eScientific Library

Volume 1 Issue 1
www.escientificlibrary.com

ES Journal of Microbiology

How the artificial intelligence tool iHyd-PseAAC is working in predicting the hydroxyproline and hydroxylysine in proteins

Short Communication

Kuo-Chen Chou*

Gordon Life Science Institute, Massachusetts, USA

Received: Mar 06, 2020; **Accepted:** Mar 20, 2020; **Published:** Mar 23, 2020

***Corresponding author:** Kuo-Chen Chou, Gordon Life Science Institute, Boston, Massachusetts 02478, United States of America

In 2014 a very powerful AI (artificial intelligence) tool has been established for identifying hydroxyproline and hydroxylysine sites in proteins, two of the important post modifications in proteins [1].

To see how the web-server is working, please do the following.

Step 1: Open the web server at the site at http://app.aporc.org/iHyd-PseAAC/ and you will see the top page of the predictor on your computer screen, as shown in Figure 1. Click on the Read Me button to see a brief introduction about **iHyd-PseAAC** predictor and the caveat when using it.

Step 2: Either type or copy/paste the query protein sequences into the input box at the center of Figure 1. The protein sequences should be in FASTA format. The input examples can be seen by clicking on the Example button right above the input box.

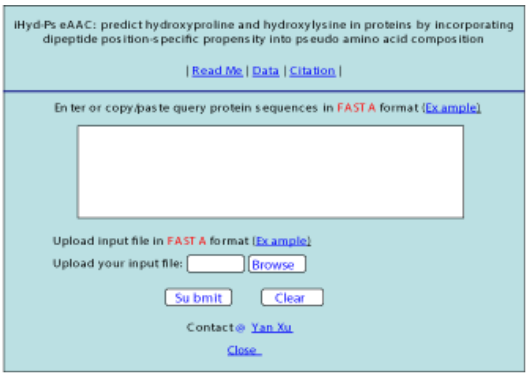

Figure 1: The top-page of the web-server iHyd-PseAAC athttp://app.aporc.org/iHyd-PseAAC/ Adapted from [1] with permission).

Step 3: Click on the Submit button to see the predicted result. For instance, if you use the protein sequences in the Example window as the input, after a few seconds, you will see the corresponding predicted results, which is fully consistent with experiment observations.

Step 4: Click on the Citation button to find the relevant paper that documents the detailed development and algorithm of **iHyd-PseAAC**.

Step 5: Click on the Data button to download the benchmark dataset used to train and test the **iHyd-PseAAC** predictor.

It is anticipated that the Web-Server will be very useful because the vast majority of biological scientists can easily get their desired results without the need to go through the complicated equations in [2] that were presented just for the integrity in developing the predictor.

Also, note that the web-server predictor has been developed by strictly observing the guidelines of "Chou's 5-steps rule" and hence have the following notable merits (see, e.g., [3,4] and three comprehensive review papers [5-7]): (1) crystal clear in logic development, (2) completely transparent in operation, (3) easily to repeat the reported results by other investigators, (4) with high potential in stimulating other sequence-analyzing methods, and (5) very convenient to be used by the majority of experimental scientists.

It has not escaped our notice that during the development of iDNA6mA-PseKNC web-server, the approach of general pseudo amino acid components [8] or PseAAC [9] had been utilized and hence its accuracy would

be much higher than its counterparts, as concurred by many investigators (see, e.g., [10-13]).

For the marvelous and awesome roles of the "5-steps rule" in driving proteome, genome analyses and drug development, see a series of recent papers [14-35] where the rule and its wide applications have been very impressively presented from various aspects or at different angles.

