

pLoc_Deep-mHum: Predict Subcellular Localization of Human Proteins by Deep Learning

Yu-Tao Shao¹, Xin-Xin Liu¹, Zhe Lu¹, Kuo-Chen Chou^{1,2}

¹Computer Science, Jingdezhen Ceramic Institute, Jingdezhen, China; ²Gordon Life Science Institute, Boston, MA 02478, USA

Correspondence to: Yu-Tao Shao, 532606318@qq.com; Xin-Xin Liu, 1332250394@qq.com; Zhe Lu, 454170054@qq.com; Kuo-Chen Chou, kcchou@gordonlifescience.org, kcchou38@gmail.com

Keywords: Coronavirus; Multi-Label System; Human Proteins; Deep Learning; Five-Steps Rule; PseAAC

Received: July 15, 2020

Accepted: July 24, 2020

Published: July 27, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

ABSTRACT

Recently, the life of human beings around the entire world has been endangering by the spreading of pneumonia-causing virus, such as Coronavirus, COVID-19, and H1N1. To develop effective drugs against Coronavirus, knowledge of protein subcellular localization is indispensable. In 2019, a predictor called “pLoc_bal-mHum” was developed for identifying the subcellular localization of human proteins. Its predicted results are significantly better than its counterparts, particularly for those proteins that may simultaneously occur or move between two or more subcellular location sites. However, more efforts are definitely needed to further improve its power since pLoc_bal-mHum was still not trained by a “deep learning”, a very powerful technique developed recently. The present study was devoted to incorporate the “deep-learning” technique and develop a new predictor called “pLoc_Deep-mHum”. The global absolute true rate achieved by the new predictor is over 81% and its local accuracy is over 90%. Both are overwhelmingly superior to its counterparts. Moreover, a user-friendly web-server for the new predictor has been well established at http://www.jci-bioinfo.cn/pLoc_Deep-mHum/, which will become a very useful tool for fighting pandemic coronavirus and save the mankind of this planet.

1. INTRODUCTION

Knowledge of the subcellular localization of proteins is crucially important for fulfilling the following two important goals: 1) revealing the intricate pathways that regulate biological processes at the cellular level [1, 2]. 2) selecting the right targets [3] for developing new drugs.

With the avalanche of protein sequences in the post-genomic age, we are challenged to develop computational tools for effectively identifying their subcellular localization purely based on the sequence in-

formation.

The development of predicting human subcellular localization can be briefly summarized as follows. The early human protein predictor HSLPred [4] can only predict four subcellular localizations. In 2006, Chou *et al.* developed Hum-PLoc [5], which can cover up to 12 subcellular localizations, but Hum-PLoc can only predict a single position. In 2011, by extracting the GO (Gene Ontology) information of the proteins [6], the same predictor can be used to deal with multiple locations proteins, achieving 76% accuracy. It is through these kinds of procedures and follow-up procedures, that the capacity in dealing with multi-site systems and raising the accuracy is further improved. Particularly for the predictor called pLoc_bal-mHum [6].

Although the pLoc_bal-mHum predictor [7] has the aforementioned merits, it has not been trained at a deeper level yet [8-11].

The present study was initiated in an attempt to address this problem. As done in pLoc_bal-mHum [12] as well as many other recent publications in developing new prediction methods (see, e.g., [9-11, 13-50]), the guidelines of the 5-step rule [51] are followed. They are about the detailed procedures for 1) benchmark dataset, 2) sample formulation, 3) operation engine or algorithm, 4) cross-validation, and 5) web-server. But here our attentions are focused on the procedures that significantly differ from those in developing the predictor pLoc_bal-mHum [12].

2. MATERIALS AND METHODS

2.1. Benchmark Dataset

The benchmark dataset used in this study is exactly the same as that in pLoc_bal-mHum [12]. It can be formulated as

$$\mathbb{S} = \mathbb{S}_1 \cup \mathbb{S}_2 \cup \dots \cup \mathbb{S}_u \cup \dots \cup \mathbb{S}_{13} \cup \mathbb{S}_{14} \quad (1)$$

where \mathbb{S}_1 only contains the protein samples from the “Centrosome” location (cf. Table 1), \mathbb{S}_2 only contains those from the “Cytoplasm” location, and so forth; \cup denotes the symbol for “union” in the set theory [52]. The detailed sequences of these proteins and their accession numbers (or ID codes) are given in Supporting Information S1 that is also available at http://www.jci-bioinfo.cn/pLoc_bal-mHum/Suppl.pdf where none of proteins included has $\geq 25\%$ sequence identity to any other in the same subset.

2.2. Proteins Sample Formulation

Now let us consider the 2nd step of the 5-step rule [51]; *i.e.*, how to formulate the biological sequence samples with an effective mathematical expression that can truly reflect their essential correlation with the target concerned. Given a protein sequence \mathbf{P} , its most straightforward expression is

Table 1. Comparison with the state-of-the-art method in predicting human protein subcellular localization^a.

Predictor	Aiming (\uparrow) ^a	Coverage (\uparrow) ^a	Accuracy (\uparrow) ^a	Absolute true (\uparrow) ^a	Absolute false (\downarrow) ^a
pLoc_bal-mHum ^b	92.06%	94.54%	92.29%	89.28%	0.48%
pLoc_Deep-Hum ^c	98.00%	97.66%	97.36%	96.47%	0.00%

^aSee Eq. 4 for the definition of the metrics. ^bSee [12], where the reported metrics rates were obtained by the jackknife test on the benchmark dataset of Supporting Information S1 that contains experiment-confirmed proteins only. ^cThe proposed predictor; to assure that the test was performed on exactly the same experimental data as reported in [12] for pLoc_bal-mHum.

$$\mathbf{P} = R_1 R_2 R_3 R_4 R_5 R_6 R_7 \cdots R_L \quad (2)$$

where L denotes the protein's length or the number of its constituent amino acid residues, R_1 is the 1st residue, R_2 the 2nd residue, R_3 the 3rd residue, and so forth. Since all the existing machine-learning algorithms can only handle vectors as elaborated in [3], one has to convert a protein sample from its sequential expression (Equation (2)) to a vector. But a vector defined in a discrete model might completely miss all the sequence-order or pattern information. To deal with this problem, the Pseudo Amino Acid Composition [53] or PseAAC [54]. Ever since then, the concept of "Pseudo Amino Acid Composition" or "PseAAC" has been widely used in nearly all the areas of computational proteomics with the aim to grasp various different sequence patterns that are essential to the targets investigated (see, e.g., [11, 55-207]).

Because it has been widely and increasingly used, recently three powerful open access soft-wares, called "PseAAC-Builder" [87], "propy" [144], and "PseAAC-General" [114], were established: the former two are for generating various modes of special PseAAC [208]; while the 3rd one for those of general PseAAC [51], including not only all the special modes of feature vectors for proteins but also the higher level feature vectors such as "Functional Domain" mode, "Gene Ontology" mode, and "Sequential Evolution" or "PSSM" mode. Encouraged by the successes of using PseAAC to deal with protein/peptide sequences, its idea and approach were extended to PseKNC (Pseudo K-tuple Nucleotide Composition) to generate various feature vectors for DNA/RNA sequences [209] that have proved very successful as well [210-221]. According to the concept of general PseAAC [51], any protein sequence can be formulated as a PseAAC vector given by

$$\mathbf{P} = [\Psi_1 \ \Psi_2 \ \cdots \ \Psi_u \ \cdots \ \Psi_\Omega]^T$$

$$\mathbf{P} = [\Psi_1 \ \Psi_2 \ \cdots \ \Psi_u \ \cdots \ \Psi_\Omega]^T \quad (3)$$

where \mathbf{T} is a transpose operator, while the integer Ω is a parameter and its value as well as the components Ψ_u ($u=1,2,\dots,\Omega$) will depend on how to extract the desired information from the amino acid sequence of \mathbf{P} , as elaborated in [12]. Thus, by following exactly the same procedures as described in the Section 2.2 of [12], each of the protein samples in the benchmark dataset can be uniquely defined as a 14-D numerical vector as given in columns 16 - 29 of [Supporting Information S2](#), which can also be directly downloaded at http://www.jci-bioinfo.cn/pLoc_bal-mHum/Supp2.pdf.

2.3. Architecture for the Novel CNN-BiLSTM Network

In this study, a novel deep CNN-BiLSTM neural network ([Figure 1](#)) model included one convolution layer and one BiLSTM block is proposed to predict subcellular localization of multi-label human proteins. The strong point of this model is that it allows extracting the maximum amount of information from human protein features using CNN convolution layers. This output becomes the BiLSTM input, which allows keeping the dependency relationship between the features in both directions.

CNN is known for its ability to extract as many features as possible from the data. In our model, 14 human protein features are used as convolutional neural network input. Then, fourteen filters of sizes 2 with a stride of 1 are applied. This gives a 14 * 14 feature map. After each filter, a layer of MaxPooling1D is applied to update and reduce the size of the data and this gives a 7 * 14 feature map.

Then, the 7 * 14 feature map is concatenated to build the BiLSTM input, which applies a BiLSTM layer to filter the information, using its three gates. The output of this step is the input of the fully connected layer which consists of 3 hidden layer. It links each piece of input information with a piece of output information. Finally, we apply the sigmoid function as an activation function in order to transform the vector into probability to define the class of each output. The label of human protein was decided by the Probability value θ . If the output is greater than 0.5, the outcome was true; otherwise, false.

The new predictor developed via the above procedures is called "pLoc_Deep-mHum", where "pLoc_Deep" stands for "predict subcellular localization by deep learning", and "mHum" for "multi-label human proteins".

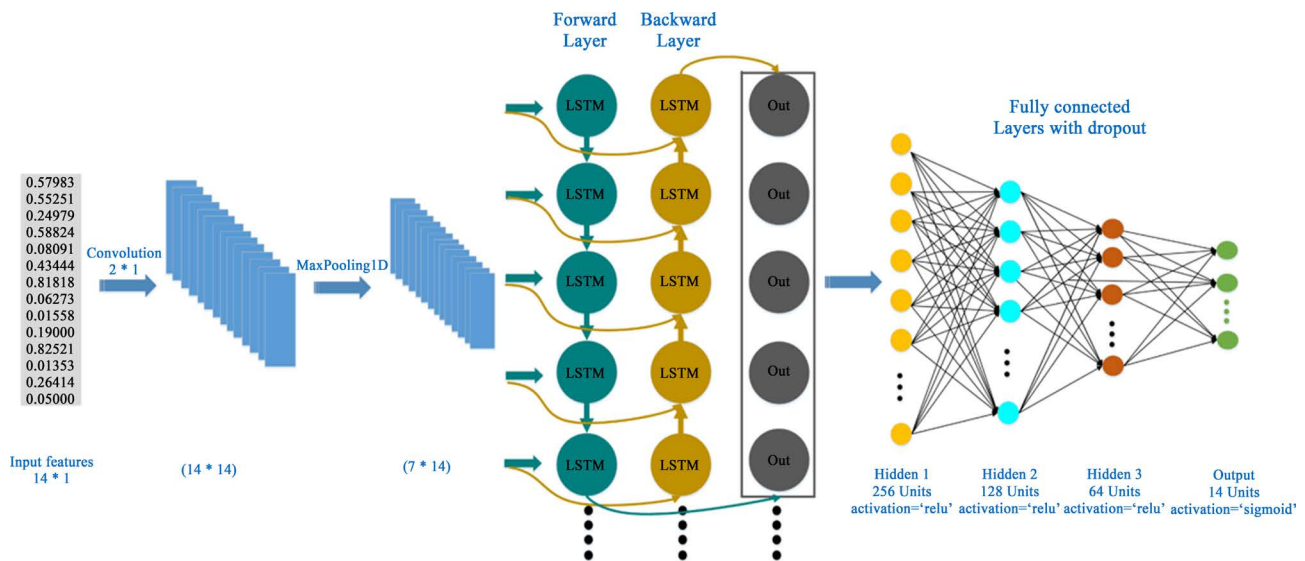


Figure 1. An illustration to show a CNN and LSTM neural network models to predict the subcellular localization of human proteins.

The new predictor developed via the above procedures is called “pLoc_Deep-mHum”, where “pLoc_Deep” stands for “predict subcellular localization by deep learning”, and “mHum” for “multi-label human proteins”.

3. RESULTS AND DISCUSSION

According to the 5-step rules [51], one of the important procedures in developing a new predictor is how to properly evaluate its anticipated accuracy. To deal with that, two issues need to be considered. 1) What metrics should be used to quantitatively reflect the predictor’s quality? 2) What test method should be applied to score the metrics?

3.1. A Set of Five Metrics for Multi-Label Systems

Different from the metrics used to measure the prediction quality of single-label systems, the metrics for the multi-label systems are much more complicated [222]. For the latter, two kinds of metrics are needed: one is the “global metrics” and the other is the “local metrics”.

3.2. Comparison with the State-of-the-Art Predictor

Listed in **Table 1** are the rates achieved by the current pLoc_Deep-mHum predictor via the cross validations on the same experiment-confirmed dataset as used in [22] according to the scores of “global metrics”. For facilitating comparison, listed there are also the corresponding results obtained by the pLoc_bal-mHum [12], the existing most powerful predictor for identifying the subcellular localization of human proteins with both single and multiple location sites. As shown in **Table 1**, the newly proposed predictor pLoc_Deep-mHum is remarkably superior to the existing state-of-the-art predictor pLoc_bal-mHum in all the five metrics. Particularly, it can be seen from the table that the absolute true rate achieved by the new predictor is over 81%, which is far beyond the reach of any other existing methods. This is because it is extremely difficult to enhance the absolute true rate of a prediction method for a multi-label system as clearly elucidated in [12]. Actually, to avoid embarrassment, many investigators even chose not to mention the metrics of absolute true rate in dealing with multi-label systems (see, e.g., [85, 141, 147]).

