BOHR International Journal of Biocomputing and Nano Technology
2020, Vol. 1, No. 1, pp. 1–4

https://doi.org/10.54646/bijbnt.001
www.bohrpub.com

The pLoc_bal-mHum is a powerful web-serve for predicting the subcellular localization of human proteins purely based on their sequence information

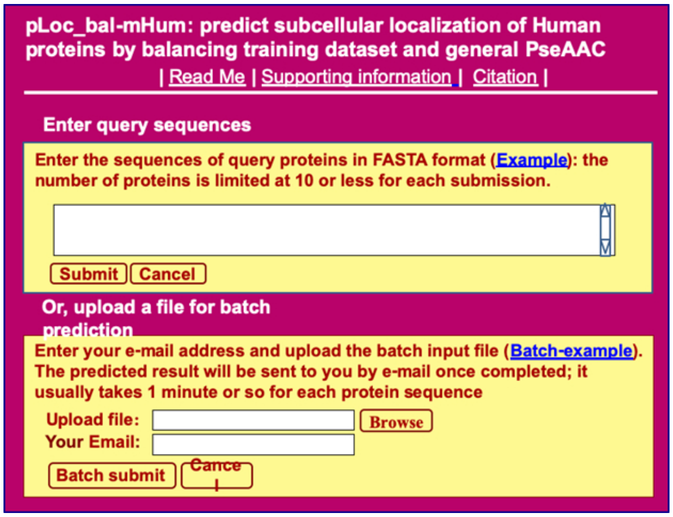
Kuo-Chen Chou

Gordon Life Science Institute, Boston, Massachusetts 02478, USA
kcchou@gordonlifescience.org or kcchou38@gmail.com

In 2019 a very powerful web-server, or AI (Artificial Intelligence) tool, has been developed for predicting the subcellular localization of human proteins purely according to their information for the multi-label systems [1], in which a same protein may appear or travel between two or more locations and hence its identification needs the multi-label mark [2].

The AI tool is named as "pLoc_bal-mEuk", where "bal" stands for that the AI tool has been further treated by balancing the training dataset [3–9], and "m" for that the AI tool can be used to cope with multi-label systems. Below, let us see how the AI tool is working.

Clicking the link at http://www.jci-bioinfo.cn/pLoc_bal-mHum/, you will see the top page of the pLoc_bal-mHum web-server appearing on your computer screen (**Figure 1**). Then by following the Step 2 and Step 3 in [5], you will see **Figure 2** on the screen of your computer. The corresponding detailed predicted results were given in ref. 5. As you can see from there: nearly all the success rates achieved by the AI tool for the human proteins in each of the 14 subcellular locations are within the range of 94–100%. Such a high prediction quality is far beyond the reach of any of its counterparts.


pLoc_bal-mHum: predict subcellular localization of Human proteins by balancing training dataset and general PseAAC

Enter query sequences

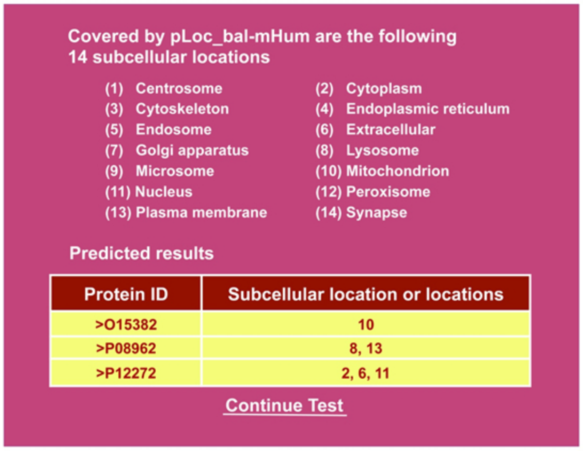
Enter the sequences of query proteins in FASTA format (Example): the number of proteins is limited at 10 or less for each submission.

Or, upload a file for batch prediction

Enter your e-mail address and upload the batch input file (Batch-example). The predicted result will be sent to you by e-mail once completed; it usually takes 1 minute or so for each protein sequence

Upload file:
Your Email:

Figure 1. A semi screenshot for the top page of pLoc_bal-mHum (Adapted from [6] with permission).


