

# A Short Survey on Genetic Sequences, Chou's Pseudo Amino Acid Composition and its Combination with Fuzzy Set Theory

D.N. Georgiou<sup>1</sup>, T.E. Karakasidis<sup>2,\*</sup> and A.C. Megaritis<sup>3</sup>

<sup>1</sup>University of Patras, Department of Mathematics, 265 00 Patras, Greece

<sup>2</sup>University of Thessaly, Department of Civil Engineering, 383 34 Volos, Greece

<sup>3</sup>Technological Educational Institute of Messolonghi, Department of Accounting, 30200 Messolonghi, Greece

**Abstract:** The study of genetic sequences is of great importance in biology and medicine. Sequence analysis and taxonomy are two major fields of application of bioinformatics. In this survey, we present results concerning genetic sequences and Chou's pseudo amino acid composition as well as methodologies developed based on this concept along with elements of fuzzy set theory, and emphasize on fuzzy clustering and its application in analysis of genetic sequences.

**Keywords:** DNA, Amino acids, genetic sequences, fuzzy clustering, bioinformatics.

## 1. INTRODUCTION

The study of genetic sequences presents particular interest since it is of great importance for diagnosis reasons and taxonomy. Several efforts have been made in this direction. The computational effort for such purposes becomes significant as the length of polynucleotides increases and there is also a difficulty when it comes to compare polynucleotides of different length. An important advancement came from the introduction of the pseudo amino acid concept by Chou in 2001 [1] which is based on the transformation of the polynucleotide to a limited size vector based on several key properties like the hydrophobicity, hydrophilicity and side chain molecular weight as well as other approaches.

This approach has been employed in a large number of studies of protein attributes, such as identifying bacterial virulent proteins [2], predicting super-secondary structure [3], predicting protein subcellular location [4-6], predicting membrane protein types [7], discriminating outer membrane proteins [8], identifying antibacterial peptides [9, 10], identifying allergenic proteins [11], predicting metalloproteinase family [12], predicting protein structural class [13], identifying GPCRs and their types [14], identifying protein quaternary structural attributes [15], predicting protein submitochondria locations [16], identifying risk type of human papillomaviruses [17], identifying cyclin proteins [18], predicting GABA(A) receptor proteins [19], among many others (see a long list of papers cited in the References section of [20]). Recently, the concept of PseAAC was further extended to represent the feature vectors of DNA and nucleotides [21, 22], as well as other biological samples (see, for example [23-25]). Because it has been widely and

increasingly used, recently two powerful softwares, specifically 'PseAAC-Builder' [26] and 'propy' [27], were introduced for generating various special Chou's pseudo amino acid compositions, in addition to the web server 'PseAAC' [28] built in 2008. Furthermore, it can also be used to study the biological systems with multi-labels as indicated in [10, 29, 30] and a recent review [30].

Many studies from various research laboratories around the world have indicated that mathematical analysis, computational modeling, and the introduction of novel physical concepts to solve important problems in genetics sequences classification can timely provide very useful information and insights for basic research and hence are widely welcomed by science community. Fuzzy theory has been used to identify G-protein-coupled receptor functional classes [31], nuclear receptor subfamilies [32,33], antimicrobial peptides and their functional types [34]. It has also been employed in identifying Nuclear Receptor Subfamilies Based on Sequence-Derived Features [35] and predicting G protein-coupled receptors and their types by hybridizing two different modes of pseudo amino acid compositions [36].

The paper is structured as follows: Section 2 presents results concerning genetic sequences and Chou's pseudo amino acid composition as well as other ideas of transformation of long polynucleotides to representations in a lower dimension space and relevant work. Section 3 presents notions and methodologies concerning fuzzy sets and fuzzy clustering and mentions work performed combining the two methodologies.

## 2. DNA, RNA, AND POLYNUCLEOTIDE SPACE [0,1]<sup>12</sup>

### 2.1. Genetic Sequences

DNA and RNA are made of triplets XYZ of codons each of them having the possibility to be one of four nucleotides

\*Address correspondence to this author at the Department of Civil Engineering, University of Thessaly, 38334 Volos, Greece;  
Tel: +302421074163; Fax: +302421074169;  
E-mail: [thkarak@uth.gr](mailto:thkarak@uth.gr)

$\{U, C, A, G\}$  in the case of DNA and  $\{T, C, A, G\}$  in the case of RNA (A = Adenine, C = Cytosine, G = Guanine, T = Thymine, U=Uracil).

In the case of RNA alphabet if U is the first letter of this alphabet one codes it as  $(1, 0, 0, 0)$ : 1 because the first letter U is present, 0 since the second letter does not appear, 0 since the third letter is not present and 0 since the fourth letter G does not appear. In a similar way C is represented as  $(0, 1, 0, 0)$ , A as  $(0, 0, 1, 0)$  and G as  $(0, 0, 0, 1)$ . So if we have a nucleotide described by the codon UCG (serine) this would be written in the  $[0, 1]^{12}$  hypercube as

$(1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1)$ .

There are cases where the exact chemical structure of the sequence is not known for the complete sequence. In this case some components of its fuzzy code being neither 0 nor 1 but a value in the interval  $(0, 1)$  are sequences not necessarily at a corner of the hypercube.

First Professor Sadegh-Zadeh [37] showed that nucleic acids (DNA and RNA) can be treated as ordered fuzzy sets in a 12-dimensional space. The genetic code can be represented in a 12-dimensional space because a triplet codon  $XYZ$  has a  $3 \times 4 = 12$  dimensional fuzzy code  $(a_1, \dots, a_{12})$  and it is a point in the 12-dimensional fuzzy polynucleotide space  $[0, 1]^{12}$  as a subspace of the real space  $[0, \infty]^{12}$ . Sadegh-Zadeh (see [37]) and Nieto *et al.* see [38-40] introduced the Fuzzy Polynucleotide Space (FPS) based on the principle of the fuzzy hypercube [41].