Moreover, to in-depth examine the prediction quality of the new predictor for the proteins in each of the subcellular locations concerned (cf. **Table 2**), we used a set of four “local metrics” to score its accuracy,

Table 2. Performance of pLoc_Deep-mHum for each of the 14 subcellular locations.

<i>i</i>	Location ^a	Sn(<i>i</i>) ^b	Sp(<i>i</i>) ^b	Acc(<i>i</i>) ^b	MCC(<i>i</i>) ^b
1	Centrosome	0.9836	0.9988	0.9976	0.9831
2	Cytoplasm	0.9338	0.9890	0.9851	0.8898
3	Cytoskeleton	0.9949	0.9992	0.9989	0.9920
4	Endoplasmic reticulum	0.9801	0.9986	0.9973	0.9805
5	Endosome	0.9960	0.9994	0.9991	0.9935
6	Extracellular	0.9850	0.9978	0.9968	0.9771
7	Golgi apparatus	0.9753	0.9982	0.9965	0.9745
8	Lysosome	0.9991	0.9994	0.9994	0.9957
9	Microsome	0.9971	0.9999	0.9997	0.9980
10	Mitochondrion	0.9845	0.9986	0.9976	0.9827
11	Nucleus	0.9409	0.9917	0.9880	0.9129
12	Peroxisome	0.9981	1.0000	0.9999	0.9990
13	Plasma membrane	0.9792	0.9981	0.9967	0.9760
14	Synapse	1.0000	0.9999	0.9999	0.9995

^aSee [Table 1](#) and the relevant context for further explanation. ^bSee Eq.6 for the metrics definition.

which were derived in [223] based on the Chou's symbols introduced for studying protein signal peptides [224] and that have ever since been widely concurred or justified (see, e.g., [209-217, 223, 225-274]).

Listed in [Table 2](#) are the rates achieved by pLoc_Deep-mHum for the human proteins in each of 22 subcellular locations. As we can see from the table, nearly all the success rates achieved by the new predictor for the human proteins in each of the 22 subcellular locations are within the range of 90% - 100%, which is once again far beyond the reach of any of its counterparts.

Meanwhile, as a byproduct, the present paper has also stimulated some very interesting or challenging papers (see, e.g., [275-280]).

3.3. Comparison with Several Classic Machine Learning Methods

In this paper, the CNN-BiLSTM neural network model is constructed through the Python language and Keras, and the model is trained through binary_crossentropy and Adam optimizers (lr = 0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon = 10⁻⁸).

In order to improve the performance of the CNN-BiLSTM neural network model, based on the above parameter settings, the hyperparameters of the constructed CNN-BiLSTM model are adjusted. Through experimental analysis, we find that four parameters have a greater impact on the performance of the deep learning model: the height of the CNN filter, the number of filters in the convolutional layer, the output dimension of BiLSTM, and Dropout rate. Finally, the four optimal parameters are set to (2, 14, 14, 0.1).

We compared the performance of the pLoc_Deep-mHum model with seven state-of-the-art methods in [Table 3](#), including Rake1D [281], LabelPowerset [282], ML-SVM [283], MLARAM{Benites, 2015 #8224}, RandomForestClassifier [284], MLKNN [285], BRkNNaClassifier{Spyromitros, 2008 #8227}. It

Table 3. Comparison with several classic machine learning methods in predicting human protein subcellular localization^a.

Predictor	Aiming (↑) ^a	Coverage (↑) ^a	Accuracy (↑) ^a	Absolute true (↑) ^a	Absolute false (↓) ^a
Rake1D	85.47%	88.61%	84.52%	79.96%	0.00%
LabelPowerset	85.44%	88.38%	84.55%	80.40%	0.00%
ML-SVM	85.19%	85.67%	84.61%	82.95%	0.00%
MLARAM	93.42%	97.13%	92.59%	87.94%	0.00%
RandomForestClassifier	90.14%	89.63%	89.58%	88.98%	0.00%
MLKNN	93.93%	93.88%	92.91%	90.98%	0.00%
BRkNNaClassifier	95.26%	94.28%	94.08%	92.79%	0.00%
Full-Dense-connected	97.52%	97.32%	96.67%	95.14%	0.00%
pLoc_Deep-Hum ^c	98.00%	97.66%	97.36%	96.47%	0.00%

can be found that the performance of the pLoc_Deep-mHum model is significantly better than its cohorts in the multi-label metrics scores.

As pointed out in [286], user-friendly and publicly accessible web-servers represent the future direction for developing practically more useful predictors. Actually, user-friendly web-servers as given in a series of recent publications (see, e.g., Song, 2018 #3338; Wang, 2018 #3353) will significantly enhance the impacts of theoretical work because they can attract the broad experimental scientists [287]. In view of this, the web-server of the current pLoc_Deep-mHum predictor has also been established at http://www.jci-bioinfo.cn/pLoc_Deep-mHum/, by which the majority of experimental scientists can easily get the data they need.

4. CONCLUSION

It is anticipated that the pLoc_Deep-Hum predictor holds very high potential to become a useful high throughput tool in identifying the subcellular localization of human proteins, particularly for finding multi-target drugs that is currently a very hot trend in drug development. Particularly, it will become a very useful tool to find novel drugs against the current pandemic of Coronavirus endangering mankind's life.

ACKNOWLEDGMENTS

This work was supported by the grants from the National Natural Science Foundation of China (No. 31560316, 61261027, 61262038, 61202313 and 31260273), the Province National Natural Science Foundation of Jiangxi (No. 20132BAB201053), the Jiangxi Provincial Foreign Scientific and Technological Cooperation Project (No.20120BDH80023), the Department of Education of Jiangxi Province (GJJ160866).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Ehrlich, J.S., Hansen, M.D. and Nelson, W.J. (2002) Spatio-Temporal Regulation of Rac1 Localization and La-

mellipodia Dynamics during Epithelial Cell-Cell Adhesion. *Developmental Cell*, **3**, 259-270.

[https://doi.org/10.1016/S1534-5807\(02\)00216-2](https://doi.org/10.1016/S1534-5807(02)00216-2)

2. Glory, E. and Murphy, R.F. (2007) Automated Subcellular Location Determination and High-Throughput Microscopy. *Developmental Cell*, **12**, 7-16. <https://doi.org/10.1016/j.devcel.2006.12.007>
3. Chou, K.C. (2015) Impacts of Bioinformatics to Medicinal Chemistry. *Medicinal Chemistry*, **11**, 218-234. <https://doi.org/10.2174/1573406411666141229162834>
4. Garg, A., Bhasin, M. and Raghava, G.P. (2005) Support Vector Machine-Based Method for Subcellular Localization of Human Proteins Using Amino Acid Compositions, Their Order, and Similarity Search. *The Journal of Biological Chemistry*, **280**, 14427-14432. <https://doi.org/10.1074/jbc.M411789200>
5. Chou, K.C. and Shen, H.B. (2006) Addendum to "Hum-PLoc: A Novel Ensemble Classifier for Predicting Human Protein Subcellular Localization". *Biochemical and Biophysical Research Communications*, **348**, 1479. <https://doi.org/10.1016/j.bbrc.2006.08.030>
6. Chou, K.C. (2019) Showcase to Illustrate How the Web-Server iNitro-Tyr Is Working. *Global Journal of Computer Science and Information Technology*, **2**, 1-16.
7. Chou, K.C. (2019) An Insightful Recollection for Predicting Protein Subcellular Locations in Multi-Label Systems. *Genomics*. <https://www.sciencedirect.com/science/article/pii/S0888754319304604?via%3Dihub>
<https://doi.org/10.1016/j.ygeno.2019.08.008>
8. Maxwell, A., Li, R., Yang, B., Weng, H., Ou, A., Hong, H., Zhou, Z., Gong, P. and Zhang, C. (2017) Deep Learning Architectures for Multi-Label Classification of Intelligent Health Risk Prediction. *BMC Bioinformatics*, **18**, 523. <https://doi.org/10.1186/s12859-017-1898-z>
9. Barukab, O., Khan, Y.D., Khan, S.A. and Chou, K.C. (2019) iSulfoTyr-PseAAC: Identify Tyrosine Sulfation Sites by Incorporating Statistical Moments via Chou's 5-Steps Rule and Pseudo Components *Current Genomics*, **20**, 306-320. <https://doi.org/10.2174/1389202920666190819091609>
10. Khan, Y.D., Amin, N., Hussain, W., Rasool, N., Khan, S.A. and Chou, K.C. (2020) iProtease-PseAAC(2L): A Two-Layer Predictor for Identifying Proteases and Their Types Using Chou's 5-Step-Rule and General PseAAC. *Analytical Biochemistry*, **588**, Article ID: 113477. <https://doi.org/10.1016/j.ab.2019.113477>
11. Nazari, I., Tahir, M., Tayari, H. and Chong, K.T. (2019) 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 (CHEMOLAB)*. <https://doi.org/10.1016/j.chemolab.2019.103811>
12. Chou, K.C., Cheng, X. and Xiao, X. (2019) pLoc_bal-mHum: Predict Subcellular Localization of Human Proteins by PseAAC and Quasi-Balancing Training Dataset. *Genomics*, **111**, 1274-1282. <https://doi.org/10.1016/j.ygeno.2018.08.007>
13. Butt, A.H. and Khan, Y.D. (2018) Prediction of S-Sulfenylation Sites Using Statistical Moments Based Features via Chou's 5-Step Rule. *International Journal of Peptide Research and Therapeutics (IJPRT)*. <https://doi.org/10.1007/s10989-019-09931-2>
14. Awais, M., Hussain, W., Khan, Y.D., Rasool, N., Khan, S.A. and Chou, K.C. (2019) 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 Transactions on Computational Biology and Bioinformatics*. <https://doi.org/10.1109/TCBB.2019.2919025>
15. Butt, A.H. and Khan, Y.D. (2019) Prediction of S-Sulfenylation Sites Using Statistical Moments Based Features via Chou's 5-Step Rule. *International Journal of Peptide Research and Therapeutics (IJPRT)*. <https://doi.org/10.1007/s10989-019-09931-2>

16. Chen, Y. and Fan, X. (2019) Use Chou's 5-Steps Rule to Reveal Active Compound and Mechanism of Shuangsheng Pingfei San on Idiopathic Pulmonary Fibrosis. *Current Molecular Medicine*, **20**, No. 3. <https://doi.org/10.2174/1566524019666191011160543>
17. Du, X., Diao, Y., Liu, H. and Li, S. (2019) MsDBP: Exploring DNA-Binding Proteins by Integrating Multi-Scale Sequence Information via Chou's 5-Steps Rule. *Journal of Proteome Research*, **18**, 3119-3132. <https://doi.org/10.1021/acs.jproteome.9b00226>
18. Dutta, A., Dalmia, A., Singh, K.K. and Anand, A. (2019) Using the Chou's 5-Steps Rule to Predict Splice Junctions with Interpretable Bidirectional Long Short-Term Memory Networks. *Computers in Biology and Medicine*, **116**, Article ID: 103558. <https://doi.org/10.1016/j.compbiomed.2019.103558>
19. Ehsan, A., Mahmood, M.K., Khan, Y.D., Barukab, O.M., Khan, S.A. and Chou, K.C. (2019) 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. *Current Genomics*, **20**, 124-133. <https://doi.org/10.2174/1389202920666190325162307>
20. Hussain, W., Khan, S.D., Rasool, N., Khan, S.A. and Chou, K.C. (2019) SPalmitoylC-PseAAC: A Sequence-Based Model Developed via Chou's 5-Steps Rule and General PseAAC for Identifying S-Palmitoylation Sites in Proteins. *Analytical Biochemistry*, **568**, 14-23. <https://doi.org/10.1016/j.ab.2018.12.019>
21. Hussain, W., Khan, Y.D., Rasool, N., Khan, S.A. and Chou, K.C. (2019) SPrenylC-PseAAC: A Sequence-Based Model Developed via Chou's 5-Steps Rule and General PseAAC for Identifying S-Prenylation Sites in Proteins. *Journal of Theoretical Biology*, **468**, 1-11. <https://doi.org/10.1016/j.jtbi.2019.02.007>
22. Jun, Z. and Wang, S.Y. (2019) Identify Lysine Neddylolation Sites Using Bi-Profile Bayes Feature Extraction via the Chou's 5-Steps Rule and General Pseudo Components. *Current Genomics*, **20**, 592-601. <https://doi.org/10.2174/1389202921666191223154629>
23. Khan, S., Khan, M., Iqbal, N., Hussain, T., Khan, S.A. and Chou, K.C. (2019) A Two-Level Computation Model Based on Deep Learning Algorithm for Identification of piRNA and Their Functions via Chou's 5-Steps Rule. *Human Genetics*. <https://doi.org/10.1007/s10989-019-09887-3>
24. Khan, Z.U., Ali, F., Khan, I.A., Hussain, Y. and Pi, D. (2019) iRSpot-SPI: Deep Learning-Based Recombination Spots Prediction by Incorporating Secondary Sequence Information Coupled with Physio-Chemical Properties via Chou's 5-Step Rule and Pseudo Components. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)*, **189**, 169-180. <https://doi.org/10.1016/j.chemolab.2019.05.003>
25. Lan, J., Liu, J., Liao, C., Merkle, D.J., Han, Q. and Li, J. (2019) A Study for Therapeutic Treatment against Parkinson's Disease via Chou's 5-Steps Rule. *Current Topics in Medicinal Chemistry*. <http://www.eurekaselect.com/175887/article>
<https://doi.org/10.2174/1568026619666191019111528>
26. Le, N.Q.K. (2019) iN6-Methylat (5-Step): Identifying DNA N(6)-Methyladenine Sites in Rice Genome Using Continuous Bag of Nucleobases via Chou's 5-Step Rule. *Molecular Genetics and Genomics. MGG*, **294**, 1173-1182. <https://doi.org/10.1007/s00438-019-01570-y>
27. Le, N.Q.K., Yapp, E.K.Y., Ho, Q.T., Nagasundaram, N., Ou, Y.Y. and Yeh, H.Y. (2019) iEnhancer-5Step: Identifying Enhancers Using Hidden Information of DNA Sequences via Chou's 5-Step Rule and Word Embedding. *Analytical Biochemistry*, **571**, 53-61. <https://doi.org/10.1016/j.ab.2019.02.017>
28. Le, N.Q.K., Yapp, E.K.Y., Ou, Y.Y. and Yeh, H.Y. (2019) iMotor-CNN: Identifying Molecular Functions of Cytoskeleton Motor Proteins Using 2D Convolutional Neural Network via Chou's 5-Step Rule. *Analytical Biochemistry*, **575**, 17-26. <https://doi.org/10.1016/j.ab.2019.03.017>
29. Liang, R., Xie, J., Zhang, C., Zhang, M., Huang, H., Huo, H., Cao, X. and Niu, B. (2019) Identifying Cancer Targets Based on Machine Learning Methods via Chou's 5-Steps Rule and General Pseudo Components. *Current*

