



# The Ploc Bal-Manimal is a Powerful Artificial Intelligence Tool for Predicting the Subcellular Localization of Animal Proteins Based on their Sequence Information Alone

**Chou KC\***

Gordon Life Science Institute, USA

**\*Corresponding author:** Kuo-Chen Chou, Gordon Life Science Institute, Boston, Massachusetts 02478, USA, Email: kcchou@gordonlifescience.org

**Short Communication**

**Volume 4 Issue 1**

**Received Date:** January 29, 2020

**Published Date:** March 09, 2020

**Keywords:** Animal Protein; Multi Label System; Sequence Analyzing Method

**Abbreviations:** AI: Artificial Intelligence; PseAAC: Pseudo Amino Acid Composition.

## Introduction

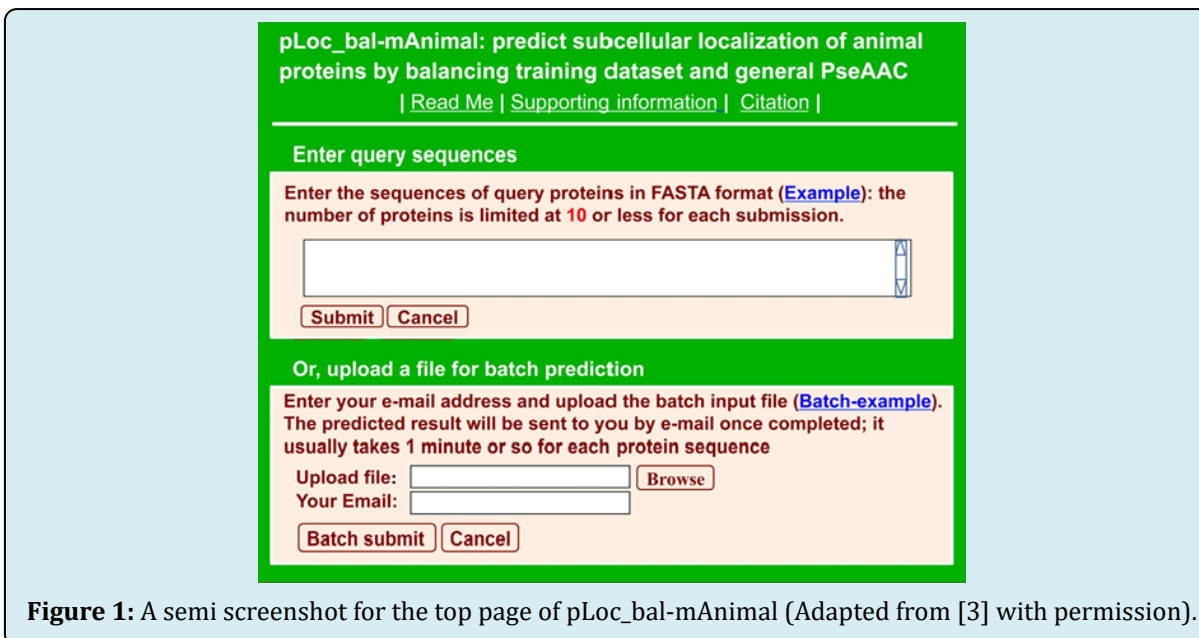
Recently a very useful web-server, or AI (Artificial Intelligence) tool, has been established for predicting the

subcellular localization of animal proteins purely according to their sequences information for the multi-label systems [1], in which a same protein may occur or travel between two or more locations and hence its ID (identification) needs two or more labels as well, namely the “multi-label mark” [2]. The AI tool is named as “pLoc\_bal-mAnimal”, where “bal” stands for that the AI tool has been treated by balancing out the training dataset [3], and “m” for that the AI tool bears the capacity to deal with the multi-label systems. Below, let us show how the AI tool is working.

