



# Artificial Intelligence (AI) Tools Constructed via the 5-Steps Rule for Predicting Post-Translational Modifications

Kuo-Chen Chou<sup>1,2\*</sup>

<sup>1</sup>Gordon Life Science Institute, Massachusetts, USA

<sup>2</sup>Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China

## Abstract

Identification of the sites of post-translational modifications (PTMs) in protein, RNA, and DNA sequences is currently a very hot topic. This is because the information thus obtained is very useful for in-depth understanding the biological processes at the cellular level and for developing effective drugs against major diseases including cancers and Alzheimer's as well. Although this can be realized by means of various experimental techniques, it is both time-consuming and costly to determine the PTM sites purely based on experiments. With the avalanche of biological sequences generated in the post-genomic age, it is highly desired to develop artificial intelligence (AI) tools for rapidly and effectively identifying the PTM sites. In the last few years, many efforts have been made in this regard, and considerable progresses have been achieved. This review is focused on those AI tools that have the following two features. (1) They have been developed by strictly observing the 5-steps rule so that they each have a user-friendly web-server for the majority of experimental scientists to easily get their desired data without the need to go through the detailed mathematics involved. (2) Their cornerstones have been based on PseAAC (Pseudo Amino Acid Composition) or PseKNC (Pseudo K-tuple Nucleotide Composition), and hence the prediction quality is generally remarkably higher than most of the other PTM prediction methods without such base.

## Keywords

Artificial intelligence (AI) tools, Five-step rules, Post-translational modifications, Absolute true rate, Web-server

## Introduction

Post-translational modification, or PTM, means the covalent and generally enzymatic modification of proteins right after they are biosynthesized. After being synthesized by ribosomes, proteins may undergo PTM to form the mature protein products. PTMs can occur on the amino acid side chains of a protein or at its C- or N- terminus. They can covalently modify the existing functional group of an amino acid and make it have other functional group. Therefore, the chemical repertoire of the 20 standard amino acids can be considerably extended via the process of PTMs.

According to their occurrence in three different types of biological sequences, PTMs can be classified into the following three different categories: (1) PTLM (post-translational modification) in proteins, (2) PTCM (post-transcriptional modification) in RNA, and (3) PTRM (post-replication modification) in DNA. PTMs play a key role in providing bio-macromolecules with structural and functional diversity, as well as in regulating cellular plasticity and dynamics. Meanwhile, PTMs are also closely associated with many major diseases including cancer, Alzheimer's, and Parkinson's. Therefore, identifying the PTM sites in biological sequences is very important for both basic research and drug development.

## Historical Reflection

Before going on, it is illuminative to make a historical reflection. For quite a long period of time, the information derived by the computational approaches were not trusted very much by most experimental scientists due to the notorious local minimum problem [1]. Actually, they only trusted the results determined by the experiments, and thought that computational results were not reliable unless they had been confirmed by experiments. This kind of situation has been changed during the last decade or so owing to the rapid development of structural bioinformatics and sequential bioinformatics. For the 3D structures of proteins, what they

**\*Corresponding author:** Kuo-Chen Chou, Gordon Life Science Institute, Boston, Massachusetts, 02478, USA; Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, 610054, China

**Accepted:** August 12, 2019

**Published online:** August 14, 2019

**Citation:** Kuo-Chen C (2019) Artificial Intelligence (AI) Tools Constructed via the 5-Steps Rule for Predicting Post-Translational Modifications. Trends Artif Intell 3(1):60-74

trusted most were those determined by the X-ray crystallography. Unfortunately, it is time-consuming and expensive, and not all proteins can be successfully crystallized. Membrane proteins are difficult to crystallize and most of them will not dissolve in normal solvents. Accordingly, so far very few membrane protein structures have been determined. NMR is indeed a very powerful tool in determining the 3D structures of membrane proteins (see, e.g., [2-19]), but it is also time-consuming and costly. In order to acquire the structural information in a timely manner, a series of 3D protein structures have been developed by means of structural bioinformatics tools (see, e.g., [20-32]) and they have been found very useful in conducting mutagenesis studies [33] for rational drug design. Meanwhile, facing the explosive growth of biological sequences discovered in the post-genomic age, to timely use them for drug development, a lot of useful information have been revealed or deduced by various AI tools via the PseAAC approach [34-36] and PseKNC approach [37-39]. Actually, this kind of AI technique has played increasingly important roles in driving the medicinal chemistry into an unprecedented revolution [40,41] by significantly speeding up the process of finding novel drugs [42-44].

As it was in the last few years that many AI tools were developed for predicting the PTM sites in biological sequences [40,45-87] in compliance with the Chou's 5-steps rule [88] by going through the following five procedures: (1) How to select or construct a valid benchmark dataset to train and test the predictor; (2) How to represent the samples with an effective formulation that can truly reflect their intrinsic correlation with the target to be predicted; (3) How to introduce or develop a powerful algorithm to conduct the prediction; (4) How to properly perform cross-validation tests to objectively evaluate the anticipated prediction accuracy; (5) How to establish a user-friendly web-server for the predictor that is accessible to the public.

The AI tools constructed thru the 5-steps rule bear the following notable merits: (1) Crystal clear in logic development, (2) Complete transparent in operation, (3) Quite easy to repeat the reported results by others, (4) Holding high potential in stimulating other sequence-analyzing methods, and (5) Very convenient to be used by broad experimental scientists.

Therefore, focused on the current review paper are only those AI tools that were born through the Chou's 5-steps rule [88]. As for the importance of the 5-steps rule and how to use it in developing new predictor for proteome and genome analyses, see an insightful Wikipedia article at [https://en.wikipedia.org/wiki/5-step\\_rules](https://en.wikipedia.org/wiki/5-step_rules).

Besides, with the avalanche of biological sequences in the post-genomic era, one of the most important but also most difficult problems in developing AI tools for investigation into biology is how to express a biological sequence with a discrete model or a vector, yet still considerably keep its sequence-order information or key pattern characteristic. This is because all the existing machine-learning algorithms (such as "Optimization" algorithm [89], "Covariance Discriminant" or "CD" algorithm [90,91], "Nearest Neighbor" or "NN" algorithm [92], and "Support Vector Machine" or "SVM" algorithm [92,93])

can only handle vectors as elaborated in a comprehensive review [40].

However, a vector defined in a discrete model may completely lose all the sequence-pattern information. To avoid completely losing the sequence-pattern information for proteins, the pseudo amino acid composition [34] or PseAAC [35] was proposed. Ever since the concept of Chou's PseAAC was proposed, it has been widely used in nearly all the areas of computational proteomics (see, e.g., [45,48,52,60,67,73,77,78,82,83,85-87,94-251] as well as a long list of references cited in [41]).

Because it has been widely and increasingly used, four powerful open access soft-wares, called 'PseAAC' [252], 'PseAAC-Builder' [128], 'propy' [146], and 'PseAAC-General' [166], were established: the former three are for generating various modes of Chou's special PseAAC [253]; while the 4<sup>th</sup> one for those of Chou's general PseAAC [88], including not only all the special modes of feature vectors for proteins but also the higher level feature vectors such as "Functional Domain" mode (see Eqs.9-10 of [88]), "Gene Ontology" mode (see Eqs.11-12 of [88]), and "Sequential Evolution" or "PSSM" mode (see Eqs.13-14 of [88]).

Meanwhile, the idea of PseAAC was extended to generate various modes of feature vectors for DNA and RNA sequences [37-39,254-258], and has been proved very useful as well.

Given an AI tool, its name can be defined as

$$\text{Name of AI tool} = \text{AI}(\mathbb{X}) \quad (1)$$

where the wildcard  $\mathbb{X}$  denotes the web-server or software based on which the AI tool has been constructed. For instance: when  $\mathbb{X} = \text{Isno-PseAAC}$ , the AI tool is for predicting cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition; when  $\mathbb{X} = \text{Irna-PseU}$ , the AI tool is for predicting RNA pseudouridine sites; when  $\mathbb{X} = \text{Idna-Methyl}$ , the AI tool is for predicting DNA methylation sites via pseudo trinucleotide composition; and so forth.

## Sixteen AI Tools for Identifying PTM or PTLM Sites in Protein Sequences

The 16 AI tools are: (1) AI SNO-PseAAC [46]; (2) AI (iS-NO-AAPair) [47]; (3) AI (iMethyl-AAC) [49]; (4) AI (iHyd-seAAC) [50]; (5) AI (iNitro-Tyr) [51]; (6) AI (iUbiq-Lys) [54]; (7) AI (iSuc-PseOpt) [56]; (8) AI (pSuc-Lys) [57]; (9) AI Car-PseCp [58]; (10) AI (pSumo-CD) [59]; (11) AI (iHyd-PseCp) [62]; (12) AI (iPTM-mLys) [63]; (13) AI (iPhos-PseEn) [64]; (14) AI (iPGK-PseAAC) [68]; (15) AI (iPhos-PseEvo) [71]; (16) AI (iPreny-PseAAC) [72]. Their functions and web-server links are each given in Table 1.