30. Liang, Y. and Zhang, S. (2019) Identifying DNase I Hypersensitive Sites Using Multi-Features Fusion and F-Score Features Selection via Chou's 5-Steps Rule. *Biophysical Chemistry*, **253**, Article ID: 106227. <https://doi.org/10.1016/j.bpc.2019.106227>
31. Liu, Z., Dong, W., Jiang, W. and He, Z. (2019) csDMA: An Improved Bioinformatics Tool for Identifying DNA 6 mA Modifications via Chou's 5-Step Rule. *Scientific Reports*, **9**, Article No. 13109. <https://doi.org/10.1038/s41598-019-49430-4>
32. Malebary, S.J., Rehman, M.S.U. and Khan, Y.D. (2019) iCrotoK-PseAAC: Identify Lysine Crotonylation Sites by Blending Position Relative Statistical Features According to the Chou's 5-Step Rule. *PLoS ONE*, **14**, e0223993. <https://doi.org/10.1371/journal.pone.0223993>
33. Ning, Q., Ma, Z. and Zhao, X. (2019) dForml(KNN)-PseAAC: Detecting Formylation Sites from Protein Sequences Using K-Nearest Neighbor Algorithm via Chou's 5-Step Rule and Pseudo Components. *Journal of Theoretical Biology*, **470**, 43-49. <https://doi.org/10.1016/j.jtbi.2019.03.011>
34. Tahir, M., Tayara, H. and Chong, K.T. (2019) iDNA6mA 5-Step Rule, Identification of DNA N6-Methyladenine Sites in the Rice Genome by Intelligent Computational Model via Chou's 5-Step Rule. *CHEMOLAB*, **189**, 96-101. <https://doi.org/10.1016/j.chemolab.2019.04.007>
35. Wiktorowicz, A., Wit, A., Dziewierz, A., Rzeszutko, L., Dudek, D. and Kleczynski, P. (2019) Calcium Pattern Assessment in Patients with Severe Aortic Stenosis via the Chou's 5-Steps Rule. *Current Pharmaceutical Design*. <https://doi.org/10.2174/1381612825666190930101258>
36. Yang, L., Lv, Y., Wang, S., Zhang, Q., Pan, Y., Su, D., Lu, Q. and Zuo, Y. (2019) 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>
37. Akbar, S., Rahman, A.U., Hayat, M., *et al.* (2020) cACP: Classifying Anticancer Peptides Using Discriminative Intelligent Model via Chou's 5-Step Rules and General Pseudo Components. *Chemometrics and Intelligent Laboratory (CHEMOLAB)*, **196**, Article ID: 103912. <https://doi.org/10.1016/j.chemolab.2019.103912>
38. Akmal, M.A., Hussain, W., Rasool, N., Khan, Y.D., Khan, S.A. and Chou, K.C. (2020) Using Chou's 5-Steps Rule to Predict O-Linked Serine Glycosylation Sites by Blending Position Relative Features and Statistical Moment. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. <https://doi.org/10.1109/TCBB.2020.2968441>
39. Charoenkwan, P., Schaduangrat, N., Nantasenamat, C., Piacham, T. and 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. *International Journal of Molecular Sciences*, **21**, 75. <https://doi.org/10.3390/ijms21010075>
40. Charoenkwan, P., Schaduangrat, N., Nantasenamat, C., Piacham, T. and Shoombuatong, W. (2020) Correction: Shoombuatong, W., *et al.* iQSP: A Sequence-Based Tool for the Prediction and Analysis of Quorum Sensing Peptides via Chou's 5-Steps Rule and Informative Physicochemical Properties. *International Journal of Molecular Sciences*, **21**, 75. <https://doi.org/10.3390/ijms21072629>
41. Chen, Y. and Fan, X. (2020) Use of Chou's 5-Steps Rule to Reveal Active Compound and Mechanism of Shuangshen Pingfei San on Idiopathic Pulmonary Fibrosis. *Current Molecular Medicine*, **20**, 220-230. <https://doi.org/10.2174/1566524019666191011160543>
42. Dobosz, R., Mucko, J. and Gawinecki, R. (2020) Using Chou's 5-Step Rule to Evaluate the Stability of Tautomers: Susceptibility of 2-[(Phenylimino)-methyl]-cyclohexane-1,3-diones to Tautomerization Based on the Calculated Gibbs Free Energies. *Energies*, **13**, 183. <https://doi.org/10.3390/en13010183>

43. Du, L., Meng, Q., Jiang, H. and Li, Y. (2020) Using Evolutionary Information and Multi-Label Linear Discriminant Analysis to Predict the Subcellular Location of Multi-Site Bacterial Proteins via Chou's 5-Steps Rule. *IEEE Access*, **8**, 56452-56461. <https://doi.org/10.1109/ACCESS.2020.2982160>
44. Dutta, A., Dalmia, A., R, A., Singh, K.K. and Anand, A. (2020) Using the Chou's 5-Steps Rule to Predict Splice Junctions with Interpretable Bidirectional Long Short-Term Memory Networks. *Computers in Biology and Medicine*, **116**, Article ID: 103558. <https://doi.org/10.1016/j.compbiomed.2019.103558>
45. Ju, Z. and Wang, S.Y. (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**, 859-866. <https://doi.org/10.1016/j.ygeno.2019.05.027>
46. Kabir, M., Ahmad, S., Iqbal, M. and 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**, 276-285. <https://doi.org/10.1016/j.ygeno.2019.02.006>
47. Lin, W., Xiao, X., Qiu, W. and Chou, K.C. (2020) Use Chou's 5-Steps Rule to Predict Remote Homology Proteins by Merging Grey Incidence Analysis and Domain Similarity Analysis. *Natural Science*, **12**, 181-198. <https://doi.org/10.4236/ns.2020.123016>
48. Vishnoi, S., Garg, P. and Arora, P. (2020) Physicochemical n-Grams Tool: A Tool for Protein Physicochemical Descriptor Generation via Chou's 5-Step Rule. *Chemical Biology & Drug Design*, **95**, 79-86. <https://doi.org/10.1111/cbdd.13617>
49. Vundavilli, H., Datta, A., Sima, C., Hua, J., Lopes, R. and Bittner, M. (2020) Using Chou's 5-Steps Rule to Model Feedback in Lung Cancer *IEEE Journal of Biomedical and Health Informatics*, **21**, 1-24. <https://doi.org/10.1109/JBHI.2019.2958042>
50. Yang, L., Lv, Y., Wang, S., Zhang, Q., Pan, Y., Su, D., Lu, Q. and Zuo, Y. (2020) Identifying FL11 Subtype by Characterizing Tumor Immune Microenvironment in Prostate Adenocarcinoma via Chou's 5-Steps Rule. *Genomics*, **112**, 1500-1515. <https://doi.org/10.1016/j.ygeno.2019.08.021>
51. Chou, K.C. (2011) Some Remarks on Protein Attribute Prediction and Pseudo Amino Acid Composition (50th Anniversary Year Review, 5-Steps Rule). *Journal of Theoretical Biology*, **273**, 236-247. <https://doi.org/10.1016/j.jtbi.2010.12.024>
52. Chou, K.C. and Shen, H.B. (2007) Recent Progresses in Protein Subcellular Location Prediction. *Analytical Biochemistry*, **370**, 1-16. <https://doi.org/10.1016/j.ab.2007.07.006>
53. Chou, K.C. (2001) Prediction of Protein Cellular Attributes Using Pseudo Amino Acid Composition. *PROTEINS: Structure, Function, and Genetics*, **43**, 246-255. (Erratum: *ibid*, 2001, Vol. 44, 60) <https://doi.org/10.1002/prot.1035>
54. Chou, K.C. (2005) Using Amphiphilic Pseudo Amino Acid Composition to Predict Enzyme Subfamily Classes. *Bioinformatics*, **21**, 10-19. <https://doi.org/10.1093/bioinformatics/bth466>
55. Ding, Y.S. and Zhang, T.L. (2008) Using Chou's Pseudo Amino Acid Composition to Predict Subcellular Localization of Apoptosis Proteins: An Approach with Immune Genetic Algorithm-Based Ensemble Classifier. *Pattern Recognition Letters*, **29**, 1887-1892. <https://doi.org/10.1016/j.patrec.2008.06.007>
56. Fang, Y., Guo, Y., Feng, Y. and Li, M. (2008) Predicting DNA-Binding Proteins: Approached from Chou's Pseudo Amino Acid Composition and Other Specific Sequence Features. *Amino Acids*, **34**, 103-109. <https://doi.org/10.1007/s00726-007-0568-2>
57. Jiang, X., Wei, R., Zhang, T.L. and Gu, Q. (2008) Using the Concept of Chou's Pseudo Amino Acid Composition to Predict Apoptosis Proteins Subcellular Location: An Approach by Approximate Entropy. *Protein & Peptide Letters*, **15**, 392-396. <https://doi.org/10.2174/092986608784246443>

58. Jiang, X., Wei, R., Zhao, Y. and Zhang, T. (2008) Using Chou's Pseudo Amino Acid Composition Based on Approximate Entropy and an Ensemble of AdaBoost Classifiers to Predict Protein Subnuclear Location. *Amino Acids*, **34**, 669-675. <https://doi.org/10.1007/s00726-008-0034-9>
59. Li, F.M. and Li, Q.Z. (2008) Predicting Protein Subcellular Location Using Chou's Pseudo Amino Acid Composition and Improved Hybrid Approach. *Protein & Peptide Letters*, **15**, 612-616. <https://doi.org/10.2174/092986608784966930>
60. Lin, H. (2008) The Modified Mahalanobis Discriminant for Predicting Outer Membrane Proteins by Using Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **252**, 350-356. <https://doi.org/10.1016/j.jtbi.2008.02.004>
61. Lin, H., Ding, H., Guo, F.B., Zhang, A.Y. and Huang, J. (2008) Predicting Subcellular Localization of Mycobacterial Proteins by Using Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **15**, 739-744. <https://doi.org/10.2174/092986608785133681>
62. Nanni, L. and Lumini, A. (2008) Genetic Programming for Creating Chou's Pseudo Amino Acid Based Features for Submitochondria Localization. *Amino Acids*, **34**, 653-660. <https://doi.org/10.1007/s00726-007-0018-1>
63. Zhang, G.Y., Li, H.C., Gao, J.Q. and Fang, B.S. (2008) Predicting Lipase Types by Improved Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **15**, 1132-1137. <https://doi.org/10.2174/092986608786071184>
64. Zhang, S.W., Chen, W., Yang, F. and Pan, Q. (2008) Using Chou's Pseudo Amino Acid Composition to Predict Protein Quaternary Structure: A Sequence-Segmented PseAAC Approach. *Amino Acids*, **35**, 591-598. <https://doi.org/10.1007/s00726-008-0086-x>
65. Zhang, S.W., Zhang, Y.L., Yang, H.F., Zhao, C.H. and Pan, Q. (2008) Using the Concept of Chou's Pseudo Amino Acid Composition to Predict Protein Subcellular Localization: An Approach by Incorporating Evolutionary Information and Von Neumann Entropies. *Amino Acids*, **34**, 565-572. <https://doi.org/10.1007/s00726-007-0010-9>
66. Chen, C., Chen, L., Zou, X. and Cai, P. (2009) Prediction of Protein Secondary Structure Content by Using the Concept of Chou's Pseudo Amino Acid Composition and Support Vector Machine. *Protein & Peptide Letters*, **16**, 27-31. <https://doi.org/10.2174/092986609787049420>
67. Georgiou, D.N., Karakasidis, T.E., Nieto, J.J. and Torres, A. (2009) Use of Fuzzy Clustering Technique and Matrices to Classify Amino Acids and Its Impact to Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **257**, 17-26. <https://doi.org/10.1016/j.jtbi.2008.11.003>
68. Li, Z.C., Zhou, X.B., Dai, Z. and Zou, X.Y. (2009) Prediction of Protein Structural Classes by Chou's Pseudo Amino Acid Composition: Approached Using Continuous Wavelet Transform and Principal Component Analysis. *Amino Acids*, **37**, 415-425. <https://doi.org/10.1007/s00726-008-0170-2>
69. Lin, H., Wang, H., Ding, H., Chen, Y.L. and Li, Q.Z. (2009) Prediction of Subcellular Localization of Apoptosis Protein Using Chou's Pseudo Amino Acid Composition. *Acta Biotheoretica*, **57**, 321-330. <https://doi.org/10.1007/s10441-008-9067-4>
70. Qiu, J.D., Huang, J.H., Liang, R.P. and Lu, X.Q. (2009) Prediction of G-Protein-Coupled Receptor Classes Based on the Concept of Chou's Pseudo Amino Acid Composition: An Approach from Discrete Wavelet Transform. *Analytical Biochemistry*, **390**, 68-73. <https://doi.org/10.1016/j.ab.2009.04.009>
71. Zeng, Y.H., Guo, Y.Z., Xiao, R.Q., Yang, L., Yu, L.Z. and Li, M.L. (2009) Using the Augmented Chou's Pseudo Amino Acid Composition for Predicting Protein Submitochondria Locations Based on Auto Covariance Approach. *Journal of Theoretical Biology*, **259**, 366-372. <https://doi.org/10.1016/j.jtbi.2009.03.028>
72. Esmaili, M., Mohabatkar, H. and Mohsenzadeh, S. (2010) Using the Concept of Chou's Pseudo Amino Acid