The situation becomes more complex when dealing with polynucleotides since a polynucleotide consisting of a sequence of  $k$  triplets  $XYZ$  can be seen as a point in a  $[0, 1]^{12 \times k}$  space. This means that as the length of polynucleotides increases the comparison and taxonomy procedure presents increased complexity [1]. In his initial approach, Chou instead of dealing with the whole polynucleotide proposed a methodology that takes into account the correlation of properties of the residues of the polynucleotide based on their physical properties such as hydrophobicity, hydrophilicity and side-chain mass [1]. This approach resulted in a representation of the amino acid with a  $(20 + \lambda)$  dimensional vector, where  $\lambda$  is the degree of correlation. In such an approach, the reduced representation carries sequence order effects. In case the  $(20 + \lambda)$  components were imposed by the normalized condition as in the case of the classical amino acid composition, the dimension-reduced operation is often needed when performing the prediction with some operation engines (such as the covariant discriminant [42]) in order to avoid the divergence problem. However, which one of these normalized components should be removed? Will the result be different by removing a different component? To address these problems, the Chou's Invariance Theorem was developed in 1995. According to the Chou's Invariance Theorem, the outcome of the covariant discriminant will remain exactly the same regardless of which one of the components is left out. In other words, any one of the constituent normalized components can be left out to

overcome the divergence problem without changing the final result. For more information about Chou's Invariance Theorem and its applications as well as the relevant references, see a Wikipedia article at <http://en.wikipedia.org/wiki/Chou>.

In a similar spirit, Torres & Nieto [38] mapped a polynucleotide on a  $[0, 1]^{12}$  space by considering the frequencies of appearance of the nucleotides at the three base sites of a codon in the coding sequence. This would result in even more reduction of the information that one should devote to represent an amino acid. In this representation, only the nature of residues plays role in the representation. In that work using a metric motivated by publications of Lin [43] and Sadegh-Zadeh [37], they calculated distances between nucleotides. They also applied their algorithm for the comparison of complete genomes (for example: *M.tuberculosis* and *E.coli*). Further work has been recently performed using the idea of Nieto *et al.* [40] in which the influence of several metrics has been examined. The advantages of this methodology are:

- a) One can compare polynucleotides of very big length in a very computationally efficient way and
- b) One can apply the algorithm in order to compare polynucleotides of different length as it is the case for genomes of different organisms.

We point out that metrics play an important role in computational biology. Different metrics have been used to study secondary structures see [44] or biopolymer contact structures (see [45]). The interest in this domain as well as many important biological and medical implications of the study of genetic sequences is reflected on the number of several works performed (see for example [46-51]).

## 2.2. Polynucleotides

The genetic code consists of strings made up from four constitutive elements which correspond to the four nucleic acids: A (adenosine), T (thymidine), C (cytidine), and G (guanosine). Amino acids (or codons) are strings consisting of three nucleic acids. Provided that we have four letters (nucleic acids) with the possibility of three positions in a codon, this results in 64 possible combinations which correspond to 64 possible amino acids (See the table, for example, in <http://psyche.uthct.edu/shaun/SBlack/geneticd.html>, Freeland and Hurst 1998 [52]). Three of these possible codons specify the termination of the polypeptide chain and thus they are called "stop codons". This leaves 61 codons to specify only 20 different amino acids.

There is an important problem which consists of how to proceed to taxonomy of polynucleotides and how similar they are. In this frame, a reduced common base to describe a polynucleotide like Chou's pseudo amino acid composition concept can be very useful. In fact in an earlier paper [53], the physicochemical distance among the 20 amino acids [54] was adopted to define PseAA. Subsequently, several researchers employed the complexity measure factor [55], results from cellular automata analysis have also been presented ([56-59]), several researchers used hydrophobicity and/or hydrophilicity properties [60-66], while others employed Fourier transform [67, 68].

The pseudo amino acid composition was originally introduced in order to improve the prediction for protein subcellular localization and membrane protein type [1], as well as for enzyme functional class [61]. The pseudo amino acid composition has the advantage that on one hand it can represent a protein sequence with a discrete model without on the other hand losing its sequence-order information [69]. For this reasons, it can be very useful in the analysis of a large number complex sequences in taxonomic studies. Later, many researchers have studied the protein and yielded interesting results. Indicative examples are presented in a number of research papers (see [1, 43, 53, 55-59, 60-62, 65-66, 69-103]). Several excellent reviews of these results can also be found in [104-105].

In the next section we present a brief description of the fuzzy sets theory.

### 3. FUZZY SET THEORY

In this section, we give basic notions of Fuzzy Set Theory which are essential and have important applications to issues associated with genetic sequences and Chou's pseudo amino acid composition.

#### 3.1. Fuzzy Sets and Similarity

Let  $X$  be a set.  $A$  is a *fuzzy subset* of  $X$  if  $A = \{(x, \mu_A(x)) : x \in X\}$ , where  $\mu_A$  is a function of  $X$  into  $[0,1] = I$ , that is  $A$  is the set of all pairs  $(x, \mu_A(x))$  such that  $x \in X$  and  $\mu_A(x)$  are the degrees of its membership in  $A$ . In what follows if  $X = \{x_1, x_2, \dots, x_n\}$  and

$$A = \{(x_1, \mu_A(x_1)), \dots, (x_n, \mu_A(x_n))\} \quad (1)$$

then we write

$$A = (\mu_A(x_1), \dots, \mu_A(x_n)) \quad (2)$$

Kosko [41] introduced a geometrical interpretation of fuzzy sets as points in a hypercube. For a given set  $X = \{x_1, x_2, \dots, x_n\}$ , the set of all fuzzy subsets (of  $X$ ) is precisely the unit hypercube  $I^n = [0,1]^n$ , since any fuzzy subset  $A$  determines a point  $P \in I^n$  given by  $P = (\mu_A(x_1), \dots, \mu_A(x_n))$ . Reciprocally, any point  $P = (a_1, \dots, a_n) \in I^n$  generates a fuzzy subset  $A$  of  $X$  defined by the map  $\mu_A : X \rightarrow [0,1]$  such that  $\mu_A(x_i) = a_i$ ,  $i = 1, 2, \dots, n$ .

Non-fuzzy or crisp subsets of  $X = \{x_1, \dots, x_n\}$  are given by mappings

$$\mu : X \rightarrow \{0,1\} \quad (3)$$

from the set  $X$  into the set  $\{0,1\}$  and they are located at the  $2^n$  corners of the  $n$ -dimensional unit hypercube  $I^n$ . So, the ground set  $X = \{x_1, \dots, x_n\}$  is itself the fuzzy set  $(1,1, \dots, 1) \in I^n$ . Also, the empty fuzzy set is the fuzzy set  $(0,0, \dots, 0) \in I^n$ , denoted by  $\emptyset$ .

