pLoc_Deep-mEuk: Predict Subcellular Localization of Eukaryotic Proteins by Deep Learning

Yutao Shao[1], Kuo-Chen Chou[1,2]

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

Correspondence to: Yutao Shao, 532606318@qq.com; Kuo-Chen Chou, kcchou@gordonlifescience.org, kcchou38@gmail.com

ABSTRACT

Recently, the life of worldwide human beings 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 prerequisite. In 2019, a predictor called "pLoc_bal-mEuk" was developed for identifying the subcellular localization of eukaryotic 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-mEuk was still not trained by a "deep learning", a very powerful technique developed recently. The present study was devoted to incorporating the "deep-learning" technique and developed a new predictor called "pLoc_Deep-mEuk". 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-mEuk/, by which the majority of experimental scientists can easily get their desired data.

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.

In 2019, a very powerful predictor, called "pLoc_bal-mEuk" [4], was developed for predicting the subcellular localization of eukaryotic proteins based on their sequence information alone. It has the following remarkable advantages. 1) Most existing protein subcellular location prediction methods were developed based on the single-label system in which it was assumed that each constituent protein had one, and only one, subcellular location (see, e.g., [5-7] and a long list of references cited in a review papers [8]). With more experimental data uncovered, however, the localization of proteins in a cell is actually a multi-label system, where some proteins may simultaneously occur in two or more different location sites. This kind of multiplex proteins often bears some exceptional functions worthy of our special notice [2]. And the pLoc_bal-mEuk predictor [4] can cover this kind of important information missed by most other methods since it was established based on the multi-label benchmark dataset and theory. 2) Although there are a few methods (see, e.g., [9, 10]) that can be used to deal with multi-label subcellular localization for eukaryotic proteins, the prediction quality achieved by pLoc_bal-mEuk [4] is overwhelmingly higher, particularly in the absolute true rate. 3) Although the pLoc_bal-mEuk predictor [4] has the aforementioned merits, it has not been trained at a deeper level yet [11-14].

The present study was initiated in an attempt to address this problem. As done in pLoc_bal-mEuk [4] as well as many other recent publications in developing new prediction methods (see, e.g., [12-57]), the guidelines of the 5-step rule [58] 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-mEuk [4].

2. MATERIALS AND METHODS

2.1. Benchmark Dataset

The benchmark dataset used in this study is exactly the same as that in pLoc_bal-mEuk [4]; *i.e.*,

$$\mathbb{S} = \mathbb{S}_1 \bigcup \mathbb{S}_2 \bigcup \cdots \bigcup \mathbb{S}_u \bigcup \cdots \bigcup \mathbb{S}_{21} \bigcup \mathbb{S}_{22} \tag{1}$$

where $\mathbb{S}_1$ only contains the protein samples from the "Acrosome" location, $\mathbb{S}_2$ only contains those from the "Cell membrane" location, and so forth; $\bigcup$ denotes the symbol for "union" in the set theory. For readers' convenience, their detailed sequences and accession numbers (or ID codes) are given in Supporting Information S1 that is also available at http://www.jci-bioinfo.cn/pLoc_bal-mEuk/Supp1.pdf, where none of proteins included has ≥25% sequence identity to any other in the same subset (subcellular location).

2.2. Proteins Sample Formulation

Now let us consider the 2nd step of the 5-step rule [58]; *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

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

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 [59] or PseAAC [60]. Ever since then, the concept of "Pseudo Amino Acid Composition" 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., [4, 10, 23, 24, 61-227]). Because it has been widely and increasingly used, recently three powerful open access soft-wares, called "PseAAC-

Builder" [93], "propy" [181], and "PseAAC-General" [120], were established: the former two are for generating various modes of special PseAAC [228]; while the 3rd one for those of general PseAAC [58], 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 [229] that have proved very successful as well [141, 146, 147, 230-238]. According to the concept of general PseAAC [58], any protein sequence can be formulated as a PseAAC vector given by

$$\mathbf{P} = \begin{bmatrix} \Psi_1 \ \Psi_2 \cdots \Psi_u \cdots \Psi_\Omega \end{bmatrix}^{\mathbf{T}} \tag{3}$$

where $\mathbf{T}$ is a transpose operator, while the integer Ω is a parameter and its value as well as the components $\Psi_u \left(u = 1, 2, \cdots, \Omega \right)$ will depend on how to extract the desired information from the amino acid sequence of $\mathbf{P}$, as elaborated in [4]. Thus, by following exactly the same procedures as described in the Section 2.2 of [4], each of the protein samples in the benchmark dataset can be uniquely defined as a 22-D numerical vector as given in columns 3 - 24 of Supporting Information S2, which can also be directly downloaded at http://www.jci-bioinfo.cn/pLoc_bal-mEuk/Supp2.pdf.

2.3. Installing Deep-Learning for Three Deeper Levels

In this study, a dense neural network with 3 fully connected layers was used to predict subcellular localization of multi-label eukaryotic proteins, as illustrated in **Figure 1**. The predicted results were decided by the output of the threshold θ. If the output is greater than 0.5, the outcome was true; otherwise, false. For more information about this, see [11], where the details have been clearly elaborated and hence there is no need to repeat here.

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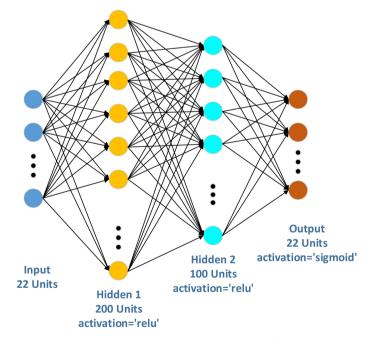

Figure 1. An illustration to show a dense neural network with 3 fully connected layers. Adapted from [11] with permission.