## Seven AI Tools for Identifying PTM or PTCM Sites in RNA Sequences

The 7 AI tools are: (1) AI (iRNA-PseU) [55]; (2) AI (pRNAm-PC) [61]; (3) AI (iRNA-PseColl) [66]; (4) AI (iRNA-methyl) [69]; (5) AI (iRNAm5C-PseDNC) [70]; (6) AI (iRNA(m6A)-PseDNC)

**Table 1:** List of 16 AI tools for identifying the PTM sites in protein sequences.

No	Function	Web-server link
1	Cysteine S-nitrosylation	<a href="http://app.aporc.org/iSNO-PseAAC/">http://app.aporc.org/iSNO-PseAAC/</a>
2	Cysteine S-nitrosylation	<a href="http://app.aporc.org/iSNO-AAPair/">http://app.aporc.org/iSNO-AAPair/</a>
3	Methylation	<a href="http://www.jci-bioinfo.cn/iMethyl-PseAAC">http://www.jci-bioinfo.cn/iMethyl-PseAAC</a>
4	Hydroxyproline and Hydroxylysine	<a href="http://app.aporc.org/iHyd-PseAAC/">http://app.aporc.org/iHyd-PseAAC/</a>
5	Nitrotyrosine	<a href="http://app.aporc.org/iNitro-Tyr/index.html">http://app.aporc.org/iNitro-Tyr/index.html</a>
6	Lysine ubiquitination	<a href="http://www.jci-bioinfo.cn/iUbiq-Lys">http://www.jci-bioinfo.cn/iUbiq-Lys</a>
7	Lysine succinylation	<a href="http://www.jci-bioinfo.cn/iSuc-PseOpt">http://www.jci-bioinfo.cn/iSuc-PseOpt</a>
8	Lysine succinylation	<a href="http://www.jci-bioinfo.cn/iSuc-PseOpt">http://www.jci-bioinfo.cn/iSuc-PseOpt</a>
9	Carbonylation	<a href="http://www.jci-bioinfo.cn/iCar-PseCp">http://www.jci-bioinfo.cn/iCar-PseCp</a>
10	Sumoylation	<a href="http://www.jci-bioinfo.cn/pSumo-CD">http://www.jci-bioinfo.cn/pSumo-CD</a>
11	Hydroxyproline and Hydroxylysine	<a href="http://www.jci-bioinfo.cn/iHyd-PseCp">http://www.jci-bioinfo.cn/iHyd-PseCp</a>
12	Lysine	<a href="http://www.jci-bioinfo.cn/iHyd-PseCp">http://www.jci-bioinfo.cn/iHyd-PseCp</a>
13	Phosphorylation	<a href="http://www.jci-bioinfo.cn/iPhos-PseEvo">http://www.jci-bioinfo.cn/iPhos-PseEvo</a>
14	Lysine phosphoglycerlation	<a href="http://app.aporc.org/iPGK-PseAAC/">http://app.aporc.org/iPGK-PseAAC/</a>
15	Human phosphorylated protein	<a href="http://www.jci-bioinfo.cn/iPhos-PseEvo">http://www.jci-bioinfo.cn/iPhos-PseEvo</a>
16	C-terminal cysteine prenylation	<a href="http://app.aporc.org/iPreny-PseAAC/">http://app.aporc.org/iPreny-PseAAC/</a>

**Table 2:** List of 7 AI tools for identifying the PTM sites in RNA sequences.

No	Function	Web-server link
1	Pseudouridine	<a href="http://app.aporc.org/iPreny-PseAAC/">http://app.aporc.org/iPreny-PseAAC/</a>
2	N-methyladenosine	<a href="http://www.jci-bioinfo.cn/pRNAm-PC">http://www.jci-bioinfo.cn/pRNAm-PC</a>
3	Different RNA modifications	<a href="http://www.jci-bioinfo.cn/iRNAm5C-PseDNC">http://www.jci-bioinfo.cn/iRNAm5C-PseDNC</a>
4	RNA 2'-O-methylation	<a href="http://www.jci-bioinfo.cn/iRNAm5C-PseDNC">http://www.jci-bioinfo.cn/iRNAm5C-PseDNC</a>
5	5-methylcytosine sites	<a href="http://www.jci-bioinfo.cn/iRNAm5C-PseDN">http://www.jci-bioinfo.cn/iRNAm5C-PseDN</a>
6	N6-methyladenosine	<a href="http://lin-group.cn/server/iRNA(m6A)-PseDNC.php">http://lin-group.cn/server/iRNA(m6A)-PseDNC.php</a>
7	3-types of adenosine	<a href="http://lin-group.cn/server/iRNA-3typeA/">http://lin-group.cn/server/iRNA-3typeA/</a>

**Table 3:** List of one AI tool for identifying the PTM sites in DNA sequences.

No	Function	Web-server link
1	Methylation	<a href="http://www.jci-bioinfo.cn/iDNA-Methyl">http://www.jci-bioinfo.cn/iDNA-Methyl</a>

[75]; (7) AI (iRNA-3typeA) [76]. Their functions and web-server links are each given in Table 2.

### One AI Tool for Identifying PTM or PTRM Sites in DNA Sequences

AI (iDNA-Methyl) is the AI tool for identifying the PTM sites in DNA sequences [259]. Its function and web-server link are given in Table 3.

### Discussions

For measuring the success rates of the AI tools, a set of four metrics [260] are usually used in literature. They are: (1) Overall accuracy or Acc, (2) Mathew's correlation coefficient or MCC, (3) Sensitivity or Sn, and (4) Specificity or Sp, as given below

$$\left\{ \begin{array}{l}
 S_n = \frac{TP}{TP + FN} \\
 S_p = \frac{TN}{TN + FP} \\
 Acc = \frac{TP + TN}{TP + TN + FP + FN} \\
 MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) (TP + FN) (TN + FP) (TN + FN)}}
 \end{array} \right. \quad (2)$$

Although the above four metrics copied from math books were often used in literature to measure the prediction quality of a prediction method, they are lacking intuitiveness and no easy-to-understand for most biologists. Particularly the MCC (the Matthews correlation coefficient), which is a very important metrics used for reflecting the stability of a prediction method. Fortunately, based on the Chou's symbols introduced for studying protein signal peptides [261,262], a set of four intuitive metrics were derived [47,263,264], as given below

$$\begin{cases}
 S_n = 1 - \frac{N^+}{N^+} & 0 \leq S_n \leq 1 \\
 S_p = 1 - \frac{N^-}{N^-} & 0 \leq S_p \leq 1 \\
 Acc = \wedge = 1 - \frac{N^+ + N^-}{N^+ + N^-} & 0 \leq Acc \leq 1 \\
 MCC = \frac{1 - (\frac{N^+}{N^+} + \frac{N^-}{N^-})}{\sqrt{\left(1 + \frac{N^- - N^+}{N^+}\right) \left(1 + \frac{N^+ - N^-}{N^-}\right)}} & -1 \leq MCC \leq 1
 \end{cases} \quad (3)$$

According to Eq.3 we can easily see the following. When  $N^+ = 0$  meaning none of the positive samples is mispredicted to be negative, we have the sensitivity  $S_n = 1$ ; while  $N^+ = N^+$  meaning that all the positive samples are mispredicted to be negative, we have the sensitivity  $S_n = 0$ . Likewise, when  $N^- = 0$  meaning none of the negative samples is incorrectly predicted to be positive, we have the specificity  $S_p = 1$ ; while  $N^- = N^-$  meaning all the negative samples are incorrectly predicted to be positive, we have the specificity  $S_p = 0$ . When  $N^+ = N^+ = 0$  meaning that none of the positive samples and none of the negative samples is incorrectly predicted, we have the overall accuracy  $Acc = 1$ ; while  $N^+ = N^+$  and  $N^- = N^-$  meaning that all the positive samples and all the negative samples are mispredicted, we have the overall accuracy  $Acc = 0$ , and ;  $MCC = 1$ : when  $N^+ = N^+ / 2$  and  $N^- = N^- / 2$  we have  $MCC = 0$  meaning no better than random prediction; when  $N^+ = N^+$  and  $N^- = N^-$  we have  $MCC = -1$  meaning total disagreement between prediction and observation. As we can see from the above discussion, it is much more intuitive and easier to understand when using Eq.3 instead of Eq.1 to examine a predictor for its four metrics, particularly for its Mathew’s correlation coefficient.

It is instructive to point out, however, that some AI tools may have the multi-label feature, such as AII (iDNA-Methyl) [63] having the capacity to identify multiple lysine PTM sites and their different types. Actually, in the real world the multi-label systems (where a sample may simultaneously belong to several classes) have become more frequent in both system biology [265-292] and system medicine [293,294].

To examine the performance of multi-label AI tools, one also needs a set of global metrics [295,296], as elaborated below.

$$\begin{cases}
 \text{Aiming} \uparrow = \frac{1}{N^q} \sum_{k=1}^{N^q} \left( \frac{\|L_k \cap L_k^*\|}{\|L_k^*\|} \right), & [0,1] \\
 \text{Coverage} \uparrow = \frac{1}{N^q} \sum_{k=1}^{N^q} \left( \frac{\|L_k \cap L_k^*\|}{\|L_k\|} \right), & [0,1] \\
 \text{Accuracy} \uparrow = \frac{1}{N^q} \sum_{k=1}^{N^q} \left( \frac{\|L_k \cap L_k^*\|}{\|L_k \cup L_k^*\|} \right), & [0,1] \\
 \text{Absolute true} \uparrow = \frac{1}{N^q} \sum_{k=1}^{N^q} \Delta(L_k, L_k^*), & [0,1] \\
 \text{Absolute false} \downarrow = \frac{1}{N^q} \sum_{k=1}^{N^q} \left( \frac{\|L_k \cup L_k^*\| - \|L_k \cap L_k^*\|}{M} \right), & [1,0]
 \end{cases} \quad (4)$$

where  $N^q$  is the total number of query or tested samples,  $M$  is the total number of different labels for the investigated system,  $\|U\|$  means the operator acting on the set therein to count the number of its elements,  $U$  means the symbol for the “union” in the set theory,  $\cap$  denotes the symbol for the “intersection”,  $L_k$  the subset that contains all the labels observed by experiments for the  $k$ -th tested sample,  $L_k^*$  represents the subset that contains all the labels predicted for the  $k$ -th sample, and

$$\Delta(L_k, L_k^*) = \begin{cases} 1, & \text{if all the labels in } L_k^* \text{ are identical to those in } L_k \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

In Eq.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 5<sup>th</sup> metrics with a down arrow  $\downarrow$  is called negative metrics, implying just the opposite meaning. As we can see from Eq.1: (1) The “Aiming” defined by the 1<sup>st</sup> 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 2<sup>nd</sup> 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 3<sup>rd</sup> 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 4<sup>th</sup> 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 5<sup>th</sup> sub-equation is for checking the ratio of the completely wrong prediction over the total prediction events.

The five metrics in Eq.4 reflect the quality of a multi-label predictor from five different angles at the global level. It is instructive to point out, however, among the five global metrics the most important one and also the most difficult to improve its success rate is the “Absolute true” or “perfectly correct” rate [295]. Why? This is because the score standard for the absolute true rate is very harsh. According to its definition, for a statistical sample that is actually simultaneously with the states (“A”, “B”, “C”). If the predicted result is not exactly the three states but (“A”, “B”) or (“A”, “B”, “C”, “D”), no score whatsoever will be given. In other words, when and only when the predicted outcome for the statistical sample is perfectly identical to its actual status, can we add one point for the absolute true rate; otherwise, zero. That is why many investigators even chose not to mention the metrics of absolute true rate; otherwise they would face the embarrassment of reporting a very low success rate for their prediction methods.

The set of metrics in Eq.4 are used to evaluate the prediction quality of a multi-label AI tool for all the samples in the entire system concerned [296], and hence is called the “set of metrics for the global accuracy” or the “set of global metrics”.