Hypercubical calculus is developed in [106], and some applications of the fuzzy unit hypercube are given in [39,107-109]. In this context, a codon corresponds to a corner of the 12-dimensional unit hypercube  $[0,1]^{12}$ . Any element of  $[0,1]^{12}$  may be viewed as a fuzzy codon.

Consider the  $n$ -dimensional unit hypercube  $I^n$ . If

$$p = (p_1, \dots, p_n), q = (q_1, \dots, q_n) \in I^n \quad (4)$$

are two different fuzzy polynucleotides, then we consider the following distance between the elements  $p$  and  $q$ :

$$d(p, q) = \frac{\sum_{i=1}^n |p_i - q_i|}{\sum_{i=1}^n \max\{p_i, q_i\}} \quad (5)$$

Also, if  $p = q = \emptyset = (0, \dots, 0)$ , then  $d(\emptyset, \emptyset) = 0$  (see [110]).

The distance  $d$  is motivated by publications [37] and [43]. We know that  $d$  is a metric [110] and has already been employed in [38] and [40]. In [111] (see, also [112]) it is proposed to call this metric as the *NTV metric*.

Let  $X = \{x_1, \dots, x_n\}$  be a set and  $A = (a_1, \dots, a_n)$ , and  $B = (b_1, \dots, b_n)$ , where  $a_i, b_i \in [0,1]$ , two fuzzy sets of  $X$ . The degree of similarity between  $A$  and  $B$  (see [40]), denoted by  $sim(A, B)$  is defined to be the number

$$sim(A, B) = \frac{\min\{a_1, b_1\} + \dots + \min\{a_n, b_n\}}{\frac{a_1 + b_1}{2} + \dots + \frac{a_n + b_n}{2}} \quad (6)$$

#### 3.2. Fuzzy Relations and Fuzzy Clustering

When performing clustering, an important parameter is the metric (see [113]) employed in the distance calculation between the elements to be classified. These elements are considered as point in a finite dimensional space.

In what follows by  $R$  we denote a *fuzzy relation* on a set  $X$ , that is a fuzzy set in the direct product  $X \times X = \{(x, y) : x, y \in X\}$  which is characterized by the membership function:

$$\mu_R : X \times X \rightarrow [0,1] \quad (7)$$

Also by  $\mathfrak{R}$  we denote the set of all real numbers and by  $\mathfrak{R}^+$  the set of all positive real numbers.

Let  $X = \{x_1, x_2, \dots, x_n\}$  be a finite set. A fuzzy relation  $R$  in  $X \times X$  can be expressed by a  $n \times n$  matrix as following:

$$R = \begin{pmatrix} \mu_R(x_1, x_1) & \mu_R(x_1, x_2) & \cdots & \mu_R(x_1, x_n) \\ \mu_R(x_2, x_1) & \mu_R(x_2, x_2) & \cdots & \mu_R(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_R(x_n, x_1) & \mu_R(x_n, x_2) & \cdots & \mu_R(x_n, x_n) \end{pmatrix} \quad (8)$$

A fuzzy relation  $R$  on  $X$  is:

- (1) *reflexive* if  $\mu_R(x, x) = 1$  for all  $x \in X$ ,  
 (2) *symmetric* if  $\mu_R(x, y) = \mu_R(y, x)$  for all  $x, y \in X$ , and  
 (3) *max-min transitive* if

$$\mu_R(x, z) \geq \sup\{\min\{\mu_R(x, y), \mu_R(y, z)\} : y \in X\} \quad (9)$$

A fuzzy relation with the above properties is called *fuzzy similarity relation* or *fuzzy equivalence relation*.

A fuzzy relation on  $X$  that is reflexive and symmetric is usually called a *compatibility relation*.

The *max-min transitive closure* of a fuzzy relation  $R$  on  $X$  is defined as the smallest max-min fuzzy transitive relation containing  $R$ .

It is known that (see, for example, [114]) if  $R$  is a fuzzy compatibility relation on a finite set  $X = \{x_1, \dots, x_n\}$ , then the max-min transitive closure  $R_T$  is the relation  $R^{(n-1)} = R \circ \dots \circ R$  i.e., relation  $R$  composed with itself  $(n-1)$  times.

We note that if  $R$  and  $S$  are two fuzzy relations on  $X$ , the composition is characterized by the membership function:

$$\mu_{R \circ S}(x, z) = \sup\{\min\{\mu_R(x, y), \mu_S(y, z)\} : y \in X\} \quad (10)$$

Several techniques exist for the extraction of genome characteristics. A common approach is the investigation of common characteristic properties between its constituent elements. Fuzzy clustering is a procedure that can be employed in this direction. There are several methods of fuzzy clustering. Two of the most commonly employed approaches in fuzzy clustering are:

a) The fuzzy c-means algorithm [115], which needs an a priori definition of the number of classes (called clusters) and its final result critically depends on this choice.

b) The fuzzy equivalence relation-based hierarchical clustering method (see for example, [116, 117]) which does not need any a priori assumption on the number of classes. This is a significant advantage since we avoid any unwanted biases on the classified data.

A simple fuzzy cluster analysis of amino acids has been introduced by Mocz [118] to recognize secondary structure in proteins. Biological distances between the twenty aminoacids can be calculated based on their properties. Since the introduction of the pseudo amino acid composition by Chou [1], a significant number of efforts have been performed in order to employ different quantities in order to create representations of the twenty aminoacids with the aim to better capture and represent the sequence order effects through the principle of pseudo amino acid composition (PseAA).