3. RESULTS AND DISCUSSION

According to the 5-step rules [58], 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. To make them more intuitive and easier to understand for most experimental scientists, here we use the following intuitive Chou's five metrics [239] that have recently been widely used for studying various multi-label systems (see, e.g., [240, 241]):

$$
\begin{cases}
\text{Aiming} \uparrow = \dfrac{1}{N^{q}} \sum_{k=1}^{N^{q}} \left(\dfrac{\left\| \mathbb{L}_{k} \cap \mathbb{L}_{k}^{*} \right\|}{\left\| \mathbb{L}_{k}^{*} \right\|} \right), [0,1] \\[4mm]
\text{Coverage} \uparrow = \dfrac{1}{N^{q}} \sum_{k=1}^{N^{q}} \left(\dfrac{\left\| \mathbb{L}_{k} \cap \mathbb{L}_{k}^{*} \right\|}{\left\| \mathbb{L}_{k} \right\|} \right), [0,1] \\[4mm]
\text{Accuracy} \uparrow = \dfrac{1}{N^{q}} \sum_{k=1}^{N^{q}} \left(\dfrac{\left\| \mathbb{L}_{k} \cap \mathbb{L}_{k}^{*} \right\|}{\left\| \mathbb{L}_{k} \cup \mathbb{L}_{k}^{*} \right\|} \right), [0,1] \\[4mm]
\text{Absolute true} \uparrow = \dfrac{1}{N^{q}} \sum_{k=1}^{N^{q}} \Delta\left(\mathbb{L}_{k}, \mathbb{L}_{k}^{*} \right), [0,1] \\[4mm]
\text{Absolute false} \downarrow = \dfrac{1}{N^{q}} \sum_{k=1}^{N^{q}} \left(\dfrac{\left\| \mathbb{L}_{k} \cup \mathbb{L}_{k}^{*} \right\| - \left\| \mathbb{L}_{k} \cap \mathbb{L}_{k}^{*} \right\|}{M} \right), [1,0]
\end{cases}
\tag{4}
$$

where N^{q} is the total number of query proteins or tested proteins, M is the total number of different labels for the investigated system (for the current study it is $L_{\text{cell}} = 22$), $\| \ \|$ means the operator acting on the set therein to count the number of its elements, $\bigcup$ means the symbol for the "union" in the set theory, $\bigcap$ denotes the symbol for the "intersection", $\mathbb{L}_{k}$ denotes the subset that contains all the labels observed by experiments for the k-th tested sample, $\mathbb{L}_{k}^{*}$ represents the subset that contains all the labels predicted for the k-th sample, and

$$
\Delta\left(\mathbb{L}_{k}, \mathbb{L}_{k}^{*} \right) =
\begin{cases}
1, \text{if all the labels in } \mathbb{L}_{k}^{*} \text{ are identical to those in } \mathbb{L}_{k} \\
0, \text{otherwise}
\end{cases}
\tag{5}
$$

In Equation (4), the first four metrics with an upper arrow $\uparrow$ are called positive metrics, meaning that the larger the rate is the better the prediction quality will be; the 5th metrics with a down arrow $\downarrow$ is called negative metrics, implying just the opposite meaning.

From Equation (4) we can see the following: 1) the "Aiming" defined by the 1st sub-equation is for checking the rate or percentage of the correctly predicted labels over the practically predicted labels; 2) the "Coverage" defined in the 2nd sub-equation is for checking the rate of the correctly predicted labels over the actual labels in the system concerned; 3) the "Accuracy" in the 3rd sub-equation is for checking the average ratio of correctly predicted labels over the total labels including correctly and incorrectly predicted labels as well as those real labels but are missed in the prediction; 4) the "Absolute true" in the 4th sub-equation is for checking the ratio of the perfectly or completely correct prediction events over the total prediction events; 5) the "Absolute false" in the 5th sub-equation is for checking the ratio of the completely wrong prediction over the total prediction events.

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

Listed in Table 1 are the rates achieved by the current pLoc_Deep-mEuk predictor via the cross validations on the same experiment-confirmed dataset as used in [4]. For facilitating comparison, listed there are also the corresponding results obtained by the pLoc_bal-mEuk [4], the existing most powerful predictor for identifying the subcellular localization of eukaryotic proteins with both single and multiple location sites. As shown in Table 1, the newly proposed predictor pLoc_Deep-mEuk is remarkably superior to the existing state-of-the-art predictor pLoc_bal-mEuk 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 [4]. 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., [91, 178, 184]).

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 intuitive metrics that were derived in [242] based on the Chou's symbols introduced for studying protein signal peptides [243] and that have ever since been widely concurred or justified (see, e.g., [242, 244]). For the current study, the set of metrics can be formulated as:

$$
\begin{cases}
Sn(i) = 1 - \dfrac{N_-^+(i)}{N^+(i)} & 0 \le Sn \le 1 \\[2mm]
Sp(i) = 1 - \dfrac{N_+^-(i)}{N^-(i)} & 0 \le Sp \le 1 \\[2mm]
Acc(i) = 1 - \dfrac{N_-^+(i) + N_+^-(i)}{N^+(i) + N^-(i)} & 0 \le Acc \le 1 \\[2mm]
MCC(i) = \dfrac{1 - \left(\dfrac{N_-^+(i)}{N^+(i)} + \dfrac{N_+^-(i)}{N^-(i)} \right)}{\sqrt{\left(1 + \dfrac{N_+^-(i) - N_-^+(i)}{N^+(i)}\right)\left(1 + \dfrac{N_-^+(i) - N_+^-(i)}{N^-(i)}\right)}} & -1 \le MCC \le 1 \\[2mm]
(i = 1, 2, \cdots, 22)
\end{cases}
\tag{6}
$$

Table 1. Comparison with the state-of-the-art method in predicting eukaryotic protein subcellular localizationa.

