

# SCIENTIFIC REPORTS



OPEN

## Identifying anticancer peptides by using improved hybrid compositions

Feng-Min Li &amp; Xiao-Qian Wang

Received: 12 July 2016  
Accepted: 02 September 2016  
Published: 27 September 2016

Cancer is one of the main causes of threats to human life. Identification of anticancer peptides is important for developing effective anticancer drugs. In this paper, we developed an improved predictor to identify the anticancer peptides. The amino acid composition (AAC), the average chemical shifts (acACS) and the reduced amino acid composition (RAAC) were selected to predict the anticancer peptides by using the support vector machine (SVM). The overall prediction accuracy reaches to 93.61% in jackknife test. The results indicated that the combined parameter was helpful to the prediction for anticancer peptides.

Cancer is the leading cause of death in economically developed countries<sup>1</sup>. The traditional methods for the treatment of cancer are surgical operation, radiotherapy and chemotherapy. However, radiotherapy and chemotherapy are expensive and cause side effect. Since Swedish scientist Boman first found and determined the antimicrobial peptides primary structure of *Hyalophora cecropia* in 1972, many studies had shown that the antimicrobial peptides have anti tumor activity<sup>2,3</sup>. The antimicrobial peptides which have anti tumor activity are called anticancer peptides (ACPs). Anticancer peptides contain 12–50 amino acid residues. Some of these peptides present in membranes with  $\alpha$ -helical or  $\beta$ -sheet structures, and the others have particular folds. The mechanism of the action of anticancer peptides includes: necrosis by the cell membrane lytic effect and non-membranolytic<sup>4–6</sup>. Advances in experimental technology have enabled to identify whether a protein belongs to anticancer peptide. However, experimental determination of the anticancer peptide remains time-consuming and laborious. Hence, developing and improving a fast and effective way to predict whether a protein is anticancer peptide would be very necessary. Recently, many methods were focused on the prediction of antimicrobial peptides from primary protein sequences<sup>7–9</sup>. However, the identification of anticancer peptides from primary protein sequences is still at the infant stage. Hajisharifi *et al.*<sup>10</sup> attempted to identify the anticancer peptide based on pseudo amino acid compositions (PseAAC) and the local alignment kernel by using of the Support Vector Machine (SVM). The prediction was made by the 5-fold cross-validation test and the overall success rate was 89.7%. Recently, Chen *et al.*<sup>11</sup> predicted the anticancer peptide with the iACP, the better predictive results were obtained.

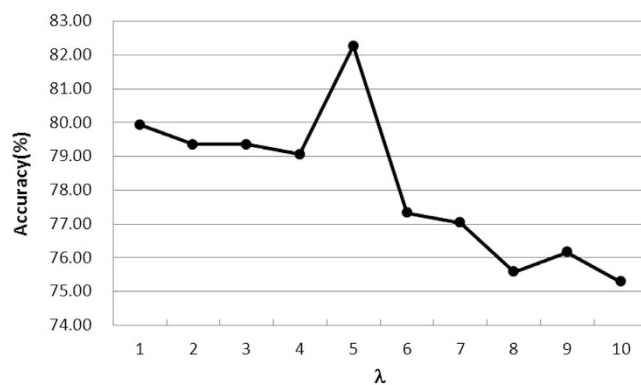
In this paper, the amino acid composition (AAC), the average chemical shifts (acACS) and the reduced amino acid composition (RAAC) were selected to predict the anticancer peptides with the same datasets as investigated by Hajisharifi *et al.*<sup>10</sup>. The overall prediction accuracy in jackknife test was 93.61% by using the combined parameter AAC + RAAC + acACS for support vector machine (SVM). The predictive results showed significant improvement compared with Hajisharifi's method.

### Results

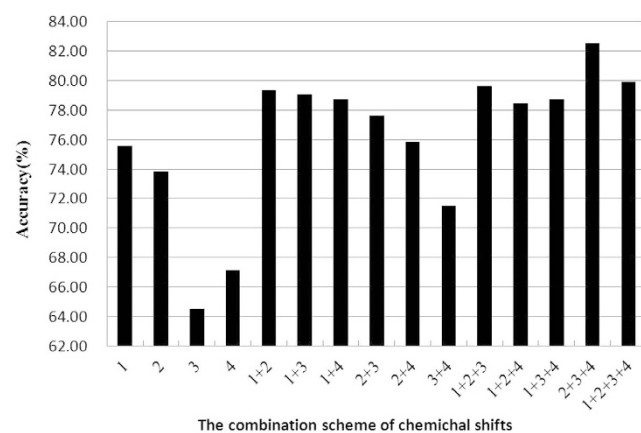
**The prediction of anticancer peptides.** In order to predict the anticancer peptides, it is very important to choose a classifier and a set of reasonable information parameters from protein sequence. In this paper, the local amino acids composition (AAC), the average chemical shift (acACS) and the reduced amino acid composition (RAAC) were selected to predict the anticancer peptides.

