Selected List of Publications
Using Features Provided by PROFEAT

Table of Contents

Feature Group 1 & 2 [G1, G2]: Amino Acid Composition, Dipeptide Composition

- Reczko, M. and Bohr, H. (1994) The DEF data base of sequence based protein fold class predictions. ***Nucleic Acids Res***, 22, 3616-3619.
- Grassmann, J., Reczko, M., Suhai, S. and Edler, L. (1999) Protein fold class prediction: new methods of statistical classification. ***Proc Int Conf Intell Syst Mol Biol***, 106-112.
- Hua, S. and Sun, Z. (2001) Support vector machine approach for protein subcellular localization prediction. ***Bioinformatics***, 17, 721-728.
- Chou, K.C. and Cai, Y.D. (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. ***J Biol Chem***, 277, 45765-45769.
- Bhasin, M. and Raghava, G.P. (2004) Classification of nuclear receptors based on amino acid composition and dipeptide composition. ***J Biol Chem***, 279, 23262-23266.

Feature Group 3 [G3]: Autocorrelation Descriptor

- Feng, Z.P. and Zhang, C.T. (2000) Prediction of membrane protein types based on the hydrophobic index of amino acids. ***J Protein Chem***, 19, 269-275.
- Lin, Z. and Pan, X.M. (2001) Accurate prediction of protein secondary structural content. ***J Protein Chem***, 20, 217-220.
- Horne, D.S. (1988) Prediction of protein helix content from an autocorrelation analysis of sequence hydrophobicities. ***Biopolymers***, 27, 451-477.
- Sokal, R.R. and Thomson, B.A. (2006) Population structure inferred by local spatial autocorrelation: an example from an Amerindian tribal population. ***Am J Phys Anthropol***, 129, 121-131.

Feature Group 4 [G4]: Composition, Transition, Distribution (CTD)

- Dubchak, I., Muchnik, I., Holbrook, S.R. and Kim, S.H. (1995) Prediction of protein folding class using global description of amino acid sequence. ***Proc Natl Acad Sci USA***, 92, 8700-8704.
- Dubchak, I., Muchnik, I., Mayor, C., Dralyuk, I. and Kim, S.H. (1999) Recognition of a protein fold in the context of the Structural Classification of Proteins (SCOP) classification. ***Proteins***, 35, 401-407.
- Bock, J.R. and Gough, D.A. (2001) Predicting protein-protein interactions from primary structure. ***Bioinformatics***, 17, 455-460.
- Cai, C.Z., Han, L.Y., Ji, Z.L., Chen, X. and Chen, Y.Z. (2003) SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. ***Nucleic Acids Res***, 31, 3692-3697.
- Cai, C.Z., Han, L.Y., Ji, Z.L. and Chen, Y.Z. (2004) Enzyme family classification by support vector machines. ***Proteins***, 55, 66-76.

- Han, L.Y., Cai, C.Z., Lo, S.L., Chung, M.C. and Chen, Y.Z. (2004) Prediction of RNA-binding proteins from primary sequence by a support vector machine approach. ***RNA***, 10, 355-368.
- Han, L.Y., Cai, C.Z., Ji, Z.L., Cao, Z.W., Cui, J. and Chen, Y.Z. (2004) Predicting functional family of novel enzymes irrespective of sequence similarity: a statistical learning approach. ***Nucleic Acids Res***, 32, 6437-6444.
- Lo, S.L., Cai, C.Z., Chen, Y.Z. and Chung, M.C. (2005) Effect of training datasets on support vector machine prediction of protein-protein interactions. ***Proteomics***, 5, 876-884.
- Lin, H.H., Han, L.Y., Cai, C.Z., Ji, Z.L. and Chen, Y.Z. (2006) Prediction of transporter family from protein sequence by support vector machine approach. ***Proteins***, 62, 218-231.
- H.H. Lin, L.Y. Han, H.L. Zhang, C.J. Zheng, B. Xie, and Y.Z. Chen. (2006) Prediction of the Functional Class of Lipid-Binding Proteins from Sequence Derived Properties Irrespective of Sequence Similarity. ***J. Lipid Res***. 47(4):824-31.
- H.H. Lin, L.Y. Han, H.L. Zhang, C.J. Zheng, B. Xie, and Y.Z. Chen. (2006) Prediction of the Functional Class of Metal-Binding Proteins from Sequence Derived Physicochemical Properties by Support Vector Machine Approach. ***BMC Bioinformatics*** 7(Suppl 5): S13.
- Cui, J., Han, L.Y., Lin, H.H., Zhang, H.L., Tang, Z.Q., Zheng, C.J., Cao, Z.W. and Chen, Y.Z. (2007) Prediction of MHC-Binding Peptides of Flexible Lengths from Sequence-Derived Structural and Physicochemical Properties. ***Mol. Immunol***. 44: 866-877.
- J. Cui, L.Y. Han, H.H. Lin, Z.Q. Tang, C.J. Zheng, Z.W. Cao, and Y.Z. Chen (2007). Computer Prediction of Allergen Proteins from Sequence-Derived Protein Structural and Physicochemical Properties. ***Mol. Immunol***. 44(4): 514-520.
- L.Y. Han, C.J. Zheng, B. Xie, J. Jia, X.H. Ma, F. Zhu, H.H. Lin, X. Chen, and Y.Z. Chen. (2007) Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. ***Drug Discovery Today*** 12(7-8): 304-313.

Feature Group 5 [G5]: Quasi-Sequence-Order (QSO) Descriptors

- Chou, K.C. (2000) Prediction of protein subcellular locations by incorporating quasi-sequence-order effect. ***Biochem Biophys Res Commun***, 278, 477-483.
- Chou, K.C. and Cai, Y.D. (2004) Prediction of protein subcellular locations by GO-FunD-PseAA predictor. ***Biochem Biophys Res Commun***, 320, 1236-1239.