Predictor	Aiming (↑)[a]	Coverage (↑)[a]	Accuracy (↑)[a]	Absolute true (↑)[a]	Absolute false (↓)[a]
pLoc_bal-mEuk[b]	88.31%	85.06%	84.34%	78.78%	0.07%
pLoc_Deep-Euk[c]	89.31%	90.82%	87.28%	81.71%	0.00%

[a]See Equation (4) for the definition of the metrics. [b]See [4], 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. [c]The proposed predictor; to assure that the test was performed on exactly the same experimental data as reported in [4] for pLoc_bal-mEuk.

Table 2. Performance of pLoc_Deep-mEuk for each of the 22 subcellular locations.

i	Location[a]	Sn(i)[b]	Sp(i)[b]	Acc(i)[b]	MCC(i)[b]
1	Acrosome	0.9000	1.000	0.9997	0.9463
2	Cell membrane	0.9977	0.9950	0.9891	0.9492
3	Cell wall	0.9460	0.9993	0.9988	0.9333
4	Centrosome	0.8880	0.9983	0.9964	0.8932
5	Chloroplast	0.9847	0.9995	0.9985	0.9879
6	Cyanelle	1.0000	1.0000	1.0000	1.0000
7	Cytoplasm	0.8808	0.9627	0.9462	0.8340
8	Cytoskeleton	0.8718	0.9979	0.9945	0.8923
9	Endoplasmic reticulum	0.9687	0.9976	0.9953	0.9679
10	Endosome	0.8640	0.9997	0.9986	0.9105
11	Extracell	0.9818	0.9955	0.9931	0.9758
12	Golgi apparatus	0.9213	0.9971	0.9935	0.9276
13	Hydrogenosome	1.0000	1.0000	1.0000	1.0000
14	Lysosome	0.9505	0.9998	0.9993	0.9675
15	Melanosome	0.9449	0.9998	0.9993	0.9632
16	Microsome	0.8114	0.9997	0.9991	0.8451
17	Mitochondrion	0.9622	0.9967	0.9932	0.9627
18	Nucleus	0.9085	0.9722	0.9590	0.9515
19	Peroxisome	0.9470	0.9992	0.9982	0.9515
20	Spindle pole body	0.9500	0.9995	0.9989	0.9530
21	Synapse	0.9423	0.9995	0.9989	0.9458
22	Vacuole	0.9510	0.9994	0.9980	0.9649

[a]See Table 1 and the relevant context for further explanation. [b]See Equation (6) for the metrics definition.

where Sn, Sp, Acc, and MCC represent the sensitivity, specificity, accuracy, and Mathew's correlation coefficient, respectively (Chen et al., 2007), and i denotes the i-th subcellular location (or subset) in the benchmark dataset. $N^+(i)$ is the total number of the samples investigated in the i-th subset, whereas $N^+_-(i)$ is the number of the samples in $N^+(i)$ that are incorrectly predicted to be of other locations; $N^-(i)$ is the total number of samples in any locations but not the i-th location, whereas $N^-_+(i)$ is the number of the samples in $N^-(i)$ that are incorrectly predicted to be of the i-th location.

Listed in Table 2 are the results achieved by pLoc_Deep-mEuk for the eukaryotic 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 eukaryotic 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.

3.3. Web Server and User Guide

As pointed out in [245], 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., [219, 220, 234, 246-300]) will significantly enhance the impacts of theoretical work because they can attract the broad experimental scientists [301]. In view of this, the web-server of the current pLoc_Deep-mEuk predictor has also been established. Moreover, to maximize users' convenience, a step-by-step guide is given below.

Step 1. Click the link at http://www.jci-bioinfo.cn/pLoc_Deep-mEuk/, the top page of the pLoc_Deep-mEukweb-server will appear on your computer screen, as shown in **Figure 2**. Click on the Read Me button to see a brief introduction about the predictor.

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

Step 3. Click on the Submit button to see the predicted result. For instance, if you use the four protein sequences in the Example window as the input, after 10 seconds or so, you will see a new screen (**Figure 3**) occurring. On its upper part are listed the names of the subcellular locations numbered from (1) to (22) covered by the current predictor. On its lower part are the predicted results: the query protein Q63564 of example-1 corresponds to "1," meaning it belonging to "Acrosome" only; the query protein P23276 of example-2 corresponds to "2, 8" meaning it belonging to "Cell membrane" and "Cytoskeleton"; the query protein Q9VVV9 of example-3 corresponds to "2, 7, 18", meaning it belonging to "Cell membrane", "Cytoplasm", and "Nucleus"; the query protein Q673G8 of example-4 corresponds to "2, 7, 10, 18", meaning it belonging to "Cell membrane", "Cytoplasm", "Endosome", and "Nucleus". All these results are perfectly consistent with experimental observations.

Step 4. As shown on the lower panel of **Figure 2**, you may also choose the batch prediction by entering your e-mail address and your desired batch input file (in FASTA format of course) via the Browse button. To see the sample of batch input file, click on the button Batch-example. After clicking the button Batch-submit, you will see "Your batch job is under computation; once the results are available, you will be notified by e-mail".

Figure 2. A semi screenshot for the top page of pLoc_Deep-mEuk.

Figure 3. A semi screenshot for the webpage obtained by following Step 2 of Section 3.3.

Step 5. Click on the Citation button to find the papers that have played the key role in developing the current predictor of pLoc_Deep-mEuk.

Step 6. Click the Supporting Information button to download the Supporting Informations mentioned in this paper.