In order to illustrate the clustering methodology based on fuzzy equivalence relations (see, for example, [53, 72-74]), let's consider a data set:

$$X = \{x_1, x_2, \dots, x_n\} \quad (11)$$

where

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in \mathfrak{R}^m \quad (12)$$

and  $i = 1, 2, \dots, n$ . Then we proceed with the following three steps:

**Step 1:** We define a fuzzy relation  $R$  on  $X$ , using the distance function of Minkowski, via the membership function:

$$\mu_R(x_i, x_k) = 1 - \frac{(\sum_{j=1}^m |x_{ij} - x_{kj}|^q)^{1/q}}{d} \quad (13)$$

for all  $(x_i, x_k) \in X \times X$ , where  $q \in \mathfrak{R}^+$  and

$$d = \max\{(\sum_{j=1}^m |x_{ij} - x_{kj}|^q)^{1/q} : i, k = 1, 2, \dots, n\} \quad (14)$$

Clearly  $R$  is a fuzzy compatibility relation but not necessarily a fuzzy equivalence relation [117].

**Remark:** Also we use (see Section 3.2) the distance function of NTV (see, [110]-[112]) via the membership function:

$$\mu_R(x_i, x_k) = 1 - \frac{\sum_{j=1}^m |x_{ij} - x_{kj}|}{\sum_{j=1}^m \max\{x_{ij} - x_{kj}\}} \quad (15)$$

for all  $(x_i, x_k) \in X \times X$  and  $x_i, x_k \neq (0, 0, \dots, 0) \in \mathfrak{R}^m$ .

**Step 2:** We find the max-min transitive closure  $R_T$ .

**Step 3:** For every  $a \in [0, 1]$ , call the *degree of similarity*, we define a new fuzzy matrix  $R_T^a$  as follows

$$\mu_{R_T^a}(x, y) = \begin{cases} 1 & \text{if } \mu_{R_T}(x, y) \geq a \\ 0 & \text{if } \mu_{R_T}(x, y) < a \end{cases} \quad (16)$$

The intervals of  $a$  that determine the partitions are derived from the values of the matrix  $R_T$ . In this way by the examination of the  $R_T$  matrix, we can determine the resulting partitions for all intervals of  $a$ -cuts.

### 3.3. Applications of Fuzzy Set Theory

Several properties can be used in the clustering methodology. Such properties can be selected from the AAindex database (see [119, 120]) and motivate selection of the considered properties. The results of such classifications can contribute to the prediction of subcellular location of proteins, which is a very hot topic of research in the domain of bioinformatics. The resulting clusters may also be very helpful in understanding the origin and emergence of the alphabet of amino acids encoded by the standard genetic code.

A clustering analysis of the twenty amino acids based on several physical properties is presented in [121]. The properties employed in the analysis are the following: the number of codons that code the protein, molecular weight, hydrophobicity, the number of atoms of different type and the corresponding number of protons as well as the number of total protons. The analysis concerned the influence of the

properties on the classification procedure and the resulting clusters as well as the effect of the metric employed in the clustering procedure.

Recently, Stepehn and Freeland [122] have presented the first quantitative exploration of nature's "choices" set against various models for plausible alternatives with the help of computational chemistry. This analysis made clear that the fuzzy approach such as fuzzy clustering and fuzzy cognitive maps can be very useful in the field of the protein content prediction and the prediction of protein structural classes.

Fuzzy set theory has also been employed using the pseudoamino acid composition concept also in other works. Representative results are that of Ding *et al.* [123] who employed fuzzy support vector machine for the prediction of protein structure classes with pseudo aminoacid composition, Shen *et al.* [124] who used Fuzzy KNN for predicting membrane protein types from pseudo aminoacid composition, Hayat *et al.* [125] who employed fuzzy K-nearest neighbor algorithms based on Chou's pseudo aminoacid composition, Shen *et al.* [126] who applied supervised fuzzy clustering to predict protein structural classes and Goergiou *et al.* [121] who applied fuzzy clustering techniques to classify aminoacids based on Chou's aminoacid composition.

#### 4. CONCLUSIONS

In a nutshell, we can say that the aminoacid composition principle introduced by Chou has greatly contributed to the taxonomy, study and prediction in bioinformatics. It has also initiated further work in this direction of studying genetic sequences in a reduced representation space by other researchers in the field. The combination of such principles with elements of fuzzy set theory seems to be very promising for further extension in the research of genetic sequences and necessitates further interdisciplinary research.

#### CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

#### ACKNOWLEDGEMENT

Declared none.