Covered by pLoc_bal-mHum are the following 14 subcellular locations

(1) Centrosome
(2) Cytoplasm
(3) Cytoskeleton
(4) Endoplasmic reticulum
(5) Endosome
(6) Extracellular
(7) Golgi apparatus
(8) Lysosome
(9) Microsome
(10) Mitochondrion
(11) Nucleus
(12) Peroxisome
(13) Plasma membrane
(14) Synapse

Predicted results

Protein ID	Subcellular location or locations
>O15382	10
>P08962	8, 13
>P12272	2, 6, 11

Figure 2. A semi screenshot for the webpage obtained by following Step 3 of Section 3.5 (Adapted from [6] with permission).

In addition to the advantages of high accuracy and easy to use, the AI tool has been built up by strictly complying with the "Chou's 5-steps rule" and hence bears the following remarkable and notable merits as concurred by many investigators (see, e.g., [10–91] as well as three comprehensive review papers [2, 92, 93]): (1) crystal clear in logic development, (2) completely transparent in operation, (3) easy to repeat the reported results by other investigators, (4) high potential in stimulating other sequence-analyzing methods, and (5) very convenient to be used by the majority of experimental scientists.

For the fantastic and awesome roles of the “5-steps rule” in driving proteome, genome analyses and drug development, see a series of recent papers [2, 93–104] where the rule and its wide applications have been very impressively presented from various aspects or at different angles.

References

[1] K.C. Chou, H.B. Shen, Recent progresses in protein subcellular location prediction. *Anal Biochem.* **370** (2007): 1–16.

[2] K.C. Chou, Advance in predicting subcellular localization of multi-label proteins and its implication for developing multi-target drugs. *Curr Med Chem.* **26** (2019): 4918–4943.

[3] X. Xiao, X. Cheng, G. Chen, Q. Mao, K.C. Chou, pLoc_bal-mVirus: Predict subcellular localization of multi-label virus proteins by Chou’s general PseAAC and IHTS treatment to balance training dataset. *Med Chem*. **15** (2019): 496–509.

[4] X. Xiao, X. Cheng, G. Chen, Q. Mao, K.C. Chou, pLoc_bal-mGpos: predict subcellular localization of Gram-positive bacterial proteins by quasi-balancing training dataset and PseAAC. *Genomics* **111** (2019): 886–892.

[5] K.C. Chou, X. Cheng, X. Xiao, pLoc_bal-mEuk: predict subcellular localization of eukaryotic proteins by general PseAAC and quasi-balancing training dataset. *Med Chem*. **15** (2019): 472–485.

[6] K.C. Chou, X. Cheng, X. Xiao, pLoc_bal-mHum: predict subcellular localization of human proteins by PseAAC and quasi-balancing training dataset. *Genomics* **111** (2019): 1274–1282.

[7] X. Cheng, W.Z. Lin, X. Xiao, K.C. Chou, pLoc_bal-mAnimal: predict subcellular localization of animal proteins by balancing training dataset and PseAAC. *Bioinformatics*. **35** (2019): 398–406.

[8] X. Cheng, X. Xiao, K.C. Chou, pLoc_bal-mPlant: predict subcellular localization of plant proteins by general PseAAC and balancing training dataset. *Curr Pharm Des*. **24** (2018): 4013–4022.

[9] X. Cheng, X. Xiao, K.C. Chou, pLoc_bal-mGneg: predict subcellular localization of Gram-negative bacterial proteins by quasi-balancing training dataset and general PseAAC. *J Theor Biol*. **458** (2018): 92–102.

[10] A.H. Butt, Y.D. Khan, Prediction of S-sulfenylation sites using statistical moments based features via Chou’s 5-step rule. *Int J Pept Res Ther.* https://doi.org/10.1007/s10989-019-09931-2 (2018).

[11] M. Awais, W. Hussain, Y.D. Khan, N. Rasool, S.A. Khan, K.C. Chou, iPhosH-PseAAC: Identify phosphohistidine sites in proteins by blending statistical moments and position relative features according to the Chou’s 5-step rule and general pseudo amino acid composition. *IEEE/ACM Trans Comput Biol Bioinform*. https://doi.org/10.1109/TCBB.2019.2919025 or https://www.ncbi.nlm.nih.gov/pubmed/31144645 (2019).