4. CONCLUSION

It is anticipated that the pLoc_Deep-Euk predictor holds very high potential to become a useful high throughput tool in identifying the subcellular localization of eukaryotic proteins, particularly for finding multi-target drugs that is currently a very hot trend in drug development.

ACKNOWLEDGEMENTS

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

croscopy. *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. Chou, K.C., Cheng, X. and Xiao, X. (2019) pLoc_bal-mEuk: Predict Subcellular Localization of Eukaryotic Proteins by General PseAAC and Quasi-Balancing Training Dataset. *Medicinal Chemistry*, **15**, 472-485. https://doi.org/10.2174/1573406415666181218102517

5. Nakai, K. and Kanehisa, M. (1992) A Knowledge Base for Predicting Protein Localization Sites in Eukaryotic Cells. *Genomics*, **14**, 897-911. https://doi.org/10.1016/S0888-7543(05)80111-9

6. Cedano, J., Aloy, P., Perez-Pons, J.A. and Querol, E. (1997) Relation between Amino Acid Composition and Cellular Location of Proteins. *Journal of Molecular Biology*, **266**, 594-600. https://doi.org/10.1006/jmbi.1996.0804

7. Reinhardt, A. and Hubbard, T. (1998) Using Neural Networks for Prediction of the Subcellular Location of Proteins. *Nucleic Acids Research*, **26**, 2230-2236. https://doi.org/10.1093/nar/26.9.2230

8. 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

9. Chou, K.C., Wu, Z.C. and Xiao, X. (2011) iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular Localization of Singleplex and Multiplex Eukaryotic Proteins. *PLoS ONE*, **6**, e18258. https://doi.org/10.1371/journal.pone.0018258

10. 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

11. 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**, Article No. 523. https://doi.org/10.1186/s12859-017-1898-z

12. 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. *International Journal of Peptide Research and Therapeutics*, **26**, 795-809. https://doi.org/10.1007/s10989-019-09887-3

13. 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*, **189**, 169-180. https://doi.org/10.1016/j.chemolab.2019.05.003

14. 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*), **193**, Article ID: 103811. https://doi.org/10.1016/j.chemolab.2019.103811

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. 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

17. 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

18. 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

19. Chen, Y. and Fan, X. (2020) Use Chou's 5-Steps Rule to Reveal Active Compound and Mechanism of Shuangsheng Pingfei San on Idiopathic Pulmonary Fibrosis. *Current Molecular Medicine*, **20**, 220-230. https://doi.org/10.2174/1566524019666191011160543

20. 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

21. Dutta, A., Dalmia, A., R, 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. *Computer in Biology Medicine*, **116**, Article ID: 103558. https://doi.org/10.1016/j.compbiomed.2019.103558

22. 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

23. 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

24. 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

25. Jun, Z. and Wang, S.Y. (2019) Identify Lysine Neddylation Sites Using Bi-Profile Bayes Feature Extraction via the Chou's 5-Steps Rule and General Pseudo Components. *Current Genomics*, **20**, 592-601. https://doi.org/10.2174/1389202921666191223154629

26. Lan, J., Liu, J., Liao, C., Merkler, 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*, **19**, 2318-2333. http://www.eurekaselect.com/175887/article https://doi.org/10.2174/1568026619666191019111528

27. 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*, **294**, 1173-1182. https://doi.org/10.1007/s00438-019-01570-y

28. 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

29. 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

30. 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 Topics in Medicinal Chemistry*, **19**, 2301-2317. https://doi.org/10.2174/1568026619666191016155543

31. 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

32. Liu, Z., Dong, W. Jiang, W. and He, Z.L. (2019) csDMA: An Improved Bioinformatics Tool for Identifying DNA 6 mA Modifications via Chou's 5-Step Rule. *Scientific Reports*, **9**, Article ID: 13109. https://doi.org/10.1038/s41598-019-49430-4

33. 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

34. 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

35. 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. *Chemometrics and Intelligent Laboratory Systems*, **189**, 96-101. https://doi.org/10.1016/j.chemolab.2019.04.007

36. 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*, **25**, 3769-3775. https://doi.org/10.2174/1381612825666190930101258

37. Yang, L., Lv, Y., Wang, S.Y., Q., Zhang, Y., Pan, D., Su, Q. and Lu, Zuo, Y. (2019) 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

38. Akbar, S., Rahman, A.U., Hayat, M. and Sohail, M. (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

39. 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*, in press. https://doi.org/10.1109/TCBB.2020.2968441

40. 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

41. 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. (Correction on 2020, Vol. 21, No. 7, 2629) https://doi.org/10.3390/ijms21010075

42. 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

43. 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

44. 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

45. Dutta, A., Dalmia, A., Athul R., 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. *Computer in Biology and Medicine*, **116**, Article ID: 103558. https://doi.org/10.1016/j.compbiomed.2019.103558

46. 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

47. 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

48. 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

49. 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

50. 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

51. Vundavilli, H., Datta, A., Sima, C., Hua, J.P. Lopes, R. and Bittner, M.L. (2020) Using Chou's 5-Steps Rule to Model Feedback in Lung Cancer. *IEEE Journal of Biomedical and Health Informatics*. https://doi.org/10.1109/JBHI.2019.2958042

52. Yang, L., Lv, Y., Wang, S., Zhang, Q., Pan, Y., Su, D., Lu, Q. and Zuo, Y.C. (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

53. Xuan, P., Cui, H., Shen, T. Sheng, N. and Zhang, T. (2019) HeteroDualNet: A Dual Convolutional Neural Network With Heterogeneous Layers for Drug-Disease Association Prediction via Chou's Five-Step Rule. *Frontiers in Pharmacology*, **10**, 1301. https://doi.org/10.3389/fphar.2019.01301

54. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mPlant Is Working.