Feature Group 6 [G6]: Pseudo-Amino Acid Composition (PAAC)

- Cai YD, Chou KC.(2005) Predicting enzyme subclass by functional domain composition and pseudo amino acid composition. ***J Proteome Res***. 4(3):967-71.
- Gao Y, Shao S, Xiao X, Ding Y, Huang Y, Huang Z, Chou KC. (2005) Using pseudo amino acid composition to predict protein subcellular location: approached with Lyapunov index, Bessel function, and Chebyshev filter. ***Amino Acids***. 28(4):373-6.
- Liu H, Yang J, Wang M, Xue L, Chou KC. (2005) Using Fourier spectrum analysis and pseudo amino acid composition for prediction of membrane protein types. ***Protein J***. 24(6):385-9.

- Shen HB, Chou KC. (2005) Predicting protein subnuclear location with optimized evidence-theoretic K-nearest classifier and pseudo amino acid composition. ***Biochem Biophys Res Commun***. 337(3):752-6.
- Xiao X, Shao S, Ding Y, Huang Z, Chou KC.(2006) Using cellular automata images and pseudo amino acid composition to predict protein subcellular location. ***Amino Acids***. 30(1):49-54.
- Cai YD, Chou KC.(2006) Predicting membrane protein type by functional domain composition and pseudo-amino acid composition. ***J Theor Biol***.;238(2):395-400.
- Shen HB, Yang J, Chou KC. (2006) Fuzzy KNN for predicting membrane protein types from pseudo-amino acid composition. ***J Theor Biol***. 240(1):9-13.
- Chou KC, Cai YD. (2006) Predicting protein-protein interactions from sequences in a hybridization space. ***J Proteome Res***. 5(2):316-22.
- Zhou GP, Cai YD. (2006) Predicting protease types by hybridizing gene ontology and pseudo amino acid composition. ***Proteins***. 63(3):681-4.
- Xiao X, Shao SH, Huang ZD, Chou KC.(2006) Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor. ***J Comput Chem***. 27(4):478-82.
- Zhang SW, Pan Q, Zhang HC, Shao ZC, Shi JY. (2006) Prediction of protein homo-oligomer types by pseudo amino acid composition: Approached with an improved feature extraction and Naive Bayes Feature Fusion. ***Amino Acids***. 30(4):461-8
- Zhang T, Ding Y, Chou KC. (2006) Prediction of protein subcellular location using hydrophobic patterns of amino acid sequence. ***Comput Biol Chem***. 30(5):367-71.
- Chen C, Zhou X, Tian Y, Zou X, Cai P. (2006) Predicting protein structural class with pseudo-amino acid composition and support vector machine fusion network. ***Anal Biochem***. 357(1):116-21.
- Chen C, Tian YX, Zou XY, Cai PX, Mo JY. (2006) Using pseudo-amino acid composition and support vector machine to predict protein structural class. ***J Theor Biol***. 7;243(3):444-8.
- Mondal S, Bhavna R, Mohan Babu R, Ramakumar S.(2006) Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification. ***J Theor Biol***. 243(2):252-60
- Shen HB, Chou KC. (2007) Using ensemble classifier to identify membrane protein types. ***Amino Acids***. 32(4):483-8.
- Lin H, Li QZ. (2007) Using pseudo amino acid composition to predict protein structural class: approached by incorporating 400 dipeptide components. ***J Comput Chem***. 28(9):1463-6.

Feature Group 7 [G7]: Amphiphilic Pseudo-Amino Acid Composition (APAAC)

- Chou KC. (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. ***Bioinformatics***, 21(1):10-19.
- Ding H, Luo L, Lin H. (2009). Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. ***Protein Pept Lett***. 16(4):351-5.

- Zhou XB, Chen C., Li,ZC., Zou XY.(2007) Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. ***Journal of Theoretical Biology*** 248(3): 546–551.
- Huang WL, Tung CW, Huang HL, Ho SY.(2009) Predicting protein subnuclear localization using GO-amino-acid composition features. ***Biosystems***. 98(2):73-9.
- Khan A, Majid A, Choi TS. (2010) Predicting protein subcellular location: exploiting amino acid based sequence of feature spaces and fusion of diverse classifiers. ***Amino Acids***. 38(1):347-50.
- Huang WL, Tung CW, Ho SW, Hwang SF, Ho SY. (2008) ProLoc-GO: utilizing informative Gene Ontology terms for sequence-based prediction of protein subcellular localization. *BMC **Bioinfo*****rmatics**. 9:80.
- Zhang GY, Fang BS.(2008) Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo-amino acid composition. ***Theor Biol.*** 253(2):310-5.
- Chou KC, Shen HB.(2006) Predicting protein subcellular location by fusing multiple classifiers. ***J Cell Biochem***.99(2):517-27.
- Chou KC, Cai YD.(2005) Prediction of membrane protein types by incorporating amphipathic effects. ***J Chem Inf Model***. 45(2):407-13.

Feature Group 8 [G8]: Topological Descriptors at Atomic Level

- Philip D.Mosier, Anne E. Counterman and Peter C. Jurs (2002) Prediction of peptide icon collision cross sections from topological molecular structure and amino acid parameters. ***Anal Chem***,74:1360-1370
- Mao S, Huo DD, Mei H,Liang GZ, Zhang M, Li ZL.(2008) New descriptors of amino acids and its applications to peptide quantitative structure-activity relationships. ***Chinese J. Struct. Chem.*** 27:1375-1383.
- Zhao C, Zhang H, Luan F, Zhang R, Liu M, Hu Z, Fan B.(2007) QSAR method for prediction of protein-peptide binding affinity: application to MHC class I molecule HLA-A*0201.***J Mol Graph Model***. 26(1):246-54.
- Todeschini R, Consonni V. Handbook of Molecular Descriptors; ***Wiley-VCH***: Weinheim, 2000.

Feature Group 9 [G9]: Total Amino Acid Properties (TAAP)

- Gromiha MM, Suwa M.(2006) Influence of amino acid properties for discriminating outer membrane proteins at better accuracy. ***Biochimica of Biophysica Acta***, 1764, 1493-1497.
- Huang LT, Gromiha MM.(2008) Analysis and Prediction of Protein Folding Rates Using Quadratic Response Surface Models. ***J Comput Ch****em* 29: 1675–1683.
- Gromiha, MM.(2003) Importance of Native-State Topology for Determining the Folding Rate of Two-State Proteins. ***J. Chem. Inf. Comput. Sci.*** 43(5):1481-1485.