#### REFERENCES

- [1] K.C. Chou, "Prediction of protein cellular attributes using pseudo amino acid composition", *PROTEINS: Structure, Function, and Genetics.*, (Erratum: *ibid.*, vol.44, 60, 2001) vol. 43, pp. 246-255, 2001.
- [2] L. Nanni, A. Lumini, D. Gupta, and A. Garg, "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 Trans. Comput. Biol. Bioinform.*, vol. 9, pp. 467-475, 2012.
- [3] D. Zou, Z. He, J. He, and Y. Xia, "Supersecondary structure prediction using Chou's pseudo amino acid composition", *J. Comput. Chem.*, vol. 32, pp. 271-278, 2011.
- [4] S. W. Zhang, Y.L. Zhang, H.F. Yang, C. H. Zhao, and Q. Pan, "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*, vol. 34, pp. 565-572, 2008.
- [5] K.K. Kandaswamy, G. Pugalenth, S. Moller, E. Hartmann, K.U. Kalies, P.N. Suganthan, and T. Martinetz, "Prediction of apoptosis protein locations with genetic algorithms and support vector machines through a new mode of pseudo amino acid composition", *Protein Peptide Lett.*, vol. 17, pp. 1473-1479, 2010.
- [6] S. Mei, "Predicting plant protein subcellular multi-localization by Chou's PseAAC formulation based multi-label homolog knowledge transfer learning", *J. Theor. Biol.*, vol. 310, pp. 80-87, 2012.
- [7] Y. K. Chen, and K. B. Li, "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.*, vol. 318, pp. 1-12, 2013.
- [8] M. Hayat, and A. Khan, "Discriminating outer membrane proteins with fuzzy k-nearest neighbor algorithms based on the general form of Chou's PseAAC", *Protein Peptide Lett.*, vol. 19, pp. 411-421, 2012.
- [9] Khosravian M., Faramarzi F. K., Beigi M. M., Behbahani M., and Mohabatkhar H., "Predicting antibacterial peptides by the concept of Chou's pseudo-amino acid composition and machine learning methods", *Protein Peptide Lett.*, vol. 20, pp. 180-186, 2012.
- [10] X. Xiao, P. Wang, W. Z. Lin, J. H. Jia, and K. C. Chou, "iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types", *Anal. Biochem.*, vol. 436, pp. 168-177, 2013.
- [11] H. Mohabatkhar, M. M. Beigi, K. Abdolahi, and S. Mohsenzadeh, "Prediction of Allergenic Proteins by Means of the Concept of Chou's Pseudo Amino Acid Composition and a Machine Learning Approach", *Med. Chem.*, vol. 9, pp. 133-137, 2013.
- [12] M. M. Beigi, M. Behjati, and H. Mohabatkhar, "Prediction of metalloproteinase family based on the concept of Chou's pseudo amino acid composition using a machine learning approach", *J. Struct. Funct. Genomics*, vol. 12, pp. 191-197, 2011.
- [13] S.S. Sahu, and G. Panda, "A novel feature representation method based on Chou's pseudo amino acid composition for protein structural class prediction", *Comput. Biol. Chem.*, 2010, vol. 34, pp. 320-327, 2010.
- [14] U. R. Zia., and A. Khan, "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 Lett.*, vol. 19, pp. 890-903 2012.
- [15] X. Y. Sun, S. P. Shi, J. D. Qiu, S. B. Suo, S.Y. Huang, and R. P. Liang, "Identifying protein quaternary structural attributes by incorporating physicochemical properties into the general form of Chou's PseAAC via discrete wavelet transform", *Mol. Bio. Syst.*, vol. 8, pp. 3178-3184, 2012.
- [16] L. Nanni, and A. Lumini, "Genetic programming for creating Chou's pseudo amino acid based features for submitochondria localization", *Amino Acids*, vol. 34, pp. 653-660, 2008.
- [17] M. Esmaili, H. Mohabatkhar, S. Mohsenzadeh, "Using the concept of Chou's pseudo amino acid composition for risk type prediction of human papillomaviruses", *J. Theor. Biol.*, vol. 263, pp. 203-209, 2010.
- [18] H. Mohabatkhar, "Prediction of cyclin proteins using Chou's pseudo amino acid composition", *Protein Peptide Lett.*, vol. 17, pp. 1207-1214 (2010).
- [19] H. Mohabatkhar, M. M. Beigi, and A. Esmaili, "Prediction of GABA(A) receptor proteins using the concept of Chou's pseudo-amino acid composition and support vector machine", *J. Theor. Biol.*, vol. 281, pp. 18-23, 2011.
- [20] K. C. Chou, "Some remarks on protein attribute prediction and pseudo amino acid composition (50th Anniversary Year Review)", *J. Theor. Biol.*, vol. 273, pp. 236-247, 2011.
- [21] W. Chen, P. M. Feng, H. Lin, and K. C. Chou, "iRSpot-PseDNC: identify recombination spots with pseudo dinucleotide composition", *Nucleic Acid Res.*, vol. 41, p. e68, 2013.
- [22] W. Chen, H. Lin, P. M. Feng, C. Ding., Y. C. Zuo, and K. C. Chou, "iNuc-PhysChem: A Sequence-Based Predictor for Identifying Nucleosomes via Physicochemical Properties", *PLoS ONE*, vol. 7, article ID e47843, 2012.
- [23] B. Q. Li, T. Huang, L. Liu, Y. D. Cai, and K. C. Chou, "Identification of colorectal cancer related genes with mRMR and shortest path in protein-protein interaction network", *PLoS ONE*, vol. 7, p. e33393, 2012.
- [24] Y. Jiang, T. Huang, C. Lei, Y. F. Gao, Y. D. Cai, K. C. Chou, "Signal propagation in protein interaction network during colorectal cancer progression", *Bio. Med. Res. Int.*, vol. 2013, pp. 1-9, 2013.