55. Zou, Y., Ding, Y.J., Tang, J.J., Guo, F. and Peng, L. (2019) FKRR-MVSF: A Fuzzy Kernel Ridge Regression Model for Identifying DNA-Binding Proteins by Multi-View Sequence Features via Chou's Five-Step Rule. *International Journal of Molecular Sciences*, **20**, 4175. https://doi.org/10.3390/ijms20174175

56. 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

57. Mohabatkar, H., Ebrahimi, S. and Moradi, M. (2020) Using Chou's Five-Steps Rule to Classify and Predict Glutathione S-Transferases with Different Machine Learning Algorithms and Pseudo Amino Acid Composition. *International Journal of Peptide Research and Therapeutics*. https://doi.org/10.1007/s10989-020-10087-7

58. Chou, K.C. (2011) Some Remarks on Protein Attribute Prediction and Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **273**, 236-247. https://doi.org/10.1016/j.jtbi.2010.12.024

59. 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

60. 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

61. 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

62. 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

63. 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

64. 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

65. 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

66. 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

67. Lin, H., Ding, F.B., 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

68. 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

69. 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

70. 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*, **355**, 591-598. https://doi.org/10.1007/s00726-008-0086-x

71. 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

72. 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

73. 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

74. 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 Anal-

ysis. *Amino Acids*, **37**, 415-425. https://doi.org/10.1007/s00726-008-0170-2

75. 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

76. 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

77. 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

78. Esmaeili, 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

79. 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

80. 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

81. 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

82. 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

83. 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

84. 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

85. 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

86. 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.

87. Mohabatkar, H., Mohammad Beigi, M. and Esmaedi, A. (2011) Esmaeili, 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

88. 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

89. 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 Com-

position. *Journal of Molecular Graphics & Modelling*, **30**, 129-134. https://doi.org/10.1016/j.jmgm.2011.06.014

90. 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

91. 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

92. 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

93. Du, P., Wang, X., Xu, C. and Gao, Y. (2012) PseAAC-Builder: A Cross-Platform Stand-Alone Program for Generating Various Special Chou's Pseudo Amino Acid Compositions. *Analytical Biochemistry*, **425**, 117-119. https://doi.org/10.1016/j.ab.2012.03.015

94. 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

95. 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

96. 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

97. 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

98. 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

99. 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

100. 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

101. 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

102. Zhao, X.W., Ma, Z.Q. and Yin, M.H. (2012) Predicting Protein-Protein Interactions by Combing 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

103. Zia-ur-Rehman and Khan, A. (2012) Identifying GPCRs and Their Types with Chou's Pseudo Amino Acid Composition: An Approach from Multi-Scale Energy Representation and Position Specific Scoring Matrix. *Protein & Peptide Letters*, **19**, 890-903. https://doi.org/10.2174/092986612801619589

104. 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

105. 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

106. 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

107. 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

108. 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. *Biosystems*, **113**, 50-57. https://doi.org/10.1016/j.biosystems.2013.04.005

109. 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

110. 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

111. 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

112. 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

113. 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

114. 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

115. 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

116. 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 and Peptide Letters*, **20**, 781-795. https://doi.org/10.2174/0929866511320070008

117. 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

118. 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

119. Niu, X.H., Li, N.N., Xia, J.B., Chen, D.Y., Peng, Y.H., Xiao, Y., Wei, W.Q., Wang, D.M. and Wang, Z.Z. (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

120. 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

121. 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

122. 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

123. 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

124. 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

125. 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

126. 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

127. 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

128. 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

129. 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

130. 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 and Peptide Letters*, **22**, 547-556. https://doi.org/10.2174/0929866522666150209151344

131. 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*, **36**, 197-203. https://doi.org/10.1016/j.jtbi.2014.10.014

132. 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

133. 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

134. 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

135. 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

136. 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

137. Huo, H., Li, T., Wang, S., Lv, Y., Zuo, Y. and Yang, L. (2017) Prediction of Presynaptic and Postsynaptic Neurotoxins by Combining Various Chou's Pseudo Components. *Scientific Reports*, **7**, Article No. 5827. https://doi.org/10.1038/s41598-017-06195-y

138. 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

139. 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

140. 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.jmgm.2017.07.012

141. Al Maruf, M.A. and Shatabda, S. (2019) iRSpot-SF: Prediction of Recombination Hotspots by Incorporating Sequence Based Features into Chou's Pseudo Components. *Genomics*, **111**, 966-972. https://doi.org/10.1016/j.ygeno.2018.06.003

142. 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

143. Cui, X., Yu, Z., Yu, B., Wang, M., Tian, B. and Ma, Q. (2018) UbiSitePred: A Novel Method for Improving the Accuracy of Ubiquitination Sites Prediction by Using LASSO to Select the Optimal Chou's Pseudo Components. *Chemometrics and Intelligent Laboratory Systems*, **184**, 28-43. https://doi.org/10.1016/j.chemolab.2018.11.012

144. 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

145. Qiu, W., Li, S., Cui, X., Yu, Z., Wang, M., Du, J., Peng, Y. and Yu, B. (2018) Predicting Protein Submitochondrial Locations by Incorporating the Pseudo-Position Specific Scoring Matrix into the General Chou's Pseudo-Amino Acid Composition. *Journal of Theoretical Biology*, **450**, 86-103. https://doi.org/10.1016/j.jtbi.2018.04.026

146. 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

147. Zhang, S., Yang, K., Lei, Y. and Song, K. (2018) iRSpot-DTS: Predict Recombination Spots by Incorporating the

Dinucleotide-Based Spare-Cross Covariance Information into Chou'S Pseudo Components. *Genomics*, **11**, 457-464.