Feature Group 10 & 11 [G10, G11]: Network Descriptors

- Barabási AL, Oltvai ZN. (2004) Network Biology: Understanding the Cell's Functional Organization. ***Nat Rev Genet.*** 5(2):101-13.
- Barabási AL, Gulbahce N, Loscalzo J. (2011) Network Medicine: a Network-Based Approach to Human Disease. ***Nat Rev Genet.*** 12(1):56-68
- Hopkins AL. (2008) Network Pharmacology: The Next Paradigm in Drug Discovery. ***Nat Chem Biol.*** 4(11):682-90
- Yildirim MA, Goh KI, *et al*. (2007) Drug-target network. ***Nat Biotechnol***. 25(10):1119-26.
- Hopkins AL. (2007) Network Pharmacology. ***Nat Biotechnol***. 25(10):1110-1.
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. (2007) The Human Disease Network. ***Proc Natl Acad Sci USA***. 104(21):8685-90.
- Stelzl U, Worm U, *et al*. (2005) A Human Protein-Protein Interaction Network: a Resource for Annotating the Proteome. ***Cell***. 122(6):957-68.
- Pujol A, Mosca R, Farrés J, Aloy P. (2010) Unveiling the Role of Network and Systems Biology in Drug Discovery. ***Trends Pharmacol Sci***. 31(3):115-23.
- Chandra N, Padiadpu J. (2013) Network Approaches to Drug Discovery. ***Expert Opin Drug Discov***. 8(1):7-20.
- Yook SH, Oltvai ZN, Barabási AL. (2004) Functional and Topological Characterization of Protein Interaction Networks. ***Proteomics***. 4(4):928-42.
- You ZH, Yin Z, *et al*. (2010) A Semi-Supervised Learning Approach to Predict Synthetic Genetic Interactions by Combining Functional and Topological Properties of Functional Gene Network. ***BMC Bioinformatics***. 11:343
- Dong J, Horvath S. (2007) Understanding Network Concepts in Modules. ***BMC Syst Biol***. 1:24.
- Rubinov M, Sporns O. (2010) Complex Network Measures of Brain Connectivity: Uses and Interpretations. ***Neuroimage***. 52(3):1059-69.
- Mueller LA, Kugler KG, Netzer M, Graber A, Dehmer M. (2011) A Network-Based Approach to Classify the Three Domains of Life. ***Biol Direct***. 6:53.
- Joyce KE, Laurienti PJ, Burdette JH, Hayasaka S. (2010) A New Measure of Centrality for Brain Networks. ***PLoS One***. 5(8):e12200.
- Pritykin Y, Singh M. (2013) Simple topological features reflect dynamics and modularity in protein interaction networks. ***PLoS Comput Biol.*** 9(10):e1003243.
- Haiyuan Yu, Philip M Kim, Mark Gerstein, *et al*. (2007) The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. ***PLoS Comput Biol***. 3(4): e59.
- Dyer MD, Murali TM, Sobral BW. (2008) The Landscape of Human Proteins Interacting With Viruses and Other Pathogens. ***PLoS Pathog***. 4(2):e32.
- Emmert-Streib F, Dehmer M. (2011) Networks for Systems Biology: Conceptual Connection of Data and Function. ***IET Syst Biol***. 5(3):185-207.

- Zhang B, Horvath S. (2005) A General Framework for Weighted Gene Co-expression Network Analysis. ***Stat Appl Genet Mol Biol***. 4: Article17.
- Emig D, Ivliev A, Pustovalova O, Nikolsky Y, Bessarabova M. (2013) Drug Target Prediction and Repositioning Using an Integrated Network-Based Approach. ***PLoS One***. 8(4):e60618.
- Hsu CL, Huang YH, Hsu CT, Yang UC. (2011) Prioritizing Disease Candidate Genes by a Gene Interconnectedness-Based Approach. ***BMC Genomics***. 12 Suppl 3:S25.
- Zhu C, Kushwaha A, Berman K, Jegga AG. (2012) A Vertex Similarity-Based Framework to Discover and Rank Orphan Disease-Related Genes. ***BMC Syst Biol***. 6 Suppl 3:S8.
- David F. Gleich. (2015) PageRank Beyond the Web. ***SIAM Rev***. 57(3), 321–363.
- Koschützki D, Schreiber F. (2008) Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. ***Gene Regul Syst Bio***. 2:193-201.
- Bánky D, Iván G, Grolmusz V. (2013) Equal Opportunity for Low-Degree Network Nodes: a PageRank-Based Method for Protein Target Identification in Metabolic Graphs. ***PLoS One***. 8(1):e54204.
- Iván G, Grolmusz V. (2010) When the Web Meets the Cell: Using Personalized PageRank for Analyzing Protein Interaction Networks. ***Bioinformatics***. 27 (3): 405-407.

List of Publications Citing PROFEAT (162 citations in Google Scholar, till Dec 2015)

2015:

