

Qiu, Zhijun; Wang, Xicheng

**Prediction of protein-protein interaction sites using patch-based residue characterization.**

(English) [Zbl 1307.92088](#)

J. Theor. Biol. 293, 143-150 (2012).

Summary: Identifying protein-protein interaction sites provides important clues to the function of a protein and is becoming increasingly relevant in topics such as systems biology and drug discovery. Using a patch-based model for residue characterization, we trained random forest classifiers for residue-based interface prediction, which was followed by a clustering procedure to produce patches for patch-based interface prediction. For residue-based interface prediction, our method achieves a specificity rate of 0.7 and a sensitivity rate of 0.78. For patch-based interface prediction, a success rate of 0.80 is achieved. Based on same datasets, we also compare it with several published methods. The results show that our method is a successful predictor for residue-based and patch-based interface prediction.

MSC:

92C40 Biochemistry, molecular biology

Cited in 4 Documents

Keywords:

random forests; multiple-patch model; residue clustering

Software:

AFP-Pred; Cell-PLoc; Euk-mPLoc; GPCR-2L; GPCR-CA; GPCR-GIA; iLoc-Euk; iLoc-Virus; NR-2L; Plant-mPLoc; PSAIA; Quat-2L; SecretP

Full Text: [DOI](#)

References:

- [1] Abagyan, R., Protein structure prediction by global energy optimization, (), 32
- [2] Andraos, J., Kinetic plasticity and the determination of product ratios for kinetic schemes leading to multiple products without rate laws: new methods based on directed graphs, Can. J. chem., 86, 342-357, (2008)
- [3] Bahar, I.; Atilgan, A.R.; Erman, B., Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Fold. des., 2, 173-181, (1997)
- [4] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E., The protein data bank, Nucleic acids res., 28, 235-242, (2000)
- [5] Bordner, A.J.; Abagyan, R., Statistical analysis and prediction of protein – protein interfaces, Proteins, 60, 353-366, (2005)
- [6] Bradford, J.R.; Needham, C.J.; Bulpitt, A.J.; Westhead, D.R., Insights into protein – protein interfaces using a Bayesian network prediction method, J. mol. biol., 362, 365-386, (2006)
- [7] Bradford, J.R.; Westhead, D.R., Improved prediction of protein – protein binding sites using a support vector machines approach, Bioinformatics, 21, 1487-1494, (2005)
- [8] Breiman, L., Random forests, Mach. learn., 45, 5-32, (2001) · [Zbl 1007.68152](#)
- [9] Chelliah, V.; Blundell, T.L.; Fernández-Recio, J., Efficient restraints for protein – protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment, J. mol. biol., 357, 1669-1682, (2006)
- [10] Chen, C.; Chen, L.; Zou, X.; Cai, P., Prediction of protein secondary structure content by using the concept of Chou's pseudo amino acid composition and support vector machine, Protein pept. lett., 16, 27-31, (2009)
- [11] Chen, H.L.; Zhou, H.X., Prediction of interface residues in protein – protein complexes by a consensus neural network method: test against NMR data, Proteins, 61, 21-35, (2005)
- [12] Chen, X.W.; Jeong, J.C., Sequence-based prediction of protein interaction sites with an integrative method, Bioinformatics, 25, 585-591, (2009)
- [13] Chou, K.C., Review: low-frequency collective motion in biomacromolecules and its biological functions, Biophys. chem., 30, 3-48, (1988)
- [14] Chou, K.C., Graphic rules in steady and non-steady enzyme kinetics, J. biol. chem., 264, 12074-12079, (1989)