- [25] T. Huang, J. Wang, Y. D. Cai, H. Yu, and K. C. Chou, "Hepatitis C virus network based classification of hepatocellular cirrhosis and carcinoma", *PLoS ONE*, vol. 7, article ID e34460, 9 pages, 2012.
- [26] P. Du, X. Wang, C. Xu, and Y. Gao, "PseAAC-BUILDER: A cross-platform stand-alone program for generating various special Chou's pseudo-amino acid compositions", *Anal. Biochem.*, vol. 425, pp. 117-119, 2012.
- [27] D. S. Cao, Q. S. Xu, and Y. Z. Liang, "propy: a tool to generate various modes of Chou's PseAAC", *Bioinformatics*, vol. 29, pp. 960-962, 2013.
- [28] H. B. Shen, and K. C. Chou, "PseAAC: a flexible web-server for generating various kinds of protein pseudo amino acid composition", *Anal. Biochem.*, vol. 373, pp. 386-388, 2008.
- [29] W. Z. Lin, J. A. Fang, X. Xiao, and K. C. Chou, "iLoc-Animal: A multi-label learning classifier for predicting subcellular localization of animal proteins", *Mol. BioSyst.*, vol. 9, pp. 634-644, 2013.
- [30] K. C. Chou, "Some Remarks on Predicting Multi-Label Attributes in Molecular Biosystems", *Mol. Biosyst.*, vol. 9, pp. 1092-1100, 2013.
- [31] X. Xiao, P. Wang, and K. C. Chou, "GPCR-2L: Predicting G protein-coupled receptors and their types by hybridizing two different modes of pseudo amino acid compositions", *Mol. Biosyst.*, vol. 7, pp. 911-919, 2011.
- [32] X. Xiao, P. Wang, and K. C. Chou, "iNR-PhysChem: A sequence-based predictor for identifying nuclear receptors and their subfamilies via physical-chemical property matrix", *PLoS ONE*, vol. 7, article ID e30869, 2012.
- [33] P. Wang, X. Xiao, and K. C. Chou, "NR-2L: A two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features", *PLoS ONE*, vol. 6, article ID e23505, 2011.
- [34] X. Xiao, P. Wang, W. Z. Lin, J. H. Jia, and K. C. Chou, "iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types", *Anal. Biochem.*, vol. 436, pp. 168-177, 2013.
- [35] P. Wang, X. Xiao, K. C. Chou, "NR-2L: A two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features", *PLoS ONE*, vol. 6, article ID e23505, 2011.
- [36] X. Xiao, P. Wang, and K. C. Chou, "GPCR-2L: Predicting G protein-coupled receptors and their types by hybridizing two different modes of pseudo amino acid compositions", *Mol. Biosyst.*, vol. 7, pp. 911-919, 2011.
- [37] K. Sadegh-Zadeh, "Fuzzy genomes", *Artif. Intell. Med.*, vol. 18, pp. 1-28, 2000.
- [38] A. Torres, and J. J. Nieto, "The fuzzy polynucleotide space: basic properties", *Bioinformatics*, vol. 19, pp. 587-592, 2003.
- [39] J. J. Nieto and A. Torres, "Midpoints for fuzzy sets and their application in medicine", *Artif. Intell. Med.*, vol. 17, pp. 81-101, 2003.
- [40] J. J. Nieto, A. Torres, D. N. Georgiou, and T. E. Karakasidis, Fuzzy Polynucleotide spaces and Metrics, *Bull. Math. Biol.*, vol. 68, pp. 703-725, 2006.
- [41] B. Kosko, *Neural Networks and Fuzzy Systems*, Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [42] K. C. Chou, and D. W. Elrod, "Protein subcellular location prediction", *Protein. Eng.*, vol. 12, pp. 107-118, 1999.
- [43] C. T. Lin, "Adaptive subsethood for radial basis fuzzy systems", in *Fuzzy Engineering* Kosko, B., Ed., Upper Saddle River, NJ: Prentice-Hall, 1997, pp. 429-464.
- [44] V. Moulton, M. Zuker, M. Steel, R. Pointon and D. Penny, "Metrics on RNA secondary structures", *J. Comput. Biol.*, vol. 7, pp. 277-292, 2000.
- [45] M. Liabres and F. Rossello, "A new family of metrics for biopolymer contact structures", *Comput. Biol. Chem.*, vol. 28, pp. 21-37, 2004.
- [46] B. DasGupta, T. Jiang, S. Kannan and E. Sweedyk, "On the complexity and approximation of synthetic distance", *Discrete. Appl. Math.*, vol. 88, pp. 59-82, 1998.
- [47] M. Foster, A. Heath and M. Afzal, "Application of distance geometry to 3D visualization of sequence relationships", *Bioinformatics*, vol. 15, pp. 89-90, 1999.
- [48] V. D. Gusev, L. A. Nemytikova and N. A. Chuzhanova, "On the complexity measures of genetic sequences", *Bioinformatics*, vol. 15, pp. 994-999, 1999.
- [49] T. Jiang, G. Lin, B. Ma and K. Zhang, "A general edit distance between RNA structures", *J. Comput. Biol.*, vol. 9, pp. 371-388, 2002.
- [50] D. Liben-Nowell, "On the structure of synthetic distance", *J. Comput. Biol.*, vol. , pp. 53-67, 2001.
- [51] M. Li, J. H. Badger, X. Chen, S. Kwong, P. Kearney and H. Zhang, "An information-based sequence distance and its application to whole mitochondrial phylogeny", *Bioinformatics*, vol. 17, pp. 149-154, 2001.
- [52] S. J. Freeland and L. D. Hurst, "The Genetic Code is one in a million", *J. Mol. Evol.*, vol. 47, pp. 238-248, 1998.
- [53] K. C. Chou, "Prediction of protein subcellular locations by incorporating quasi - sequence - order effect", *Biochem. Biophys. Res. Commun.*, vol. 278, pp. 477-483, 2000.
- [54] G. Schneider, and P. Wrede, "The rational design of amino acid sequences by artificial neural networks and simulated molecular evolution: de novo design of an idealized leader peptidase cleavage site", *Biophys. J.*, vol. 66, pp. 335-344, 1994.
- [55] X. Xiao, S. Shao, Y. Ding, Z. Huang, Y. Huang, and K.C. Chou, "Using complexity measure factor to predict protein subcellular location", *Amino Acids.*, vol. 28, pp. 57-61, 2005.
- [56] X. Xiao, S. H. Shao, Y. S. Ding, Z. D. Huang, and K. C. Chou, "Using cellular automata images and pseudo amino acid composition to predict protein sub-cellular location", *Amino Acids.*, vol. 30, pp. 49-54, 2006.
- [57] X. Xiao, S. H. Shao, Z. D. Huang, and K. C. Chou, "Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor", *J. Comput. Chem.*, vol. 27, pp. 478-482, 2006.
- [58] X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, and K. C. Chou, "Using cellular automata to generate Image representation for biological sequences", *Amino Acids*, vol. 28, pp. 29-35, 2005.
- [59] X. Xiao, S. Shao, Y. Ding, Z. Huang, X. Chen, and K. C. Chou, "An Application of Gene Comparative Image for Predicting the Effect on Replication Ratio by HBV Virus Gene Missense Mutation", *J. Theor. Biol.*, vol. 235, pp. 555-565, 2005.
- [60] C. Chen, X. Zhou, Y. Tian, X. Zou, and P. Cai, "Predicting protein structural class with pseudo-amino acid composition and support vector machine fusion network", *Anal. Biochem.*, vol. 357, pp.116-121, 2006.
- [61] K. C. Chou, "Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes", *Bioinformatics*, vol. 21, pp. 10-19, 2005.
- [62] K. C. Chou, Y. D. Cai, "Prediction of membrane protein types by incorporating amphipathic effects", *J. Chem. Inf. Model.*, vol. 45, pp. 407-413, 2005.
- [63] Z. P. Feng, "An overview on predicting the subcellular location of a protein", *In Silico. Biol.* vol. 2, pp. 291-303, 2002.
- [64] S. Mondal, R. Bhavna, R. Mohan Babu, and S. Ramakumar, "Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification", *J. Theor. Biol.*, vol. 243, pp. 252-260, 2006.
- [65] M. Wang, J. Yang, G. P. Liu, Z. J. Xu, and K. C. Chou, "Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition, Protein Engineering", *Des. Select.*, vol. 17, pp. 509-516, 2004.
- [66] S. Q. Wang, J. Yang, and K. C. Chou, "Using stacked generalization to predict membrane protein types based on pseudo amino acid composition", *J. Theor. Biol.*, vol. 242, pp. 941-946, 2006.
- [67] Y. Z. Guo, M. Li, M. Lu, Z. Wen, K. Wang, G. Li, and J. Wu, "Classifying G protein-coupled receptors and nuclear receptors based on protein power spectrum from fast Fourier transform", *Amino Acids*, vol. 30, pp. 397-402, 2006.
- [68] H. Liu, M. Wang, and K. C. Chou, "Low-frequency Fourier spectrum for predicting membrane protein types", *Biochem. Biophys. Res. Commun.*, vol. 336, pp. 737-739, 2005.
- [69] K. C. Chou, and H. B. Shen, "Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites", *J. Proteome. Res.*, vol. 6, pp. 1728-1734, 2007.
- [70] C. Chen, Y. X. Tian, X. Y. Zou, P. X. Cai, and J. Y. Mo, "Using pseudo-amino acid composition and support vector machine to predict protein structural class", *J. Theor. Biol.*, vol. 243, pp. 444-448, 2006.
- [71] Y. L. Chen, and Q. Z. Li, "Prediction of apoptosis protein subcellular location using improved hybrid approach and pseudo amino acid composition", *J. Theor. Biol.*, vol. 248, pp. 377-381, 2007.