1. L Olsen, C Oostenbrink, FS Jorgensen. Prediction of cytochrome P450 mediated metabolism. ***Advanced Drug Delivery Reviews***, 2015
2. N Xiao, DS Cao, MF Zhu, QS Xu. protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequences. ***Bioinformatics***, 2015
3. YB Ruiz-Blanco, W Paz, J Green. ProtDCal: A program to compute general-purpose-numerical descriptors for sequences and 3D-structures of proteins. ***BMC Bioinformatics***, 2015
4. Z Zinati, F Zamansani, AH KayvanJoo. New layers in understanding and predicting α-linolenic acid content in plants using amino acid characteristics of omega-3 fatty acid desaturase. ***Computers in Biology and Medicine***, 2014
5. Y Liang, S Liu, S Zhang. Prediction of Protein Structural Class Based on Different Autocorrelation Descriptors of Position-Specific Scoring Matrix. ***MATCH Commun. Math. Comput. Chem***. 2015
6. A Yousef, NM Charkari. SFM: A novel sequence-based fusion method for disease genes identification and prioritization. ***Journal of Theoretical Biology***, 2015
7. W Zhang, L Ji, Y Chen, K Tang, H Wang. When drug discovery meets web search: Learning to Rank for ligand-based virtual screening. ***Journal of Cheminformatics***, 2015
8. W Bao, D Wang, F Kong, R Han, Y Chen. Prediction of Protein Structure Classes. ***Intelligent Computing Theories and Methodologies***, 2015
9. Y Wang, Y Guo, Q Kuang, X Pu, Y Ji, Z Zhang. A comparative study of family-specific protein-ligand complex affinity prediction based on random forest approach. ***Journal of Computer-Aided Molecular Design***, 2015
10. D Andreu, M Torrent. Prediction of Bioactive Peptides Using Artificial Neural Networks. ***Artificial Neural Networks***, 2015
11. S Li, J Li, L Ning, S Wang, Y Niu, N Jin. In Silico Identification of Protein S-Palmitoylation Sites and Their Involvement in Human Inherited Disease. ***Journal of Chemical Information and Modeling***, 2015
12. SI Al Gharabli, S Al Agtash. Artificial neural networks for dihedral angles prediction in enzyme loops: a novel approach. ***International Journal of Bioinformatics Research and Applications***, 2015
13. Y Marrero-Ponce, E Contreras-Torres. Novel 3D bio-macromolecular bilinear descriptors for protein science: Predicting protein structural classes. ***Journal of theoretical biology***, 2015
14. G Govindan, AS Nair. Sequence Features and Subset Selection Technique for the Prediction of Protein Trafficking Phenomenon in Eukaryotic Non Membrane Proteins. ***Biomedical Data Mining***, 2015
15. S Banerjee, PSS Gupta, A Nayek, S Das, VP Sur. PHYSICO2: an UNIX based standalone procedure for computation of physicochemical, window-dependent and substitution based evolutionary properties of protein sequences along with automated block preparation tool, version 2. ***Bioinformation***, 2015
16. AK Tiwari, R Srivastava. An efficient approach for the prediction of ion channels and their subfamilies. ***Computational Biology and Chemistry***, 2015
17. Y Hu, Y Guo, Y Shi, M Li, X Pu. A consensus subunit-specific model for annotation of substrate specificity for ABC transporters. ***RSC Advances***, 2015
18. A Yousef, NM Charkari. A novel method based on physicochemical properties of amino acids and one class classification algorithm for disease gene identification. ***Journal of Biomedical Informatics***, 2015

19. LP Eng, TW Tan, JC Tong. Building MHC Class II Epitope Predictor Using Machine Learning Approaches. ***Computational Peptidology***, 2015
20. F Camacho, R Torres, R Ramos-Pollan. Feature Learning Using Stacked Autoencoders to Predict the Activity of Antimicrobial Peptides. ***Computational Methods in Systems Biology***, 2015
21. A Duardo-Sanchez, H Gonzalez-Diaz. MI-NODES Multiscale Models of Metabolic Reactions, Brain Connectome, Ecological, Epidemic, World Trade, and Legal-Social Networks. ***Current Bioinformatics***, 2015
22. I Cortes-Ciriano, QU Ain, V Subramanian. Polypharmacology modelling using proteochemometrics (PCM): recent methodological developments, applications to target families, and future prospects. ***MedChemComm***, 2015
23. D Ofer, N Rappoport, M Linial. The Little Known Universe of Short Proteins in Insects: A Machine Learning Approach. ***Short Views on Insect Genomics and Proteomics***, 2015
24. L Li, J Li, W Xiao, Y Li, Y Qin, S Zhou, H Yang. Prediction the substrate specificities of membrane transport proteins based on support vector machine and hybrid features. ***Computational Biology and Bioinformatics, IEEE/ACM Transactions***, 2015
25. AK Tiwari, R Srivastava. Feature based classification of nuclear receptors and their subfamilies using fuzzy K-nearest neighbor. ***Computer Engineering and Applications (ICACEA) International Conference on Advances***, 2015
26. AK Tiwari, R Srivastava, S Srivastava. An Efficient Approach for the Prediction of G-Protein Coupled Receptors and Their Subfamilies. ***Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics***, 2015
27. SK Yadav, A Bhola, AK Tiwari. Classification of enzyme functional classes and subclasses using support vector machine. ***Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE) International Conference***, 2015
28. D Wang, W Bao, S Han, Y Chen. Prediction of protein structure classes. ***Informative and Cybernetics for Computational Social Systems (ICCSS) International Conference***, 2015

2014:

29. H Wang, M Wang, H Tan, Y Li, Z Zhang, J Song. PredPPCrys: Accurate Prediction of Sequence Cloning, Protein Production, Purification and Crystallization Propensity from Protein Sequences Using Multi-Step Heterogeneous Feature Fusion and Selection. ***PLOS one***, 2014
30. L Li, X Cui, S Yu, Y Zhang, Z Luo, H Yang, Y Zhou. PSSP-RFE: Accurate prediction of protein structural class by recursive feature extraction from PSI-BLAST profile, physical-chemical property and functional annotations. ***PLOS One***, 2014
31. GS Han, ZG Yu, V Anh. A two-stage SVM method to predict membrane protein types by incorporating amino acid classifications and physicochemical properties into a general form of Chou's PseAAC. ***Journal of Theoretical Biology***, 2014
32. L Li, S Yu, W Xiao, Y Li, M Li, L Huang, X Zheng. Prediction of bacterial protein subcellular localization by incorporating various features into Chou's PseAAC and a backward feature selection approach. ***Biochimie***, 2014
33. MR Bakhtiarizadeh, M Moradi-Shahrbabak. Neural network and SVM classifiers accurately predict lipid binding proteins, irrespective of sequence homology. ***Journal of Theoretical Biology***, 2014