- [15] Chou, K.C., Low-frequency resonance and cooperativity of hemoglobin, *Trends biochem. sci.*, 14, 212, (1989)
- [16] Chou, K.C., A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space, *Proteins*, 21, 319-344, (1995)
- [17] Chou, K.C., Prediction of protein cellular attributes using pseudo amino acid composition, *Proteins*, 43, 246-255, (2001), (Erratum: *Proteins*, 2001, 44, 60)
- [18] Chou, K.C., Graphic rule for drug metabolism systems, *Curr. drug metab.*, 11, 369-378, (2010)
- [19] Chou, K.C., Some remarks on protein attribute prediction and pseudo amino acid composition (50th anniversary year review), *J. theor. biol.*, 273, 236-247, (2011) · [Zbl 1405.92212](#)
- [20] Chou, K.C.; Cai, Y.D., Predicting protein – protein interactions from sequences in a hybridization space, *J. proteome res.*, 5, 316-322, (2006)
- [21] Chou, K.C.; Shen, H.B., Review: recent progresses in protein subcellular location prediction, *Anal. biochem.*, 370, 1-16, (2007)
- [22] Chou, K.C.; Shen, H.B., Euk-mploc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites, *J. proteome res.*, 6, 1728-1734, (2007)
- [23] Chou, K.C.; Shen, H.B., Cell-ploc: a package of web servers for predicting subcellular localization of proteins in various organisms (updated version: cell-ploc 2.0: an improved package of web-servers for predicting subcellular localization of proteins in various organisms. *natural science*, 2010, 2, 1090-1103), *Nat. protocols*, 3, 153-162, (2008)
- [24] Chou, K.C.; Shen, H.B., Review: recent advances in developing web-servers for predicting protein attributes, *Nat. sci.*, 2, 63-92, (2009)
- [25] Chou, K.C.; Shen, H.B., Plant-mploc: a top – down strategy to augment the power for predicting plant protein subcellular localization, *Plos one*, 5, e11335, (2010)
- [26] Chou, K.C.; Wu, Z.C.; Xiao, X., Iloc-euk: a multi-label classifier for predicting the subcellular localization of singleplex and multiplex eukaryotic proteins, *Plos one*, 6, e18258, (2011)
- [27] Chou, K.C.; Zhang, C.T., Review: prediction of protein structural classes, *Crit. rev. biochem. mol. biol.*, 30, 275-349, (1995)
- [28] Chou, K.C.; Zhou, G.P., Role of the protein outside active site on the diffusion-controlled reaction of enzyme, *J. am. chem. soc.*, 104, 1409-1413, (1982)
- [29] Chung, J.L.; Wang, W.; Bourne, P.E., Exploiting sequence and structure homologs to identify protein – protein binding sites, *Proteins*, 62, 630-640, (2006)
- [30] de Vries, S.J.; Bonvin, A.M.J.J., How proteins get in touch: interface prediction in the study of biomolecular complexes, *Curr. protein pept. sci.*, 9, 394-406, (2008)
- [31] de Vries, S.J.; van Dijk, A.D.; Bonvin, A.M., WHISCY: what information does surface conservation yield? application to data-driven docking, *Proteins*, 63, 479-489, (2006)
- [32] Ding, H.; Luo, L.; Lin, H., Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition, *Protein pept. lett.*, 16, 351-355, (2009)
- [33] Dong, Q.W.; Wang, X.L.; Lin, L.; Guan, Y., Exploiting residue-level and profile-level interface propensities for usage in binding sites prediction of proteins, *BMC bioinf.*, 8, 147, (2007)
- [34] Esmaceli, M.; Mohabatkar, H.; Mohsenzadeh, S., Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses, *J. theor. biol.*, 263, 203-209, (2010)
- [35] Fariselli, P.; Pazos, F.; Valencia, A.; Casadio, R., Prediction of protein – protein interaction sites in heterocomplexes with neural networks, *Eur. J. biochem.*, 269, 1356-1361, (2002)
- [36] Fauchere, J.L.; Pliska, V., Hydrophobic parameters- $\pi$  of amino-acid side-chains from the partitioning of n-acetyl-amino-acid amides, *Eur. J. med. chem.*, 18, 369-375, (1983)
- [37] Fernandez-Recio, J.; Totrov, M.; Abagyan, R., Identification of protein – protein interaction sites from docking energy landscapes, *J. mol. biol.*, 335, 843-865, (2004)
- [38] Georgiou, D.N.; Karakasidis, T.E.; Nieto, J.J.; Torres, A., Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition, *J. theor. biol.*, 257, 17-26, (2009)
- [39] Gu, Q.; Ding, Y.S.; Zhang, T.L., Prediction of G-protein-coupled receptor classes in low homology using Chou's pseudo amino acid composition with approximate entropy and hydrophobicity patterns, *Protein pept. lett.*, 17, 559-567, (2010)
- [40] Heuser, P.; Bau, D.; Benkert, P.; Schomburg, D., Refinement of unbound protein docking studies using biological knowledge, *Proteins*, 61, 1059-1067, (2005)
- [41] Higa, R.H.; Tozzi, C.L., A simple and efficient method for predicting protein – protein interaction sites, *Genet. mol. res.*, 7, 898-909, (2008)
- [42] He, Z.; Zhang, J.; Shi, X.H.; Hu, L.L.; Kong, X.; Cai, Y.D.; Chou, K.C., Predicting drug – target interaction networks based on functional groups and biological features, *Plos one*, 5, e9603, (2010)
- [43] Hu, L.; Huang, T.; Shi, X.; Lu, W.C.; Cai, Y.D.; Chou, K.C., Predicting functions of proteins in mouse based on weighted protein – protein interaction network and protein hybrid properties, *Plos one*, 6, e14556, (2011)
- [44] Huang, T.; Shi, X.H.; Wang, P.; He, Z.; Feng, K.Y.; Hu, L.; Kong, X.; Li, Y.X.; Cai, Y.D.; Chou, K.C., Analysis and prediction of the metabolic stability of proteins based on their sequential features, subcellular locations and interaction networks, *Plos one*, 5, e10972, (2010)
- [45] Jia, S.C.; Hu, X.Z., Using random forest algorithm to predict beta-hairpin motifs, *Protein pept. lett.*, 18, 609-617, (2011)