- Composition for Risk Type Prediction of Human Papillomaviruses. *Journal of Theoretical Biology*, **263**, 203-209. <https://doi.org/10.1016/j.jtbi.2009.11.016>
73. Gu, Q., Ding, Y.S. and Zhang, T.L. (2010) Prediction of G-Protein-Coupled Receptor Classes in Low Homology Using Chou's Pseudo Amino Acid Composition with Approximate Entropy and Hydrophobicity Patterns. *Protein & Peptide Letters*, **17**, 559-567. <https://doi.org/10.2174/092986610791112693>
 74. Mohabatkar, H. (2010) Prediction of Cyclin Proteins Using Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **17**, 1207-1214. <https://doi.org/10.2174/092986610792231564>
 75. Qiu, J.D., Huang, J.H., Shi, S.P. and Liang, R.P. (2010) 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 & Peptide Letters*, **17**, 715-722. <https://doi.org/10.2174/092986610791190372>
 76. Sahu, S.S. and Panda, G. (2010) A Novel Feature Representation Method Based on Chou's Pseudo Amino Acid Composition for Protein Structural Class Prediction. *Computational Biology and Chemistry*, **34**, 320-327. <https://doi.org/10.1016/j.compbiolchem.2010.09.002>
 77. Yu, L., Guo, Y., Li, Y., Li, G., Li, M., Luo, J., Xiong, W. and Qin, W. (2010) SecretP: Identifying Bacterial Secreted Proteins by Fusing New Features into Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **267**, 1-6. <https://doi.org/10.1016/j.jtbi.2010.08.001>
 78. Guo, J., Rao, N., Liu, G., Yang, Y. and Wang, G. (2011) Predicting Protein Folding Rates Using the Concept of Chou's Pseudo Amino Acid Composition. *Journal of Computational Chemistry*, **32**, 1612-1617. <https://doi.org/10.1002/jcc.21740>
 79. Lin, J. and Wang, Y. (2011) Using a Novel AdaBoost Algorithm and Chou's Pseudo Amino Acid Composition for Predicting Protein Subcellular Localization. *Protein & Peptide Letters*, **18**, 1219-1225. <https://doi.org/10.2174/092986611797642797>
 80. Lin, J., Wang, Y. and Xu, X. (2011) A Novel Ensemble and Composite Approach for Classifying Proteins Based on Chou's Pseudo Amino Acid Composition. *African Journal of Biotechnology*, **10**, 16963-16968. <https://doi.org/10.5897/AJB11.429>
 81. Mohabatkar, H., Mohammad Beigi, M. and Esmaili, A. (2011) Prediction of GABA(A) Receptor Proteins Using the Concept of Chou's Pseudo Amino Acid Composition and Support Vector Machine. *Journal of Theoretical Biology*, **281**, 18-23. <https://doi.org/10.1016/j.jtbi.2011.04.017>
 82. Mohammad, B.M., Behjati, M. and Mohabatkar, H. (2011) Prediction of Metalloproteinase Family Based on the Concept of Chou's Pseudo Amino Acid Composition Using a Machine Learning Approach. *Journal of Structural and Functional Genomics*, **12**, 191-197. <https://doi.org/10.1007/s10969-011-9120-4>
 83. Qiu, J.D., Suo, S.B., Sun, X.Y., Shi, S.P. and Liang, R.P. (2011) OligoPred: A Web-Server for Predicting Homo-Oligomeric Proteins by Incorporating Discrete Wavelet Transform into Chou's Pseudo Amino Acid Composition. *Journal of Molecular Graphics & Modelling*, **30**, 129-134. <https://doi.org/10.1016/j.jmglm.2011.06.014>
 84. Zou, D., He, Z., He, J. and Xia, Y. (2011) Supersecondary Structure Prediction Using Chou's Pseudo Amino Acid Composition. *Journal of Computational Chemistry*, **32**, 271-278. <https://doi.org/10.1002/jcc.21616>
 85. Cao, J.Z., Liu, W.Q. and Gu, H. (2012) Predicting Viral Protein Subcellular Localization with Chou's Pseudo Amino Acid Composition and Imbalance-Weighted Multi-Label K-Nearest Neighbor Algorithm. *Protein and Peptide Letters*, **19**, 1163-1169. <https://doi.org/10.2174/092986612803216999>
 86. Chen, C., Shen, Z.B. and Zou, X.Y. (2012) Dual-Layer Wavelet SVM for Predicting Protein Structural Class via the General Form of Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **19**, 422-429. <https://doi.org/10.2174/092986612799789332>
 87. Du, P., Wang, X., Xu, C. and Gao, Y. (2012) PseAAC-Builder: A Cross-Platform Stand-Alone Program for Ge-

nerating Various Special Chou's Pseudo Amino Acid Compositions. *Analytical Biochemistry*, **425**, 117-119. <https://doi.org/10.1016/j.ab.2012.03.015>

88. Fan, G.L. and Li, Q.Z. (2012) Predict Mycobacterial Proteins Subcellular Locations by Incorporating Pseudo-Average Chemical Shift into the General form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **304**, 88-95. <https://doi.org/10.1016/j.jtbi.2012.03.017>
89. Fan, G.L. and Li, Q.Z. (2012) Predicting Protein Submitochondria Locations by Combining Different Descriptors into the General Form of Chou's Pseudo Amino Acid Composition. *Amino Acids*, **43**, 545-555. <https://doi.org/10.1007/s00726-011-1143-4>
90. Li, L.Q., Zhang, Y., Zou, L.Y., Zhou, Y. and Zheng, X.Q. (2012) Prediction of Protein Subcellular Multi-Localization Based on the General form of Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **19**, 375-387. <https://doi.org/10.2174/092986612799789369>
91. Liu, L., Hu, X.Z., Liu, X.X., Wang, Y. and Li, S.B. (2012) Predicting Protein Fold Types by the General Form of Chou's Pseudo Amino Acid Composition: Approached from Optimal Feature Extractions. *Protein & Peptide Letters*, **19**, 439-449. <https://doi.org/10.2174/092986612799789378>
92. Nanni, L., Brahnam, S. and Lumini, A. (2012) Wavelet Images and Chou's Pseudo Amino Acid Composition for Protein Classification. *Amino Acids*, **43**, 657-665. <https://doi.org/10.1007/s00726-011-1114-9>
93. Nanni, L., Lumini, A., Gupta, D. and 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**, 467-475. <https://doi.org/10.1109/TCBB.2011.117>
94. Niu, X.H., Hu, X.H., Shi, F. and Xia, J.B. (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**, 940-948. <https://doi.org/10.2174/092986612802084492>
95. Ren, L.Y., Zhang, Y.S. and 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 & Peptide Letters*, **19**, 1170-1176. <https://doi.org/10.2174/092986612803217088>
96. Zhao, X.W., Ma, Z.Q. and Yin, M.H. (2012) Predicting Protein-Protein Interactions by Combining Various Sequence-Derived Features into the General Form of Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **19**, 492-500. <https://doi.org/10.2174/092986612800191080>
97. Zia-ur-Rehman, K.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 & Peptide Letters*, **19**, 890-903. <https://doi.org/10.2174/092986612801619589>
98. Chen, Y.K. and Li, K.B. (2013) Predicting Membrane Protein Types by Incorporating Protein Topology, Domains, Signal Peptides, and Physicochemical Properties into the General Form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **318**, 1-12. <https://doi.org/10.1016/j.jtbi.2012.10.033>
99. Fan, G.L. and Li, Q.Z. (2013) Discriminating Bioluminescent Proteins by Incorporating Average Chemical Shift and Evolutionary Information into the General Form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **334**, 45-51. <https://doi.org/10.1016/j.jtbi.2013.06.003>
100. Georgiou, D.N., Karakasidis, T.E. and Megaritis, A.C. (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**, 41-48. <https://doi.org/10.2174/1875036201307010041>
101. Gupta, M.K., Niyogi, R. and 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 and QSAR in Environmental*

Research, **24**, 597-609. <https://doi.org/10.1080/1062936X.2013.773378>

102. Huang, C. and Yuan, J. (2013) Using Radial Basis Function on the General Form of Chou's Pseudo Amino Acid Composition and PSSM to Predict Subcellular Locations of Proteins with Both Single and Multiple Sites. *Bio-systems*, **113**, 50-57. <https://doi.org/10.1016/j.biosystems.2013.04.005>
103. Huang, C. and Yuan, J.Q. (2013) A Multilabel Model Based on Chou's Pseudo Amino Acid Composition for Identifying Membrane Proteins with Both Single and Multiple Functional Types. *The Journal of Membrane Biology*, **246**, 327-334. <https://doi.org/10.1007/s00232-013-9536-9>
104. Huang, C. and Yuan, J.Q. (2013) Predicting Protein Subchloroplast Locations with Both Single and Multiple Sites via Three Different Modes of Chou's Pseudo Amino Acid Compositions. *Journal of Theoretical Biology*, **335**, 205-212. <https://doi.org/10.1016/j.jtbi.2013.06.034>
105. Khosravian, M., Faramarzi, F.K., Beigi, M.M., Behbahani, M. and Mohabatkar, H. (2013) Predicting Antibacterial Peptides by the Concept of Chou's Pseudo Amino Acid Composition and Machine Learning Methods. *Protein & Peptide Letters*, **20**, 180-186. <https://doi.org/10.2174/092986613804725307>
106. Lin, H., Ding, C., Yuan, L.-F., Chen, W., Ding, H., Li, Z.-Q., Guo, F.-B., Huang, J. and Rao, N.-N. (2013) Predicting Subchloroplast Locations of Proteins Based on the General Form of Chou's Pseudo Amino Acid Composition: Approached from Optimal Tripeptide Composition. *International Journal of Biomathematics*, **6**, Article ID: 1350003. <https://doi.org/10.1142/S1793524513500034>
107. Liu, B., Wang, X., Zou, Q., Dong, Q. and Chen, Q. (2013) Protein Remote Homology Detection by Combining Chou's Pseudo Amino Acid Composition and Profile-Based Protein Representation. *Molecular Informatics*, **32**, 775-782. <https://doi.org/10.1002/minf.201300084>
108. Mohabatkar, H., Beigi, M.M., Abdolahi, K. and Mohsenzadeh, S. (2013) Prediction of Allergenic Proteins by Means of the Concept of Chou's Pseudo Amino Acid Composition and a Machine Learning Approach. *Medicinal Chemistry*, **9**, 133-137. <https://doi.org/10.2174/157340613804488341>
109. Qin, Y.F., Zheng, L. and Huang, J. (2013) Locating Apoptosis Proteins by Incorporating the Signal Peptide Cleavage Sites into the General Form of Chou's Pseudo Amino Acid Composition. *International Journal of Quantum Chemistry*, **113**, 1660-1667. <https://doi.org/10.1002/qua.24383>
110. Sarangi, A.N., Lohani, M. and Aggarwal, R. (2013) Prediction of Essential Proteins in Prokaryotes by Incorporating Various Physico-Chemical Features into the General Form of Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **20**, 781-795. <https://doi.org/10.2174/0929866511320070008>
111. Wan, S., Mak, M.W. and Kung, S.Y. (2013) GOASVM: A Subcellular Location Predictor by Incorporating Term-Frequency Gene Ontology into the General Form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **323**, 40-48. <https://doi.org/10.1016/j.jtbi.2013.01.012>
112. Wang, X., Li, G.Z. and Lu, W.C. (2013) Virus-ECC-mPLoc: A Multi-Label Predictor for Predicting the Subcellular Localization of Virus Proteins with Both Single and Multiple Sites Based on a General Form of Chou's Pseudo Amino Acid Composition. *Protein & Peptide Letters*, **20**, 309-317. <https://doi.org/10.2174/092986613804910608>
113. Niu, X.H., *et al.* (2013) Using the Concept of Chou's Pseudo Amino Acid Composition to Predict Protein Solubility: An Approach with Entropies in Information Theory. *Journal of Theoretical Biology*, **332**, 211-217. <https://doi.org/10.1016/j.jtbi.2013.03.010>
114. Du, P., Gu, S. and Jiao, Y. (2014) PseAAC-General: Fast Building Various Modes of General Form of Chou's Pseudo Amino Acid Composition for Large-Scale Protein Datasets. *International Journal of Molecular Sciences*, **15**, 3495-3506. <https://doi.org/10.3390/ijms15033495>
115. Hajisharifi, Z., Piryaiee, M., Mohammad Beigi, M., Behbahani, M. and Mohabatkar, H. (2014) Predicting Anti-