34. DS Cao, N Xiao, QS Xu, AF Chen. Rcpi: R/Bioconductor package to generate various descriptors of proteins, compounds, and their interactions. ***Bioinformatics***, 2014
35. CLP Eng, JC Tong, TW Tan. Predicting host tropism of influenza A virus proteins using random forest. ***BMC Medical Genomics***, 2014
36. QU Ain, O Mendez-Lucio, IC Ciriano, T Malliavin. Modelling ligand selectivity of serine proteases using integrative proteochemometric approaches improves model performance and allows the multi-target dependent interpretation of features. ***Integrative Biology***, 2014
37. BA van den Berg, MJT Reinders, JA Roubos. SPiCE: a web-based tool for sequence-based protein classification and exploration. ***BMC Bioinformatics***, 2014
38. CO Sakar, O Kursun, H Seker. Combining multiple clusterings for protein structure prediction. ***International Journal of Data Mining and Bioinformatics***, 2014
39. X Du, J Cheng, T Zheng, Z Duan, F Qian. A Novel Feature Extraction Scheme with Ensemble Coding for Protein-Protein Interaction Prediction. ***International Journal of Molecular Sciences***, 2014
40. S Datta, S Mukhopadhyay. An ensemble method approach to investigate kinase-specific phosphorylation sites. ***International journal of Nanomedicine***, 2014
41. N Habibi, SZM Hashim, A Norouzi. A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli. ***BMC Bioinformatics***, 2014
42. RD Roy, M Bhardwaj, V Bhatnagar, K Chakraborty. How do eubacterial organisms manage aggregation-prone proteome? ***F1000Research***, 2014
43. AN Mbah, ISRN Computational Biology, 2014. Application of Hybrid Functional Group to ATP binding Protein Prediction. ***ISRN Computational Biology***, 2014
44. W Huang, G Yang, X Zhao, Z Li. Prediction of HLA-DRB1 0401 binding peptides using support vector machine. ***International Journal of Data Mining and Bioinformatics***, 2014
45. B Ma, AO Charkowski, JD Glasner, NT Perna. Identification of host-microbe interaction factors in the genomes of soft rot-associated pathogens Dickeya dadantii 3937 and Pectobacterium carotovorum WPP14 with supervised machine learning. ***BMC Genomics***, 2014
46. CO Sakar, O Kursun, H Seker, F Gurgen. Combining multiple views: Case studies on protein and arrhythmia features. ***Engineering Applications of Artificial Intelligence***, 2014
47. N Xiao, Q Xu, D Cao. protr: Protein Sequence Descriptor Calculation and Similarity Computation with R. ***R Package***, 2014
48. Y Liu, Y Li, Z Huang, Z Xu, Z Yang, Z Chen. Multi-algorithm and multi-model based drug target prediction and web server. ***Acta Pharmacologica Sinica***, 2014
49. SE Richards, E Holmes. Transcriptomics, proteomics, metabolomics, and metagenomics datasets. ***Metabolomics as a Tool in Nutrition Research***, 2014
50. JJ Valdes, IH Moal. Prediction of Kunitz ion channel effectors and protease inhibitors from the Ixodes ricinus sialome. ***Ticks and Tick-Borne Diseases***, 2014
51. P Klus, B Bolognesi, F Agostini, D Marchese. The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. ***Bioinformatics***, 2014
52. L Li, S Yu, W Xiao, Y Li, W Hu, L Huang, X Zheng. Protein submitochondria localization from integrated sequence repesentation and SVM-based backward feature extraction. ***Molecular BioSystems***, 2014

53. JB Brown, Y Okuno, G Marcou, A Varnek. Computational chemogenomics: Is it more than inductive transfer? ***Journal of Computer-Aided Molecular Design***, 2014
54. RD Roy, D Dash. Selection of relevant features from amino acids enables development of robust classifiers. ***Amino Acids***, 2014
55. W Zhang, M Ke. Protein Encoding: A Matlab toolbox of representing or encoding protein sequences as numerical vectors for bioinformatics. ***Journal of Chemical & Pharmaceutical Research***, 2014

2013:

56. C Mathe, S Devineau, JC Aude, G Lagniel, S Chedin. Structural determinants for protein adsorption/non-adsorption to silica surface. ***PLOS One***, 2013
57. GS Han, ZG Yu, V Anh, AP Krishnajith, YC Tian. An ensemble method for predicting subnuclear localizations from primary protein structures. ***PLOS One***, 2013
58. M Lapins, A Worachartcheewan, O Spjuth, V Georgiev. A unified proteochemometric model for prediction of inhibition of cytochrome P450 isoforms. ***PLOS One***, 2013
59. RG Ramani, SG Jacob. Improved classification of lung cancer tumors Based on structural and physicochemical properties of proteins using data mining models. ***PLOS One***, 2013
60. DS Cao, QS Xu, YZ Liang. propy: a tool to generate various modes of Chou's PseAAC. ***Bioinformatics***, 2013
61. TH Chang, LC Wu, TY Lee, SP Chen. EuLoc: a web-server for accurately predict protein subcellular localization in eukaryotes by incorporating various features of sequence segments into the general form of Chou's PseAAC. ***Journal of Computer-Aided Molecular Design***, 2013
62. GT Valente, ML Acencio, C Martins, N Lemke. The development of a universal in silico predictor of protein-protein interactions. ***PLOS One***, 2013
63. GJP van Westen, RF Swier, I Cortes-Ciriano. Benchmarking of protein descriptor sets in proteochemometric modeling (part 2): modeling performance of 13 amino acid descriptor sets. ***Journal of Cheminformatics***, 2013
64. F Hosseinzadeh, AH KayvanJoo, M Ebrahimi. Prediction of lung tumor types based on protein attributes by machine learning algorithms. ***Springerplus***, 2013
65. DS Cao, YZ Liang, J Yan, GS Tan. PyDPI: freely available Python package for chemoinformatics, bioinformatics, and chemogenomics studies. ***Journal of Chemical Information and Modeling***, 2013
66. AN Sarangi, M Lohani. Prediction of essential proteins in prokaryotes by incorporating various physico-chemical features into the general form of Chou's pseudo amino acid composition. ***Protein and Peptide Letters***, 2013
67. A Shiraishi, S Niijima, JB Brown. Chemical Genomics Approach for GPCR-Ligand Interaction Prediction and Extraction of Ligand Binding Determinants. ***Journal of Chemical Information and Modeling***, 2013
68. M Wang, Z Sun, T Akutsu, J Song. Recent advances in predicting functional impact of single amino acid polymorphisms: a review of useful features, computational methods and available tools. ***Current Bioinformatics***, 2013
69. G Govindan, AS Nair. Bagging with CTD - A Novel Signature for the Hierarchical Prediction of Secreted Protein Trafficking in Eukaryotes. ***Genomics, proteomics & bioinformatics***, 2013
70. J Wang, W Yu, Z Chen, H Li. Predicting Drug-Target Interactions of Nuclear Receptors Based on Molecular Descriptors Information. ***Letters in Drug Design & Discovery***, 2013