## Concluding Remarks and Perspectives

The AI tools introduced in this review paper for predicting PTM sites have been all established by following the 5-steps

rule [88], and hence they each have a user-friendly web server for the majority of experimental scientists to easily get their desired data. Also, their cornerstones are based on PseAAC [34-36,88,253] or PseKNC [37,254,256,257,264,297], and hence their prediction quality is usually higher than the other PTM prediction methods without using the PseAAC or PseKNC approach.

As we can see from the Sections 3, 4, and 5, the most web-servers available are for the AI tools aimed at identifying the PTM sites in protein sequences, the next are at DNA sequences, and the least at RNA sequences. It is anticipated, however, that with more experimental data available in the future, the benchmark datasets for the PTM sites in RNA and DNA sequences will be enriched as well. The existing AI tools will not only be easily extended to cover more RNA and DNA sequences, but also further improve the prediction quality in all kinds of biological sequences.

It is worthy of noting that recently the 5-steps rule has also been used in many different areas [84,298-314].

Meanwhile it has not escaped our notice that using graphic approaches to study biological and medical systems can provide an intuitive vision and useful insights for helping analyze complicated relations therein as shown in the systems of enzyme fast reaction [315-317], graphical rules in molecular biology [318-321], and low-frequency internal motion in biomacromolecules (such as protein and DNA) [322]. Particularly, what happened is that this kind of insightful implication has also been demonstrated in [323] and many follow-up publications [324-339].

## Acknowledgement

The author wishes to thank Dr. Michelle Claus for the invitation to write this paper.

## References

1. KC Chou, CT Zhang (1995) Prediction of protein structural classes. *Crit Rev Biochem Mol Biol* 30: 275-349.
2. JJ Chou, H Matsuo, H Duan, et al. (1998) Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. *Cell* 171:180.
3. K Oxenoid, YS Dong, C Cao, et al. (2016) Architecture of the mitochondrial calcium uniporter. *Nature* 533: 269-273.
4. J Dev, D Park, Q Fu, et al. (2016) Structural basis for membrane anchoring of HIV-1 envelope spike. *Science* 353: 172-175.
5. JR Schnell, JJ Chou (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451: 591-595.
6. MJ Berardi, WM Shih, SC Harrison, et al. (2011) Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. *Nature* 476: 109-113.
7. JJ Chou, S Li, C B Klee, et al. (2001) Solution structure of Ca<sup>2+</sup>-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol* 990-997.
8. B OuYang, S Xie, MJ Berardi, et al. (2013) Unusual architecture of the p7 channel from hepatitis C virus. *Nature* 498: 521-525.
9. J Wang, RM Pielak, MA McClintock, et al. (2009) Solution structure and functional analysis of the influenza B proton channel. *Nat Struct Mol Biol* 161: 267-271.
10. Q Fu, TM Fu, AC Cruz, et al. (2016) Structural Basis and Functional role of intramembrane trimerization of the fas/cd95 death receptor. *Mol Cell* 61: 602-613.
11. JJ Chou, H Li, GS Salvessen, et al. (1999) Solution structure of BID, an intracellular amplifier of apoptotic signalling. *Cell* 96: 615-624.
12. K Oxenoid, JJ Chou (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci U S A* 102: 10870-10875.
13. ME Call, JR Schnell, C Xu, et al. (2006) Wucherpfennig, The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor. *Cell* 127: 355-368.
14. ME Call, KW Wucherpfennig, JJ Chou (2010) The structural basis for intramembrane assembly of an activating immunoreceptor complex. *Nat Immunol* 11: 1023-1029.
15. E Gagnon, C Xu, W Yang, et al. (2010) Response multilayered control of T cell receptor phosphorylation. *Cell* 142: 669-671.
16. S Bruschweiler, Q Yang, C Run, et al. (2015) Substrate-modulated ADP/ATP-transporter dynamics revealed by NMR relaxation dispersion. *Nat Struct Mol Biol* 22: 636-641.
17. C Cao, S Wang, T Cui, et al. (2017) Ion and inhibitor binding of the double-ring ion selectivity filter of the mitochondrial calcium uniporter. *Proc Natl Acad Sci U S A* 114: 2846-2851.
18. A Piai, J Dev, Q Fu, et al. (2017) Stability and water accessibility of the trimeric membrane anchors of the hiv-1 envelope spikes. *J Am Chem Soc* 139: 18432-18435.
19. L Pan, TM Fu, W Zhao, et al. (2019) Higher-order clustering of the transmembrane anchor of dr5 drives signaling. *Cell* 176: 1477-1489.
20. KC Chou, AG Tomasselli, RL Heinrikson (2000) Prediction of the tertiary structure of a caspase-9/inhibitor complex. *FEBS Lett* 470: 249-256.
21. KC Chou, D Jones, RL Heinrikson (1997) Prediction of the tertiary structure and substrate binding site of caspase-8. *FEBS Letters* 419: 49-54.
22. KC Chou (2004) Insights from modelling the 3D structure of the extracellular domain of alpha7 nicotinic acetylcholine receptor. *Biochem Biophys Res Commun (BBRC)* 319: 433-438.
23. KC Chou (2005) Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein. *J Proteome Res* 4: 1681-1686.
24. KC Chou, WJ Howe (2002) Prediction of the tertiary structure of the beta-secretase zymogen. *Biochem Biophys Res Commun (BBRC)* 292: 702-708.
25. KC Chou (2004) Insights from modelling the tertiary structure of BACE2. *Journal of Proteome Research* 3: 1069-1072.
26. KC Chou (2004) Insights from modelling three-dimensional structures of the human potassium and sodium channels. *J Proteome Res* 3: 856-861.
27. KC Chou (2005) Modeling the tertiary structure of human cathepsin-E. *Biochem. Biophys Res Commun (BBRC)* 331: 56-60.
28. KC Chou (2005) Insights from modeling the 3D structure of DNA-CBF3b complex. *J Proteome Res* 4: 1657-1660.

29. SQ Wang, QS Du (2007) Study of drug resistance of chicken influenza A virus (H5N1) from homology-modeled 3D structures of neuraminidases. *Biochem Biophys Res Comm (BBRC)* 354: 634-640.
30. SQ Wang, QS Du, RB Huang, et al. (2009) Insights from investigating the interaction of oseltamivir (Tamiflu) with neuraminidase of the H1N1 swine flu virus. *Biochem Biophys Res Commun (BBRC)* 386: 432-436.
31. XB Li, SQ Wang, WR Xu, et al. (2011) Novel inhibitor design for hemagglutinin against H1N1 influenza virus by core hopping method. *PLoS One* 6: e28111.
32. Y Ma, SQ Wang, WR Xu, et al. (2012) Design novel dual agonists for treating type-2 diabetes by targeting peroxisome proliferator-activated receptors with core hopping approach. *PLoS One* 7: e38546.
33. KC Chou (2004) Structural bioinformatics and its impact to biomedical science. *Curr Med Chem* 11: 2105-2134.
34. KC Chou (2001) Prediction of protein cellular attributes using pseudo amino acid composition. *PROTEINS: Structure, Function, and Genetics* 43: 246-255.
35. KC Chou (2005) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics* 21: 10-19.
36. KC Chou (2005) Progress in protein structural class prediction and its impact to bioinformatics and proteomics. *Curr Protein Pept Sci* 6: 423-436.
37. W Chen, TY Lei, DC Jin, et al. (2014) PseKNC: A flexible web-server for generating pseudo K-tuple nucleotide composition. *Anal Biochem* 456: 53-60.
38. SH Guo, EZ Deng, LQ Xu, et al. (2014) iNuc-PseKNC: A sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. *Bioinformatics* 30: 1522-1529.
39. W Chen, X Zhang, J Brooker, et al. (2015) PseKNC-General: A cross-platform package for generating various modes of pseudo nucleotide compositions. *Bioinformatics* 31: 119-120.
40. KC Chou (2015) Impacts of bioinformatics to medicinal chemistry. *Med Chem* 11: 218-234.
41. KC Chou (2017) An unprecedented revolution in medicinal chemistry driven by the progress of biological science. *Curr Top Med Chem* 17: 2337-2358.
42. WZ Zhong, SF Zhou (2014) Molecular science for drug development and biomedicine. *Int J Mol Sci* 15: 20072-20078.
43. GP Zhou, WZ Zhong (2016) Perspectives in Medicinal Chemistry. *Current Topics in Medicinal Chemistry* 16: 381-382.
44. SF Zhou, WZ Zhong (2017) Drug design and discovery: Principles and applications. *Molecules*.
45. HL Xie, L Fu, XD Nie (2013) Using ensemble SVM to identify human GPCRs N-linked glycosylation sites based on the general form of Chou's PseAAC. *Protein Eng Des Sel* 26: 735-742.
46. Y Xu, J Ding, LY Wu (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.
47. Y Xu, XJ Shao, LY Wu, et al. (2013) iSNO-AAPair: Incorporating amino acid pairwise coupling into PseAAC for predicting cysteine S-nitrosylation sites in proteins. *PeerJ* 1: e171.
48. C Jia, X Lin, Z Wang (2014) Prediction of Protein s-nitrosylation sites based on adapted normal distribution bi-profile bayes and chou's pseudo amino acid composition. *Int J Mol Sci* 15: 10410-10423.
49. WR Qiu, X Xiao, WZ Lin (2014) iMethyl-PseAAC: Identification of protein methylation sites via a pseudo amino acid composition approach. *Biomed Res Int (BMRI)*.
50. Y Xu, X Wen, XJ Shao (2014) iHyd-PseAAC: Predicting hydroxyproline and hydroxylysine in proteins by incorporating dipeptide position-specific propensity into pseudo amino acid composition. *Int J Mol Sci (IJMS)* 15: 7594-7610.
51. Y Xu, X Wen, LS Wen, et al. (2014) iNitro-Tyr: Prediction of nitrotyrosine sites in proteins with general pseudo amino acid composition. *PLoS ONE* 9: e105018.
52. J Zhang, X Zhao, P Sun, et al. (2014) PSNO: Predicting cysteine s-nitrosylation sites by incorporating various sequence-derived features into the general form of chou's PSEAAC. *Int J Mol Sci* 15: 11204-11219.
53. W Chen, P Feng, H Ding, et al. (2015) iRNA-Methyl: Identifying N6-methyladenosine sites using pseudo nucleotide composition. *Anal Biochem* 490: 26-33.
54. WR Qiu, X Xiao, WZ Lin (2015) iUbiq-Lys: Prediction of lysine ubiquitination sites in proteins by extracting sequence evolution information via a grey system model. *Journal of Biomolecular Structure and Dynamics (JBSD)* 33: 1731-1742.
55. W Chen, H Tang, J Ye, et al. (2016) iRNA-PseU: Identifying RNA pseudouridine sites. *Mol Ther Nucleic Acids* 5: e332.
56. J Jia, Z Liu, X Xiao, et al. (2016) iSuc-PseOpt: Identifying lysine succinylation sites in proteins by incorporating sequence-coupling effects into pseudo components and optimizing imbalanced training dataset. *Anal Biochem* 497: 48-56.
57. J Jia, Z Liu, X Xiao, et al. (2016) pSuc-Lys: Predict lysine succinylation sites in proteins with PseAAC and ensemble random forest approach. *Journal of Theoretical Biology* 394: 223-230.
58. J Jia, Z Liu, X Xiao, et al. (2016) iCar-PseCp: Identify carbonylation sites in proteins by Monto Carlo sampling and incorporating sequence coupled effects into general PseAAC. *Oncotarget* 7: 34558-34570.
59. J Jia, L Zhang, Z Liu, et al. (2016) pSumo-CD: Predicting sumoylation sites in proteins with covariance discriminant algorithm by incorporating sequence-coupled effects into general PseAAC. *Bioinformatics* 32: 3133-3141.
60. Z Ju, JZ Cao, H Gu (2016) Predicting lysine phosphoglycerylation with fuzzy SVM by incorporating k-spaced amino acid pairs into Chou's general PseAAC. *J Theor Biol* 397: 145-150.
61. Z Liu, X Xiao, DJ Yu, et al. (2016) pRNA-PC: Predicting N-methyladenosine sites in RNA sequences via physical-chemical properties. *Anal Biochem* 497: 60-67.
62. WR Qiu, BQ Sun, X Xiao, et al. (2016) iHyd-PseCp: Identify hydroxyproline and hydroxylysine in proteins by incorporating sequence-coupled effects into general PseAAC. *Oncotarget* 7: 44310-44321.
63. WR Qiu, BQ Sun, X Xiao, et al. (2016) iPTM-mLys: Identifying multiple lysine PTM sites and their different types. *Bioinformatics* 32: 3116-3123.
64. WR Qiu, X Xiao, ZC Xu (2016) iPhos-PseEn: Identifying