148. Zhao, W., Wang, L., Zhang, T.X., Zhao, Z.N. and Du, P.F. (2018) A Brief Review on Software Tools in Generating Chou's Pseudo-Factor Representations for All Types of Biological Sequences. *Protein and Peptide Letters*, **25**, 822-829. https://doi.org/10.2174/0929866525666180905111124

149. Al Maruf, M.A. and Shatabda, S. (2019) iRSpot-SF: Prediction of Recombination Hotspots by Incorporating Sequence Based Features into Chou's Pseudo Components. *Genomics*, **111**, 966-972. https://doi.org/10.1016/j.ygeno.2018.06.003

150. 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://www.sciencedirect.com/science/article/pii/S1567576919321277 https://doi.org/10.1016/j.intimp.2019.106020

151. Pan, Y., Wang, S., Zhang, Q., Lu, Q., Su, D., Zuo, Y. and Yang, L. (2019) Analysis and Prediction of Animal Toxins by Various Chou's Pseudo Components and Reduced Amino Acid Compositions. *Journal of Theoretical Biology*, **462**, 221-229. https://doi.org/10.1016/j.jtbi.2018.11.010

152. Tahir, M., Tayara, H. and Chong, K.T. (2019) iRNA-PseKNC(2methyl): Identify RNA 2'-O-Methylation Sites by Convolution Neural Network and Chou's Pseudo Components. *Journal of Theoretical Biology*, **465**, 1-6. https://doi.org/10.1016/j.jtbi.2018.12.034

153. Tian, B., Wu, X., Chen, C., Qiu, W., Ma, Q. and Yu, B. (2019) Predicting Protein-Protein Interactions by Fusing Various Chou's Pseudo Components and Using Wavelet Denoising Approach. *Journal of Theoretical Biology*, **462**, 329-346. https://doi.org/10.1016/j.jtbi.2018.11.011

154. Zhang, L. and Kong, L. (2019) iRSpot-PDI: Identification of Recombination Spots by Incorporating Dinucleotide Property Diversity Information into Chou's Pseudo Components. *Genomics*, **111**, 457-464. https://doi.org/10.1016/j.ygeno.2018.03.003

155. Zhang, S., Yang, K., Lei, Y. and Song, K. (2019) iRSpot-DTS: Predict Recombination Spots by Incorporating the Dinucleotide-Based Spare-Cross Covariance Information into Chou's Pseudo Components. *Genomics*, **111**, 1760-1770. https://doi.org/10.1016/j.ygeno.2018.11.031

156. 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

157. 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*, **294**, 199-210. https://doi.org/10.1007/s00438-018-1498-2

158. 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

159. 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

160. 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

161. 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

162. 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

163. Javed, F. and Hayat M. (2019) Predicting Subcellular Localizations 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

164. 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*, **6**, 66545-66556.

165. 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

166. 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

167. 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

168. 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

169. 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

170. 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

171. 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

172. 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

173. 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

174. 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

175. 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

176. Han, G.S., Yu, Z.G. and Anh, V. (2014) A Two-Stage SVM Method to Predict Membrane Protein Types by In-

corporating 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

177. 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

178. 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

179. 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

180. 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

181. 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

182. 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

183. 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

184. 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

185. Mei, S. (2012) Multi-Kernel Transfer Learning Based on Chou's PseAAC Formulation for Protein Submitochondria Localization. *Journal of Theoretical Biology*, **293**, 121-130. https://doi.org/10.1016/j.jtbi.2011.10.015

186. 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

187. 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

188. 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

189. 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

190. Ilyas, S., Hussain, W., Ashraf, A., Khan, Y.D., Khan, S.A. and Chou, K.C. (2019) iMethylK-PseAAC: Improving Accuracy for Lysine Methylation Sites Identification by Incorporating Statistical Moments and Position Relative Features into General PseAAC via Chou's 5-Steps Rule. *Current Genomics*, **20**, 275-292.

https://doi.org/10.2174/1389202920666190809095206

191. 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

192. 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

193. 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

194. 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

195. 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

196. 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

197. 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

198. 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

199. 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

200. Cheng, X., Xiao, X. and Chou, K.C. (2018) pLoc_bal-mPlant: Predict Subcellular Localization of Plant Proteins by General PseAAC and Balancing Training Dataset. *Current Pharmaceutical Design*, **24**, 4013-4022. https://doi.org/10.2174/1381612824666181119145030

201. Cheng, X., Xiao, X. and Chou, K.C. (2018) pLoc_bal-mGneg: Predict Subcellular Localization of Gram-Negative Bacterial Proteins by Quasi-Balancing Training Dataset and General PseAAC. *Journal of Theoretical Biology*, **458**, 92-102. https://doi.org/10.1016/j.jtbi.2018.09.005

202. Cheng, X., Xiao, X. and Chou, K.C. (2018) pLoc-mHum: Predict Subcellular Localization of Multi-Location Human Proteins via General PseAAC to Winnow out the Crucial GO Information. *Bioinformatics*, **34**, 1448-1456. https://doi.org/10.1093/bioinformatics/btx711

203. Cheng, X., Xiao, X. and Chou, K.C. (2018) pLoc-mGneg: Predict Subcellular Localization of Gram-Negative Bacterial Proteins by Deep Gene Ontology Learning via General PseAAC. *Genomics*, **110**, 231-239. https://doi.org/10.1016/j.ygeno.2017.10.002

204. Cheng, X., Xiao, X. and Chou, K.C. (2018) pLoc-mEuk: Predict Subcellular Localization of Multi-Label Eukaryotic Proteins by Extracting the Key GO Information into General PseAAC. *Genomics*, **110**, 50-58. https://doi.org/10.1016/j.ygeno.2017.08.005