71. L Pollo-Oliveira, H Post, ML Acencio. Unravelling the Neospora caninum secretome through the secreted fraction (ESA) and quantification of the discharged tachyzoite using high-resolution mass spectrometry-based proteomics. ***Parasites & Vectors***, 2013
72. D Rognan. Novel chemogenomic approaches to drug design. ***Chemical Genomics and Proteomics***, 2013
73. V Subramanian, P Prusis, LO Pietila. Visually Interpretable Models of Kinase Selectivity Related Features Derived from Field-Based Proteochemometrics. ***Journal of Chemical Information and Modeling***, 2013
74. L Xi, S Li, Y Wei, X Wu, H Liu, X Yao. Recognition of protein folding kinetics pathways based on amino acid properties information derived from primary sequence. ***Chemometrics and Intelligent Laboratory Systems***, 2013
75. AS Hall, J Huan. KUChemBio: A repository of computational chemical biology data sets. ***Big Data IEEE International Conference***, 2013
76. Z Mao, GS Han, TT Wang. Effects of amino acid classification on prediction of protein structural classes. ***Fuzzy Systems and Knowledge Discovery (FSKD), 10th International Conference***, 2013
77. S Nanuwa. Investigation into the role of sequence-driven-features and amino acid indices for the prediction of structural classes of proteins. ***BioInformatics and BioEngineering (BIBE) 8th IEEE International Conference***, 2013

2012:

78. F Cheng, Y Zhou, J Li, W Li, G Liu, Y Tang. Prediction of chemical-protein interactions: multitarget-QSAR versus computational chemogenomic methods. ***Molecular BioSystems***, 2012
79. H Yu, J Chen, X Xu, Y Li, H Zhao, Y Fang, X Li. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. ***PLOS One***, 2012
80. F Hosseinzadeh, M Ebrahimi, B Goliaei, N Shamabadi. Classification of lung cancer tumors based on structural and physicochemical properties of proteins by bioinformatics models. ***PLOS One***, 2012
81. PVPS Arun, RK Bakku, M Subhashini, P Singh. CyanoPhyChe: A Database for Physico-Chemical Properties, Structure and Biochemical Pathway Information of Cyanobacterial Proteins. ***PLOS One***, 2012
82. L Li, Y Zhang, L Zou, C Li, B Yu, X Zheng, Y Zhou. An ensemble classifier for eukaryotic protein subcellular location prediction using gene ontology categories and amino acid hydrophobicity. ***PLOS One***, 2012
83. XW Zhao, ZQ Ma, MH Yin. Predicting protein-protein interactions by combing various sequence-derived features into the general form of Chou's pseudo amino acid composition. ***Protein and peptide letters***, 2012
84. S Niijima, A Shiraishi, Y Okuno. Dissecting kinase profiling data to predict activity and understand cross-reactivity of kinase inhibitors. ***Journal of Chemical Information and Modeling***, 2012
85. E Gumus, O Kursun, A Sertbas, D Ustek. Application of canonical correlation analysis for identifying viral integration preferences. ***Bioinformatics***, 2012
86. M Hayat, A Khan. Mem-PHybrid: Hybrid features-based prediction system for classifying membrane protein types. ***Analytical Biochemistry***, 2012
87. H Gonzalez-Diaz. New Markov-autocorrelation indices for re-evaluation of links in chemical and biological complex networks used in metabolomics, parasitology, neurosciences, and epidemiology. ***Journal of Chemical Information and Modeling***, 2012

88. ZC Li, YH Lai, LL Chen, X Zhou, Z Dai, XY Zou. Identification of human protein complexes from local sub-graphs of protein-protein interaction network based on random forest with topological structure features. ***Analytica Chimica Acta***, 2012
89. S Sunil Kumaran Nair, NV Subba Reddy. Machine learning study of Classifiers trained with Biophysiochemical properties of amino acids to predict fibril forming peptide motifs. ***Protein and Peptide Letters***, 2012
90. R Tao, Z Jiang, W Yu, J Wang. Predicting Coupling Specificity of GPCRs Based on the Optimization of the Coupling Regions. ***Combinatorial Chemistry & High Throughput Screening***, 2012
91. M Anurag, GP Singh, D Dash. Location of disorder in coiled coil proteins is influenced by its biological role and subcellular localization: a GO-based study on human proteome. ***Molecular BioSystems***, 2012
92. Y He, Z Xiang. Bioinformatics analysis of bacterial protective antigens in manually curated Protegen database. ***Procedia in Vaccinology***, 2012
93. A Schuffenhauer. Computational methods for scaffold hopping. ***Wiley Interdisciplinary Reviews: Computational Molecular Science***, 2012
94. LQ Li, Y Zhang, LY Zou, Y Zhou. Prediction of protein subcellular multi-localization based on the general form of Chou's pseudo amino acid composition. ***Protein and Peptide Letters***, 2012
95. S Joseph, S Karnik, P Nilawe, VK Jayaraman. ClassAMP: a prediction tool for classification of antimicrobial peptides. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, 2012
96. B Karacali. Hierarchical motif vectors for prediction of functional sites in amino acid sequences using quasi-supervised learning. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, 2012
97. KY Chen, CB Yang, KS Huang. Prediction of protein backbone structure by preference classification with SVM. ***Proceedings of the 9th International Conference on Information Systems and Technology Management***, 2012
98. JB Brown, S Niijima, A Shiraishi. Chemogenomic approach to comprehensive predictions of ligand-target interactions: A comparative study. ***Bioinformatics and Biomedicine Workshops (BIBMW) IEEE International Conference***, 2012