- phosphorylation sites in proteins by fusing different pseudo components into an ensemble classifier. *Oncotarget* 7: 51270-51283.
65. Y Xu (2016) Recent progress in predicting posttranslational modification sites in proteins. *Curr Top Med Chem* 16: 591-603.
66. P Feng, H Ding, H Yang, et al. (2017) iRNA-PseColl: Identifying the occurrence sites of different RNA modifications by incorporating collective effects of nucleotides into PseKNC. *Mol Ther Nucleic Acids* 7: 155-163.
67. Z Ju, JJ He (2017) Prediction of lysine crotonylation sites by incorporating the composition of k-spaced amino acid pairs into Chou's general PseAAC. *J Mol Graph Model* 77: 200-204.
68. LM Liu, Y Xu (2017) iPGK-PseAAC: Identify lysine phosphoglycerylation sites in proteins by incorporating four different tiers of amino acid pairwise coupling information into the general PseAAC. *Med Chem* 13: 552-559.
69. WR Qiu, SY Jiang, BQ Sun, et al. (2017) iRNA-2methyl: Identify RNA 2'-O-methylation sites by incorporating sequence-coupled effects into general PseKNC and ensemble classifier. *Med Chem* 13: 734-743.
70. WR Qiu, SY Jiang, ZC Xu, et al. (2017) iRNAm5C-PseDNC: Identifying RNA 5-methylcytosine sites by incorporating physical-chemical properties into pseudo dinucleotide composition. *Oncotarget* 8: 41178-41188.
71. WR Qiu, BQ Sun, X Xiao (2017) iPhos-PseEvo: Identifying human phosphorylated proteins by incorporating evolutionary information into general PseAAC via grey system theory. *Mol Inform* 36.
72. Y Xu, C Li (2017) iPreny-PseAAC: Identify C-terminal cysteine prenylation sites in proteins by incorporating two tiers of sequence couplings into PseAAC. *Med Chem* 13: 544-551.
73. S Akbar, M Hayat (2018) iMethyl-STTNC: Identification of N(6)-methyladenosine sites by extending the Idea of SAAC into Chou's PseAAC to formulate RNA sequences. *J Theor Biol* 455: 205-211.
74. A Chandra, A Sharma, A Dehzangi, et al. (2018) PhoglyStruct: Prediction of phosphoglycerylated lysine residues using structural properties of amino acids. *Sci Rep* 8: 17923.
75. W Chen, H Ding, X Zhou (2018) iRNA(m6A)-PseDNC: Identifying N6-methyladenosine sites using pseudo dinucleotide composition. *Anal Biochem* 561-562: 59-65.
76. W Chen, P Feng, H Yang, et al. (2018) iRNA-3typeA: Identifying 3-types of modification at RNA's adenosine sites. *Molecular Therapy: Nucleic Acid* 11: 468-474.
77. AW Ghauri, YD Khan, N Rasool, et al. (2018) pNitro-Tyr-PseAAC: Predict nitrotyrosine sites in proteins by incorporating five features into Chou's general PseAAC. *Curr Pharm Des* 24 : 4034-4043.
78. Z Ju, SY Wang (2018) Prediction of citrullination sites by incorporating k-spaced amino acid pairs into Chou's general pseudo amino acid composition. *Gene* 664: 78-83.
79. YD Khan, N Rasool, W Hussain (2018) iPhosT-PseAAC: Identify phosphothreonine sites by incorporating sequence statistical moments into PseAAC. *Anal Biochem* 550: 109-116.
80. YD Khan, N Rasool, W Hussain, et al. (2018) iPhosY-PseAAC: Identify phosphotyrosine sites by incorporating sequence statistical moments into PseAAC. *Mol Biol Rep* 45: 2501-2509.
81. WR Qiu, BQ Sun, X Xiao, et al. (2018) iKcr-PseEns: Identify lysine crotonylation sites in histone proteins with pseudo components and ensemble classifier. *Genomics* 110: 239-246.
82. MF Sabooh, N Iqbal, M Khan, et al. (2018) Identifying 5-methylcytosine sites in RNA sequence using composite encoding feature into Chou's PseKNC. *J Theor Biol* 452: 1-9.
83. W Hussain, SD Khan, N Rasool (2019) PalmitoylC-PseAAC: A sequence-based model developed via Chou's 5-steps rule and general PseAAC for identifying S-palmitoylation sites in proteins. *Anal Biochem* 568: 14-23.
84. F Li, Y Zhang, AW Purcell, et al. (2019) Positive-unlabelled learning of glycosylation sites in the human proteome. *BMC Bioinformatics* 20: 112.
85. L Wang, R Zhang, Y Mu (2019) Fu-SulfPred: Identification of Protein S-sulfenylation Sites by Fusing Forests via Chou's General PseAAC. *J Theor Biol* 461: 51-58.
86. Q Ning, Z Ma, X Zhao (2019) dForml(KNN)-PseAAC: Detecting formylation sites from protein sequences using K-nearest neighbor algorithm via Chou's 5-step rule and pseudo components. *J Theor Biol* 470: 43-49.
87. A Ehsan, MK Mahmood, Y Khan, et al. (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.
88. KC Chou (2011) Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review, 5-steps rule). *J Theor Biol* 273: 236-247.
89. CT Zhang (1992) An optimization approach to predicting protein structural class from amino acid composition. *Protein Sci* 1: 401-408.
90. KC Chou, DW (2002) Elrod bioinformatical analysis of g-protein-coupled receptors. *J Proteome Res* 1: 429-433.
91. KC Chou, YD Cai (2003) Prediction and classification of protein subcellular location: Sequence-order effect and pseudo amino acid composition. *J Cell Biochem* 90: 1250-1260.
92. L Hu, T Huang, X Shi, et al. (2011) Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. *PLoS ONE* 6: e14556.
93. YD Cai, KY Feng, WC Lu (2006) Using LogitBoost classifier to predict protein structural classes. *J Theor Biol* 238: 172-176.
94. XB Zhou, C Chen, ZC Li, et al. (2007) Using Chou's amphiphilic pseudo amino acid composition and support vector machine for prediction of enzyme subfamily classes. *J Theor Biol* 248: 546-551.
95. YS Ding, TL Zhang (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.
96. Y Fang, Y Guo, Y Feng, et al. (2008) Predicting DNA-binding proteins: Approached from Chou's pseudo amino acid composition and other specific sequence features. *Amino Acids* 34: 103-109.
97. X Jiang, R Wei, TL Zhang, et al. (2008) Using the concept of Chou's pseudo amino acid composition to predict apoptosis proteins subcellular location: an approach by approximate entropy. *Protein Pept Lett* 15: 392-396.
98. X Jiang, R Wei, Y Zhao, et al. (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.
99. FM Li, QZ Li (2008) Predicting protein subcellular location using Chou's pseudo amino acid composition and improved hybrid approach. *Protein Pept Lett* 15: 612-616.
100. H Lin (2008) The modified Mahalanobis discriminant for predicting outer membrane proteins by using Chou's pseudo amino acid composition. *J Theor Biol* 252: 350-356.
101. H Lin, H Ding, FB Feng-Biao Guo, et al. (2008) Predicting subcellular localization of mycobacterial proteins by using Chou's pseudo amino acid composition. *Protein Pept Lett* 15: 739-744.
102. L Nanni, A Lumini (2008) Genetic programming for creating Chou's pseudo amino acid based features for submitochondria localization. *Amino Acids* 34: 653-660.
103. GY Zhang, BS Fang (2008) Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo amino acid composition. *J Theor Biol* 253: 310-315.
104. GY Zhang, HC Li, JQ Gao, et al. (2008) Predicting lipase types by improved Chou's pseudo amino acid composition. *Protein Pept Lett* 15: 1132-1137.
105. SW Zhang, W Chen, F Yang, et al. (2008) Using Chou's pseudo amino acid composition to predict protein quaternary structure: A sequence-segmented PseAAC approach. *Amino Acids* 35: 591-598.
106. SW Zhang, YL Zhang, HF Yang, et al. (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.
107. C Chen, L Chen, X Zou, et al. (2009) Prediction of protein secondary structure content by using the concept of Chou's pseudo amino acid composition and support vector machine. *Protein Pept Lett* 16: 27-31.
108. H Ding, L Luo, H Lin (2009) Prediction of cell wall lytic enzymes using Chou's amphiphilic pseudo amino acid composition. *Protein Pept Lett* 16: 351-355.
109. DN Georgiou, TE Karakasidis, JJ Nieto, et al. (2009) Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition. *J Theor Biol* 257: 17-26.
110. ZC Li, XB Zhou, Z Dai, et al. (2009) Prediction of protein structural classes by Chou's pseudo amino acid composition: Approached using continuous wavelet transform and principal component analysis. *Amino Acids* 37: 415-425.
111. H Lin, H Wang, H Ding, et al. (2009) Prediction of Subcellular Localization of Apoptosis Protein Using Chou's Pseudo Amino Acid Composition. *Acta Biotheoretica* 57: 321-330.
112. JD Qiu, JH Huang, RP Liang, et al. (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. *Anal Biochem* 390: 68-73.
113. YH Zeng, YZ Guo, RQ Xiao, et al. (2009) Using the augmented Chou's pseudo amino acid composition for predicting protein submitochondria locations based on auto covariance approach. *J Theor Biol* 259: 366-372.
114. M Esmaeili, H Mohabatkar, S Mohsenzadeh (2010) Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses. *J Theor Biol* 263: 203-209.
115. Q Gu, YS Ding, TL Zhang (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 Pept Lett* 17: 559-567.
116. H. Mohabatkar (2017) Prediction of cyclin proteins using Chou's pseudo amino acid composition. *Protein Pept Lett* 17: 1207-1214.
117. JD Qiu, JH Huang, SP Shi, et al. (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 Pept Lett* 17: 715-722.
118. SS Sahu, G Panda (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.
119. L Yu, Y Guo, Y Li, et al. (2010) Secret P: Identifying bacterial secreted proteins by fusing new features into Chou's pseudo amino acid composition. *J Theor Biol* 267: 1-6.
120. J Guo, N Rao, G Liu, et al. (2011) Predicting protein folding rates using the concept of Chou's pseudo amino acid composition. *J Comput Chem* 32: 1612-1617.
121. J Lin, Y Wang (2011) Using a novel AdaBoost algorithm and Chou's pseudo amino acid composition for predicting protein subcellular localization. *Protein Pept Lett* 18: 1219-1225.
122. H Mohabatkar, M Mohammad Beigi, A Esmaeili (2011) Prediction of GABAA receptor proteins using the concept of Chou's pseudo amino acid composition and support vector machine. *J Theor Biol* 281: 18-23.
123. BM Mohammad, M Behjati, H Mohabatkar (2011) Prediction of metalloproteinase family based on the concept of Chou's pseudo amino acid composition using a machine learning approach. *J Struct Funct Genomics* 12: 191-197.
124. JD Qiu, SB Suo, XY Sun, et al. (2011) OligoPred: A web-server for predicting homo-oligomeric proteins by incorporating discrete wavelet transform into Chou's pseudo amino acid composition. *J Mol Graph Model* 30: 129-134.
125. D Zou, Z He, J He, et al. (2011) Supersecondary structure prediction using Chou's pseudo amino acid composition. *J Comput Chem* 32: 271-278.
126. JZ Cao, WQ Liu, H Gu (2012) Predicting viral protein subcellular localization with Chou's pseudo amino acid composition and imbalance-weighted multi-label k-nearest neighbor algorithm. *Protein Pept Lett* 19: 1163-1169.
127. C Chen, ZB Shen, XY Zou (2012) Dual-layer wavelet svm for predicting protein structural class via the general form of Chou's pseudo amino acid composition. *Protein Pept Lett* 19: 422-429.
128. P Du, X Wang, C Xu, et al. (2012) PseAAC-Builder: A cross-platform stand-alone program for generating various special Chou's pseudo amino acid compositions. *Anal Biochem* 425: 117-119.
129. GL. Fan, QZ Li (2012) Predict mycobacterial proteins subcellular locations by incorporating pseudo-average chemical shift into the general form of Chou's pseudo amino acid composition. *J*