- [72] Y. L. Chen, and Q. Z. Li, "Prediction of the subcellular location of apoptosis proteins", *J. Theor. Biol.*, vol. 245, pp. 775-783, 2007.
- [73] K. C. Chou, "A novel approach to predicting protein structural classes in a (20- $\lambda$ )-D amino acid composition space", *Proteins*, vol. 21, pp. 319-344, 1995.
- [74] K. C. Chou, "Review: Prediction of protein structural classes and subcellular locations", *Curr. Protein Peptide Sci.*, vol. 1, pp. 171-208, 2000.
- [75] K. C. Chou, "Prediction of G-protein-coupled receptor classes", *J. Proteome Res.*, vol. 4, pp. 1413-1418, 2005.
- [76] K. C. Chou, and Y. D. Cai, "Predicting protein-protein interactions from sequences in a hybridization space", *J. Proteome Res.*, vol. 5, pp. 316-322, 2006.
- [77] K. C. Chou, Y. D. Cai, and W. Z. Zhong, "Predicting networking couples for metabolic pathways of Arabidopsis", *EXCLI J.*, vol. 5, pp. 55-65, 2006.
- [78] K. C. Chou, and D. W. Elrod, "Bioinformatical analysis of G-protein-coupled receptors", *J. Proteome Res.*, vol. 1, pp. 429-433, 2002.
- [79] K.C. Chou, D. W. Elrod, "Protein subcellular location prediction", *Protein Eng.*, vol. 12, pp. 107-118, 1999.
- [80] K. C. Chou, and D. W. Elrod, "Prediction of membrane protein types and subcellular locations", *Proteins*, vol. 34, pp. 137-153, 1999.
- [81] K. C. Chou, and D. W. Elrod, "Prediction of enzyme family classes", *J. Proteome Res.*, vol. 2, pp. 183-190, 2003.
- [82] K. C. Chou, and C. T. Zhang, "Predicting protein folding types by distance functions that make allowances for amino acid interactions", *J. Biol. Chem.*, vol. 269, pp. 22014-22020, 1994.
- [83] K. C. Chou, and C. T. Zhang, "Review: Prediction of protein structural classes", *Crit. Rev. Biochem. Mol. Biol.*, vol. 30, pp. 275-349, 1995.
- [84] K. C. Chou, and Y.D. Cai, "Predicting protein quaternary structure by pseudo amino acid composition", *Proteins*, vol. 53, pp. 282-289, 2003.
- [85] K. C. Chou and Y. D. Cai, "Predicting enzyme family class in a hybridization space", *Protein Sci.*, vol. 13, pp. 2857-2863, 2004.
- [86] K. C. Chou and H. B. Shen, "Review: Recent progresses in protein subcellular location prediction", *Anal. Biochem.*, vol. 370, pp. 1-16, 2007.
- [87] K. C. Chou and H. B. Shen, "Large-scale plant protein subcellular location prediction", *J. Cell. Biochem.*, vol. 100, pp. 665-678, 2007.
- [88] K. C. Chou and H. B. Shen, "MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM", *Biochem. Biophys. Res. Commun.*, vol. 360, pp. 339-345, 2007.
- [89] K.C. Chou and H. B. Shen, "Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides", *Biochem. Biophys. Res. Commun.*, vol. 357, pp. 633-640, 2007.
- [90] K. C. Chou and H. B. Shen, "Cell-PLoc: A package of web-servers for predicting subcellular localization of proteins in various organisms", *Nat. Protoc.*, vol. 3, pp. 153-162, 2008.
- [91] K. C. Chou, "Prediction of protein cellular attributes using pseudo amino acid composition", *Proteins*, vol. 43, pp. 246-255, 2001.
- [92] K. C. Chou, "Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes", *Bioinformatics*, vol. 21, pp. 10-19, 2005.
- [93] P. Du and Y. Li, "Prediction of protein submitochondria locations by hybridizing pseudo-amino acid composition with various physicochemical features of segmented sequence", *BMC Bioinformatics*, vol. 7, p. 518, 2006.
- [94] P. Mundra, M. Kumar, K. K. Kumar, V. K. Jayaraman and B. D. Kulkarni, "Using pseudo amino acid composition to predict protein subnuclear localization: Approached with PSSM", *Patt. Recognit. Lett.*, vol. 28, pp. 1610-1615, 2007.
- [95] H. B. Shen and K. C. Chou, "Using optimized evidence-theoretic K-nearest neighbor classifier and pseudo amino acid composition to predict membrane protein types", *Biochem. Biophys. Res. Commun.*, vol. 334, pp. 288-292, 2005.
- [96] H. B. Shen and K. C. Chou, "Ensemble classifier for protein fold pattern recognition", *Bioinformatics*, vol. 22, pp. 1717-1722, 2006.
- [97] H. B. Shen and K.C. Chou, "Hum-mPLoc: An ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites", *Biochem. Biophys. Res. Commun.*, vol. 355, pp. 1006-1011, 2007.
- [98] H.B. Shen and K. C. Chou, "EzyPred: A top-down approach for predicting enzyme functional classes and subclasses", *Biochem. Biophys. Res. Commun.*, vol. 364, pp. 53-59, 2007.
- [99] H. B. Shen and K. C. Chou, "Signal-3L: a 3-layer approach for predicting signal peptide", *Biochem. Biophys. Res. Commun.*, vol. 363, pp. 297-303, 2007.
- [100] H. B. Shen, J. Yang and K. C. Chou, "Fuzzy KNN for predicting membrane protein types from pseudo amino acid composition", *J. Theor. Biol.*, vol. 240, pp. 9-13, 2006.