- [46] Jones, S.; Thornton, J.M., Prediction of protein – protein interaction sites using patch analysis, *J. mol. biol.*, 272, 133-143, (1997)
- [47] Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 22, 2577-2637, (1983)
- [48] Kandaswamy, K.K.; Chou, K.C.; Martinetz, T.; Moller, S.; Suganthan, P.N.; Sridharan, S.; Pugalenti, G., AFP-pred: a random forest approach for predicting antifreeze proteins from sequence-derived properties, *J. theor. biol.*, 270, 56-62, (2011)
- [49] Koike, A.; Takagi, T., Prediction of protein – protein interaction sites using support vector machines, *Protein eng. des. sel.*, 17, 165-173, (2004)
- [50] Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M.; Abagyan, R., PIER: protein interface recognition for structural proteomics, *Proteins*, 67, 400-417, (2007)
- [51] Kyte, J.; Doolittle, R.F., A simple method for displaying the hydropathic character of a protein, *J. mol. biol.*, 157, 105-132, (1982)
- [52] Li, J.J.; Huang, D.S.; Wang, B.; Chen, P., Identifying protein – protein interfacial residues in heterocomplexes using residue conservation scores, *Int. J. biol. macromol.*, 38, 241-247, (2006)
- [53] Liang, S.D.; Zhang, C.; Liu, S.; Zhou, Y., Protein binding site prediction using an empirical scoring function, *Nucleic acids res.*, 34, 3698-3707, (2006)
- [54] Lin, H., The modified Mahalanobis discriminant for predicting outer membrane proteins by using Chou's pseudo amino acid composition, *J. theor. biol.*, 252, 350-356, (2008)
- [55] Lin, W.Z.; Xiao, X.; Chou, K.C., GPCR-GIA: a web-server for identifying G-protein coupled receptors and their families with grey incidence analysis, *Protein eng. des. sel.*, 22, 699-705, (2009)
- [56] Madkan, A.; Blank, M.; Elson, E.; Chou, K.C.; Geddis, M.S.; Goodman, R., Steps to the clinic with ELF EMF, *Nat. sci.*, 1, 157-165, (2009)
- [57] Mihel, J.; Sikic, M.; Tomić, S.; Jeren, B.; Vlahovick, K., PSAIA—protein structure and interaction analyzer, *BMC struct. biol.*, 8, 21, (2008)
- [58] Mohabatkar, H., Prediction of cyclin proteins using Chou's pseudo amino acid composition, *Protein pept. lett.*, 17, 1207-1214, (2010)
- [59] Mohabatkar, H.; Mohammad Beigi, M.; Esmaili, A., Prediction of GABA(A) receptor proteins using the concept of Chou's pseudo-amino acid composition and support vector machine, *J. theor. biol.*, 281, 18-23, (2011)
- [60] Neuvirth, H.; Raz, R.; Schreiber, G., Promate: a structure based prediction program to identify the location of protein – protein binding sites, *J. mol. biol.*, 338, 181-199, (2004)
- [61] Ofran, Y.; Rost, B., Predicted protein – protein interaction sites from local sequence information, *FEBS lett.*, 544, 236-239, (2003)