2011:

99. X Shao, CSH Tan, C Voss, SSC Li, N Deng. A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain-peptide interaction from primary sequences. ***Bioinformatics***, 2011
100. K Kavousi, B Moshiri, M Sadeghi, BN Araabi. A protein fold classifier formed by fusing different modes of pseudo amino acid composition via PSSM. ***Computational Biology and Chemistry***, 2011
101. DG Udatha, I Kouskoumvekaki, L Olsson. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. ***Biotechnology Advances***, 2011
102. H Yabuuchi, S Niijima, H Takematsu. Analysis of multiple compound-protein interactions reveals novel bioactive molecules. ***Molecular Systems Biology***, 2011
103. CR Andersson, MG Gustafsson. Quantitative chemogenomics: machine-learning models of protein-ligand interaction. ***Current Topics in Medicinal Chemistry***, 2011
104. SSK Nair, NVS Reddy, KS Hareesha. Exploiting heterogeneous features to improve in silico prediction of peptide status-amyloidogenic or non-amyloidogenic. ***BMC Bioinformatics***, 2011

105. S Li, L Xi, J Li, C Wang, B Lei, Y Shen. In silico prediction of deleterious single amino acid polymorphisms from amino acid sequence. ***Journal of Computational Chemistry***, 2011
106. D Restrepo-Montoya, C Pino, LF Nino. NClassG+: A classifier for non-classically secreted Gram-positive bacterial proteins. ***BMC Bioinformatics***, 2011
107. C Wang, S Li, L Xi, H Liu, X Yao. Accurate prediction of the burial status of transmembrane residues of α-helix membrane protein by incorporating the structural and physicochemical features. ***Amino Acids***, 2011
108. W Yu, Z Jiang, J Wang, R Tao. Using feature selection technique for drug-target interaction networks prediction. ***Current Medicinal Chemistry***, 2011
109. C Chrysostomou, H Seker, N Aydin. Investigation into the effects of an individual amino acid on protein function by means of a resonant recognition model. ***Convergence and Hybrid Information Technology***, 2011
110. M Singh, G Singh. Development of Predictor for Sequence Derived Features from Amino Acid Sequence using Associate Rule Mining. ***International Journal of Computer Science and Security***, 2011
111. G Govindan, AS Nair. Composition, Transition and Distribution (CTD) - A dynamic feature for predictions based on hierarchical structure of cellular sorting. ***India Conference (INDICON) Annual IEEE***, 2011
112. G Govindan, AS Nair. Classification of proteins in intracellular and secretory pathway using global descriptors of amino acid sequence. ***Information and Communication Technologies (WICT) World Congress***, 2011
113. C Chrysostomou, H Seker. Effects of windowing and zero-padding on complex resonant recognition model for protein sequence analysis. ***Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE***, 2011

2010:

114. M Lapins, JES Wikberg. Kinome-wide interaction modelling using alignment-based and alignment-independent approaches for kinase description and linear and non-linear data analysis techniques. ***BMC Bioinformatics***, 2010
115. QB Gao, XF Ye, ZC Jin, J He. Improving discrimination of outer membrane proteins by fusing different forms of pseudo amino acid composition. ***Analytical Biochemistry***, 2010
116. ZL Peng, JY Yang, X Chen. An improved classification of G-protein-coupled receptors using sequence-derived features. ***BMC Bioinformatics***, 2010
117. X Zhou, Z Li, Z Dai, X Zou. QSAR modeling of peptide biological activity by coupling support vector machine with particle swarm optimization algorithm and genetic algorithm. ***Journal of Molecular Graphics and Modelling***, 2010
118. N Ahmad, D Alahakoon, R Chau. Cluster identification and separation in the growing self-organizing map: application in protein sequence classification. ***Neural Computing and Applications***, 2010
119. L Xi, S Li, H Liu, J Li, B Lei, X Yao. Global and local prediction of protein folding rates based on sequence autocorrelation information. ***Journal of Theoretical Biology***, 2010
120. H Strombergsson, M Lapins, GJ Kleywegt. Towards Proteome-Wide Interaction Models Using the Proteochemometrics Approach. ***Molecular Informatics***, 2010
121. I Valavanis, G Spyrou, K Nikita. A similarity network approach for the analysis and comparison of protein sequence/structure sets. ***Journal of Biomedical Informatics***, 2010
122. M Li, Z Li, Z Jiang, D Li. Prediction of disease-related genes based on hybrid features. ***Current Proteomics***, 2010

123. S Niijima, H Yabuuchi, Y Okuno. Cross-target view to feature selection: identification of molecular interaction features in ligand-target space. ***Journal of Chemical Information and Modeling***, 2010
124. XM Ren, JF Xia. Prediction of protein-protein interaction sites by using autocorrelation descriptor and support vector machine. ***Advanced Intelligent Computing Theories and Applications***, 2010
125. M Gorania, H Seker, P Haris. Predicting a protein's melting temperature from its amino acid sequence. ***Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE***, 2010
126. Sakar, O Kursun, H Seker. Prediction of protein sub-nuclear location by clustering mRMR ensemble feature selection. ***Pattern Recognition (ICPR), 20th International Conference***, 2010
127. N Cai, Y Chen. Predicting protein structural class with Ensemble of Flexible Neural Tree. ***Natural Computation (ICNC), 6th International Conference***, 2010
128. J Zhang, J Huan. Novel biological network features discovery for in silico identification of drug targets. ***Proceedings of the 1st ACM International Health Informatics Symposium***, 2010
129. J Zhang, J Huan. Analysis of network topological features for identifying potential drug targets. ***Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining***, 2010
130. S Kocbek, G Stiglic, I Pernek, P Kokol. Stability of different feature selection methods for selecting protein sequence descriptors in protein solubility classification problem. ***Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Syst***ems, 2010
131. J Cui, JD Puett, S Hong, Y Xu. Protein markers identification for gastric cancer diagnosis. ***US Patent US20120053080 A1***. 2010