[12] O. Barukab, Y.D. Khan, S.A. Khan, K.C. Chou, iSulfoTyr-PseAAC: Identify tyrosine sulfation sites by incorporating statistical moments via Chou’s 5-steps rule and pseudo components. *Curr Genet*. https:/doi.org/10.2174/1389202920666190819091609 or http://www.eurekaselect.com/174277/article (2019).

[13] A.H. Butt, Y.D. Khan, Prediction of S-sulfenylation sites using statistical moments based features via Chou’s 5-step rule. *Int J Pept Res Ther*. htpps://doi.org/10.1007/s10989-019-09931-2 (2019).

[14] Y. Chen, X. Fan, Use of Chou’s 5-steps rule to reveal active compound and mechanism of Shuangshen Pingfei San on idiopathic pulmonary fibrosis. *Curr Mol Med*. https://doi.org/10.2174/1566524019666191011160543 (2019).

[15] X. Du, Y. Diao, H. Liu, S. Li, MsDBP: Exploring DNA-binding proteins by integrating multi-scale sequence information via Chou’s 5-steps rule. *J Proteome Res*. **18** (2019): 3119–3132.

[16] A. Ehsan, M.K. Mahmood, Y.D. Khan, O.M. Barukab, S.A. Khan, K.C. Chou, iHyd-PseAAC (EPSV): Identify hydroxylation sites in proteins by extracting enhanced position and sequence variant feature via Chou’s 5-step rule and general pseudo amino acid composition. *Curr Genomics* **20** (2019): 124–133.

[17] W. Hussain, S.D. Khan, N. Rasool, S.A. Khan, K.C. Chou, SPalmitoylC-PseAAC: A sequence-based model developed via Chou’s 5-steps rule and general PseAAC for identifying S-palmitoylation sites in proteins. *Anal Biochem*. **568** (2019): 14–23.

[18] W. Hussain, Y.D. Khan, N. Rasool, S.A. Khan, K.C. Chou, SPrenylC-PseAAC: A sequence-based model developed via Chou’s 5-steps rule and general PseAAC for identifying S-prenylation sites in proteins. *J Theor Biol*. **468** (2019): 1–11.

[19] Z. Ju, S.Y. Wang, Prediction of lysine formylation sites using the composition of k-spaced amino acid pairs via Chou’s 5-steps rule and general pseudo components. *Genomics*, htpps://doi.org/10.1016/j.ygeno.2019.05.027 or https://www.ncbi.nlm.nih.gov/pubmed/31175975 (2019).

[20] M. Kabir, S. Ahmad, M. Iqbal, M. Hayat, iNR-2L: A two-level sequence-based predictor developed via Chou’s 5-steps rule and general PseAAC for identifying nuclear receptors and their families. *Genomics*, https://www.ncbi.nlm.nih.gov/pubmed/30779939 (2019).

[21] Z.U. Khan, F. Ali, I.A. Khan, Y. Hussain, D. Pi, iRSpot-SPI: Deep learning-based recombination spots prediction byincorporating secondary sequence information coupled withphysio-chemical properties via Chou’s 5-step rule and pseudo components. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)*. **189** (2019): 169–180.

[22] J. Lan, J. Liu, C. Liao, D.J. Merkler, Q. Han, J. Li, A study for therapeutic treatment against Parkinson’s disease via Chou’s 5-steps rule. *Curr Top Med Chem*. https://doi.org/10.2174/1568026619666191019111528 or http://www.eurekaselect.com/175887/article (2019).

[23] N.Q.K. Le, iN6-methylat (5-step): identifying DNA N(6)-methyladenine sites in rice genome using continuous bag of nucleobases via Chou’s 5-step rule. *Mol Genet Genomics* **294** (2019): 1173–1182.

[24] N.Q.K. Le, E.K.Y. Yapp, Q.T. Ho, N. Nagasundaram, Y.Y. Ou, H.Y. Yeh, iEnhancer-5Step: Identifying enhancers using hidden information of DNA sequences via Chou’s 5-step rule and word embedding. *Anal Biochem*. **571** (2019): 53–61.