205. 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*, **45**, 2295-2306. https://doi.org/10.1007/s11033-018-4391-5

206. Xu, C., Ge, L., Zhang, Y., Dehmer, M. and Gutman, I. (2017) Computational Prediction of Therapeutic Peptides Based on Graph Index. *Journal of Biomedical Informatics*, **75**, 63-69. https://doi.org/10.1016/j.jbi.2017.09.011

207. Xiao, X., Cheng, X., Su, S., Nao, Q. and Chou, K.C. (2017) pLoc-mGpos: Incorporate Key Gene Ontology Information into General PseAAC for Predicting Subcellular Localization of Gram-Positive Bacterial Proteins. *Natural Science*, **9**, 331-349. https://doi.org/10.4236/ns.2017.99032

208. 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

209. Qiu, W.R., Sun, B.Q., Xiao, X., Xu, D. and Chou, K.C. (2017) iPhos-PseEvo: Identifying Human Phosphorylated Proteins by Incorporating Evolutionary Information into General PseAAC via Grey System Theory. *Molecular Informatics*, **36**, UNSP 1600010. https://doi.org/10.1002/minf.201600010

210. 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

211. 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. *Medicinal Chemistry*, **13**, 552-559. https://doi.org/10.2174/1573406413666170515120507

212. 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

213. 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

214. Cheng, X., Xiao, X. and Chou, K.C. (2017) pLoc-mVirus: Predict Subcellular Localization of Multi-Location Virus Proteins via Incorporating the Optimal GO Information into General PseAAC. *Gene*, **628**, 315-321. (Erratum: ibid., 2018, Vol. 644, 156-156) https://doi.org/10.1016/j.gene.2017.07.036

215. Cheng, X., Xiao, X. and Chou, K.C. (2017) pLoc-mPlant: Predict Subcellular Localization of Multi-Location Plant Proteins via Incorporating the Optimal GO Information into General PseAAC. *Molecular BioSystems*, **13**, 1722-1727. https://doi.org/10.1039/C7MB00267J

216. 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

217. Qiu, W.R., Sun, B.Q., Xiao, X., Xu, Z.C. and Chou, K.C. (2016) iHyd-PseCp: Identify Hydroxyproline and Hydroxylysine in Proteins by Incorporating Sequence-Coupled Effects into General PseAAC. *Oncotarget*, **7**, 44310-44321. https://doi.org/10.18632/oncotarget.10027

218. 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

219. 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

220. Jia, J., Liu, Z., Xiao, X., Liu, B. and Chou, K.C. (2016) iCar-PseCp: Identify Carbonylation Sites in Proteins by Monto Carlo Sampling and Incorporating Sequence Coupled Effects into General PseAAC. *Oncotarget*, **7**, 34558-34570. https://doi.org/10.18632/oncotarget.9148

221. 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*, **142**, 28-35. https://doi.org/10.1016/j.chemolab.2015.01.004

222. 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*, **146**, 232-240. https://doi.org/10.1016/j.chemolab.2015.05.028

223. 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

224. Ju, Z., Cao, J.Z. and Gu, H. (2015) iLM-2L: A Two-Level Predictor for Identifying Protein Lysine Methylation Sites and Their Methylation Degrees by Incorporating K-Gap Amino Acid Pairs into Chous General PseAAC. *Journal of Theoretical Biology*, **385**, 50-57. https://doi.org/10.1016/j.jtbi.2015.07.030

225. 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

226. 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

227. 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

228. 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

229. 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

230. 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

231. Chen, W., Feng, P.M., Lin, H. and Chou, K.C. (2014) iSS-PseDNC: Identifying Splicing Sites Using Pseudo Dinucleotide Composition. *Biomed Research International*, **2014**, Article ID: 623149. https://doi.org/10.1155/2014/623149

232. 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.

233. 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

234. Liu, B., Long, R. and Chou, K.C. (2016) iDHS-EL: Identifying DNase I Hypersensi-Tivesites 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

235. 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

236. 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

237. 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

238. 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

239. Chou, K.C. (2013) Some Remarks on Predicting Multi-Label Attributes in Molecular Biosystems. *Molecular Biosystems*, **9**, 1092-1100. https://doi.org/10.1039/c3mb25555g

240. Cheng, X., Zhao, S.G., Xiao, X. and Chou, K.C. (2017) iATC-mISF: A Multi-Label Classifier for Predicting the Classes of Anatomical Therapeutic Chemicals. *Bioinformatics*, **33**, 341-346. (Corrigendum, ibid., 2017, Vol. 33, 2610) https://doi.org/10.1093/bioinformatics/btx387

241. Cheng, X., Zhao, S.G., Xiao, X. and Chou, K.C. (2017) iATC-mHyb: A Hybrid Multi-Label Classifier for Predicting the Classification of Anatomical Therapeutic Chemicals. *Oncotarget*, **8**, 58494-58503. https://doi.org/10.18632/oncotarget.17028

242. 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

243. Chou, K.C. (2001) Using Subsite Coupling to Predict Signal Peptides. *Protein Engineering, Design and Selection*, **14**, 75-79. https://doi.org/10.1093/protein/14.2.75

244. 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*, **34**, 1946-1961. https://doi.org/10.1080/07391102.2015.1095116

245. 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

246. Song, J., Li, F., Leier, A., Marquez-Lago, T.T., Akutsu, T., Haffari, G., Chou, K.C., Webb, G.I. and Pike, R.N. (2018) PROSPERous: High-Throughput Prediction of Substrate Cleavage Sites for 90 Proteases with Improved Accuracy. *Bioinformatics*, **34**, 684-687. https://doi.org/10.1093/bioinformatics/btx670

247. 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

248. 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. *Medicinal Chemistry*, **13**, 544-551. https://doi.org/10.2174/1573406413666170419150052