- Theor Biol 304: 88-95.
130. GL Fan, QZ Li (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.
  131. M Hayat, A Khan (2012) Discriminating outer membrane proteins with fuzzy k-nearest neighbor algorithms based on the general form of Chou's PseAAC. *Protein Pept Lett* 19: 411-421.
  132. LQ Li, Y Zhang, LY Zou, et al. (2012) Prediction of protein subcellular multi-localization based on the general form of Chou's pseudo amino acid composition. *Protein Pept Lett* 19: 375-387.
  133. B Liao, Q Xiang, D Li (2012) Incorporating secondary features into the general form of Chou's PseAAC for predicting protein structural class. *Protein Pept Lett* 19: 1133-1138.
  134. L Liu, XZ Hu, XX Liu, et al. (2012) Predicting protein fold types by the general form of Chou's pseudo amino acid composition: Approached from optimal feature extractions. *Protein Pept Lett* 19: 439-449.
  135. S Mei (2012) Multi-kernel transfer learning based on Chou's PseAAC formulation for protein submitochondria localization. *J Theor Biol* 293: 121-130.
  136. S Mei (2012) Predicting plant protein subcellular multi-localization by Chou's PseAAC formulation based multi-label homolog knowledge transfer learning. *J Theor Biol* 310: 80-87.
  137. Nanni L, Brahnam S, Lumini A (2012) Wavelet images and Chou's pseudo amino acid composition for protein classification. *Amino Acids* 43: 657-665.
  138. Nanni L, Lumini A, Gupta D, et al. (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.
  139. Niu XH, Hu XH, Shi F, et al. (2012) Predicting protein solubility by the general form of Chou's Pseudo amino acid composition: Approached from chaos game representation and fractal dimension. *Protein Pept Lett* 19: 940-948.
  140. Qin YF, Wang CH, Yu XQ, et al. (2012) Predicting protein structural class by incorporating patterns of over-represented k-mers into the general form of Chou's PseAAC. *Protein Pept Lett* 19: 388-397.
  141. Ren LY, Zhang YS, Gutman I (2012) Predicting the classification of transcription factors by incorporating their binding site properties into a novel mode of Chou's Pseudo amino acid composition. *Protein Pept Lett* 19: 1170-1176.
  142. Sun XY, Shi SP, Qiu JD, et al. (2012) Identifying protein quaternary structural attributes by incorporating physicochemical properties into the general form of Chou's PseAAC via discrete wavelet transform. *Mol Biosyst* 8: 3178-3184.
  143. Zhao XW, Li XT, Ma ZQ, et al. (2012) Identify DNA-Binding proteins with optimal Chou's amino acid composition. *Protein Pept Lett* 19: 398-405.
  144. Zhao XW, Ma ZQ, Yin (2012) MH Predicting protein-protein interactions by combining various sequence-derived features into the general form of Chou's Pseudo amino acid composition. *Protein Pept Lett* 19: 492-500.
  145. Zia-ur-Rehman, Khan A (2012) Identifying GPCRs and their types with Chou's Pseudo amino acid composition: An approach from multi-scale energy representation and position specific scoring matrix. *Protein & Peptide Letters* 19: 890-903.
  146. Cao DS, Xu QS, Liang YZ (2013) Propy: A tool to generate various modes of Chou's PseAAC. *Bioinformatics* 29: 960-962.
  147. Chang TH, Wu LC, Lee TY, et al. (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. *J Comput Aided Mol Des* 27: 91-103.
  148. Chen YK, Li KB (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. *J Theor Biol* 318: 1-12.
  149. Fan GL, Li QZ, Zuo YC (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.
  150. Fan GL, Li QZ (2013) Discriminating bioluminescent proteins by incorporating average chemical shift and evolutionary information into the general form of Chou's pseudo amino acid composition. *J Theor Biol* 334: 45-51.
  151. Georgiou DN, Karakasidis TE, Megaritis AC (2013) A short survey on genetic sequences, Chou's pseudo amino acid composition and its combination with fuzzy set theory. *The Open Bioinformatics Journal* 7: 41-48.
  152. Gupta MK, Niyogi R, Misra M (2013) An alignment-free method to find similarity among protein sequences via the general form of Chou's pseudo amino acid composition. *SAR QSAR Environ Res* 24: 597-609.
  153. Huang C, 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.
  154. Huang C, Yuan JQ (2013) A multilabel model based on Chou's pseudo amino acid composition for identifying membrane proteins with both single and multiple functional types. *J Membr Biol* 246: 327-334.
  155. Huang C, Yuan JQ (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.
  156. Khosravian M, Faramarzi FK, Beigi MM, et al. (2013) Predicting antibacterial peptides by the concept of Chou's pseudo amino acid composition and machine learning methods. *Protein & Peptide Letters* 20: 180-186.
  157. Lin H, Ding C, Yuan LF, et al. (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: 1350003.
  158. B Liu, X Wang, Q Zou, et al. (2013) Protein remote homology detection by combining Chou's pseudo amino acid composition and profile-based protein representation. *Molecular Informatics* 32: 775-782.
  159. H Mohabatkar, MM Beigi, K Abdolahi, et al. (2013) Prediction of allergenic proteins by means of the concept of Chou's pseudo amino acid composition and a machine learning approach.