[25] N.Q.K. Le, E.K.Y. Yapp, Y.Y. Ou, H.Y. Yeh, iMotor-CNN: Identifying molecular functions of cytoskeleton motor proteins using 2D convolutional neural network via Chou’s 5-step rule. *Anal Biochem*. **575** (2019): 17–26.

[26] R. Liang, J. Xie, C. Zhang, M. Zhang, H. Huang, H. Huo, X. Cao, B. Niu, Identifying cancer targets based on machine learning methods via Chou’s 5-steps rule and general pseudo components. *Curr Top Med Chem*. htpps://doi.org/10.2174/1568026619666191016155543 (2019).

[27] Y. Liang, S. Zhang, Identifying DNase I hypersensitive sites using multi-features fusion and F-score features selection via Chou’s 5-steps rule. *Biophys Chem*. 253 (2019) 106227.

[28] S.J. Malebary, M.S.U. Rehman, Y.D. Khan, iCrotoK-PseAAC: Identify lysine crotonylation sites by blending position relative statistical features according to the Chou’s 5-step rule. *PLoS One* **14** (2019): e0223993.

[29] I. Nazari, M. Tahir, H. Tayari, K.T. Chong, iN6-Methyl (5-step): Identifying RNA N6-methyladenosine sites using deep learning mode via Chou’s 5-step rules and Chou’s general PseKNC. *Chemometrics and Intelligent Laboratory Systems*, htpps://doi.org/10.1016/j.chemolab.2019.103811 (2019).

[30] Q. Ning, Z. Ma, X. Zhao, dForml(KNN)-PseAAC: Detecting formylation sites from protein sequences using K-nearest neighbor algorithm via Chou’s 5-step rule and pseudo components. *J Theor Biol*. **470** (2019): 43–49.

[31] Salman, M. Khan, N. Iqbal, T. Hussain, S. Afzal, K.C. Chou, A two-level computation model based on deep learning algorithm for identification of piRNA and their functions via Chou’s 5-steps rule. *Int J Pept Res Ther*. htpps:/doi.org/10.1007/s10989-019-09887-3 (2019).

[32] M. Tahir, H. Tayara, K.T. Chong, iDNA6mA (5-step rule): Identification of DNA N6-methyladenine sites in the rice genome by intelligent computational model via Chou's 5-step rule. *Chemometrics and Intelligent Laboratory Systems* **189** (2019): 96–101.

[33] S. Vishnoi, P. Garg, P. Arora, Physicochemical n-Grams Tool: A tool for protein physicochemical descriptor generation via Chou's 5-steps rule. *Chem Biol Drug Des*. htpps://doi.org/10.1111/cbdd.13617 or https://www.ncbi.nlm.nih.gov/pubmed/31483930 (2019).

[34] A. Wiktorowicz, A. Wit, A. Dziewierz, L. Rzeszutko, D. Dudek, P. Kleczynski, Calcium Pattern Assessment in Patients with Severe Aortic Stenosis Via the Chou's 5-Steps Rule. *Curr Pharm Des*. https://doi.org/10.2174/1381612825666190930101258 (2019).

[35] L. Yang, Y. Lv, S. Wang, Q. Zhang, Y. Pan, D. Su, Q. Lu, Y. Zuo, Identifying FL11 subtype by characterizing tumor immune microenvironment in prostate adenocarcinoma via Chou's 5-steps rule. *Genomics*, https://doi.org/10.1016/j.ygeno.2019.08.021 (2019).

[36] L. Yang, Y. Lv, S. Wang, Q. Zhang, Y. Pan, D. Su, Q. Lu, Y. Zuo, Identifying FL11 subtype by characterizing tumor immune microenvironment in prostate adenocarcinoma via Chou's 5-steps rule. *Genomics* (2019).

[37] Y.D. Khan, N. Amin, W. Hussain, N. Rasool, S.A. Khan, K.C. Chou, iProtease-PseAAC(2L): A two-layer predictor for identifying proteases and their types using Chou's 5-step-rule and general PseAAC. *Anal Biochem*. **588** (2020): 113477.

[38] Y. Xu, J. Ding, L.Y. Wu, K.C. Chou, iSNO-PseAAC: Predict cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition. *PLoS ONE* **8** (2013): e55844.