- [62] Ofran, Y.; Rost, B., ISIS: interaction sites identified from sequence, *Bioinformatics*, 23, E13-E16, (2007)
- [63] Pettit, F.K.; Bare, E.; Tsai, A.; Bowie, J.U., Hotpatch: a statistical approach to finding biologically relevant features on protein surfaces, *J. mol. biol.*, 369, 863-879, (2007)
- [64] Porollo, A.; Meller, J., Prediction-based fingerprints of protein – protein interactions, *Proteins*, 66, 630-645, (2007)
- [65] Qin, S.B.; Zhou, H.X., A holistic approach to protein docking, *Proteins*, 69, 743-749, (2007)
- [66] Qiu, J.D.; Huang, J.H.; Shi, S.P.; Liang, R.P., Using the concept of Chou's pseudo amino acid composition to predict enzyme family classes: an approach with support vector machine based on discrete wavelet transform, *Protein pept. lett.*, 17, 715-722, (2010)
- [67] Qiu, Z.J.; Wang, X.C., Identification of ligand-binding pockets in proteins using residue preference methods, *Protein pept. lett.*, 16, 984-990, (2009)
- [68] Ren, L.H.; Shen, Y.Z.; Ding, Y.S.; Chou, K.C., Bio-entity network for analysis of protein – protein interaction networks, *Asian J. control*, 13, 726-737, (2011) · [Zbl 1303.93030](#)
- [69] Res, I.; Mihalek, I.; Lichtarge, O., An evolution based classifier for prediction of protein interfaces without using protein structures, *Bioinformatics*, 21, 2496-2501, (2005)
- [70] Sanner, M.F.; Olson, A.J.; Spehner, J.C., Reduced surface: an efficient way to compute molecular surfaces, *Biopolymers*, 38, 305-320, (1996)
- [71] Sikic, M.; Tomic, S.; Vlahovick, K., Prediction of protein – protein interaction sites in sequences and 3D structures by random forests, *Plos comput. biol.*, 5, e1000278, (2009)
- [72] Tjong, H.; Qin, S.; Zhou, H.X., PI2PE: protein interface/interior prediction engine, *Nucleic acids res.*, 35, W357-W362, (2007)
- [73] Tress, M.; de Juan, D.; Graña, O.; Gómez, M.J.; Gómez-Puertas, P.; González, J.M.; López, G.; Valencia, A., Scoring docking models with evolutionary information, *Proteins*, 60, 275-280, (2005)
- [74] van Dijk, A.D.J.; de Vries, S.J.; Dominguez, C.; Chen, H.; Zhou, H.X.; Bonvin, A.M.J.J., Data-driven docking: HADDOCK's adventures in CAPRI, *Proteins*, 60, 232-238, (2005)
- [75] Wang, B.; Chen, P.; Huang, D.S.; Li, J.J.; Lok, T.M.; Lyu, M.R., Predicting protein interaction sites from residue spatial sequence profile and evolution rate, *FEBS lett.*, 580, 380-384, (2006)
- [76] Wang, B.; Wong, H.S.; Huang, D.S., Inferring protein – protein interacting sites using residue conservation and evolutionary information, *Protein pept. lett.*, 13, 999-1005, (2006)