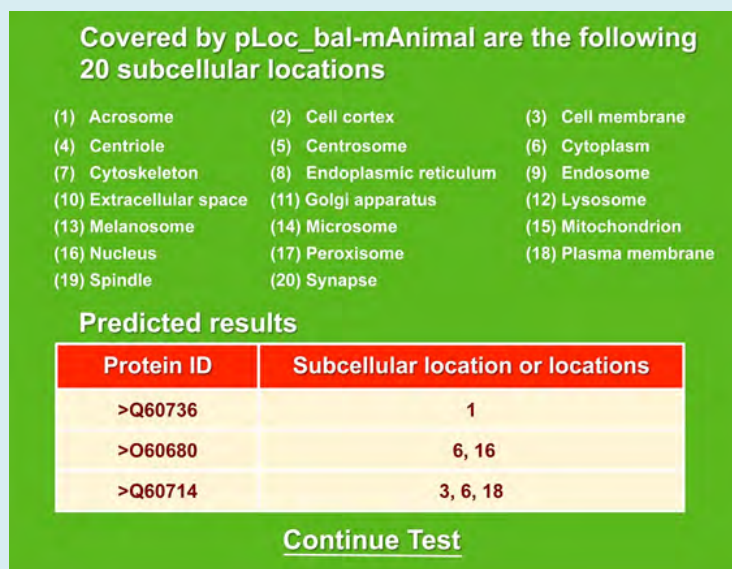
Predictor	Aiming (↑) <sup>a</sup>	Coverage(↑) <sup>a</sup>	Accuracy(↑) <sup>a</sup>	Absolute true (↑) <sup>a</sup>	Absolute false (↓) <sup>a</sup>
pLoc-mAnimal <sup>b</sup>	87.96%	85.33%	84.64%	73.11%	1.65%
pLoc_bal-Animal <sup>c</sup>	93.22%	96.54%	93.00.%	88.70%	0.57%

<sup>a</sup> Taken from Ref 7.

**Table 1:** Comparison with the state-of-the-art method in predicting animal protein subcellular localization.



**Figure 1:** A semi screenshot for the top page of pLoc\_bal-mAnimal (Adapted from [3] with permission).



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

Clicking the link at [http://www.jci-bioinfo.cn/pLoc\\_bal-mAnimal/](http://www.jci-bioinfo.cn/pLoc_bal-mAnimal/), you will see the top page of the pLoc\_bal-mAnimal web-server prompted on your computer's screen (Figure 1). Then, following the commands given in the Step 2 and Step 3 of [3], you will see (Figure 2) on the screen of your computer. The corresponding reports were detailed in Table 3 of [3]. You can see from there: nearly all the success rates achieved by the AI tool for the animal proteins in each of the 20 subcellular locations are within the range of 94-100%. Such a high prediction quality is far beyond the reach of any of its counterparts.

In addition to the advantages of high accuracy and easy to use, the AI tool has been constructed by strictly complying with the "Chou's 5-steps rule" and hence possesses the following terrific merits as concurred by many investigators (see eg. [4-22]) as well as three comprehensive review papers [2,23,24]):

- Crystal clear in logic development
- Completely transparent in operation
- Easily to repeat the reported results by other investigators,
- With high potential in stimulating other sequence-analyzing methods, and
- Very convenient to be used by the majority of experimental scientists.

Besides, the approach [25-27] of PseAAC (Pseudo Amino Acid Composition) has also been used during the development of the AI tool. It is a very powerful approach for formulating the samples of proteins by catching their special features, as done by many investigators (see, eg., [28,29,30-

45]).

For the wonderful and awesome roles of the "5-steps rule" in driving proteome, genome analyses and drug development, see a series of recent papers [46-65] where the rule and its wide applications have been very impressively presented from various aspects or at different angles.

## References

1. Chou KC, Shen HB (2007) Recent progresses in protein subcellular location prediction. *Anal Biochem* 370(1): 1-16.
2. Chou KC (2019) Advance in predicting subcellular localization of multi-label proteins and its implication for developing multi-target drugs. *Curr Med Chem*
3. Cheng X, Lin WZ, Xiao X, Chou KC (2019) pLoc\_bal-mAnimal: predict subcellular localization of animal proteins by balancing training dataset and PseAAC. *Bioinformatics* 35(3): 398-406.
4. Barukab O, Khan YD, Khan SA, Chou KC (2019) iSulfoTyr-PseAAC: Identify tyrosine sulfation sites by incorporating statistical moments via Chou's 5-steps rule and pseudo components. *Curr Genomics* 20(4): 306-320.
5. Chen Y, Fan X (2020) Use of Chou's 5-Steps Rule to Reveal Active Compound and Mechanism of Shuangsheng Pingfei San on Idiopathic Pulmonary Fibrosis. *Curr Mol Med* 20(3): 220-230.