[39] Y. Xu, X.J. Shao, L.Y. Wu, N.Y. Deng, K.C. Chou, iSNO-AAPair: incorporating amino acid pairwise coupling into PseAAC for predicting cysteine S-nitrosylation sites in proteins. *PeerJ*. **1** (2013): e171.

[40] Y. Xu, X. Wen, X.J. Shao, N.Y. Deng, K.C. Chou, iHyd-PseAAC: Predicting hydroxyproline and hydroxylysine in proteins by incorporating dipeptide position-specific propensity into pseudo amino acid composition. *Int J Mol Sci*. **15** (2014): 7594–7610.

[41] Y. Xu, X. Wen, L.S. Wen, L.Y. Wu, N.Y. Deng, K.C. Chou, iNitro-Tyr: Prediction of nitrotyrosine sites in proteins with general pseudo amino acid composition. *PLoS ONE* **9** (2014): e105018.

[42] Y. Xu, K.C. Chou, Recent progress in predicting posttranslational modification sites in proteins. *Curr Top Med Chem*. **16** (2016): 591–603.

[43] L.M. Liu, Y. Xu, K.C. Chou, iPGK-PseAAC: identify lysine phosphoglycerylation sites in proteins by incorporating four different tiers of amino acid pairwise coupling information into the general PseAAC. *Med Chem*. **13** (2017): 552–559.

[44] Y. Xu, C. Li, K.C. Chou, iPreny-PseAAC: identify C-terminal cysteine prenylation sites in proteins by incorporating two tiers of sequence couplings into PseAAC. *Med Chem*. **13** (2017): 544–551.

[45] L. Cai, C.L. Wan, L. He, S. Jong, K.C. Chou, Gestational influenza increases the risk of psychosis in adults. *Med Chem*. **11** (2015): 676–682.

[46] J. Liu, J. Song, M.Y. Wang, L. He, L. Cai, K.C. Chou, Association of EGF rs4444903 and XPD rs13181 polymorphisms with cutaneous melanoma in Caucasians. *Med Chem*. **11** (2015): 551–559.

[47] L. Cai, Y.H. Yang, L. He, K.C. Chou, Modulation of cytokine network in the comorbidity of schizophrenia and tuberculosis. *Curr Top Med Chem*. **16** (2016): 655–665.

[48] L. Cai, W. Yuan, Z. Zhang, L. He, K.C. Chou, In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Scientific Reports* **6** (2016): 36540.

[49] Y. Zhu, Q.W. Cong, Y. Liu, C.L. Wan, T. Yu, G. He, L. He, L. Cai, K.C. Chou, Antithrombin is an importantly inhibitory role against blood clots. *Curr Top Med Chem*. **16** (2016): 666–674.

[50] Z.D. Zhang, K. Liang, K. Li, G.Q. Wang, K.W. Zhang, L. Cai, S.T. Zha, K.C. Chou, Chlorella vulgaris induces apoptosis of human non-small cell lung carcinoma (NSCLC) cells. *Med Chem*. **13** (2017): 560–568.

[51] L. Cai, T. Huang, J. Su, X. Zhang, W. Chen, F. Zhang, L. He, K.C. Chou, Implications of newly identified brain eQTL genes and their interactors in Schizophrenia. *Mol Ther-Nucl Acids*. **12** (2018): 433–442.

[52] B. Niu, M. Zhang, P. Du, L. Jiang, R. Qin, Q. Su, F. Chen, D. Du, Y. Shu, K.C. Chou, Small molecular floribundiquinone B derived from medicinal plants inhibits acetylcholinesterase activity. *Oncotarget*. **8** (2017): 57149–57162.

[53] Q. Su, W. Lu, D. Du, F. Chen, B. Niu, K.C. Chou, Prediction of the aquatic toxicity of aromatic compounds to tetrahymena pyriformis through support vector regression. *Oncotarget*. **8** (2017): 49359–49369.

[54] Y. Lu, S. Wang, J. Wang, G. Zhou, Q. Zhang, X. Zhou, B. Niu, Q. Chen, K.C. Chou, An epidemic avian influenza prediction model based on Google trends. *Lett Org Chem*. **16** (2019): 303–310.

