSION

The accurate prediction of protein subchloroplast locations using interpretable features is important to better understand protein sorting mechanism and help to annotate proteins of unknown functions and locations. This study proposed a Random Forest based method named ChloroRF to predict subchloroplast locations using interpretable physicochemical properties. The ChloroRF with a slightly better overall accuracy of 67.43% are comparable with a nearest neighbor-based method SubChlo. However, compared to the pseudo-amino acid compositions used by SubChlo, the human interpretable physicochemical properties used by ChloroRF can provide insights into the underlying mechanism of protein sorting.

By using the Random Forests to identify important physicochemical properties, seven important properties for protein locations can be identified consisting of four hydrophobicity-related and three membrane localization-related properties. Finally, the comparison of property sets selected by mean of accuracy and Gini results in a set of 12 important physicochemical properties. The future works include the collection of more dataset and dealing with proteins annotated with multi-locations.

ACKNOWLEDGMENT

The authors would like to thank the National Science Council of Taiwan for financially supporting this research under the contract numbers NSC 96-2628-E-009-141-MY3 and NSC 98-2627-B-009-004-.

REFERENCES

[1] W. Martin and R. G. Herrmann, "Gene transfer from organelles to the nucleus: how much, what happens, and Why?," *Plant Physiol*, vol. 118, pp. 9-17, Sep 1998.

[2] J. B. Peltier, G. Friso, D. E. Kalume, P. Roepstorff, F. Nilsson, I. Adamska, and K. J. van Wijk, "Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins," *Plant Cell*, vol. 12, pp. 319-41, Mar 2000.

[3] J. B. Peltier, O. Emanuelsson, D. E. Kalume, J. Ytterberg, G. Friso, A. Rudella, D. A. Liberles, L. Soderberg, P. Roepstorff, G. von Heijne, and K. J. van Wijk, "Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction," *Plant Cell*, vol. 14, pp. 211-36, Jan 2002.