6. Du X, Diao Y, Liu H, Li S (2019) MsDBP: Exploring DNA-binding Proteins by Integrating Multi-scale Sequence Information via Chou's 5-steps Rule. *J Proteome Res* 18(8): 3119-3132.
7. Dutta A, Dalmia A, RA, Singh KK, Anand A (2020) A Using the Chou's 5-steps rule to predict splice junctions with interpretable bidirectional long short-term memory networks. *Comput Biol Med* 116: 103558.
8. Hussain W, Khan YD, Rasool N, Khan SA, Chou KC (2019) 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: 14-23.
9. Zhe Ju, Wang SY (2019) Identify Lysine Neddylation Sites Using Bi-profile Bayes Feature Extraction via the Chou's 5-steps Rule and General Pseudo Components. *Current Genomics* 20(8): 592-601.
10. Khan S, Khan M, Iqbal N, Hussain T, Khan SA, et al. (2019) 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.*
11. Lan J, Liu Z, Liao C, Merkler DJ, Han Q, et al. (2019) A Study for Therapeutic Treatment against Parkinson's Disease via Chou's 5-steps Rule. *Curr Top Med Chem* 19(25): 2318-2333.
12. Liang R, Xie J, Zhang C, Zhang M, Huang H, et al. (2019) Identifying Cancer Targets Based on Machine Learning Methods via Chou's 5-steps Rule and General Pseudo Components. *Curr Top Med Chem* 19(25): 2301-2317.
13. Liang Y, Zhang S (2019) Identifying DNase I hypersensitive sites using multi-features fusion and F-score features selection via Chou's 5-steps rule. *Biophys Chem* 253: 106227.
14. Wiktorowicz A, Wit A, Dziewierz A, Rzeszutko L, Dudek D, et al. (2019) Calcium Pattern Assessment in Patients with Severe Aortic Stenosis Via the Chou's 5-Steps Rule. *Curr Pharm Des* 25(35): 3769-3775.
15. Yang L, Lv Y, Wang S, Zhang Q, Pan Y, et al. (2020) Identifying FL11 subtype by characterizing tumor immune microenvironment in prostate adenocarcinoma via Chou's 5-steps rule. *Genomics* 112(2): 1500-1515.
16. Akmal MA, Hussain W, Rasool N, Khan YD, Khan SA, et al. (2020) Using Chou's 5-steps rule to predict O-linked serine glycosylation sites by blending position relative features and statistical moment. *IEEE/ACM Trans Comput Biol Bioinform.*
17. Charoenkwan P, Schaduangrat N, Nantasenamat C, Piacham T, Shoombuatong W (2020) iQSP: A Sequence-Based Tool for the Prediction and Analysis of Quorum Sensing Peptides via Chou's 5-Steps Rule and Informative Physicochemical Properties. *Int J Mol Sci* 21(1): 75.
18. Ju Z, Wang SY (2020) 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* 112(1): 859-866.
19. Kabir M, Ahmad S, Iqbal M, Hayat M (2020) 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* 112(1): 276-285.
20. Vundavilli H, Datta A, Sima C, Hua J, Lopes R, et al. (2019) Using Chou's 5-steps rule to Model Feedback in Lung Cancer. *IEEE J Biomed Health Inform*
21. Chou KC (2011) Some remarks on protein attribute prediction and pseudo amino acid composition. *J Theor Biol* 273(1): 236-247.
22. Chou KC (2019) Impacts of pseudo amino acid components and 5-steps rule to proteomics and proteome Analysis. *Curr Top Med Chem* 19(25): 2283-2300.
23. Chou KC (2001) Prediction of protein cellular attributes using pseudo amino acid composition. *Proteins* 43(3): 246-255.
24. Chou KC (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics* 21(1): 10-19.
25. Chou KC (2009) Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. *Current Proteomics* 6(4): 262-274.
26. Mohabatkar H, Mohammad Beigi M, Esmaeili (2011) A Prediction of GABAA receptor proteins using the concept of Chou's pseudo amino acid composition and support vector machine. *J Theor Biol* 281(1): 18-23.
27. Mohammad BM, Behjati M, Mohabatkar H (2011) Prediction of metalloproteinase family based on the concept of Chou's pseudo amino acid composition using a machine learning approach. *J Struct Funct Genomics* 12(4): 191-197.
28. Hayat M, Khan A (2012) Discriminating Outer Membrane Proteins with Fuzzy K-Nearest Neighbor Algorithms Based on the General Form of Chou's PseAAC. *Protein Pept Lett* 19(4): 411-421.

