BOHR International Journal of Biocomputing and Nano Technology
2020, Vol. 1, No. 1, pp. 5–6

https://doi.org/10.54646/bijbnt.002
www.bohrpub.com

How the artificial intelligence tool iPGK-PseAAC is working in predicting lysine phosphoglycerylation sites in proteins

Kuo-Chen Chou

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

In 2017 a very powerful AI (artificial intelligence) tool has been established for predicting lysine phosphoglycerylation sites in proteins, one of the most important post modifications in proteins [1].

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

Step 1. Opening the web-server at http://app.aporc.org/iPGK-PseAAC/, you will see the top page of iPGK-PseAAC on your computer screen, as shown in **Figure 1**. Click on the Read Me button to see a brief introduction about this predictor.

Step 2. Either type or copy/paste your query protein sequences into the input box at the center of **Figure 1.** The input sequences should be in the FASTA format. For the examples of sequences in FASTA format, click the Example button right above the input box.

Step 3. Click on the Submit button to see the predicted result. For example, if you use the 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.

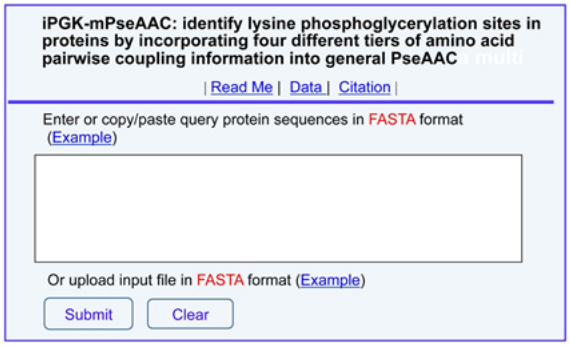

Figure 1. A semi-screenshot for the top-page of the **iPGK-PseAAC** web-server at http://app.aporc.org/iPGK-PseAAC/ (Adapted from [1] with permission).

Step 4. Click the Data button to download the benchmark dataset used in this study.

Step 5. Click the Citation button to find the relevant papers that document the detailed development and algorithm for **iPGK-PseAAC**.

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 [1] 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., [2–4] 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., [10–12]).

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 [13–34] where the rule and its wide applications have been very impressively presented from various aspects or at different angles.

References

[1] 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.

[2] 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 Genomics*, https:/doi.org/10.2174/1389202920666190819091609 or http://www.eurekaselect.com/174277/article (2019).

[3] 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).

[4] S. Vishnoi, P. Garg, P. Arora, Physicochemical n-grams tool: A tool for protein physicochemical descriptor generation via Chou's 5-step rule. *Chem Biol Drug Des*. **95** (2020): 79–86.

[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. *Curr Med Chem*. https://doi.org/10.2174/0929867326666190507082559 or http://www.eurekaselect.com/172010/article **26** (2019): 4918–4943.

[7] 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).

[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] K.K. Kandaswamy, G. Pugalenthi, S. Moller, E. Hartmann, K.U. Kalies, P.N. Suganthan, T. Martinetz, Prediction of apoptosis protein locations with genetic algorithms and support vector machines through a new mode of pseudo amino acid composition. *Protein Pept Lett*. **17** (2010): 1473–1479.

[11] H. Mohabatkar, Prediction of cyclin proteins using Chou's pseudo amino acid composition. *Protein Pept Lett*. **17** (2010): 1207–1214.

[12] L. Nanni, S. Brahnam, A. Lumini, High performance set of PseAAC and sequence based descriptors for protein classification. *J Theor Biol.* **266** (2010): 1–10.

[13] K.C. Chou, The cradle of Gordon Life Science Institute and its development and driving force. *Int J Biol Genetics* **1** (2019): 1–28.

[14] K.C. Chou, Showcase to illustrate how the web-server iDNA6mA-PseKNC is working. *J Pathol Res Rev Rep*. **1** (2019): 1–15.

[15] 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.

[16] 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.

[17] 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.

[18] K.C. Chou, Gordon Life Science Institute: Its philosophy, achievements, and perspective, *Annals Cancer Ther Pharm*. https://onomyscience.com/onomy/cancer_archive_volume2_issue2.html **2** (2019): 001–026.

[19] K.C. Chou, Showcase to Illustrate how the webserver pLoc_bal-mEuk Is working. *Biomed J Sci Tech Res*. htpps://doi.org/10.26717/BJSTR.2020.24.004033, **24** (2020).

[20] 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. *Int J Sci Comput Intel*. **9** (2020): 27–34.

[21] K.C. Chou, How the artificial intelligence tool iRNA-2methyl is working for RNA 2′-Omethylation sites. *J Med Care Res Rev.* **3** (2020): 348–366.

[22] K.-C. Chou, Showcase to illustrate how the web-server iKcr-PseEns is working. *J Med Care Res Rev*. **3** (2020): 331–347.

[23] 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).

[24] 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.

[25] K.C. Chou, Showcase to illustrate how the web-server iRNA-Methyl is working. *J Mol Genet*. **3** (2020): 1–7.

[26] K.C. Chou, How the artificial intelligence tool iRNA-PseU is working in predicting the RNA pseudouridine sites. *Biomed J Sci & Tech Res*. https://doi.org/10.26717/BJSTR.2020.24.004016, **24** (2020).

[27] K.C. Chou, Showcase to illustrate how the web-server iSNO-AAPair is working. *J Gent & Genome*. **4** (2020).

[28] 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.

[29] K.C. Chou, Showcase to illustrate how the web-server iPTM-mLys is working. *Int J Infect Dis Ther*. **1** (2020): 1–16.

[30] 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 Com Sci Infor Tec*. **2** (2020): 1–13.

[31] K.C. Chou, Showcase to illustrate how the web-server iPreny-PseAAC is working. *Glo J of Com Sci and Infor Tec.* **2** (2020): 1–15.

[32] K.C. Chou, Some illuminating remarks on molecular genetics and genomics as well as drug development. *Mol Genet Genomics*. **295** (2020): 261–274.

[33] K.C. Chou, The problem of Elsevier series journals online submission by using artificial intelligence. *Nat Sci*. **12** (2020): 37–38.

[34] K.C. Chou, The most important ethical concerns in science. *Nat Sci*. **12** (2020): 35–36.