[4] M. Ferro, D. Salvi, H. Riviere-Rolland, T. Vermat, D. Seigneurin-Berny, D. Grunwald, J. Garin, J. Joyard, and N. Rolland, "Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters," *Proc Natl Acad Sci U S A*, vol. 99, pp. 11487-92, Aug 20 2002.
[5] M. Ferro, D. Salvi, S. Brugiere, S. Miras, S. Kowalski, M. Louwagie, J. Garin, J. Joyard, and N. Rolland, "Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana," *Mol Cell Proteomics*, vol. 2, pp. 325-45, May 2003.
[6] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence," *J Mol Biol*, vol. 300, pp. 1005-16, Jul 21 2000.
[7] O. Emanuelsson, H. Nielsen, and G. von Heijne, "ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites," *Protein Sci*, vol. 8, pp. 978-84, May 1999.
[8] F. Abdallah, F. Salamini, and D. Leister, "A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis," *Trends Plant Sci*, vol. 5, pp. 141-2, Apr 2000.
[9] W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny, "Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus," *Proc Natl Acad Sci U S A*, vol. 99, pp. 12246-51, Sep 17 2002.
[10] D. Leister, "Chloroplast research in the genomic age," *Trends Genet*, vol. 19, pp. 47-56, Jan 2003.
[11] P. Du, S. Cao, and Y. Li, "SubChlo: predicting protein subchloroplast locations with pseudo-amino acid composition and the evidence-theoretic K-nearest neighbor (ET-KNN) algorithm," *J Theor Biol*, vol. 261, pp. 330-5, Nov 21 2009.
[12] C.-W. Tung and S.-Y. Ho, "POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties," *Bioinformatics*, vol. 23, pp. 942-9, Apr 15 2007.
[13] C.-W. Tung and S.-Y. Ho, "Computational identification of ubiquitylation sites from protein sequences," *BMC Bioinformatics*, vol. 9, p. 310, 2008.
[14] K.-T. Hsu, H.-L. Huang, C.-W. Tung, Y.-H. Chen, and S.-Y. Ho, "Analysis of physicochemical properties on prediction of R5, X4, and R5X4 HIV-1 coreceptor usage," *Int J Biol Life Sci*, vol. 5, pp. 208-15, 2009.
[15] W.-L. Huang, C.-W. Tung, H.-L. Huang, S.-F. Hwang, and S.-Y. Ho, "ProLoc: Prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features," *Biosystems*, Jan 4 2007.
[16] D. Sarda, G. H. Chua, K. B. Li, and A. Krishnan, "pSLIP: SVM based protein subcellular localization prediction using multiple physicochemical properties," *BMC Bioinformatics*, vol. 6, p. 152, 2005.
[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, Oct 2001.
[18] S. Kawashima, P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, and M. Kanehisa, "AAindex: amino acid index database, progress report 2008," *Nucleic Acids Res*, vol. 36, pp. D202-5, Jan 2008.
[19] N. Lin, B. Wu, R. Jansen, M. Gerstein, and H. Zhao, "Information assessment on predicting protein-protein interactions," *BMC Bioinformatics*, vol. 5, p. 154, Oct 18 2004.
[20] D. Amaratunga, J. Cabrera, and Y. S. Lee, "Enriched random forests," *Bioinformatics*, vol. 24, pp. 2010-4, Sep 15 2008.
[21] "The Universal Protein Resource (UniProt) 2009," *Nucleic Acids Res*, vol. 37, pp. D169-74, Jan 2009.
[22] W. Li and A. Godzik, "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences," *Bioinformatics*, vol. 22, pp. 1658-9, Jul 1 2006.
[23] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, "CD-HIT Suite: a web server for clustering and comparing biological sequences," *Bioinformatics*, vol. 26, pp. 680-2, Mar 1 2010.
[24] L. Breiman, *Classification and regression trees*: Chapman & Hall/CRC, 1984.
[25] S. Rackovsky and H. Scheraga, "Differential geometry and polymer conformation. 4. Conformational and nucleation properties of individual amino acids," *Macromolecules*, vol. 15, pp. 1340-1346, 1982.
[26] R. Grantham, "Amino acid difference formula to help explain protein evolution," *Science*, vol. 185, pp. 862-4, Sep 6 1974.
[27] M. Wilce, M. Aguilar, and M. Hearn, "Physicochemical basis of amino acid hydrophobicity scales: Evaluation of four new scales of amino acid hydrophobicity coefficients derived from RP-HPLC of peptides," *Analytical chemistry*, vol. 67, pp. 1210-1219, 1995.
[28] L. Kuhn, C. Swanson, M. Pique, J. Tainer, and E. Getzoff, "Atomic and residue hydrophilicity in the context of folded protein structures," *Proteins*, vol. 23, p. 536, 1995.
[29] P. K. Ponnuswamy, M. Prabhakaran, and P. Manavalan, "Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins," *Biochim Biophys Acta*, vol. 623, pp. 301-16, Jun 26 1980.
[30] D. Eisenberg and A. D. McLachlan, "Solvation energy in protein folding and binding," *Nature*, vol. 319, pp. 199-203, Jan 16-22 1986.
[31] P. Argos, J. K. Rao, and P. A. Hargrave, "Structural prediction of membrane-bound proteins," *Eur J Biochem*, vol. 128, pp. 565-75, Nov 15 1982.
[32] H. Nakashima and K. Nishikawa, "The amino acid composition is different between the cytoplasmic and extracellular sides in membrane proteins," *FEBS letters*, vol. 303, pp. 141-146, 1992.
[33] J. Cornette, K. Cease, H. Margalit, J. Spouge, J. Berzofsky, and C. DeLisi, "Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins," *J Mol Biol*, vol. 195, pp. 659-685, 1987.
[34] S. Fukuchi and K. Nishikawa, "Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria1," *J Mol Biol*, vol. 309, pp. 835-843, 2001.
[35] S. Kumar, C. Tsai, and R. Nussinov, "Factors enhancing protein thermostability," *Protein Eng Des Sel*, vol. 13, p. 179, 2000.

Chun-Wei Tung received the BS degree in Biology, National Cheng Kung University, Tainan, Taiwan, in 2005. He is currently a PhD candidate at the Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan. His research interests include immunoinformatics, bioinformatics, machine learning and data mining.

Chyn Liaw received the BS degree in Applied Mathematics, Tatung University, Taipei, Taiwan, in 2007. She is currently a direct PhD student at the Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan. Her research interests include bioinformatics, machine learning and data mining.

Shinn-Jang Ho received the B.S. degree in power mechanic engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1983 and the M.S. and Ph.D. degrees in mechanical engineering from National Sun-Yat-Sen University, Kaohsiung, Taiwan, in 1985 and 1992, respectively. He is currently an Associate Professor in the Department of Automation Engineering at National Huwei Institute of Technology, Huwei, Yulin, Taiwan. His research interests include optimal control, fuzzy systems, genetic algorithms, and system optimization.

Shinn-Ying Ho received the BS, MS, and PhD degrees in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1984, 1986, and 1992, respectively. From 1992 to 2004, he was with the Department of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan. He is currently the vice dean of College of Biological Science and Technology, National Chiao Tong University, and a professor in the Department of Biological Science and Technology and the Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan. He serves on the Editorial Boards of International Journal of Applied Metaheuristic Computing (IJAMC), Theoretical Biology Insights and Biomedical Engineering and Computational Biology. His research interests include evolutionary algorithms, soft computing, image processing, pattern recognition, bioinformatics, data mining, machine learning, computer vision, fuzzy classifier, large-scale parameter optimization problems, and system optimization. He is a member of the IEEE.