29. Li LQ, Zhang Y, Zou LY, Zhou Y, Zheng XQ (2012) Prediction of Protein Subcellular Multi-Localization Based on the General form of Chou's Pseudo Amino Acid Composition. *Protein Pept Lett* 19(4): 375-387.
30. Liao B, Xiang Q, Li D (2012) Incorporating Secondary Features into the General form of Chou's PseAAC for Predicting Protein Structural Class. *Protein Pept Lett* 19(11): 1133-1138.
31. Liu L, Hu XZ, Liu XX, Wang Y, Li SB (2012) Predicting Protein Fold Types by the General Form of Chou's Pseudo Amino Acid Composition: Approached from Optimal Feature Extractions. *Protein Pept Lett* 19(4): 439-449.
32. Mei S (2012) Multi-kernel transfer learning based on Chou's PseAAC formulation for protein submitochondria localization. *J Theor Biol* 293: 121-130.
33. Mei S (2012) Predicting plant protein subcellular multi-localization by Chou's PseAAC formulation based multi-label homolog knowledge transfer learning. *J Theor Biol* 310: 80-87.
34. Nanni L, Brahnam S, Lumini A (2012) A Wavelet images and Chou's pseudo amino acid composition for protein classification. *Amino Acids* 43(2): 657-665.
35. Nanni L, Lumini A, Gupta D, Garg A (2012) Identifying bacterial virulent proteins by fusing a set of classifiers based on variants of Chou's pseudo amino acid composition and on evolutionary information. *IEEE-ACM Transaction on Computational Biology and Bioinformatics* 9(2): 467-475.
36. Niu XH, Hu XH, Shi F, Xia JB (2012) Predicting Protein Solubility by the General Form of Chou's Pseudo Amino Acid Composition: Approached from Chaos Game Representation and Fractal Dimension. *Protein & Peptide Letters*, 19(9): 940-948.
37. Qin YF, Wang CH, Yu XQ, Zhu J, Liu TG, et al. (2012) Predicting Protein Structural Class by Incorporating Patterns of Over- Represented k-mers into the General form of Chou's PseAAC. *Protein Pept Lett* 19(4): 388-397.
38. Ren LY, Zhang YS, Gutman I (2012) Predicting the Classification of Transcription Factors by Incorporating their Binding Site Properties into a Novel Mode of Chou's Pseudo Amino Acid Composition. *Protein Pept Lett* 19(11): 1170-1176.
39. Sun XY, Shi SP, Qiu JD, Suo SB, Huang SY, et al. (2012) Identifying protein quaternary structural attributes by incorporating physicochemical properties into the general form of Chou's PseAAC via discrete wavelet transform. *Mol BioSyst* 8(12): 3178-3184.
40. Zhao XW, Ma ZQ, Yin MH (2012) Predicting protein-protein interactions by combing various sequence-derived features into the general form of Chou's Pseudo amino acid composition. *Protein Pept Lett* 19(5): 492-500.
41. Zia-ur-Rehman, Khan A (2012) Identifying GPCRs and their Types with Chou's Pseudo Amino Acid Composition: An Approach from Multi-scale Energy Representation and Position Specific Scoring Matrix. *Protein Pept Lett* 19(8): 890-903.
42. Georgiou DN, Karakasidis TE, Megaritis AC (2013) A short survey on genetic sequences, Chou's pseudo amino acid composition and its combination with fuzzy set theory. *The Open Bioinformatics Journal* 7(S-1, M4): 41-48.
43. Gupta MK, Niyogi R, Misra M (2013) An alignment-free method to find similarity among protein sequences via the general form of Chou's pseudo amino acid composition. *SAR QSAR Environ Res* 24(7): 597-609.
44. Chou KC (2019) The cradle of Gordon Life Science Institute and its development and driving force. *Biomed J Sci & Tech Res* 23(5): 17848-17863.
45. Chou KC (2019) Showcase to illustrate how the web-server iDNA6mA-PseKNC is working. *Journal of Pathology Research Reviews & Reports* 1(1): 1-15.
46. Chou KC (2019) The pLoc\_bal-mPlant is a Powerful Artificial Intelligence Tool for Predicting the Subcellular Localization of Plant Proteins Purely based on their Sequence Information. *Int J Nutr Sci* 4(2): 1-4.
47. Chou KC, Cheng X, Xiao X (2019) pLoc\_bal-mEuk: predict subcellular localization of eukaryotic proteins by general PseAAC and quasi-balancing training dataset. *Med Chem* 15(5): 472-485.
48. Chou KC (2019) Showcase to illustrate how the web-server iNitro-Tyr is working. *Glo J of Com Sci and Infor Tec* 2: 1-16.
49. Chou KC (2019) Gordon Life Science Institute: Its philosophy, achievements, and perspective. *Annals of Cancer Therapy and Pharmacology* 2(2): 1-26.
50. Chou KC (2020) Showcase to Illustrate how the webserver pLoc\_bal-meuk Is working. *Biomed J Sci & Tech Res* 24(2): 18156-18160.
51. Chou KC (2020) The pLoc\_bal-mGneg Predictor is a