[55] B. Niu, C. Liang, Y. Lu, M. Zhao, Q. Chen, Y. Zhang, L. Zheng, K.C. Chou, Glioma stages prediction based on machine learning algorithm combined with protein-protein interaction networks. *Genomics*, doi: 10.1016/j.ygeno.2019.05.024Get (2019).

[56] J. Jia, Z. Liu, X. Xiao, B. Liu, K.C. Chou, Identification of protein-protein binding sites by incorporating the physicochemical properties and stationary wavelet transforms into pseudo amino acid composition (iPPBS-PseAAC). *J Biomol Struct Dyn*. **34** (2016): 1946–1961.

[57] J. Jia, Z. Liu, X. Xiao, B. Liu, K.C. Chou, iSuc-PseOpt: Identifying lysine succinylation sites in proteins by incorporating sequence-coupling effects into pseudo components and optimizing imbalanced training dataset. *Anal Biochem*. **497** (2016): 48–56.

[58] J. Jia, Z. Liu, X. Xiao, B. Liu, K.C. Chou, pSuc-Lys: Predict lysine succinylation sites in proteins with PseAAC and ensemble random forest approach. *J Theor Bio*. **394** (2016): 223–230.

[59] J. Jia, Z. Liu, X. Xiao, B. Liu, K.C. Chou, iCar-PseCp: identify carbonylation sites in proteins by Monto Carlo sampling and incorporating sequence coupled effects into general PseAAC. *Oncotarget* **7** (2016): 34558–34570.

[60] J. Jia, Z. Liu, X. Xiao, B. Liu, K.C. Chou, iPPBS-Opt: A sequence-based ensemble classifier for identifying protein-protein binding sites by optimizing imbalanced training datasets. *Molecules* **21** (2016): E95.

[61] J. Jia, L. Zhang, Z. Liu, X. Xiao, K.C. Chou, pSumo-CD: Predicting sumoylation sites in proteins with covariance discriminant algorithm by incorporating sequence-coupled effects into general PseAAC. *Bioinformatics*. **32** (2016): 3133–3141.

[62] Z. Liu, X. Xiao, D.J. Yu, J. Jia, W.R. Qiu, K.C. Chou, pRNAm-PC: Predicting N-methyladenosine sites in RNA sequences via physical-chemical properties. *Anal Biochem*. **497** (2016): 60–67.

[63] X. Xiao, H.X. Ye, Z. Liu, J.H. Jia, K.C. Chou, iROS-gPseKNC: predicting replication origin sites in DNA by incorporating dinucleotide position-specific propensity into general pseudo nucleotide composition. *Oncotarget* **7** (2016): 34180–34189.

[64] W.R. Qiu, B.Q. Sun, X. Xiao, Z.C. Xu, J.H. Jia, K.C. Chou, iKcr-PseEns: Identify lysine crotonylation sites in histone proteins with pseudo components and ensemble classifier. *Genomics* **110** (2018): 239–246.

[65] J. Jia, X. Li, W. Qiu, X. Xiao, K.C. Chou, iPPI-PseAAC(CGR): Identify protein-protein interactions by incorporating chaos game representation into PseAAC. *J Theor Bio*. **460** (2019): 195–203.

[66] W. Chen, H. Ding, P. Feng, H. Lin, K.C. Chou, iACP: a sequence-based tool for identifying anticancer peptides. *Oncotarget* **7** (2016): 16895–16909.

[67] W. Chen, P. Feng, H. Ding, H. Lin, K.C. Chou, Using deformation energy to analyze nucleosome positioning in genomes. *Genomics* **107** (2016): 69–75.

[68] W. Chen, H. Tang, J. Ye, H. Lin, K.C. Chou, iRNA-PseU: Identifying RNA pseudouridine sites. *Mol Ther-Nucl Acids*. **5** (2016): e332.

[69] C.J. Zhang, H. Tang, W.C. Li, H. Lin, W. Chen, K.C. Chou, iOri-Human: identify human origin of replication by incorporating dinucleotide physicochemical properties into pseudo nucleotide composition. *Oncotarget* **7** (2016): 69783–69793.