References

1. 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.
2. 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.
3. 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 Current Genomics.
4. 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, 112 (2020) 276-285.
5. K.C. Chou, Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review, 5-steps rule), J. Theor. Biol., 273 (2011) 236-247.
6. K.C. Chou, Advance in predicting subcellular localization of multi-label proteins and its implication for developing multi-target drugs, Current Medicinal Chemistry, 26 (2019) 4918-4943.
7. K.C. Chou, Impacts of pseudo amino acid components and 5-steps rule to proteomics and proteome analysis, Current Topics in Medicinak Chemistry (CTMC) (Special Issue ed. G.P Zhou).
8. K.C. Chou, Prediction of protein cellular attributes using pseudo amino acid composition, PROTEINS: Structure, Function, and Genetics (Erratum: ibid., 2001, Vol.44, 60), 43 (2001) 246-255.
9. K.C. Chou, Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes, Bioinformatics, 21 (2005) 10-19.
10. 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.
11. 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.
12. S. Ilyas, W. Hussain, A. Ashraf, Y.D. Khan, S.A. Khan, K.C. Chou, iMethylK-PseAAC: Improving accuracy for lysine methylation sites identification by incorporating statistical moments and position relative features into general PseAAC via Chou's 5-steps rule, Current Genomics, doi: 10.2174/1389202920666190809095206 (2019).
13. F. Javed, M. Hayat, Predicting subcellular localization of multi-label proteins by incorporating the sequence features into Chou's PseAAC, Genomics, 111 (2019) 1325-1332.
14. K.C. Chou, The cradle of Gordon Life Science Institute and its development and driving force, Int J Biol Genetics, 1 (2019) 1-28.
15. K.C. Chou, Showcase to illustrate how the web-server iDNA6mA-PseKNC is working, Journal of Pathology Research Reviews & Reports, 1 (2019) 1-15.
16. K.C. Chou, 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 (2019) 1-4.
17. 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.
18. K.C. Chou, Showcase to illustrate how the web-server iNitro-Tyr is working, Glo J of Com Sci and Infor Tec. , 2 (2019) 1-16.
19. K.C. Chou, Gordon Life Science Institute: Its philosophy, achievements, and perspective, Annals of Cancer Therapy and Pharmacology 2(2019) 001-026.
20. K.C. Chou, Showcase to Illustrate how the webserver pLoc_bal-mEuk Is working, Biomed J Sci & Tech Res.
21. K.C. Chou, 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, ijSci, 9 (2020) 27-34.
22. K.C. Chou, How the artificial intelligence tool iRNA-2methyl is working for RNA 2'-Omethylation sites, Journal of Medical Care Research and Review, 3 (2020) 348-366.
23. K.-C. Chou, Showcase to illustrate how the web-server iKcr-PseEns is working, Journal of Medical Care Research and Review, 3 (2020) 331-347.
24. K.C. Chou, 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 (2020).
25. K.C. Chou, How the artificial intelligence tool iSNO-PseAAC is working in predicting the cysteine S-nitrosylation sites in proteins, J Stem Cell Res Med, 4 (2019) 1-9.
26. K.C. Chou, Showcase to illustrate how the web-server iRNA-Methyl is working, J Mol Genet, 3 (2020) 1-7.
27. K.C. Chou, How the Artificial Intelligence Tool iRNA-PseU is Working in Predicting the RNA Pseudouridine Sites, Biomed J Sci & Tech Res.
28. K.C. Chou, Showcase to illustrate how the web-server iSNO-AAPair is working, J Gent & Genome, 4 (2020).
29. K.C. Chou, 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 (2020) 1-5.
30. K.C. Chou, Showcase to Illustrate How the Web-server iPTM-mLys is working, Infotext Journal of Infectious Diseases and Therapy [IJID], 1 (2020) 1-16.
31. K.C. Chou, 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 (2020) 01-13.

Citation: Chou K-C. How the artificial intelligence tool iHyd-PseAAC is working in predicting the hydroxyproline and hydroxylysine in proteins. ES J Microbiol. 2020; 1(1): 1004.

32. K.C. Chou, Showcase to illustrate how the web-server iPreny-PseAAC is working, Glo J of Com Sci and Infor Tec., 2 (2020) 01-15.

33. K.C. Chou, Some illuminating remarks on molecular genetics and genomics as well as drug development, Molecular Genetics and Genomics, 295 (2020) 261-274.

34. K.C. Chou, The Problem of Elsevier Series Journals Online Submission by Using Artificial Intelligence, Natural Science, 12 (2020) 37-38.

35. K.C. Chou, The Most Important Ethical Concerns in Science, Natural Science, 12 (2020) 35-36.