- [101] H. B. Shen and K. C. Chou, "Predicting protein subnuclear location with optimized evidence-theoretic K-nearest classifier and pseudo amino acid composition", *Biochem. Biophys. Res. Commun.*, vol. 337, pp. 752-756, 2005.
- [102] X. Xiao, S. H. Shao, Z.D. Huang and K. C. Chou, "Using pseudo amino acid composition to predict protein structural classes: approached with complexity measure factor", *J. Comput. Chem.*, vol. 27, pp. 478-482, 2006.
- [103] X. B. Zhou, C. Chen, Z. C. Li and X. Y. Zou, "Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes", *J. Theor. Biol.*, vol. 248, pp. 546-551, 2007.
- [104] G. Agüero-Chapin, H. Gonzalez-Diaz, R. Molina, J. Varona-Santos, E. Uriarte, and Y. Gonzalez-Diaz, "Novel 2D maps and coupling numbers for protein sequences. The first QSAR study of polygalacturonases; isolation and prediction of a novel sequence from *Psidium guajava*", *L. FEBS. Lett.*, vol. 580, pp. 723-730, 2006.
- [105] G. Agüero-Chapin, H. Gonzalez-Diaz, D. Riva G, E. Rodriguez, A. Sanchez-Rodriguez, G. Podda, and R. I. Vazquez-Padron, "MMM-QSAR recognition of ribonucleases without alignment: comparison with an hmm model and isolation from *Schizosaccharomyces pombe*, prediction, and experimental assay of a new sequence", *J. Chem. Inf. Model.*, vol. 48, pp. 434-448, 2008.
- [106] M. Zaus, *Crisp and Soft Computing with Hypercubical Calculus*, Heideberg: Physica-Verlag, 1999.
- [107] D. N. Georgiou, T. E. Karakasidis, J. J. Nieto, and A. A. Torres, "A study of entropy/clarity of genetic sequences using metric spaces and fuzzy sets", *J. Theor. Biol.*, vol. 267, pp. 95-105, 2010.
- [108] C. M. Hegalson and T. H. Jobe, "The fuzzy cube and causal efficacy: Representation of concomitant mechanisms in stroke", *Neural Netw.*, vol. 11, pp. 549-555, 1998.
- [109] K. Sadegh-Zadeh, "Fundamentals of clinical methodology: 3. Nosology", *Artif. Intell. Med.*, vol. 17, pp. 87-108, 1999.
- [110] J. J. Nieto, A. Torres and M. M. Vazquez-Trasande, "A metric space to study differences between polynucleotides", *Appl. Math. Lett.*, vol. 16, pp. 1289-1294, 2003.
- [111] A. Dress, T. Lokot, and L. D. Pustyl'nikov, "A new scale-invariant Geometry of  $L_1$  space", *Appl. Math. Lett.*, vol. 17, pp. 815-820, 2004.
- [112] Dress and T. Lokot, "A simple proof of the triangle inequality for the NTV metric", *Appl. Math. Lett.*, vol. 16, pp. 809-813, 2003.
- [113] R. Engelking, *General Topology*, Warszawa, 1977.
- [114] H. Hashimoto, "Szpilrajn's theorem on fuzzy orderings", *Fuzzy. Sets. Syst.*, vol. 10, pp. 101-108, 1983.
- [115] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- [116] P. Samaras, A. Kungolos, T. E. Karakasidis, D. Georgiou and K. Perakis, "Statistical Evaluation of PCDD/F Emission Data During Solid Waste Combustion by Fuzzy Clustering Techniques", *J. Environ. Sci. Health.*, vol. 36, pp. 153-161, 2001.
- [117] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, New Jersey: Prentice Hall PRT, 1995.
- [118] Gabor Mocz, "Fuzzy cluster analysis of simple physicochemical properties of amino acids for recognizing secondary structure in proteins", *Protein. Sci.*, vol. 4, pp. 1178-1187, 1995.
- [119] S. Kawashima and M. Kanehisa, "AAindex: amino acid index database", *Nucleic. Acids. Res.*, vol. 28, pp. 374, 2008.
- [120] S. Kawashima, H. Ogata, and M. Kanehisa, "AAindex: amino acid index database", *Nucleic. Acids. Res.*, vol. 27, pp. 368-369, 1999.
- [121] D. N. Georgiou, and T. E. Karakasidis, J. J. Nieto, A. Torres, "Use of fuzzy clustering technique and matrices to classify amino acids and its impact to Chou's pseudo amino acid composition", *J. Theor. Biol.*, vol. 257, pp. 17-26, 2009.
- [122] Y. L. Stephen, and J. Freeland, "A quantitative investigation of the chemical space surrounding amino acid alphabet formation", *J. Theor. Biol.*, vol. 250, pp. 349-361, 2008.

- [123] Y. S. Ding, T. L. Zhang, and K. C. Chou, "Prediction of protein structure classes with pseudo aminoacid composition and fuzzy support vector machine network", *Protein Peptide Lett.*, vol. 14, pp. 811-815, 2007.
- [124] H. B. Shen, J. Yang, and K. C. Chou, "Fuzzy KNN for predicting membrane protein types from pseudo aminoacid composition. *J. Theor. Biol.*, vol.240, 2006, pp. 9-13.
- [125] M. Hayat, A. Khan, "Discriminating outer membrane proteins with fuzzy K-nearest neighbor algorithms based on the general form of Chou's PseAAC", *Protein Peptide Lett.*, vol. 19, pp. 411-421, 2012.
- [126] H. B. Shen, J. Yang, X. J. Liu, and J. C. Chou, "Using supervised fuzzy clustering to predict protein structural classes", *Biochem. Biophys. Res. Commun.*, vol. 334, pp. 577-581, 2005.

---

Received: August 06, 2013

Revised: September 06, 2013

Accepted: September 15, 2013

© Georgiou et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.