- Med Chem 9: 133-137.
160. E Pacharawongsakda, T Theeramunkong (2013) Predict subcellular locations of singleplex and multiplex proteins by semi-supervised learning and dimension-reducing general mode of Chou's PseAAC. IEEE Trans Nanobioscience 12: 311-320.
161. YF Qin, L Zheng, J Huang (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.
162. AN Sarangi, M Lohani, R Aggarwal (2013) Prediction of essential proteins in prokaryotes by incorporating various physicochemical features into the general form of Chou's pseudo amino acid composition. Protein Pept Lett 20: 781-795.
163. S Wan, MW Mak, SY Kung, et al. (2013) GOASVM: A subcellular location predictor by incorporating term-frequency gene ontology into the general form of Chou's pseudo amino acid composition. J Theor Biol 323: 40-48.
164. X Wang, GZ Li, WC Lu (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 Pept Lett 20: 309-317.
165. N Xiaohui, L Nana, X Jingbo, et al. (2013) Using the concept of Chou's pseudo amino acid composition to predict protein solubility: An approach with entropies in information theory. J Theor Biol 332: 211-217.
166. P Du, S Gu, Y Jiao (2014) PseAAC-General: Fast building various modes of general form of Chou's pseudo amino acid composition for large-scale protein datasets. Int J Mol Sci 15: 3495-3506.
167. Z Hajisharifi, M Piryaiee, M Mohammad Beigi, et al. (2014) Predicting anticancer peptides with Chou's pseudo amino acid composition and investigating their mutagenicity via Ames test. J Theor Biol 341: 34-40.
168. GS Han, ZG Yu, V Anh, et al. (2014) A two-stage SVM method to predict membrane protein types by incorporating amino acid classifications and physicochemical properties into a general form of Chou's PseAAC. J Theor Biol 344: 31-39.
169. M Hayat, N Iqbal (2014) Discriminating protein structure classes by incorporating Pseudo Average Chemical Shift to Chou's general PseAAC and Support Vector Machine. Comput Methods Programs Biomed 116: 184-192.
170. L Kong, L Zhang, J Lv (2014) Accurate prediction of protein structural classes by incorporating predicted secondary structure information into the general form of Chou's pseudo amino acid composition. J Theor Biol 344: 12-18.
171. L Li, S Yu, W Xiao, et al. (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.
172. S Mondal, PP Pai (2014) Chou's pseudo amino acid composition improves sequence-based antifreeze protein prediction. J Theor Biol 356: 30-35.
173. L Nanni, S Brahnam, A Lumini (2014) Prediction of protein structure classes by incorporating different protein descriptors into general Chou's pseudo amino acid composition. J Theor Biol 360: 109-116.
174. J Zhang, P Sun, X Zhao, et al. (2014) PECM: Prediction of extracellular matrix proteins using the concept of Chou's pseudo amino acid composition. J Theor Biol 363: 412-418.
175. L Zhang, X Zhao, L Kong (2014) Predict protein structural class for low-similarity sequences by evolutionary difference information into the general form of Chou's pseudo amino acid composition. J Theor Biol 355: 105-110.
176. YC Zuo, Y Peng, L Liu, et al. (2014) Predicting peroxidase subcellular location by hybridizing different descriptors of Chou's pseudo amino acid patterns. Anal Biochem 458: 14-19.
177. S Ahmad, M Kabir, M Hayat (2015) Identification of Heat Shock Protein families and J-protein types by incorporating Dipeptide Composition into Chou's general PseAAC. Comput Methods Programs Biomed 122: 165-174.
178. F Ali, M Hayat (2015) Classification of membrane protein types using Voting Feature Interval in combination with Chou's Pseudo Amino Acid Composition. J Theor Biol 384: 78-83.
179. A Dehzangi, R Heffernan, A Sharma, et al. (2015) Gram-positive and Gram-negative protein subcellular localization by incorporating evolutionary-based descriptors into Chou's general PseAAC. J Theor Biol 364: 284-294.
180. GL Fan, XY Zhang, YL Liu, et al. (2015) DSPMP: Discriminating secretory proteins of malaria parasite by hybridizing different descriptors of Chou's pseudo amino acid patterns. J Comput Chem 36: 2317-2327.
181. C Huang, JQ Yuan (2015) Simultaneously identify three different attributes of proteins by fusing their three different modes of Chou's pseudo amino acid compositions. Protein Pept Lett 22: 547-556.
182. ZU Khan, M Hayat, MA Khan (2015) Discrimination of acidic and alkaline enzyme using Chou's pseudo amino acid composition in conjunction with probabilistic neural network model. J Theor Biol 365: 197-203.
183. R Kumar, A Srivastava, B Kumari, et al. (2015) Prediction of beta-lactamase and its class by Chou's pseudo amino acid composition and support vector machine. J Theor Biol 365: 96-103.
184. B Liu, J Chen, X Wang (2015) Protein remote homology detection by combining Chou's distance-pair pseudo amino acid composition and principal component analysis. Mol Genet Genomics 290: 1919-1931.
185. B Liu, J Xu, S Fan, et al. (2015) PseDNA-Pro: DNA-binding protein identification by combining Chou's PseAAC and physicochemical distance transformation. Mol Inform 34: 8-17.
186. M Mandal, A Mukhopadhyay, U Maulik (2015) Prediction of protein subcellular localization by incorporating multiobjective PSO-based feature subset selection into the general form of Chou's PseAAC. Med Biol Eng Comput 53: 331-344.
187. V Sanchez, AM Peinado, JL Perez-Cordoba, et al. (2015) A new signal characterization and signal-based Chou's PseAAC representation of protein sequences. J Bioinform Comput Biol 13: 1550024.
188. R Sharma, A Dehzangi, J Lyons, et al. (2015) Predict gram-positive and gram-negative subcellular localization via incorporating evolutionary information and physicochemical features into Chou's general PseAAC. IEEE Trans Nanobioscience 14: 915-926.
189. X Wang, W Zhang, Q Zhang, et al. (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.
190. M Zhang, B Zhao, X Liu (2015) Predicting industrial polymer melt index via incorporating chaotic characters into Chou's general PseAAC. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)* 146: 232-240.
191. SL Zhang (2015) Accurate prediction of protein structural classes by incorporating PSSS and PSSM into Chou's general PseAAC. *Chemometrics and Intelligent Laboratory Systems (CHEMOLAB)* 142: 28-35.
192. K Ahmad, M Waris, M Hayat (2016) Prediction of Protein Submitochondrial Locations by Incorporating Dipeptide Composition into Chou's General Pseudo Amino Acid Composition. *J Membr Biol* 249: 293-304.
193. M Behbahani, H Mohabtkar, M Nosrati (2016) Analysis and comparison of lignin peroxidases between fungi and bacteria using three different modes of Chou's general pseudo amino acid composition. *J Theor Biol* 411: 1-5.
194. GL Fan, YL Liu, H Wang (2016) Identification of thermophilic proteins by incorporating evolutionary and acid dissociation information into Chou's general pseudo amino acid composition. *J Theor Biol* 407: 138-142.
195. YS Jiao, PF Du (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. *J Theor Biol* 402: 38-44.
196. M Kabir, M Hayat (2016) iRSpot-GAEnsC: Identifying recombination spots via ensemble classifier and extending the concept of Chou's PseAAC to formulate DNA samples. *Mol Genet Genomics* 291: 285-296.
197. M Tahir, M Hayat (2016) iNuc-STNC: A sequence-based predictor for identification of nucleosome positioning in genomes by extending the concept of SAAC and Chou's PseAAC. *Mol Biosyst* 12: 2587-2593.
198. H Tang, W Chen, H Lin (2016) Identification of immunoglobulins using Chou's pseudo amino acid composition with feature selection technique. *Mol Biosyst* 12: 1269-1275.
199. AK Tiwari (2016) Prediction of G-protein coupled receptors and their subfamilies by incorporating various sequence features into Chou's general PseAAC. *Comput Methods Programs Biomed* 134: 197-213.
200. C Xu, D Sun, S Liu, et al. (2016) Protein sequence analysis by incorporating modified chaos game and physicochemical properties into Chou's general pseudo amino acid composition. *J Theor Biol* 406: 105-115.
201. HL Zou, X Xiao (2016) Predicting the functional types of singleplex and multiplex eukaryotic membrane proteins via different models of Chou's pseudo amino acid compositions. *J Membr Biol* 249: 23-29.
202. HL Zou, X Xiao (2016) Classifying Multifunctional Enzymes by Incorporating Three Different Models into Chou's General Pseudo Amino Acid Composition. *J Membr Biol* 249: 561-567.
203. PK Meher, TK Sahu, V Saini, et al. (2017) Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Sci Rep* 7: 42362.
204. H Huo, T Li, S Wang, et al. (2017) Prediction of presynaptic and postsynaptic neurotoxins by combining various Chou's pseudo components. *Sci Rep* 7: 5827.
205. YS Jiao, PF Du (2017) Predicting protein submitochondrial locations by incorporating the positional-specific physicochemical properties into Chou's general pseudo-amino acid compositions. *J Theor Biol* 416: 81-87.
206. Z Ju, JJ He (2017) Prediction of lysine propionylation sites using biased SVM and incorporating four different sequence features into Chou's PseAAC. *J Mol Graph Model* 76: 356-363.
207. M Khan, M Hayat, SA Khan, et al. (2017) Unb-DPC: Identify mycobacterial membrane protein types by incorporating unbiased dipeptide composition into Chou's general PseAAC. *J Theor Biol* 415: 13-19.
208. Y Liang, S Zhang (2017) Predict protein structural class by incorporating two different modes of evolutionary information into Chou's general pseudo amino acid composition. *J Mol Graph Model* 78: 110-117.
209. WR Qiu, QS Zheng, BQ Sun, et al. (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. *Mol Inform* 36: 1600085.
210. M Rahimi, MR Bakhtiarizadeh, A Mohammadi-Sangcheshmeh (2017) Ogenesis\_Pred: A sequence-based method for predicting oogenesis proteins by six different modes of Chou's pseudo amino acid composition. *J Theor Biol* 414: 128-136.
211. M Tahir, M Hayat, M Kabir (2017) Sequence based predictor for discrimination of enhancer and their types by applying general form of Chou's trinucleotide composition. *Comput Methods Programs Biomed* 146: 69-75.
212. P Tripathi, PN Pandey (2017) A novel alignment-free method to classify protein folding types by combining spectral graph clustering with Chou's pseudo amino acid composition. *J Theor Biol* 424: 49-54.
213. C Xu, L Ge, Y Zhang, et al. (2017) Prediction of therapeutic peptides by incorporating q-Wiener index into Chou's general PseAAC. *J Biomed Inform.*
214. B Yu, S Li, WY Qiu, et al. (2017) Accurate prediction of subcellular location of apoptosis proteins combining Chou's PseAAC and PsePSSM based on wavelet denoising. *Oncotarget* 8: 107640-107665.
215. B Yu, L Lou, S Li, et al. (2017) Prediction of protein structural class for low-similarity sequences using Chou's pseudo amino acid composition and wavelet denoising. *J Mol Graph Model* 76: 260-273.
216. J Ahmad, M Hayat (2018) MFSC: Multi-voting based feature selection for classification of golgi proteins by adopting the general form of Chou's pseaac components. *J Theor Biol* 463: 99-109.
217. MA Al Maruf, S Shatabda (2018) iRSpot-SF: Prediction of recombination hotspots by incorporating sequence based features into Chou's Pseudo components. *Genomics.*
218. M Arif, M Hayat, Z Jan (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. *J Theor Biol* 442: 11-21.
219. AH Butt, N Rasool, YD Khan (2018) Predicting membrane proteins and their types by extracting various sequence features into Chou's general PseAAC. *Mol Biol Rep.*