249. 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

250. 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

251. 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

252. 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

253. 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

254. Xu, Y., Ding, J., Huang, Q. and Deng, N.Y. (2013) Prediction of Protein Methylation Sites Using Conditional Random Field. *Protein & Peptide Letters*, **20**, 71-77. https://doi.org/10.2174/092986613804096865

255. 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

256. 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

257. 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

258. 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**, 95. https://doi.org/10.3390/molecules21010095

259. 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

260. 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

261. 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

262. 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

263. 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

264. 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

265. 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

266. Liu, B., Fang, L., Liu, F., Wang, X. and Chou, K.C. (2016) iMiRNA-PseDPC: microRNA Precursor Identification with a Pseudo Distance-Pair Composition Approach. *Journal of Biomolecular Structure and Dynamics*, **34**, 223-235. https://doi.org/10.1080/07391102.2015.1014422

267. 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

268. 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

269. 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

270. Liu, B. (2019) BioSeq-Analysis: A Platform for DNA, RNA, and Protein Sequence Analysis Based on Machine Learning Approaches. *Briefings in Bioinformatics*, **20**, 1280-1294. https://doi.org/10.1093/bib/bbx165

271. 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

272. 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

273. 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

274. Liu, B., Gao, X. and Zhang, H. (2019) BioSeq-Analysis2.0: An Updated Platform for Analyzing DNA, RNA and Protein Sequences at Sequence Level and Residue Level Based on Machine Learning Approaches. *Nucleic Acids Research*, **47**, e127. https://doi.org/10.1093/nar/gkz740

275. 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

276. 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

277. 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

278. 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

279. 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

280. 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

281. Lin, H. and Li, Q.Z. (2007) Predicting Conotoxin Superfamily and Family by Using Pseudo Amino Acid Composition and Modified Mahalanobis Discriminant. *Biochemical and Biophysical Research Communications*, **354**, 548-551. https://doi.org/10.1016/j.bbrc.2007.01.011

282. Lin, H. and Li, Q.Z. (2007) Using Pseudo Amino Acid Composition to Predict Protein Structural Class: Ap-

proached by Incorporating 400 Dipeptide Components. *Journal of Computational Chemistry*, **28**, 1463-1466. https://doi.org/10.1002/jcc.20554

283. Lin, H. and Ding, H. (2011) Predicting Ion Channels and Their Types by the Dipeptide Mode of Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **269**, 64-69. https://doi.org/10.1016/j.jtbi.2010.10.019

284. Lin, H., Chen, W. and Ding, H. (2013) AcalPred: A Sequence-Based Tool for Discriminating between Acidic and Alkaline Enzymes. *PLoS ONE*, **8**, e75726. https://doi.org/10.1371/journal.pone.0075726

285. Lin, H., Ding, C., Yuan, L.F., Chen, W., Ding, H., Li, Z.Q., Guo, F.B., Hung, 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 No. 1350003. https://doi.org/10.1142/S1793524513500034

286. 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

287. Shao, Y.T. and Chou, K.C. (2020) pLoc_Deep-mAnimal: A Novel Deep CNN-BLSTM Network to Predict Subcellular Localization of Animal Proteins. *Natural Science*, **12**, 281-291. https://doi.org/10.4236/ns.2020.125024

288. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mAnimal Is Working. *American Journal of Virology & Disease*, **2**, 1-2.

289. Shao, Y.T., Liu, X.X., Lu, Z. and Chou, K.C. (2020) pLoc_Deep-mPlant: Predict Subcellular Localization of Plant Proteins by Deep Learning. *Natural Science*, **12**, 237-247. https://doi.org/10.4236/ns.2020.125021

290. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mPlant Is Working. *Growth Journal*, **1**, 1-2.

291. Lu, Z. and Chou, K.C. (2020) iATC_Deep-mISF: A Multi-Label Classifier for Predicting the Classes of Anatomical Therapeutic Chemicals by Deep Learning. *Advances in Bioscience and Biotechnology*, **11**, 153-159. https://doi.org/10.4236/abb.2020.115012

292. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server iATC_Deep-mISF Is Working. *Global Journal of Science Frontier Research*: *G Bio-Tech & Genetics*, **20**, 1-3.

293. Liu, X.X. and Chou, K.C. (2020) pLoc_Deep-mGneg: Predict Subcellular Localization of Gram Negative Bacterial Proteins by Deep Learning. *Advances in Bioscience and Biotechnology*, **11**, 141-152. https://doi.org/10.4236/abb.2020.115011

294. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mGneg Is Working. *Clin Med Case Rep J*, **1**, 1-2.

295. Shao, Y.T., Cheng, X. and Chou, K.C. (2020) pLoc_Deep-mVirus: A CNN Model for Predicting Subcellular Localization of Virus Proteins by Deep Learning. *Molecular Therapy—Nucleic Acids* (*MTNA*), in press.

296. Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mVirus Is Working. *Clinical Research and Trials*, **6**, 1-2.

297. Lu, Z. and Chou, K.C. (2020) Showcase to Illustrate How the Web-Server pLoc_Deep-mGpos Is Working. *Journal of Biomedical Science and Engineering*, **13**, 55-65. https://doi.org/10.4236/jbise.2020.135005

298. Chou, K.C. (2020) Showcase to Illustrate How the Webserver pLoc_Deep-mGpos Is Working. *Open Acc J Bio Sci*, **2**, 345-346.

299. Chou, K.C. (2020) How the Artificial Intelligence Tool iSuc-PseOpt Is Working for Predicting Lysine Succinylation Sites in Proteins. *Biomedical Research and Clinical Reviews*, **1**, 1-2.

300. 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

301. 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