- Powerful Web-Server for Identifying the Subcellular Localization of Gram-Negative Bacterial Proteins based on their Sequences Information Alone. *Int j Sci* 9(1): 27-34.
52. Chou KC (2020) How the artificial intelligence tool iRNA-2methyl is working for RNA 2'-O-methylation sites. *Journal of Medical Care Research and Review* 3: 348-366.
53. Chou KC (2020) Showcase to illustrate how the web-server iKcr-PseEns is working. *International Journal of Sciences* 9(1): 85-95.
54. Chou KC (2020) The pLoc\_bal-mVirus is a powerful artificial intelligence tool for predicting the subcellular localization of virus proteins according to their sequence information alone. *J Gent & Genome* 4.
55. Chou KC (2019) How the artificial intelligence tool iSNO-PseAAC is working in predicting the cysteine S-nitrosylation sites in proteins. *J Stem Cell Res Med* 4: 1-9.
56. Chou KC (2020) Showcase to illustrate how the web-server iRNA-Methyl is working. *J Mol Genet* 3: 1-7.
57. Chou KC (2020) How the Artificial Intelligence Tool iRNA-PseU is Working in Predicting the RNA Pseudouridine Sites. *Biomed J Sci & Tech Res* 24(2): 18055-18064.
58. Chou KC (2020) Showcase to illustrate how the web-server iSNO-AAPair is working. *J Gent & Genome* 4.
59. Chou KC (2020) The pLoc\_bal-mHum is a Powerful Web-Serve for Predicting the Subcellular Localization of Human Proteins Purely Based on Their Sequence Information. *Adv Bioeng Biomed Sci Res* 3(1): 21-25.
60. Chou KC (2020) Showcase to Illustrate How the Web-server iPTM-mLys is working. *Infotext Journal of Infectious Diseases and Therapy* 1: 1-16.
61. Chou KC (2020) The pLoc\_bal-mGpos is a powerful artificial intelligence tool for predicting the subcellular localization of Gram-positive bacterial proteins according to their sequence information alone. *Glo J of Com Sci and Infor Tec* 2: 1-13.
62. Chou KC (2020) Showcase to illustrate how the web-server iPreny-PseAAC is working. *Glo J Com Sci Infor Tec* 2: 1-15.
63. Chou KC (2020) Some illuminating remarks on molecular genetics and genomics as well as drug development. *Mol Genet Genomics* 295(2): 261-274.
64. Chou KC (2020) The Problem of Elsevier Series Journals Online Submission by Using Artificial Intelligence. *Natural Science* 12: 37-38.
65. Chou KC (2020) The Most Important Ethical Concerns in Science. *Natural Science* 12: 35-36.