[70] W. Chen, P. Feng, H. Yang, H. Ding, H. Lin, K.C. Chou, iRNA-AI: identifying the adenosine to inosine editing sites in RNA sequences. *Oncotarget* **8** (2017): 4208–4217.

[71] P. Feng, H. Ding, H. Yang, W. Chen, H. Lin, K.C. Chou, iRNA-PseColl: Identifying the occurrence sites of different RNA modifications by incorporating collective effects of nucleotides into PseKNC. *Mol Ther-Nucl Acids*. **7** (2017): 155–163.

[72] W. Chen, H. Ding, X. Zhou, H. Lin, K.C. Chou, iRNA(m6A)-PseDNC: Identifying N6-methyladenosine sites using pseudo dinucleotide composition. *Anal Biochem*. **561–562** (2018): 59–65.

[73] W. Chen, P. Feng, H. Yang, H. Ding, H. Lin, K.C. Chou, iRNA-3typeA: identifying 3-types of modification at RNA's adenosine sites. *Mol Ther-Nucl Acid*. **11** (2018): 468–474.

[74] Z.D. Su, Y. Huang, Z.Y. Zhang, Y.W. Zhao, D. Wang, W. Chen, K.C. Chou, H. Lin, iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*. **34** (2018): 4196–4204.

[75] H. Yang, W.R. Qiu, G. Liu, F.B. Guo, W. Chen, K.C. Chou, H. Lin, iRSpot-Pse6NC: Identifying recombination spots in Saccharomyces cerevisiae by incorporating hexamer composition into general PseKNC. *Int J Bio Sci*. **14** (2018): 883–891.

[76] P. Feng, H. Yang, H. Ding, H. Lin, W. Chen, K.C. Chou, iDNA6mA-PseKNC: Identifying DNA N(6)-methyladenosine sites by incorporating nucleotide physicochemical properties into PseKNC. *Genomics* **111** (2019): 96–102.

[77] Q.S. Du, S.Q. Wang, N.Z. Xie, Q.Y. Wang, R.B. Huang, K.C. Chou, 2L-PCA: A two-level principal component analyzer for quantitative drug design and its applications. *Oncotarget* **8** (2017): 70564–70578.

[78] B. Liu, L. Fang, R. Long, X. Lan, K.C. Chou, iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition. *Bioinformatics* **32** (2016): 362–369.

[79] B. Liu, R. Long, K.C. Chou, iDHS-EL: Identifying DNase I hypersensi-tivesites by fusing three different modes of pseudo nucleotide composition into an ensemble learning framework. *Bioinformatics* **32** (2016): 2411–2418.

[80] B. Liu, S. Wang, R. Long, K.C. Chou, iRSpot-EL: identify recombination spots with an ensemble learning approach. *Bioinformatics* **33** (2017): 35–41.

[81] B. Liu, F. Yang, K.C. Chou, 2L-piRNA: A two-layer ensemble classifier for identifying piwi-interacting RNAs and their function. *Mol Ther-Nucl Acids*. **7** (2017): 267–277.

[82] B. Liu, K. Li, D.S. Huang, K.C. Chou, iEnhancer-EL: Identifying enhancers and their strength with ensemble learning approach. *Bioinformatics* **34** (2018): 3835–3842.

[83] B. Liu, F. Weng, D.S. Huang, K.C. Chou, iRO-3wPseKNC: Identify DNA replication origins by three-window-based PseKNC. *Bioinformatics* **34** (2018): 3086–3093.

[84] B. Liu, F. Yang, D.S. Huang, K.C. Chou, iPromoter-2L: a two-layer predictor for identifying promoters and their types by multi-window-based PseKNC. *Bioinformatics* **34** (2018): 33–40.

[85] W.R. Qiu, B.Q. Sun, X. Xiao, Z.C. Xu, K.C. Chou, iHyd-PseCp: Identify hydroxyproline and hydroxylysine in proteins by incorporating sequence-coupled effects into general PseAAC. *Oncotarget* **7** (2016): 44310–44321.

[86] W.R. Qiu, B.Q. Sun, X. Xiao, Z.C. Xu, K.C. Chou, iPTM-mLys: identifying multiple lysine PTM sites and their different types. *Bioinformatics* **32** (2016): 3116–3123.