- cancer Peptides with Chou's Pseudo Amino Acid Composition and Investigating Their Mutagenicity via Ames Test. *Journal of Theoretical Biology*, **341**, 34-40. <https://doi.org/10.1016/j.jtbi.2013.08.037>
116. Jia, C., Lin, X. and Wang, Z. (2014) Prediction of Protein S-Nitrosylation Sites Based on Adapted Normal Distribution Bi-Profile Bayes and Chou's Pseudo Amino Acid Composition. *International Journal of Molecular Sciences*, **15**, 10410-10423. <https://doi.org/10.3390/ijms150610410>
117. Kong, L., Zhang, L. and Lv, J. (2014) Accurate Prediction of Protein Structural Classes by Incorporating Predicted Secondary Structure Information into the General form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **344**, 12-18. <https://doi.org/10.1016/j.jtbi.2013.11.021>
118. Nanni, L., Brahnam, S. and Lumini, A. (2014) Prediction of Protein Structure Classes by Incorporating Different Protein Descriptors into General Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **360**, 109-116. <https://doi.org/10.1016/j.jtbi.2014.07.003>
119. Zhang, J., Sun, P., Zhao, X. and Ma, Z. (2014) PECM: Prediction of Extracellular Matrix Proteins Using the Concept of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **363**, 412-418. <https://doi.org/10.1016/j.jtbi.2014.08.002>
120. Zhang, L., Zhao, X. and Kong, L. (2014) Predict Protein Structural Class for Low-Similarity Sequences by Evolutionary Difference Information into the General Form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **355**, 105-110. <https://doi.org/10.1016/j.jtbi.2014.04.008>
121. Zuo, Y.C., Peng, Y., Liu, L., Chen, W., Yang, L. and Fan, G.L. (2014) Predicting Peroxidase Subcellular Location by Hybridizing Different Descriptors of Chou's Pseudo Amino Acid Patterns. *Analytical Biochemistry*, **458**, 14-19. <https://doi.org/10.1016/j.ab.2014.04.032>
122. Ali, F. and Hayat, M. (2015) Classification of Membrane Protein Types Using Voting Feature Interval in Combination with Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **384**, 78-83. <https://doi.org/10.1016/j.jtbi.2015.07.034>
123. Fan, G.L., Zhang, X.Y., Liu, Y.L., Nang, Y. and Wang, H. (2015) DSPMP: Discriminating Secretory Proteins of Malaria Parasite by Hybridizing Different Descriptors of Chou's Pseudo Amino Acid Patterns. *Journal of Computational Chemistry*, **36**, 2317-2327. <https://doi.org/10.1002/jcc.24210>
124. Huang, C. and Yuan, J.Q. (2015) Simultaneously Identify Three Different Attributes of Proteins by Fusing Their Three Different Modes of Chou's Pseudo Amino Acid Compositions. *Protein & Peptide Letters*, **22**, 547-556. <https://doi.org/10.2174/0929866522666150209151344>
125. Khan, Z.U., Hayat, M. and Khan, M.A. (2015) Discrimination of Acidic and Alkaline Enzyme Using Chou's Pseudo Amino Acid Composition in Conjunction with Probabilistic Neural Network Model. *Journal of Theoretical Biology*, **365**, 197-203. <https://doi.org/10.1016/j.jtbi.2014.10.014>
126. Kumar, R., Srivastava, A., Kumari, B. and Kumar, M. (2015) Prediction of Beta-Lactamase and Its Class by Chou's Pseudo Amino Acid Composition and Support Vector Machine. *Journal of Theoretical Biology*, **365**, 96-103. <https://doi.org/10.1016/j.jtbi.2014.10.008>
127. Wang, X., Zhang, W., Zhang, Q. and Li, G.Z. (2015) MultiP-SChlo: Multi-Label Protein Subchloroplast Localization Prediction with Chou's Pseudo Amino Acid Composition and a Novel Multi-Label Classifier. *Bioinformatics*, **31**, 2639-2645. <https://doi.org/10.1093/bioinformatics/btv212>
128. Jiao, Y.S. and Du, P.F. (2016) Prediction of Golgi-Resident Protein Types Using General Form of Chou's Pseudo Amino Acid Compositions: Approaches with Minimal Redundancy Maximal Relevance Feature Selection. *Journal of Theoretical Biology*, **402**, 38-44. <https://doi.org/10.1016/j.jtbi.2016.04.032>
129. Tang, H., Chen, W. and Lin, H. (2016) Identification of Immunoglobulins Using Chou's Pseudo Amino Acid Composition with Feature Selection Technique. *Molecular BioSystems*, **12**, 1269-1275.

<https://doi.org/10.1039/C5MB00883B>

130. Zou, H.L. and Xiao, X. (2016) Predicting the Functional Types of Singleplex and Multiplex Eukaryotic Membrane Proteins via Different Models of Chou's Pseudo Amino Acid Compositions. *The Journal of Membrane Biology*, **249**, 23-29. <https://doi.org/10.1007/s00232-015-9830-9>
131. Rahimi, M., Bakhtiarizadeh, M.R. and Mohammadi-Sangcheshmeh, A. (2017) OOgenesis_Pred: A Sequence-Based Method for Predicting Oogenesis Proteins by Six Different Modes of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **414**, 128-136. <https://doi.org/10.1016/j.jtbi.2016.11.028>
132. Tripathi, P. and Pandey, P.N. (2017) A Novel Alignment-Free Method to Classify Protein Folding Types by Combining Spectral Graph Clustering with Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **424**, 49-54. <https://doi.org/10.1016/j.jtbi.2017.04.027>
133. Yu, B., Lou, L., Li, S., Zhang, Y., Qiu, W., Wu, X., Wang, M. and Tian, B. (2017) Prediction of Protein Structural Class for Low-Similarity Sequences Using Chou's Pseudo Amino Acid Composition and Wavelet Denoising. *Journal of Molecular Graphics and Modelling*, **76**, 260-273. <https://doi.org/10.1016/j.jmkgm.2017.07.012>
134. Arif, M., Hayat, M. and Jan, Z. (2018) iMem-2LSAAC: A Two-Level Model for Discrimination of Membrane Proteins and Their Types by Extending the Notion of SAAC into Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **442**, 11-21. <https://doi.org/10.1016/j.jtbi.2018.01.008>
135. Mei, J. and Zhao, J. (2018) Prediction of HIV-1 and HIV-2 Proteins by Using Chou's Pseudo Amino Acid Compositions and Different Classifiers. *Scientific Reports*, **8**, Article No. 2359. <https://doi.org/10.1038/s41598-018-20819-x>
136. Nosrati, M., Mohabatkar, H. and Behbahani, M. (2019) Introducing of an Integrated Artificial Neural Network and Chou's Pseudo Amino Acid Composition Approach for Computational Epitope-Mapping of Crimean-Congo Haemorrhagic Fever Virus Antigens. *International Immunopharmacology*. <https://doi.org/10.1016/j.intimp.2019.106020>
<https://www.sciencedirect.com/science/article/pii/S1567576919321277>
137. Nosrati, M., Mohabatkar, H. and Behbahani, M. (2020) Introducing of an Integrated Artificial Neural Network and Chou's Pseudo Amino Acid Composition Approach for Computational Epitope-Mapping of Crimean-Congo Haemorrhagic Fever Virus Antigens. *International Immunopharmacology*, **78**, Article ID: 106020. <https://doi.org/10.1016/j.intimp.2019.106020>
138. Hayat, M. and Khan, A. (2012) Discriminating Outer Membrane Proteins with Fuzzy K-Nearest Neighbor Algorithms Based on the General Form of Chou's PseAAC. *Protein & Peptide Letters*, **19**, 411-421. <https://doi.org/10.2174/092986612799789387>
139. Liao, B., Xiang, Q. and Li, D. (2012) Incorporating Secondary Features into the General Form of Chou's PseAAC for Predicting Protein Structural Class. *Protein & Peptide Letters*, **19**, 1133-1138. <https://doi.org/10.2174/092986612803217051>
140. Mei, S. (2012) Multi-Kernel Transfer Learning Based on Chou's PseAAC Formulation for Protein Mitochondria Localization. *Journal of Theoretical Biology*, **293**, 121-130. <https://doi.org/10.1016/j.jtbi.2011.10.015>
141. Mei, S. (2012) Predicting Plant Protein Subcellular Multi-Localization by Chou's PseAAC Formulation Based Multi-Label Homolog Knowledge Transfer Learning. *Journal of Theoretical Biology*, **310**, 80-87. <https://doi.org/10.1016/j.jtbi.2012.06.028>
142. Qin, Y.F., Wang, C.H., Yu, X.Q., Zhu, J., Liu, T.G. and Zheng, X.Q. (2012) Predicting Protein Structural Class by Incorporating Patterns of Over-Represented k-Mers into the General Form of Chou's PseAAC. *Protein & Peptide Letters*, **19**, 388-397. <https://doi.org/10.2174/092986612799789350>
143. Sun, X.Y., Shi, S.P., Qiu, J.D., Suo, S.B., Huang, S.Y. and Liang, R.P. (2012) Identifying Protein Quaternary

Structural Attributes by Incorporating Physicochemical Properties into the General Form of Chou's PseAAC via Discrete Wavelet Transform. *Molecular BioSystems*, **8**, 3178-3184. <https://doi.org/10.1039/c2mb25280e>

144. Cao, D.S., Xu, Q.S. and Liang, Y.Z. (2013) Propy: A Tool to Generate Various Modes of Chou's PseAAC. *Bioinformatics*, **29**, 960-962. <https://doi.org/10.1093/bioinformatics/btt072>
145. Chang, T.H., Wu, L.C., Lee, T.Y., Chen, S.P., Huang, H.D. and Horng, J.T. (2013) EuLoc: A Web-Server for Accurately Predict Protein Subcellular Localization in Eukaryotes by Incorporating Various Features of Sequence Segments into the General Form of Chou's PseAAC. *Journal of Computer-Aided Molecular Design*, **27**, 91-103. <https://doi.org/10.1007/s10822-012-9628-0>
146. Fan, G.-L., Li, Q.-Z. and Zuo, Y.-C. (2013) Predicting Acidic and Alkaline Enzymes by Incorporating the Average Chemical Shift and Gene Ontology Informations into the General Form of Chou's PseAAC. *Process Biochemistry*, **48**, 1048-1053. <https://doi.org/10.1016/j.procbio.2013.05.012>
147. Pacharawongsakda, E. and Theeramunkong, T. (2013) Predict Subcellular Locations of Singleplex and Multiplex Proteins by Semi-Supervised Learning and Dimension-Reducing General Mode of Chou's PseAAC. *IEEE Transactions on Nanobioscience*, **12**, 311-320. <https://doi.org/10.1109/TNB.2013.2272014>
148. Xie, H.L., Fu, L. and Nie, X.D. (2013) Using Ensemble SVM to Identify Human GPCRs N-Linked Glycosylation Sites Based on the General Form of Chou's PseAAC. *Protein Engineering, Design and Selection*, **26**, 735-742. <https://doi.org/10.1093/protein/gzt042>
149. Han, G.S., Yu, Z.G. and Anh, V. (2014) A Two-Stage SVM Method to Predict Membrane Protein Types by Incorporating Amino Acid Classifications and Physicochemical Properties into a General Form of Chou's PseAAC. *Journal of Theoretical Biology*, **344**, 31-39. <https://doi.org/10.1016/j.jtbi.2013.11.017>
150. Li, L., Yu, S., Xiao, W., Li, Y., Li, M., Huang, L., Zheng, X., Zhou, S. and Yang, H. (2014) Prediction of Bacterial Protein Subcellular Localization by Incorporating Various Features into Chou's PseAAC and a Backward Feature Selection Approach. *Biochimie*, **104**, 100-107. <https://doi.org/10.1016/j.biochi.2014.06.001>
151. Zhang, J., Zhao, X., Sun, P. and Ma, Z. (2014) PSNO: Predicting Cysteine S-Nitrosylation Sites by Incorporating Various Sequence-Derived Features into the General Form of Chou's PseAAC. *International Journal of Molecular Sciences*, **15**, 11204-11219. <https://doi.org/10.3390/ijms150711204>
152. Liu, B., Xu, J., Fan, S., Xu, R., Zhou, J. and Wang, X. (2015) PseDNA-Pro: DNA-Binding Protein Identification by Combining Chou's PseAAC and Physicochemical Distance Transformation. *Molecular Informatics*, **34**, 8-17. <https://doi.org/10.1002/minf.201400025>
153. Mandal, M., Mukhopadhyay, A. and Maulik, U. (2015) Prediction of Protein Subcellular Localization by Incorporating Multiobjective PSO-Based Feature Subset Selection into the General Form of Chou's PseAAC. *Medical & Biological Engineering & Computing*, **53**, 331-344. <https://doi.org/10.1007/s11517-014-1238-7>
154. Sanchez, V., Peinado, A.M., Perez-Cordoba, J.L. and Gomez, A.M. (2015) A New Signal Characterization and Signal-Based Chou's PseAAC Representation of Protein Sequences. *Journal of Bioinformatics and Computational Biology*, **13**, Article ID: 1550024. <https://doi.org/10.1142/S0219720015500249>
155. Kabir, M. and Hayat, M. (2016) iRSpot-GAEnsC: Identifying Recombination Spots via Ensemble Classifier and Extending the Concept of Chou's PseAAC to Formulate DNA Samples. *Molecular Genetics and Genomics*, **291**, 285-296. <https://doi.org/10.1007/s00438-015-1108-5>
156. Tahir, M. and Hayat, M. (2016) iNuc-STNC: A Sequence-Based Predictor for Identification of Nucleosome Positioning in Genomes by Extending the Concept of SAAC and Chou's PseAAC. *Molecular BioSystems*, **12**, 2587-2593. <https://doi.org/10.1039/C6MB00221H>
157. Ju, Z. and He, J.J. (2017) Prediction of Lysine Propionylation Sites Using Biased SVM and Incorporating Four Different Sequence Features into Chou's PseAAC. *Journal of Molecular Graphics and Modelling*, **76**, 356-363.