220. E Contreras-Torres (2018) Predicting structural classes of proteins by incorporating their global and local physicochemical and conformational properties into general Chou's PseAAC. *J Theor Biol* 454: 139-145.
221. X Cui, Z Yu, B Yu (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 (CHEMOLAB)*.
222. X Fu, W Zhu, B Liso et al. (2018) Improved DNA-binding protein identification by incorporating evolutionary information into the Chou's PseAAC. *IEEE Access* 6: 66545-66556.
223. F Javed, M Hayat (2018) Predicting subcellular localizations of multi-label proteins by incorporating the sequence features into Chou's PseAAC. *Genomics*.
224. MS Krishnan (2018) Using Chou's general PseAAC to analyze the evolutionary relationship of receptor associated proteins (RAP) with various folding patterns of protein domains. *J Theor Biol* 445: 62-74.
225. Y Liang, S Zhang (2018) Identify Gram-negative bacterial secreted protein types by incorporating different modes of PSSM into Chou's general PseAAC via Kullback-Leibler divergence. *J Theor Biol* 454: 22-29.
226. J Mei, Y Fu, J Zhao (2018) Analysis and prediction of ion channel inhibitors by using feature selection and Chou's general pseudo amino acid composition. *J Theor Biol* 456: 41-48.
227. J Mei, J Zhao (2018) Prediction of HIV-1 and HIV-2 proteins by using Chou's pseudo amino acid compositions and different classifiers. *Scientific Reports* 8: 2359.
228. J Mei, J Zhao (2018) Analysis and prediction of presynaptic and postsynaptic neurotoxins by Chou's general pseudo amino acid composition and motif features. *J Theor Biol* 447: 147-153.
229. M Mousavizadegan, H Mohabatkar (2018) Computational prediction of antifungal peptides via Chou's PseAAC and SVM. *J Bioinform Comput Biol* 16.
230. W Qiu, S Li, X Cui, et al. (2018) Predicting protein submitochondrial locations by incorporating the pseudo-position specific scoring matrix into the general Chou's pseudo-amino acid composition. *J Theor Biol* 450: 86-103.
231. SM Rahman, S Shatabda, S Saha, et al. (2018) DPP-PseAAC: A dna-binding protein prediction model using chou's general PseAAC. *J Theor Biol* 452: 22-34.
232. ES Sankari, DD Manimegalai (2018) Predicting membrane protein types by incorporating a novel feature set into Chou's general PseAAC. *J Theor Biol* 455: 319-328.
233. A Srivastava, R Kumar, M Kumar, et al. (2018) predicting and classifying beta-lactamase using a 3-tier prediction system via Chou's general PseAAC. *J Theor Biol* 457: 29-36.
234. M Tahir, H Tayara, KT Chong (2019) iRNA-PseKNC(2methyl): Identify RNA 2'-O-methylation sites by convolution neural network and Chou's pseudo components. *J Theor Biol* 465: 1-6.
235. L Zhang, L Kong (2018) iRSpot-ADPM: Identify recombination spots by incorporating the associated dinucleotide product model into Chou's pseudo components. *J Theor Biol* 441: 1-8.
236. L Zhang, L Kong (2019) iRSpot-PDI: Identification of recombination spots by incorporating dinucleotide property diversity information into Chou's pseudo components. *Genomics* 111: 457-464.
237. S Zhang, X Duan (2018) Prediction of protein subcellular localization with oversampling approach and Chou's general PseAAC. *J Theor Biol* 437: 239-250.
238. S Zhang, Y Liang (2018) Predicting apoptosis protein subcellular localization by integrating auto-cross correlation and PSSM into Chou's PseAAC. *J Theor Biol* 457: 163-169.
239. S Zhang, K Yang, Y Lei, et al. (2018) iRSpot-DTS: Predict recombination spots by incorporating the dinucleotide-based spare-cross covariance information into Chou's pseudo components. *Genomics*.
240. W Zhao, L Wang, TX Zhang, et al. (2018) A brief review on software tools in generating Chou's pseudo-factor representations for all types of biological sequences. *Protein Pept Lett* 25: 822-829.
241. S Adilina, DM Farid, S Shatabda (2019) Effective DNA binding protein prediction by using key features via Chou's general PseAAC. *J Theor Biol* 460: 64-78.
242. J Ahmad, M Hayat (2019) MFSC: Multi-voting based feature selection for classification of Golgi proteins by adopting the general form of Chou's PseAAC components. *J Theor Biol* 463: 99-109.
243. G Chen, M Cao, J Yu, et al. (2019) Prediction and functional analysis of prokaryote lysine acetylation site by incorporating six types of features into Chou's general PseAAC. *J Theor Biol* 461: 92-101.
244. W Hussain, YD Khan, N Rasool, et al. (2019) SPrenylC-PseAAC: A sequence-based model developed via Chou's 5-steps rule and general PseAAC for identifying S-prenylation sites in proteins. *J Theor Biol* 468: 1-11.
245. M Kabir, S Ahmad, M Iqbal, et al. (2019) 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*.
246. NQK Le, EKY Yapp, QT Ho, et al. (2019) iEnhancer-5Step: Identifying enhancers using hidden information of DNA sequences via Chou's 5-step rule and word embedding. *Anal Biochem* 571: 53-61.
247. Y Pan, S Wang, Q Zhang, et al. (2019) Analysis and prediction of animal toxins by various Chou's pseudo components and reduced amino acid compositions. *J Theor Biol* 462: 221-229.
248. Y Shen, J Tang, F Guo (2019) Identification of protein subcellular localization via integrating evolutionary and physicochemical information into Chou's general PseAAC. *J Theor Biol* 462: 230-239.
249. M Tahir, M Hayat, SA Khan (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. *Mol Genet Genomics* 294: 199-210.
250. B Tian, X Wu, C Chen, et al. (2019) Predicting protein-protein interactions by fusing various Chou's pseudo components and using wavelet denoising approach. *J Theor Biol* 462: 329-346.
251. AH Butt, N Rasool, YD Khan (2019) Prediction of antioxidant proteins by incorporating statistical moments based features into Chou's PseAAC. *J Theor Biol* 473: 1-8.
252. HB Shen, Chou KC (2008) PseAAC: A flexible web-server for generating various kinds of protein pseudo amino acid composition. *Anal Biochem* 373: 386-388.

253. KC Chou (2009) Pseudo amino acid composition and its applications in bioinformatics, proteomics and system biology. *Current Proteomics* 6: 262-274.
254. W Chen, H Lin, KC Chou (2015) Pseudo nucleotide composition or PseKNC: An effective formulation for analyzing genomic sequences. *Mol Biosyst* 11: 2620-2634.
255. B Liu, F Liu, X Wang, et al. (2015) Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res* 43: 65-71.
256. B Liu, F Liu, L Fang, et al. (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.
257. B Liu, F Liu, L Fang, et al. (2016) repRNA: A web server for generating various feature vectors of RNA sequences. *Mol Genet Genomics* 291: 473-481.
258. B Liu, H Wu, Wang X, et al. (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. *Nucleic Acids Res*: 67-91.
259. Z Liu, X Xiao, WR Qiu, et al. (2015) iDNA-Methyl: Identifying DNA methylation sites via pseudo trinucleotide composition. *Anal Biochem* 474: 69-77.
260. J Chen, H Liu, J Yang, et al. (2007) Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 33: 423-428.
261. KC Chou (2001) Using subsite coupling to predict signal peptides. *Protein Eng* 14: 75-79.
262. KC Chou (2001) Prediction of signal peptides using scaled window. *Peptides* 22: 1973-1979.
263. KC Chou (2001) Prediction of protein signal sequences and their cleavage sites. *Proteins* 42: 136-139.
264. W Chen, PM Feng, H Lin, et al. (2013) iRSpot-PseDNC: Identify recombination spots with pseudo dinucleotide composition. *Nucleic Acids Res* 41: e68.
265. KC Chou, HB Shen (2006) Hum-PLoc: A novel ensemble classifier for predicting human protein subcellular localization. *Biochem Biophys Res Commun* 347: 150-157.
266. KC Chou, HB Shen (2006) Addendum to "Hum-PLoc: A novel ensemble classifier for predicting human protein subcellular localization". *Biochem Biophys Res Commun* 348: 1479.
267. HB Shen, KC Chou (2007) Gpos-PLoc: An ensemble classifier for predicting subcellular localization of Gram-positive bacterial proteins *Protein Eng Des Sel* 20: 39-46.
268. HB Shen, KC Chou (2007) Virus-PLoc: A fusion classifier for predicting the subcellular localization of viral proteins within host and virus-infected cells. *Biopolymers* 85: 233-240.
269. HB Shen, KC Chou (2007) Nuc-PLoc: A new web-server for predicting protein subnuclear localization by fusing PseAA composition and PsePSSM. *Protein Eng Des Sel* 20: 561-567.
270. HB Shen, J Yang, KC Chou (2007) Euk-PLoc: An ensemble classifier for large-scale eukaryotic protein subcellular location prediction. *Amino Acids* 33: 57-67.
271. KC Chou, HB Shen (2008) Cell-PLoc: A package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat Protoc* 3: 153-162.
272. KC Chou, HB Shen (2010) Cell-PLoc 2.0: An improved package of web-servers for predicting subcellular localization of proteins in various organisms. *Nat Sci* 2: 1090-1103.
273. KC Chou, ZC Wu, X Xiao (2011) iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular Localization of Singleplex and Multiplex Eukaryotic Proteins *PLoS One* 3: 18258.
274. ZC Wu, X Xiao, KC Chou (2011) iLoc-Plant: A multi-label classifier for predicting the subcellular localization of plant proteins with both single and multiple sites. *Mol Biosyst* 7: 3287-3297.
275. X Xiao, ZC Wu, KC Chou (2011) iLoc-Virus: A multi-label learning classifier for identifying the subcellular localization of virus proteins with both single and multiple sites. *J Theor Biol* 284: 42-51.
276. KC Chou, ZC Wu, X Xiao (2012) iLoc-Hum: Using accumulation-label scale to predict subcellular locations of human proteins with both single and multiple sites. *Mol Biosyst* 8: 629-641.
277. ZC Wu, X Xiao (2012) iLoc-Gpos: A multi-layer classifier for predicting the subcellular localization of singleplex and multiplex gram-positive bacterial proteins. *Protein Pept Lett* 19: 4-14.
278. WZ Lin, JA Fang, X Xiao (2013) iLoc-Animal: A multi-label learning classifier for predicting subcellular localization of animal proteins. *Mol Biosyst* 9: 634-644.
279. ZD Su, Y Huang, ZY Zhang, et al. (2018) iLoc-lncRNA: Predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics* 34: 4196-4204.
280. X Cheng, X Xiao, Chou KC (2017) pLoc-mPlant: Predict subcellular localization of multi-location plant proteins via incorporating the optimal GO information into general PseAAC. *Mol Biosyst* 13: 1722-1727.
281. X Cheng, X Xiao, Chou KC (2017) pLoc-mVirus: Predict subcellular localization of multi-location virus proteins via incorporating the optimal GO information into general PseAAC. *Gene* 628: 315-321.
282. X Cheng, SG Zhao, WZ Lin, et al. (2017) pLoc-mAnimal: Predict subcellular localization of animal proteins with both single and multiple sites. *Bioinformatics* 33: 3524-3531.
283. X Xiao, X Cheng, S Su, et al. (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.
284. X Cheng, X Xiao (2018) pLoc-mEuk: Predict subcellular localization of multi-label eukaryotic proteins by extracting the key GO information into general PseAAC. *Genomics* 110: 50-58.
285. X Cheng, X Xiao (2018) pLoc-mGneg: Predict subcellular localization of Gram-negative bacterial proteins by deep gene ontology learning via general PseAAC. *Genomics* 110: 231-239.
286. X Cheng, X Xiao (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.
287. X Cheng, X Xiao (2018) pLoc\_bal-mGneg: Predict subcellular localization of Gram-negative bacterial proteins by quasi-balancing training dataset and general PseAAC. *J Theor Biol* 458: 92-102.
288. X Cheng, X Xiao (2018) pLoc\_bal-mPlant: Predict subcellular