[87] W.R. Qiu, X. Xiao, Z.C. Xu, K.C. Chou, iPhos-PseEn: identifying phosphorylation sites in proteins by fusing different pseudo components into an ensemble classifier. *Oncotarget* **7** (2016): 51270–51283.

[88] W.R. Qiu, S.Y. Jiang, B.Q. Sun, X. Xiao, X. Cheng, K.C. Chou, iRNA-2methyl: identify RNA 2′-O-methylation sites by incorporating sequence-coupled effects into general PseKNC and ensemble classifier. *Med Chem*. **13** (2017): 734–743.

[89] W.R. Qiu, S.Y. Jiang, Z.C. Xu, X. Xiao, K.C. Chou, iRNAm5C-PseDNC: identifying RNA 5-methylcytosine sites by incorporating physical-chemical properties into pseudo dinucleotide composition. *Oncotarget* **8** (2017): 41178–41188.

[90] W.R. Qiu, B.Q. Sun, X. Xiao, D. Xu, K.C. Chou, iPhos-PseEvo: Identifying human phosphorylated proteins by incorporating evolutionary information into general PseAAC via grey system theory. *Mol Inform*. **36** (2017) UNSP 1600010.

[91] X. Zhai, M. Chen, W. Lu, Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods. *Comp Mater Sci*. **151** (2018): 41–48.

[92] K.C. Chou, Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review, 5-steps rule). *J Theor Bio*. **273** (2011): 236–247.

[93] K.C. Chou, Impacts of pseudo amino acid components and 5-steps rule to proteomics and proteome analysis. *Curr Top Med Chem.* (Special Issue ed. G.P Zhou) https://doi.org/10.2174/1568026619666191018100141 or http://www.eurekaselect.com/175823/article (2019).

[94] K.C. Chou, Two kinds of metrics for computational biology. *Genomics*, https://www.sciencedirect.com/science/article/pii/S0888754319304604?via%3Dihub (2019).

[95] K.C. Chou, Proposing pseudo amino acid components is an important milestone for proteome and genome analyses. *Int J Pept Res Ther*. https://doi.org/10.1007/s10989-019-09910-7 or https://link.springer.com/article/10.1007%2Fs10989-019-09910-7 (2019).

[96] K.C. Chou, An insightful recollection for predicting protein subcellular locations in multi-label systems. *Genomics*, http://doi.org/10.1016/j.ygeno.2019.08.008 or https://www.sciencedirect.com/science/article/pii/S0888754319304604?via%3Dihub (2019).

[97] K.C. Chou, Progresses in predicting post-translational modification. *Int J Pept Res Ther*. https:/doi.org/10.1007/s10989-019-09893-5 or https://link.springer.com/article/10.1007%2Fs10989-019-09893-5 (2019).

[98] K.C. Chou, Recent progresses in predicting protein subcellular localization with artificial intelligence (AI) tools developed Via the 5-steps rule. *Jpn J Gastroen Hepatol*. https:/doi.org/www.jjgastrohepto.org or https://www.jjgastrohepto.org/ (2019).

[99] K.C. Chou, An insightful recollection since the distorted key theory was born about 23 years ago. *Genomics*, https://doi.org/10.1016/j.ygeno.2019.09.001 or https://www.sciencedirect.com/science/article/pii/S0888754319305543?via%3Dihub (2019).

[100] K.C. Chou, Artificial intelligence (AI) tools constructed via the 5-steps rule for predicting post-translational modifications. *Trends in Artif Intel*. **3** (2019): 60–74.

[101] K.C. Chou, Distorted key theory and its implication for drug development. *Curr Genet*. http://www.eurekaselect.com/175823/article or http://www.eurekaselect.com/175823/article (2020).

[102] Chou, K.C. An insightful 10-year recollection since the emergence of the 5-steps rule. *Curr Pharm Des*. **25** (2019): 4223–4234.

[103] K.C. Chou, An insightful recollection since the birth of Gordon Life Science Institute about 17 years ago. *Adv Sci Eng Res*. **4** (2019): 31–36.

[104] K.C. Chou, Gordon Life Science Institute: Its philosophy, achievements, and perspective. *Annals Cancer Ther Pharmacol*. **2** (2019): 001–26.