<https://doi.org/10.1016/j.jmgm.2017.07.022>

158. Yu, B., Li, S., Qiu, W.Y., Chen, C., Chen, R.X., Wang, L., Wang, M.H. and Zhang, Y. (2017) Accurate Prediction of Subcellular Location of Apoptosis Proteins Combining Chou's PseAAC and PsePSSM Based on Wavelet Denoising. *Oncotarget*, **8**, 107640-107665. <https://doi.org/10.18632/oncotarget.22585>
159. Ahmad, J. and Hayat, M. (2018) MFSC: Multi-Voting Based Feature Selection for Classification of Golgi Proteins by Adopting the General Form of Chou's PseAAC Components. *Journal of Theoretical Biology*, **463**, 99-109. <https://doi.org/10.1016/j.jtbi.2018.12.017>
160. Akbar, S. and Hayat, M. (2018) iMethyl-STTNC: Identification of N(6)-methyladenosine Sites by Extending the Idea of SAAC into Chou's PseAAC to Formulate RNA Sequences. *Journal of Theoretical Biology*, **455**, 205-211. <https://doi.org/10.1016/j.jtbi.2018.07.018>
161. Contreras-Torres, E. (2018) Predicting Structural Classes of Proteins by Incorporating Their Global and Local Physicochemical and Conformational Properties into General Chou's PseAAC. *Journal of Theoretical Biology*, **454**, 139-145. <https://doi.org/10.1016/j.jtbi.2018.05.033>
162. Fu, X., Zhu, W., Liso, B., Cai, L., Peng, L. and Yang, J. (2018) Improved DNA-Binding Protein Identification by Incorporating Evolutionary Information into the Chou's PseAAC. *IEEE Access*, **20**. <https://doi.org/10.1109/ACCESS.2018.2876656>
163. Javed, F. and Hayat, M. (2018) Predicting Subcellular Localizations of Multi-Label Proteins by Incorporating the Sequence Features into Chou's PseAAC. *Genomics*. <https://doi.org/10.1016/j.ygeno.2018.09.004>
164. Mousavizadegan, M. and Mohabatkar, H. (2018) Computational Prediction of Antifungal Peptides via Chou's PseAAC and SVM. *Journal of Bioinformatics and Computational Biology*, **16**, Article ID: 1850016. <https://doi.org/10.1142/S0219720018500166>
165. Zhang, S. and Liang, Y. (2018) Predicting Apoptosis Protein Subcellular Localization by Integrating Auto-Cross Correlation and PSSM into Chou's PseAAC. *Journal of Theoretical Biology*, **457**, 163-169. <https://doi.org/10.1016/j.jtbi.2018.08.042>
166. Ahmad, J. and Hayat, M. (2019) MFSC: Multi-Voting Based Feature Selection for Classification of Golgi Proteins by Adopting the General Form of Chou's PseAAC Components. *Journal of Theoretical Biology*, **463**, 99-109. <https://doi.org/10.1016/j.jtbi.2018.12.017>
167. Butt, A.H., Rasool, N. and Khan, Y.D. (2019) Prediction of Antioxidant Proteins by Incorporating Statistical Moments Based Features into Chou's PseAAC. *Journal of Theoretical Biology*, **473**, 1-8. <https://doi.org/10.1016/j.jtbi.2019.04.019>
168. Javed, F. and Hayat, M. (2019) Predicting Subcellular Localization of Multi-Label Proteins by Incorporating the Sequence Features into Chou's PseAAC. *Genomics*, **111**, 1325-1332. <https://doi.org/10.1016/j.ygeno.2018.09.004>
169. Tahir, M., Hayat, M. and Khan, S.A. (2019) iNuc-ext-PseTNC: An Efficient Ensemble Model for Identification of Nucleosome Positioning by Extending the Concept of Chou's PseAAC to Pseudo-Tri-Nucleotide Composition. *Molecular Genetics and Genomics: MGG*, **294**, 199-210. <https://doi.org/10.1007/s00438-018-1498-2>
170. Hayat, M. and Iqbal, N. (2014) Discriminating Protein Structure Classes by Incorporating Pseudo Average Chemical Shift to Chou's General PseAAC and Support Vector Machine. *Computer Methods and Programs in Biomedicine*, **116**, 184-192. <https://doi.org/10.1016/j.cmpb.2014.06.007>
171. Ahmad, S., Kabir, M. and Hayat, M. (2015) Identification of Heat Shock Protein Families and J-Protein Types by Incorporating Dipeptide Composition into Chou's General PseAAC. *Computer Methods and Programs in Biomedicine*, **122**, 165-174. <https://doi.org/10.1016/j.cmpb.2015.07.005>
172. Dehzangi, A., Heffernan, R., Sharma, A., Lyons, J., Paliwal, K. and Sattar, A. (2015) Gram-Positive and Gram-Negative Protein Subcellular Localization by Incorporating Evolutionary-Based Descriptors into Chou's

General PseAAC. *Journal of Theoretical Biology*, **364**, 284-294. <https://doi.org/10.1016/j.jtbi.2014.09.029>

173. Sharma, R., Dehzangi, A., Lyons, J., Paliwal, K., Tsunoda, T. and Sharma, A. (2015) Predict Gram-Positive and Gram-Negative Subcellular Localization via Incorporating Evolutionary Information and Physicochemical Features Into Chou's General PseAAC. *IEEE Transactions on NanoBioscience*, **14**, 915-926. <https://doi.org/10.1109/TNB.2015.2500186>
174. Zhang, M., Zhao, B. and Liu, X. (2015) Predicting Industrial Polymer Melt Index via Incorporating Chaotic Characters into Chou's General PseAAC. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)*, **146**, 232-240. <https://doi.org/10.1016/j.chemolab.2015.05.028>
175. Zhang, S.L. (2015) Accurate Prediction of Protein Structural Classes by Incorporating PSSS and PSSM into Chou's General PseAAC. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)*, **142**, 28-35. <https://doi.org/10.1016/j.chemolab.2015.01.004>
176. Ahmad, K., Waris, M. and Hayat, M. (2016) Prediction of Protein Submitochondrial Locations by Incorporating Dipeptide Composition into Chou's General Pseudo Amino Acid Composition. *The Journal of Membrane Biology*, **249**, 293-304. <https://doi.org/10.1007/s00232-015-9868-8>
177. Behbahani, M., Mohabatkar, H. and Nosrati, M. (2016) Analysis and Comparison of Lignin Peroxidases between Fungi and Bacteria Using Three Different Modes of Chou's General Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **411**, 1-5. <https://doi.org/10.1016/j.jtbi.2016.09.001>
178. Fan, G.L., Liu, Y.L. and Wang, H. (2016) Identification of Thermophilic Proteins by Incorporating Evolutionary and Acid Dissociation Information into Chou's General Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **407**, 138-142. <https://doi.org/10.1016/j.jtbi.2016.07.010>
179. Ju, Z., Cao, J.Z. and Gu, H. (2016) Predicting Lysine Phosphoglycerylation with Fuzzy SVM by Incorporating k-Spaced Amino Acid Pairs into Chou's General PseAAC. *Journal of Theoretical Biology*, **397**, 145-150. <https://doi.org/10.1016/j.jtbi.2016.02.020>
180. Tiwari, A.K. (2016) Prediction of G-Protein Coupled Receptors and Their Subfamilies by Incorporating Various Sequence Features into Chou's General PseAAC. *Computer Methods and Programs in Biomedicine*, **134**, 197-213. <https://doi.org/10.1016/j.cmpb.2016.07.004>
181. Xu, C., Sun, D., Liu, S. and Zhang, Y. (2016) Protein Sequence Analysis by Incorporating Modified Chaos Game and Physicochemical Properties into Chou's General Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **406**, 105-115. <https://doi.org/10.1016/j.jtbi.2016.06.034>
182. Zou, H.L. and Xiao, X. (2016) Classifying Multifunctional Enzymes by Incorporating Three Different Models into Chou's General Pseudo Amino Acid Composition. *The Journal of Membrane Biology*, **249**, 561-567. <https://doi.org/10.1007/s00232-016-9904-3>
183. Jiao, Y.S. and Du, P.F. (2017) Predicting Protein Submitochondrial Locations by Incorporating the Positional-Specific Physicochemical Properties into Chou's General Pseudo-Amino Acid Compositions. *Journal of Theoretical Biology*, **416**, 81-87. <https://doi.org/10.1016/j.jtbi.2016.12.026>
184. Ju, Z. and He, J.J. (2017) Prediction of Lysine Crotonylation Sites by Incorporating the Composition of k-Spaced Amino Acid Pairs into Chou's General PseAAC. *Journal of Molecular Graphics and Modelling*, **77**, 200-204. <https://doi.org/10.1016/j.jmgm.2017.08.020>
185. Khan, M., Hayat, M., Khan, S.A. and Iqbal, N. (2017) Unb-DPC: Identify Mycobacterial Membrane Protein Types by Incorporating Un-Biased Dipeptide Composition into Chou's General PseAAC. *Journal of Theoretical Biology*, **415**, 13-19. <https://doi.org/10.1016/j.jtbi.2016.12.004>
186. Liang, Y. and Zhang, S. (2017) Predict Protein Structural Class by Incorporating Two Different Modes of Evolutionary Information into Chou's General Pseudo Amino Acid Composition. *Journal of Molecular Graphics and*

Modelling, **78**, 110-117. <https://doi.org/10.1016/j.jmgm.2017.10.003>

187. Meher, P.K., Sahu, T.K., Saini, V. and Rao, A.R. (2017) Predicting Antimicrobial Peptides with Improved Accuracy by Incorporating the Compositional, Physico-Chemical and Structural Features into Chou's General PseAAC. *Scientific Reports*, **7**, Article No. 42362. <https://doi.org/10.1038/srep42362>
188. Qiu, W.R., Zheng, Q.S., Sun, B.Q. and Xiao, X. (2017) Multi-iPPseEvo: A Multi-Label Classifier for Identifying Human Phosphorylated Proteins by Incorporating Evolutionary Information into Chou's General PseAAC via Grey System Theory. *Molecular Informatics*, **36**, UNSP 1600085. <https://doi.org/10.1002/minf.201600085>
189. Xu, C., Ge, L., Zhang, Y., Dehmer, M. and Gutman, I. (2017) Prediction of Therapeutic Peptides by Incorporating q-Wiener Index into Chou's General PseAAC. *Journal of Biomedical Informatics*. <https://doi.org/10.1016/j.jbi.2017.09.011>
190. Butt, A.H., Rasool, N. and Khan, Y.D. (2018) Predicting Membrane Proteins and Their Types by Extracting Various Sequence Features into Chou's General PseAAC. *Molecular Biology Reports*. <https://doi.org/10.1007/s11033-018-4391-5>
191. Ghauri, A.W., Khan, Y.D., Rasool, N., Khan, S.A. and Chou, K.C. (2018) pNitro-Tyr-PseAAC: Predict Nitrotyrosine Sites in Proteins by Incorporating Five Features into Chou's General PseAAC. *Current Pharmaceutical Design*, **24**, 4034-4043. <https://doi.org/10.2174/1381612825666181127101039>
192. Ju, Z. and Wang, S.Y. (2018) Prediction of Citrullination Sites by Incorporating k-Spaced Amino Acid Pairs into Chou's General Pseudo Amino Acid Composition. *Gene*, **664**, 78-83. <https://doi.org/10.1016/j.gene.2018.04.055>
193. Krishnan, M.S. (2018) Using Chou's General PseAAC to Analyze the Evolutionary Relationship of Receptor Associated Proteins (RAP) with Various Folding Patterns of Protein Domains. *Journal of Theoretical Biology*, **445**, 62-74. <https://doi.org/10.1016/j.jtbi.2018.02.008>
194. Liang, Y. and Zhang, S. (2018) Identify Gram-Negative Bacterial Secreted Protein Types by Incorporating Different Modes of PSSM into Chou's General PseAAC via Kullback-Leibler Divergence. *Journal of Theoretical Biology*, **454**, 22-29. <https://doi.org/10.1016/j.jtbi.2018.05.035>
195. Mei, J., Fu, Y. and Zhao, J. (2018) Analysis and Prediction of Ion Channel Inhibitors by Using Feature Selection and Chou's General Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **456**, 41-48. <https://doi.org/10.1016/j.jtbi.2018.07.040>
196. Mei, J. and Zhao, J. (2018) Analysis and Prediction of Presynaptic and Postsynaptic Neurotoxins by Chou's General Pseudo Amino Acid Composition and Motif Features. *Journal of Theoretical Biology*, **427**, 147-153. <https://doi.org/10.1016/j.jtbi.2018.03.034>
197. Rahman, S.M., Shatabda, S., Saha, S., Kaykobad, M. and Sohel Rahman, M. (2018) DPP-PseAAC: A DNA-Binding Protein Prediction Model Using Chou's General PseAAC. *Journal of Theoretical Biology*, **452**, 22-34. <https://doi.org/10.1016/j.jtbi.2018.05.006>
198. Sankari, E.S. and Manimegalai, D.D. (2018) Predicting Membrane Protein Types by Incorporating a Novel Feature Set into Chou's GENERAL PseAAC. *Journal of Theoretical Biology*, **455**, 319-328. <https://doi.org/10.1016/j.jtbi.2018.07.032>
199. Srivastava, A., Kumar, R. and Kumar, M. (2018) BlaPred: Predicting and Classifying Beta-Lactamase Using a 3-Tier Prediction System via Chou's General PseAAC. *Journal of Theoretical Biology*, **457**, 29-36. <https://doi.org/10.1016/j.jtbi.2018.08.030>
200. Zhang, S. and Duan, X. (2018) Prediction of Protein Subcellular Localization with Oversampling Approach and Chou's General PseAAC. *Journal of Theoretical Biology*, **437**, 239-250. <https://doi.org/10.1016/j.jtbi.2017.10.030>
201. Adilina, S., Farid, D.M. and Shatabda, S. (2019) Effective DNA Binding Protein Prediction by Using Key Features via Chou's General PseAAC. *Journal of Theoretical Biology*, **460**, 64-78.