- [77] Wang, P.; Xiao, X.; Chou, K.C., NR-2L: a two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features, *Plos one*, 6, e23505, (2011)
- [78] Wesson, L.; Eisenberg, D., Atomic solvation parameters applied to molecular-dynamics of proteins in solution, *Protein sci.*, 1, 227-235, (1992)
- [79] Xia, J.F.; Han, K.; Huang, D.S., Sequence-based prediction of protein – protein interactions by means of rotation forest and autocorrelation descriptor, *Protein pept. lett.*, 17, 137-145, (2010)
- [80] Xiao, X.; Wang, P.; Chou, K.C., Predicting protein structural classes with pseudo amino acid composition: an approach using geometric moments of cellular automaton image, *J. theor. biol.*, 254, 691-696, (2008)
- [81] Xiao, X.; Lin, W.Z.; Chou, K.C., Using grey dynamic modeling and pseudo amino acid composition to predict protein structural classes, *J. comput. chem.*, 29, 2018-2024, (2008)
- [82] Xiao, X.; Wang, P.; Chou, K.C., GPCR-CA: a cellular automaton image approach for predicting G-protein-coupled receptor functional classes, *J. comput. chem.*, 30, 1414-1423, (2009)
- [83] Xiao, X.; Chou, K.C., Using pseudo amino acid composition to predict protein attributes via cellular automata and others approaches, *Curr. bioinf.*, 2011, 6, 251-260, (2011)
- [84] Xiao, X.; Wang, P.; Chou, K.C., GPCR-2L: predicting G protein-coupled receptors and their types by hybridizing two different modes of pseudo amino acid compositions, *Mol. biosyst.*, 7, 911-919, (2011)
- [85] Xiao, X.; Wang, P.; Chou, K.C., Quat-2L: a web-server for predicting protein quaternary structural attributes, *Mol. divers.*, 15, 149-155, (2011)
- [86] Xiao, X.; Wu, Z.C.; Chou, K.C., A multi-label classifier for predicting the subcellular localization of Gram-negative bacterial proteins with both single and multiple sites, *Plos one*, 6, e20592, (2011)
- [87] Xiao, X.; Wu, Z.C.; Chou, K.C., Iloc-virus: a multi-label learning classifier for identifying the subcellular localization of virus proteins with both single and multiple sites, *J. theor. biol.*, 284, 42-51, (2011)
- [88] Yan, C., Dobbs, D., Honavar, V., 2003. Identification of surface residues involved in protein – protein interaction—a support vector machine approach. In: *Proceedings of the Conference on Intelligence System Design Application*, pp. 53-62.
- [89] Yang, J.; Jiang, X.F., A novel approach to predict protein – protein interactions related to Alzheimer’s disease based on complex network, *Protein pept. lett.*, 17, 356-366, (2010)
- [90] Yu, L.; Guo, Y.; Li, Y.; Li, G.; Li, M.; Luo, J.; Xiong, W.; Qin, W., Secretp: identifying bacterial secreted proteins by fusing new features into Chou’s pseudo-amino acid composition, *J. theor. biol.*, 267, 1-6, (2010)
- [91] Zeng, Y.H.; Guo, Y.Z.; Xiao, R.Q.; Yang, L.; Yu, L.Z.; Li, M.L., Using the augmented Chou’s pseudo amino acid composition for predicting protein submitochondria locations based on auto covariance approach, *J. theor. biol.*, 259, 366-372, (2009)
- [92] Zhang, G.Y.; Fang, B.S., Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou’s amphiphilic pseudo amino acid composition, *J. theor. biol.*, 253, 310-315, (2008)
- [93] Zhou, G.P., The disposition of the LZCC protein residues in wenxiang diagram provides new insights into the protein – protein interaction mechanism, *J. theor. biol.*, 284, 142-148, (2011)
- [94] Zhou, G.P.; Deng, M.H., An extension of Chou’s graphical rules for deriving enzyme kinetic equations to system involving parallel reaction pathways, *Biochem. J.*, 222, 169-176, (1984)
- [95] Zhou, H.X.; Qin, S.B., Interaction-site prediction for protein complexes: a critical assessment, *Bioinformatics*, 23, 2203-2209, (2007)
- [96] Zhou, H.X.; Shan, Y.B., Prediction of protein interaction sites from sequence profile and residue neighbor List, *Proteins*, 44, 336-343, (2001)
- [97] Zhou, X.B.; Chen, C.; Li, Z.C.; Zou, X.Y., Using Chou’s amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes, *J. theor. biol.*, 248, 546-551, (2007)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.