- localization of plant proteins by general PseAAC and balancing training dataset. *Curr Pharm Des* 24: 4013-4022.
289. KC Chou, X Cheng, Xiao (2018) pLoc\_bal-mHum: Predict subcellular localization of human proteins by PseAAC and quasi-balancing training dataset. *Genomics*.
290. X Xiao, X Cheng, G Chen, et al. (2019) pLoc\_bal-mGpos: Predict subcellular localization of Gram-positive bacterial proteins by quasi-balancing training dataset and PseAAC. *Genomics* 111: 886-892.
291. X Cheng, WZ Lin, X Xiao (2019) pLoc\_bal-mAnimal: Predict subcellular localization of animal proteins by balancing training dataset and PseAAC. *Bioinformatics* 35: 398-406.
292. KC Chou (2019) Progresses in predicting post-translational modification. *International Journal of Peptide Research and Therapeutics*.
293. X Cheng, SG Zhao, X Xiao (2017) iATC-mISF: A multi-label classifier for predicting the classes of anatomical therapeutic chemicals. *Bioinformatics* 34: 341-346.
294. X Cheng, SG Zhao, X Xiao (2017) iATC-mHyb: A hybrid multi-label classifier for predicting the classification of anatomical therapeutic chemicals. *Oncotarget* 8: 58494-58503.
295. KC Chou (2013) Some remarks on predicting multi-label attributes in molecular biosystems. *Mol Biosyst* 9: 1092-1100.
296. KC Chou (2019) Advance in predicting subcellular localization of multi-label proteins and its implication for developing multi-target drugs. *Curr Med Chem*.
297. H Lin, EZ Deng, H Ding, et al. (2014) iPro54-PseKNC: A sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition. *Nucleic Acids Res* 42: 12961-12972.
298. X Zhai, M Chen, W Lu (2018) Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods. *Computational Materials Science* 151: 41-48.
299. J Jia, X Li, W Qiu, et al. (2019) iPPI-PseAAC(CGR): Identify protein-protein interactions by incorporating chaos game representation into PseAAC. *J Theor Biol* 460: 195-203.
300. YD Khan, M Jamil, W Hussain, et al. (2019) pSSbond-PseAAC: Prediction of disulfide bonding sites by integration of PseAAC and statistical moments. *J Theor Biol* 463: 47-55.
301. VK Shyamili, A Vellaichamy (2019) Sequence and structure-based characterization of human and yeast ubiquitination sites by using Chou's sample formulation. *Proteins: Structure, function and bioinformatics*.
302. M Awais, W Hussain, YD Khan, et al. (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 Trans Comput Biol Bioinform*.
303. P Feng, H Yang, H Ding, et al. (2019) iDNA6mA-PseKNC: Identifying DNA N(6)-methyladenosine sites by incorporating nucleotide physicochemical properties into PseKNC. *Genomics* 111: 96-102.
304. Lu F, Zhu M, Lin Y, et al. (2019) The preliminary efficacy evaluation of the CTLA-4-Ig treatment against Lupus nephritis through in-silico analyses. *J Theor Biol* 471: 74-81.
305. Yi Lu, Shuo Wang, Jianying Wang, et al. (2019) An ep10. idemic avian influenza prediction model based on google trends. *Letters in Organic Chemistry* 16: 303.
306. Niu B, Liang C, Lu Y, et al. (2019) Glioma stages prediction based on machine learning algorithm combined with protein-protein interaction networks. *Genomics*.
307. KC Chou (2019) An insightful 20-year recollection since the birth of pseudo amino acid components. *Protein & Peptide Letters in press*.
308. KC Chou (2019) Artificial intelligence (AI) tools constructed via the 5-steps rule for predicting post-translational modifications. *Artificial Intelligence in press*.
309. KC Chou (2019) Recent progresses in predicting protein subcellular localization with artificial intelligence (AI) tools developed via the 5-steps rule. *Artificial Intelligence in press*.
310. KC Chou (2019) An insightful 10-year recollection since the emergence of the 5-steps rule. *Current Pharmaceutical Design in press*.
311. KC Chou (2019) Impacts of pseudo amino acid components and 5-steps rule to proteomics and proteome analysis. *Proteomics in press*.
312. KC Chou (2019) Distorted key theory and its implication for drug development. *Current Proteomics in press*.
313. KC Chou (2019) Proposing pseudo amino acid components is an important milestone for proteome and genome analyses. *International Journal for Peptide Research and Therapeutics (IJPT)* in press.
314. KC Chou (2019) Two kinds of metrics for computational biology. *Genomics in press*.
315. KC Chou, Forsen S (1980) Diffusion-controlled effects in reversible enzymatic fast reaction system - critical spherical shell and proximity rate constants. *Biophysical Chemistry* 12: 255-263.
316. KC Chou, Li TT, Forsen S (1980) The critical spherical shell in enzymatic fast reaction systems. *Biophysical Chemistry* 12: 265-269.
317. Li TT, Forsen S (1980) The flow of substrate molecules in fast enzyme-catalyzed reaction systems. *Chemica Scripta* 16: 192-196.
318. KC Chou, Forsen S (1980) Graphical rules for enzyme-catalyzed rate laws. *Biochem J* 187: 829-835.
319. KC Chou, Forsen S, Zhou GQ (1980) Three schematic rules for deriving apparent rate constants. *Chemica Scripta* 16: 109-113.
320. KC Chou, Carter RE, Forsen S (1981) A new graphical method for deriving rate equations for complicated mechanisms. *Chemica Scripta* 18: 82-86.
321. KC Chou, Forsen S (1981) Graphical rules of steady-state reaction systems. *Canadian Journal of Chemistry* 59: 737-755.
322. KC Chou, Chen NY, Forsen S (1981) The biological functions of low-frequency phonons: 2. Cooperative effects. *Chemica Scripta* 18: 126-132.
323. KC Chou, Jiang SP, Liu WM, et al. (1979) Graph theory of enzyme kinetics: 1. Steady-state reaction system. *Scientia Sinica* 22: 341-358.
324. Zhou GP, Deng MH (1984) An extension of Chou's graphic rules for deriving enzyme kinetic equations to systems involving parallel reaction pathways. *Biochem J* 222: 169-176.

325. KC Chou (1989) Graphic rules in steady and non-steady enzyme kinetics. *Journal of Biological Chemistry* 264: 12074-12079.
326. KC Chou (1990) Applications of graph theory to enzyme kinetics and protein folding kinetics: Steady and non-steady state systems. *Biophysical Chemistry* 35: 1-24.
327. Althaus IW, Chou JJ, Gonzales AJ, et al. (1993) Steady-state kinetic studies with the non-nucleoside HIV-1 reverse transcriptase inhibitor U-87201E. *J Biol Chem* 268: 6119-6124.
328. Althaus IW, Gonzales AJ, Chou JJ, et al. (1993) The quinoline U-78036 is a potent inhibitor of HIV-1 reverse transcriptase. *J Biol Chem* 268: 14875-14880.
329. Althaus IW, Chou JJ, Gonzales AJ, et al. (1993) Kinetic studies with the nonnucleoside HIV-1 reverse transcriptase inhibitor U-88204E. *J Biol Chem* 32: 6548-6554.
330. Althaus IW, Chou JJ, Gonzales AJ, et al. (1994) Steady-state kinetic studies with the polysulfonate U-9843, an HIV reverse transcriptase inhibitor. *Cellular and Molecular Life Science (Experientia)* 50: 23-28.
331. Althaus IW, Chou JJ, Gonzales AJ, et al. (1994) Kinetic studies with the non-nucleoside human immunodeficiency virus type-1 reverse transcriptase inhibitor U-90152E. *Biochem Pharmacol* 47: 2017-2028.
332. KC Chou, Kezdy FJ, Reusser F (1994) Kinetics of processive nucleic acid polymerases and nucleases. *Analytical Biochemistry* 221: 217-230.
333. Althaus IW, Franks KM, Diebel MR, et al. (1996) The benzylthio-pyrididine U-31,355, a potent inhibitor of HIV-1 reverse transcriptase. *Biochem Pharmacol* 51: 743-750.
334. Andraos J (2008) Kinetic plasticity and the determination of product ratios for kinetic schemes leading to multiple products without rate laws: New methods based on directed graphs. *Canadian Journal of Chemistry* 86: 342-357.
335. KC Chou, Shen HB (2009) FoldRate: A web-server for predicting protein folding rates from primary sequence. *The Open Bioinformatics Journal* 3: 31-50.
336. Shen HB, Song JN, Chou K (2009) Prediction of protein folding rates from primary sequence by fusing multiple sequential features. *Journal of Biomedical Science and Engineering (JBISE)* 2: 136-143.
337. KC Chou (2010) Graphic rule for drug metabolism systems. *Curr Drug Metab* 11: 369-378.
338. KC Chou, Lin WZ, Xiao X (2011) Wenxiang: A web-server for drawing wenxiang diagrams. *Natural Science* 3: 862-865.
339. Zhou GP (2011) The disposition of the LZCC protein residues in wenxiang diagram provides new insights into the protein-protein interaction mechanism. *J Theor Biol* 284: 142-148.

**DOI: 10.36959/643/304**