<https://doi.org/10.1016/j.jtbi.2018.10.027>

202. Behbahani, M., Nosrati, M., Moradi, M. and Mohabatkar, H. (2019) Using Chou's General Pseudo Amino Acid Composition to Classify Laccases from Bacterial and Fungal Sources via Chou's Five-Step Rule. *Applied Biochemistry and Biotechnology*, **190**, 1035-1048. <https://doi.org/10.1007/s12010-019-03141-8>
203. Chen, G., Cao, M., Yu, J., Guo, X. and Shi, S. (2019) Prediction and Functional Analysis of Prokaryote Lysine Acetylation Site by Incorporating Six Types of Features into Chou's General PseAAC. *Journal of Theoretical Biology*, **461**, 92-101. <https://doi.org/10.1016/j.jtbi.2018.10.047>
204. Shen, Y., Tang, J. and Guo, F. (2019) Identification of Protein Subcellular Localization via Integrating Evolutionary and Physicochemical Information into Chou's General PseAAC. *Journal of Theoretical Biology*, **462**, 230-239. <https://doi.org/10.1016/j.jtbi.2018.11.012>
205. Wang, L., Zhang, R. and Mu, Y. (2019) Fu-SulfPred: Identification of Protein S-Sulfenylation Sites by Fusing Forests via Chou's General PseAAC. *Journal of Theoretical Biology*, **461**, 51-58. <https://doi.org/10.1016/j.jtbi.2018.10.046>
206. Xiao, X., Cheng, X., Chen, G., Mao, Q. and Chou, K.C. (2019) pLoc_bal-mVirus: Predict Subcellular Localization of Multi-Label Virus Proteins by Chou's General PseAAC and IHTS Treatment to Balance Training Dataset. *Journal of Medicinal Chemistry*, **15**, 496-509. <https://doi.org/10.2174/1573406415666181217114710>
207. Behbahani, M., Nosrati, M., Moradi, M. and Mohabatkar, H. (2020) Using Chou's General Pseudo Amino Acid Composition to Classify Laccases from Bacterial and Fungal Sources via Chou's Five-Step Rule. *Applied Biochemistry and Biotechnology*, **190**, 1035-1048. <https://doi.org/10.1007/s12010-019-03141-8>
208. Chou, K.C. (2009) Pseudo Amino Acid Composition and Its Applications in Bioinformatics, Proteomics and System Biology. *Current Proteomics*, **6**, 262-274. <https://doi.org/10.2174/157016409789973707>
209. Chen, W., Lei, T.Y., Jin, D.C., Lin, H. and Chou, K.C. (2014) PseKNC: A Flexible Web-Server for Generating Pseudo K-Tuple Nucleotide Composition. *Analytical Biochemistry*, **456**, 53-60. <https://doi.org/10.1016/j.ab.2014.04.001>
210. Chen, W., Lin, H. and Chou, K.C. (2015) Pseudo Nucleotide Composition or PseKNC: An Effective Formulation for Analyzing Genomic Sequences. *Molecular BioSystems*, **11**, 2620-2634. <https://doi.org/10.1039/C5MB00155B>
211. Chen, W., Feng, P.M., Lin, H. and Chou, K.C. (2014) iSS-PseDNC: Identifying Splicing Sites Using Pseudo Dinucleotide Composition. *Biomed Research International (BMRJ)*, **2014**, Article ID: 623149. <https://doi.org/10.1155/2014/623149>
212. Chen, W., Tang, H., Ye, J., Lin, H. and Chou, K.C. (2016) iRNA-PseU: Identifying RNA Pseudouridine Sites. *Molecular Therapy—Nucleic Acids*, **5**, e332.
213. Liu, B., Fang, L., Long, R., Lan, X. and Chou, K.C. (2016) iEnhancer-2L: A Two-Layer Predictor for Identifying Enhancers and Their Strength by Pseudo k-Tuple Nucleotide Composition. *Bioinformatics*, **32**, 362-369. <https://doi.org/10.1093/bioinformatics/btv604>
214. Liu, B., Long, R. and Chou, K.C. (2016) iDHS-EL: Identifying DNase I Hypersensitive Sites by Fusing Three Different Modes of Pseudo Nucleotide Composition into an Ensemble Learning Framework. *Bioinformatics*, **32**, 2411-2418. <https://doi.org/10.1093/bioinformatics/btw186>
215. Feng, P., Ding, H., Yang, H., Chen, W., Lin, H. and Chou, K.C. (2017) iRNA-PseColl: Identifying the Occurrence Sites of Different RNA Modifications by Incorporating Collective Effects of Nucleotides into PseKNC. *Molecular Therapy—Nucleic Acids*, **7**, 155-163. <https://doi.org/10.1016/j.omtn.2017.03.006>
216. Liu, B., Wang, S., Long, R. and Chou, K.C. (2017) iRSpot-EL: Identify Recombination Spots with an Ensemble Learning Approach. *Bioinformatics*, **33**, 35-41. <https://doi.org/10.1093/bioinformatics/btw539>

217. Liu, B., Yang, F. and Chou, K.C. (2017) 2L-piRNA: A Two-Layer Ensemble Classifier for Identifying Piwi-Interacting RNAs and Their Function. *Molecular Therapy—Nucleic Acids*, **7**, 267-277. <https://doi.org/10.1016/j.omtn.2017.04.008>
218. Al Maruf, M.A. and Shatabda, S. (2018) iRSpot-SF: Prediction of Recombination Hotspots by Incorporating Sequence Based Features into Chou's Pseudo Components. *Genomics*. <https://doi.org/10.1016/j.ygeno.2018.06.003>
219. Sabooh, M.F., Iqbal, N., Khan, M., Khan, M. and Maqbool, H.F. (2018) Identifying 5-Methylcytosine Sites in RNA Sequence Using Composite Encoding Feature into Chou's PseKNC. *Journal of Theoretical Biology*, **452**, 1-9. <https://doi.org/10.1016/j.jtbi.2018.04.037>
220. Zhang, L. and Kong, L. (2018) iRSpot-ADPM: Identify Recombination Spots by Incorporating the Associated Dinucleotide Product Model into Chou's Pseudo Components. *Journal of Theoretical Biology*, **441**, 1-8. <https://doi.org/10.1016/j.jtbi.2017.12.025>
221. Zhang, S., Yang, K., Lei, Y. and Song, K. (2018) iRSpot-DTS: Predict Recombination Spots by Incorporating the Dinucleotide-Based Sparse-Cross Covariance Information into Chou's Pseudo Components. *Genomics*, **11**, 457-464.
222. Chou, K.C. (2019) Two Kinds of Metrics for Computational Biology. *Genomics*. <https://www.sciencedirect.com/science/article/pii/S0888754319304604?via%3Dihub>
223. Chen, W., Feng, P.M., Lin, H. and Chou, K.C. (2013) iRSpot-PseDNC: Identify Recombination Spots with Pseudo Dinucleotide Composition. *Nucleic Acids Research*, **41**, e68. <https://doi.org/10.1093/nar/gks1450>
224. Chou, K.C. (2001) Using Subsite Coupling to Predict Signal Peptides. *Protein Engineering*, **14**, 75-79. <https://doi.org/10.1093/protein/14.2.75>
225. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) Identification of Protein-Protein Binding Sites by Incorporating the Physicochemical Properties and Stationary Wavelet Transforms into Pseudo Amino Acid Composition (iPPBS-PseAAC). *Journal of Biomolecular Structure and Dynamics (JBSD)*, **34**, 1946-1961. <https://doi.org/10.1080/07391102.2015.1095116>
226. Xiao, X., Wang, P., Lin, W.Z., Jia, J.H. and Chou, K.C. (2013) iAMP-2L: A Two-Level Multi-Label Classifier for Identifying Antimicrobial Peptides and Their Functional Types. *Analytical Biochemistry*, **436**, 168-177. <https://doi.org/10.1016/j.ab.2013.01.019>
227. Jia, J., Liu, Z., Xiao, X. and Chou, K.C. (2015) iPPI-Esml: An Ensemble Classifier for Identifying the Interactions of Proteins by Incorporating Their Physicochemical Properties and Wavelet Transforms into PseAAC. *Journal of Theoretical Biology*, **377**, 47-56. <https://doi.org/10.1016/j.jtbi.2015.04.011>
228. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) iSuc-PseOpt: Identifying Lysine Succinylation Sites in Proteins by Incorporating Sequence-Coupling Effects into Pseudo Components and Optimizing Imbalanced Training Dataset. *Analytical Biochemistry*, **497**, 48-56. <https://doi.org/10.1016/j.ab.2015.12.009>
229. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) pSuc-Lys: Predict Lysine Succinylation Sites in Proteins with PseAAC and Ensemble Random Forest Approach. *Journal of Theoretical Biology*, **394**, 223-230. <https://doi.org/10.1016/j.jtbi.2016.01.020>
230. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) iCar-PseCp: Identify Carbonylation Sites in Proteins by Monte Carlo Sampling and Incorporating Sequence Coupled Effects into General PseAAC. *Oncotarget*, **7**, 34558-34570. <https://doi.org/10.18632/oncotarget.9148>
231. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) iPPBS-Opt: A Sequence-Based Ensemble Classifier for Identifying Protein-Protein Binding Sites by Optimizing Imbalanced Training Datasets. *Molecules*, **21**, E95. <https://doi.org/10.3390/molecules21010095>

232. Jia, J., Zhang, L., Liu, Z., Xiao, X. and Chou, K.C. (2016) pSumo-CD: Predicting Sumoylation Sites in Proteins with Covariance Discriminant Algorithm by Incorporating Sequence-Coupled Effects into General PseAAC. *Bioinformatics*, **32**, 3133-3141. <https://doi.org/10.1093/bioinformatics/btw387>
233. Liu, Z., Xiao, X., Yu, D.J., Jia, J., Qiu, W.R. and Chou, K.C. (2016) pRNA-PC: Predicting N-Methyladenosine Sites in RNA Sequences via Physical-Chemical Properties. *Analytical Biochemistry*, **497**, 60-67. <https://doi.org/10.1016/j.ab.2015.12.017>
234. Xiao, X., Ye, H.X., Liu, Z., Jia, J.H. and Chou, K.C. (2016) iROS-gPseKNC: Predicting Replication Origin Sites in DNA by Incorporating Dinucleotide Position-Specific Propensity into General Pseudo Nucleotide Composition. *Oncotarget*, **7**, 34180-34189. <https://doi.org/10.18632/oncotarget.9057>
235. Qiu, W.R., Sun, B.Q., Xiao, X., Xu, Z.C., Jia, J.H. and Chou, K.C. (2018) iKcr-PseEns: Identify Lysine Crotonylation Sites in Histone Proteins with Pseudo Components and Ensemble Classifier. *Genomics*, **110**, 239-246. <https://doi.org/10.1016/j.ygeno.2017.10.008>
236. Jia, J., Li, X., Qiu, W., Xiao, X. and Chou, K.C. (2019) iPPI-PseAAC(CGR): Identify Protein-Protein Interactions by Incorporating Chaos Game Representation into PseAAC. *Journal of Theoretical Biology*, **460**, 195-203. <https://doi.org/10.1016/j.jtbi.2018.10.021>
237. Xu, Y., Ding, J., Wu, L.Y. and Chou, K.C. (2013) iSNO-PseAAC: Predict Cysteine S-Nitrosylation Sites in Proteins by Incorporating Position Specific Amino Acid Propensity into Pseudo Amino Acid Composition *PLoS ONE*, **8**, e55844. <https://doi.org/10.1371/journal.pone.0055844>
238. Xu, Y., Shao, X.J., Wu, L.Y., Deng, N.Y. and Chou, K.C. (2013) iSNO-AAPair: Incorporating Amino Acid Pairwise Coupling into PseAAC for Predicting Cysteine S-Nitrosylation Sites in Proteins. *PeerJ*, **1**, e171. <https://doi.org/10.7717/peerj.171>
239. Xu, Y., Wen, X., Shao, X.J., Deng, N.Y. and Chou, K.C. (2014) iHyd-PseAAC: Predicting Hydroxyproline and Hydroxylysine in Proteins by Incorporating Dipeptide Position-Specific Propensity into Pseudo Amino Acid Composition. *International Journal of Molecular Sciences*, **15**, 7594-7610. <https://doi.org/10.3390/ijms15057594>
240. Xu, Y., Wen, X., Wen, L.S., Wu, L.Y., Deng, N.Y. and Chou, K.C. (2014) iNitro-Tyr: Prediction of Nitrotyrosine Sites in Proteins with General Pseudo Amino Acid Composition. *PLoS ONE*, **9**, e105018. <https://doi.org/10.1371/journal.pone.0105018>
241. Xu, Y. and Chou, K.C. (2016) Recent Progress in Predicting Posttranslational Modification Sites in Proteins. *Current Topics in Medicinal Chemistry*, **16**, 591-603. <https://doi.org/10.2174/1568026615666150819110421>
242. Liu, L.M., Xu, Y. and Chou, K.C. (2017) iPGK-PseAAC: Identify Lysine Phosphoglycerylation Sites in Proteins by Incorporating Four Different Tiers of Amino Acid Pairwise Coupling Information into the General PseAAC. *Journal of Medicinal Chemistry*, **13**, 552-559. <https://doi.org/10.2174/1573406413666170515120507>
243. Xu, Y., Li, C. and Chou, K.C. (2017) iPreny-PseAAC: Identify C-Terminal Cysteine Prenylation Sites in Proteins by Incorporating Two Tiers of Sequence Couplings into PseAAC. *Journal of Medicinal Chemistry*, **13**, 544-551. <https://doi.org/10.2174/1573406413666170419150052>
244. Liu, B., Zhang, D., Xu, R., Xu, J., Wang, X., Chen, Q., Dong, Q. and Chou, K.C. (2014) Combining Evolutionary Information Extracted from Frequency Profiles with Sequence-Based Kernels for Protein Remote Homology Detection. *Bioinformatics*, **30**, 472-479. <https://doi.org/10.1093/bioinformatics/btt709>
245. Chen, J., Long, R., Wang, X.L., Liu, B. and Chou, K.C. (2016) dRHP-PseRA: Detecting Remote Homology Proteins Using Profile-Based Pseudo Protein Sequence and Rank Aggregation. *Scientific Reports*, **6**, Article No. 32333. <https://doi.org/10.1038/srep32333>
246. Xu, R., Zhou, J., Liu, B., He, Y.A., Zou, Q., Wang, X. and Chou, K.C. (2015) Identification of DNA-Binding Proteins by Incorporating Evolutionary Information into Pseudo Amino Acid Composition via the Top-N-Gram

Approach. *Journal of Biomolecular Structure & Dynamics (JBSD)*, **33**, 1720-1730.