2009:

132. S Thomas, S Karnik, RS Barai. CAMP: a useful resource for research on antimicrobial peptides. ***Nucleic Acids***, 2009
133. H Strombergsson, GJ Kleywegt. A chemogenomics view on protein-ligand spaces. ***BMC bioinformatics***, 2009
134. AM Wassermann, H Geppert. Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects. ***Journal of Chemical Information and Modeling***, 2009
135. X Shao, Y Tian, L Wu, Y Wang, L Jing. Predicting DNA-and RNA-binding proteins from sequences with kernel methods. ***Journal of Theoretical Biology***, 2009
136. S Li, L Xi, C Wang, J Li, B Lei, H Liu. A novel method for protein-ligand binding affinity prediction and the related descriptors exploration. ***Journal of Computational Chemistry***, 2009
137. G Zhang, H Li, B Fang. Discriminating acidic and alkaline enzymes using a random forest model with secondary structure amino acid composition. ***Process Biochemistry***, 2009
138. D Faria, AEN Ferreira, AO Falcao. Enzyme classification with peptide programs: a comparative study. ***BMC Bioinformatics***, 2009
139. S Karnik, J Mitra, A Singh, BD Kulkarni. Identification of n-glycosylation sites with sequence and structural features employing random forests. ***Pattern Recognition and Machine Intelligence***, 2009
140. A Lindstrom, F Pettersson, A Linusson. Quantitative protein descriptors for secondary structure characterization and protein classification. ***Chemometrics and Intelligent Laboratory Systems***, 2009

141. J Hu, F Zhang. Improving protein localization prediction using amino acid group based physiochemical encoding. ***Bioinformatics and Computational Biology***, 2009
142. Y Chen, W Li, N Cai. Predicting protein structural class based on ensemble binary classification. ***Natural Computation (ICNC), 5th International Conference***, 2009
143. CO Sakar, O Kursun, H Seker, F Gurgen. Parallel interacting multiview learning: An application to prediction of protein sub-nuclear location. ***Information Technology and Applications in Biomedicine (ITAB), 9th International Conference***, 2009
144. C Newby, Y Yang, H Seker. Investigation into effectiveness of rough sets in prediction of enzyme and protein structure classes. ***Neural Networks (IJCNN) International Joint Conference***, 2009
145. GS Han, ZG Yu, V Anh, RH Chan. Distinguishing coding from non-coding sequences in a prokaryote complete genome based on the global descriptor. ***Fuzzy Systems and Knowledge Discovery (FSKD), 6th International Conference***, 2009

2008:

146. GY Zhang, BS Fang. Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo-amino acid composition. ***Journal of Theoretical Biology***, 2008
147. GY Zhang, HC Li, JQ Gao. Predicting lipase types by improved Chou's pseudo-amino acid composition. ***Protein and Peptide Letters***, 2008
148. C Chen, LX Chen, XY Zou, PX Cai. Predicting protein structural class based on multi-features fusion. ***Journal of Theoretical Biology***, 2008
149. P Minkiewicz, J Dziuba, A Iwaniak. BIOPEP database and other programs for processing bioactive peptide sequences. ***Journal of AOAC International***, 2008
150. ZC Li, XB Zhou, YR Lin, XY Zou. Prediction of protein structure class by coupling improved genetic algorithm and support vector machine. ***Amino Acids***, 2008
151. JY Yang, Y Zhou, ZG Yu, V Anh, LQ Zhou. Human Pol II promoter recognition based on primary sequences and free energy of dinucleotides. ***BMC bioinformatics***, 2008
152. YZ Chen, F Zhu, LY Han, X Chen. Homology-free prediction of functional class of proteins and peptides by support vector machines. ***Current Protein and Peptide Science***, 2008
153. P Jain, P Wadhwa, R Aygun, G Podila. Vector-G: multi-modular SVM-based heterotrimeric g protein prediction. ***In silico Biology***, 2008
154. B Senliol, G Gulgezen, L Yu, Z Cataltepe. Fast Correlation Based Filter (FCBF) with a different search strategy. ***Computer and Information Sciences (ISCIS), 23rd International Symposium***, 2008
155. JK Valavanis, GM Spyrou. Protein similarity networks and genetic algorithm driven feature selection for fold recognition. ***BioInformatics and BioEngineering (BIBE), 8th IEEE International Conference***, 2008
156. SA Rahman, ZAM Hussein. Data mining framework for protein function prediction. ***Information Technology (ITSim), International Symposium***, 2008
157. H Seker. Novel weighted amino acid composition for prediction of protein structural classes within the context of multi-sensor data fusion approach. ***BioInformatics and BioEngineering (BIBE), 8th IEEE International Conference***, 2008

2007:

158. SAK Ong, HH Lin, YZ Chen, ZR Li, Z Cao. Efficacy of different protein descriptors in predicting protein functional families. ***BMC Bioinformatics***, 2007
159. LY Han, CJ Zheng, B Xie, J Jia, XH Ma, F Zhu. Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. ***Drug Discovery Today***, 2007
160. S Ray, TB Kepler. Amino acid biophysical properties in the statistical prediction of peptide-MHC class I binding. ***Immunome Research***, 2007
161. J Cui, L Han, H Lin, Z Tang, Z Ji, Z Cao. Advances in exploration of machine learning methods for predicting functional class and interaction profiles of proteins and peptides irrespective of sequence Homology. ***Current Bioinformatics***, 2007
162. ZQ Tang, HH Lin, HL Zhang, LY Han. Prediction of functional class of proteins and peptides irrespective of sequence homology by support vector machines. ***Bioinformatics and Biology Insights***, 2007