<https://doi.org/10.1080/07391102.2014.968624>

247. Liu, B., Fang, L., Wang, S., Wang, X., Li, H. and Chou, K.C. (2015) Identification of microRNA Precursor with the Degenerate K-tuple or Kmer Strategy. *Journal of Theoretical Biology*, **385**, 153-159.
<https://doi.org/10.1016/j.jtbi.2015.08.025>
248. Liu, B., Fang, L., Liu, F., Wang, X., Chen, J. and Chou, K.C. (2015) Identification of Real microRNA Precursors with a Pseudo Structure Status Composition Approach. *PLoS ONE*, **10**, e0121501.
<https://doi.org/10.1371/journal.pone.0121501>
249. Liu, B., Xu, J., Lan, X., Xu, R., Zhou, J., Wang, X. and Chou, K.C. (2014) iDNA-Prot|dis: Identifying DNA-Binding Proteins by Incorporating Amino Acid Distance-Pairs and Reduced Alphabet Profile into the General Pseudo Amino Acid Composition. *PLoS ONE*, **9**, e106691.
<https://doi.org/10.1371/journal.pone.0106691>
250. Liu, B., Li, K., Huang, D.S. and Chou, K.C. (2018) iEnhancer-EL: Identifying Enhancers and Their Strength with Ensemble Learning Approach. *Bioinformatics*, **34**, 3835-3842. <https://doi.org/10.1093/bioinformatics/bty458>
251. Liu, B., Yang, F., Huang, D.S. and Chou, K.C. (2018) iPromoter-2L: A Two-Layer Predictor for Identifying Promoters and Their Types by Multi-Window-Based PseKNC. *Bioinformatics*, **34**, 33-40.
<https://doi.org/10.1093/bioinformatics/btx579>
252. Liu, B., Weng, F., Huang, D.S. and Chou, K.C. (2018) iRO-3wPseKNC: Identify DNA Replication Origins by Three-Window-Based PseKNC. *Bioinformatics*, **34**, 3086-3093. <https://doi.org/10.1093/bioinformatics/bty312>
253. Liu, B., Wu, H., Zhang, D., Wang, X. and Chou, K.C. (2017) Pse-Analysis: A Python Package for DNA/RNA and Protein/Peptide Sequence Analysis Based on Pseudo Components and Kernel Methods. *Oncotarget*, **8**, 13338-13343. <https://doi.org/10.18632/oncotarget.14524>
254. Liu, B., Wu, H. and Chou, K.C. (2017) Pse-in-One 2.0: An Improved Package of Web Servers for Generating Various Modes of Pseudo Components of DNA, RNA, and Protein Sequences. *Natural Science*, **9**, 67-91.
<https://doi.org/10.4236/ns.2017.94007>
255. Liu, B., Liu, F., Wang, X., Chen, J., Fang, L. and Chou, K.C. (2015) Pse-in-One: A Web Server for Generating Various Modes of Pseudo Components of DNA, RNA, and Protein Sequences. *Nucleic Acids Research*, **43**, W65-W71. <https://doi.org/10.1093/nar/gkv458>
256. Liu, B., Liu, F., Fang, L., Wang, X. and Chou, K.C. (2015) repDNA: A Python Package to Generate Various Modes of Feature Vectors for DNA Sequences by Incorporating User-Defined Physicochemical Properties and Sequence-Order Effects. *Bioinformatics*, **31**, 1307-1309. <https://doi.org/10.1093/bioinformatics/btu820>
257. Liu, B., Liu, F., Fang, L., Wang, X. and Chou, K.C. (2016) repRNA: A Web Server for Generating Various Feature Vectors of RNA Sequences. *Molecular Genetics and Genomics*, **291**, 473-481.
<https://doi.org/10.1007/s00438-015-1078-7>
258. Chen, W., Feng, P., Ding, H., Lin, H. and Chou, K.C. (2015) Benchmark Data for Identifying N6-Methyladenosine Sites in the *Saccharomyces cerevisiae* Genome. *Data in Brief*, **5**, 376-378.
<https://doi.org/10.1016/j.dib.2015.09.008>
259. Chen, W., Ding, H., Feng, P., Lin, H. and Chou, K.C. (2016) iACP: A Sequence-Based Tool for Identifying Anticancer Peptides. *Oncotarget*, **7**, 16895-16909. <https://doi.org/10.18632/oncotarget.7815>
260. Feng, P., Yang, H., Ding, H., Lin, H., Chen, W. and Chou, K.C. (2019) iDNA6mA-PseKNC: Identifying DNA N(6)-methyladenosine Sites by Incorporating Nucleotide Physicochemical Properties into PseKNC. *Genomics*, **111**, 96-102. <https://doi.org/10.1016/j.ygeno.2018.01.005>
261. Feng, P.M., Chen, W., Lin, H. and Chou, K.C. (2013) iHSP-PseRAAAC: Identifying the Heat Shock Protein

Families Using Pseudo Reduced Amino Acid Alphabet Composition. *Analytical Biochemistry*, **442**, 118-125.
<https://doi.org/10.1016/j.ab.2013.05.024>

262. Su, Z.D., Huang, Y., Zhang, Z.Y., Zhao, Y.W., Wang, D., Chen, W., Chou, K.C. and Lin, H. (2018) iLoc-lncRNA: Predict the Subcellular Location of lncRNAs by Incorporating Octamer Composition into General PseKNC. *Bioinformatics*, **34**, 4196-4204.
263. Chen, W., Lin, H., Feng, P.M., Ding, C., Zuo, Y.C. and Chou, K.C. (2012) iNuc-PhysChem: A Sequence-Based Predictor for Identifying Nucleosomes via Physicochemical Properties. *PLoS ONE*, **7**, e47843.
<https://doi.org/10.1371/journal.pone.0047843>
264. Guo, S.H., Deng, E.Z., Xu, L.Q., Ding, H., Lin, H., Chen, W. and Chou, K.C. (2014) iNuc-PseKNC: A Sequence-Based Predictor for Predicting Nucleosome Positioning in Genomes with Pseudo k-tuple Nucleotide Composition. *Bioinformatics*, **30**, 1522-1529. <https://doi.org/10.1093/bioinformatics/btu083>
265. Zhang, C.J., Tang, H., Li, W.C., Lin, H., Chen, W. and Chou, K.C. (2016) iOri-Human: Identify Human Origin of Replication by Incorporating Dinucleotide Physicochemical Properties into Pseudo Nucleotide Composition. *Oncotarget*, **7**, 69783-69793. <https://doi.org/10.18632/oncotarget.11975>
266. Lin, H., Deng, E.Z., Ding, H., Chen, W. and Chou, K.C. (2014) iPro54-PseKNC: A Sequence-Based Predictor for Identifying Sigma-54 Promoters in Prokaryote with Pseudo k-tuple Nucleotide Composition. *Nucleic Acids Research*, **42**, 12961-12972. <https://doi.org/10.1093/nar/gku1019>
267. Chen, W., Feng, P., Yang, H., Ding, H., Lin, H. and Chou, K.C. (2018) iRNA-3typeA: Identifying 3-Types of Modification at RNA's Adenosine Sites. *Molecular Therapy: Nucleic Acid*, **11**, 468-474.
<https://doi.org/10.1016/j.omtn.2018.03.012>
268. Chen, W., Feng, P., Yang, H., Ding, H., Lin, H. and Chou, K.C. (2017) iRNA-AI: Identifying the Adenosine to Inosine Editing Sites in RNA Sequences. *Oncotarget*, **8**, 4208-4217. <https://doi.org/10.18632/oncotarget.13758>
269. Chen, W., Feng, P., Ding, H., Lin, H. and Chou, K.C. (2015) iRNA-Methyl: Identifying N6-methyladenosine Sites Using Pseudo Nucleotide Composition. *Analytical Biochemistry*, **490**, 26-33.
<https://doi.org/10.1016/j.ab.2015.08.021>
270. Chen, W., Ding, H., Zhou, X., Lin, H. and Chou, K.C. (2018) iRNA(m6A)-PseDNC: Identifying N6-methyladenosine Sites Using Pseudo Dinucleotide Composition. *Analytical Biochemistry*, **561-562**, 59-65.
<https://doi.org/10.1016/j.ab.2018.09.002>
271. Yang, H., Qiu, W.R., Liu, G., Guo, F.B., Chen, W., Chou, K.C. and Lin, H. (2018) iRSpot-Pse6NC: Identifying Recombination Spots in *Saccharomyces cerevisiae* by Incorporating Hexamer Composition into General PseKNC. *International Journal of Biological Sciences*, **14**, 883-891. <https://doi.org/10.7150/ijbs.24616>
272. Chen, W., Feng, P.M., Deng, E.Z., Lin, H. and Chou, K.C. (2014) iTIS-PseTNC: A Sequence-Based Predictor for Identifying Translation Initiation Site in Human Genes Using Pseudo Trinucleotide Composition. *Analytical Biochemistry*, **462**, 76-83. <https://doi.org/10.1016/j.ab.2014.06.022>
273. Chen, W., Zhang, X., Brooker, J., Lin, H., Zhang, L. and Chou, K.C. (2015) PseKNC-General: A Cross-Platform Package for Generating Various Modes of Pseudo Nucleotide Compositions. *Bioinformatics*, **31**, 119-120.
<https://doi.org/10.1093/bioinformatics/btu602>
274. Chen, W., Feng, P., Ding, H., Lin, H. and Chou, K.C. (2016) Using Deformation Energy to Analyze Nucleosome Positioning in Genomes. *Genomics*, **107**, 69-75. <https://doi.org/10.1016/j.ygeno.2015.12.005>
275. Chou, K.C. (2020) The Development of Gordon Life Science Institute: Its Driving Force and Accomplishments. *Natural Science*, **12**, 202-217. <https://doi.org/10.4236/ns.2020.124018>
276. Chou, K.C. (2020) The Most Important Ethical Concerns in Science. *Natural Science*, **12**, 35-36.
<https://doi.org/10.4236/ns.2020.122005>

277. Chou, K.C. (2020) Other Mountain Stones Can Attack Jade: The 5-Steps Rule. *Natural Science*, **12**, 59-64. <https://doi.org/10.4236/ns.2020.123011>
278. Chou, K.C. (2020) The Problem of Elsevier Series Journals Online Submission by Using Artificial Intelligence. *Natural Science*, **12**, 37-38. <https://doi.org/10.4236/ns.2020.122006>
279. Chou, K.C. (2020) Proposing 5-Steps Rule Is a Notable Milestone for Studying Molecular Biology. *Natural Science*, **12**, 74-79. <https://doi.org/10.4236/ns.2020.123011>
280. Chou, K.C. (2020) Using Similarity Software to Evaluate Scientific Paper Quality Is a Big Mistake. *Natural Science*, **12**, 42-58. <https://doi.org/10.4236/ns.2020.123008>
281. Tsoumakas, G., Katakis, I. and Vlahavas, I. (2010) Random k-Labelsets for Multilabel Classification *IEEE Transactions on Knowledge and Data Engineering*, **23**, 1079-1089. <https://doi.org/10.1109/TKDE.2010.164>
282. Tsoumakas, G. and Katakis, L. (2007) Multi-Label Classification: An Overview *International Journal of Data Warehousing and Mining*, **3**, 13. <https://doi.org/10.4018/jdwm.2007070101>
283. Abdul Rahman, M.B., Karjiban, R.A., Salleh, A.B., Jacobs, D., Basri, M., Thean Chor, A.L., Abdul Wahab, H. and Rahman, R.N. (2009) Deciphering the Flexibility and Dynamics of Geobacillus Zalihae Strain T1 Lipase at High Temperatures by Molecular Dynamics Simulation. *Protein & Peptide Letters*, **16**, 1360-1370. <https://doi.org/10.2174/092986609789353763>
284. Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/A:1010933404324>
285. Zhang, T.L. and Ding, Y.S. (2007) Using Pseudo Amino Acid Composition and Binary-Tree Support Vector Machines to Predict Protein Structural Classes. *Amino Acids*, **33**, 623-629. <https://doi.org/10.1007/s00726-007-0496-1>
286. Chou, K.C. and Shen, H.B. (2009) Recent Advances in Developing Web-Servers for Predicting Protein Attributes. *Natural Science*, **1**, 63-92. <https://doi.org/10.4236/ns.2009.12011>
287. Chou, K.C. (2017) An Unprecedented Revolution in Medicinal Chemistry Driven by the Progress of Biological Science. *Current Topics in Medicinal Chemistry*, **17**, 2337-2358. <https://doi.org/10.2174/1568026617666170414145508>