The acACS vectors were formed based on protein sequence, and then the best  $\lambda$  and  $i$  were selected. In order to obtain the best performance of predicting anticancer peptides, the combined scheme of chemically shifted atoms and the best  $\lambda$  were optimized with the maximum accuracy. Results in Fig. 1 showed that the accuracy was the best when  $\lambda = 5$  and in Fig. 2 showed that the prediction result was the best when the combination mode of chemically shifted atoms was  $^1\text{H}_N + ^{13}\text{C}_\alpha + ^{15}\text{N}$ . Therefore, the combination mode chemically shifted  $^1\text{H}_N + ^{13}\text{C}_\alpha + ^{15}\text{N}$  was selected and the correlation factor  $\lambda$  was set to 5 for generating the acACS feature vectors.

College of Science, Inner Mongolia Agricultural University, Hohhot, 010018, China. Correspondence and requests for materials should be addressed to F.-M.L. (email: fmli@imau.edu.cn)



**Figure 1.** prediction results with respect to the correlation factor  $\lambda$  of the acACS based on the jackknife test. The triangle indicates the best results obtained with  $\lambda = 5$ .



**Figure 2.** Prediction results of different combination schemes of chemical shifts. Numbers denote the chemical shifts of atoms: 1 for  $^1\text{H}_\alpha$ , 2 for  $^1\text{H}_\text{N}$ , 3 for  $^{15}\text{N}$ , 4 for  $^{13}\text{C}_\alpha$ .

Features	$s_n$ (%)	$s_p$ (%)	$Q_A$ (%)	$MCC$
AAC	87.68	94.66	91.86	0.830
RACC	77.54	88.35	84.01	0.642
acACS	71.02	90.29	82.56	0.633
AAC + RACC	86.96	97.57	93.31	0.861
AAC + acACS	86.23	97.57	93.02	0.856
RACC + acACS	77.54	91.74	86.05	0.707
AAC + RAAC + acACS	89.86	96.12	93.61	0.867

**Table 1.** The predictive results of SVM by jackknife test for anticancer peptides.

For facilitating comparison, the benchmark dataset (see Equation (5)) generated by Hajisharifi *et al.*<sup>10</sup> was employed. The predictive results of anticancer peptides based on AAC, acACS, RAAC, AAC + RAAC, AAC + acACS, RAAC + acACS and AAC + RAAC + acACS by using SVM with jackknife test were recorded in Table 1. The results showed that the combined parameter of AAC + RAAC + acACS was better than other parameters. The overall predictive accuracy ( $Q_A$ ) and Matthew's correlation coefficient ( $MCC$ ) in jackknife test were 93.61% and 0.867 with the combined parameter of AAC + RAAC + acACS by using of the SVM, respectively. The results indicated that the combined parameter was helpful to the prediction for anticancer peptides.

In order to estimate the effectiveness of the new prediction method, an independent dataset (see Equation (6)) generated by Chen *et al.*<sup>11</sup> was employed. The independent dataset is not absolutely needed for validating a predictor via the jackknife or K-fold cross-validation, since the outcome obtained via the jackknife or K-fold cross-validation with benchmark dataset is actually from a combination of many different independent dataset tests. The combined parameter of AAC + RAAC + acACS was selected to identify the anticancer peptides in the independent dataset. The overall predictive accuracy ( $Q_A$ ) and Matthew's correlation coefficient ( $MCC$ ) in

	Validation method	$s_n$ (%)	$s_p$ (%)	$Q_A$ (%)	$MCC$
This paper	Jackknife test	89.86	96.12	93.61	0.867
	5-fold cross-validation	90.58	96.60	94.19	0.879
Hajisharifi <i>et al.</i> <sup>a</sup>	5-fold cross-validation	85.18	92.68	89.70	0.784
iACP <sup>b</sup>	5-fold cross-validation	88.40	99.02	94.77	0.893

**Table 2.** The comparison of the predictive results between this paper and the other methods. <sup>a</sup>See ref. 10. <sup>b</sup>See ref. 11.

jackknife test were 89.33% and 0.787 by using of the SVM, respectively. This results showed that the new predictive method was not only able to achieve higher overall success rates, but also more stable.

**Evaluation of the predictive performances.** In order to evaluate the predictive capability and reliability of the algorithm, the sensitivity ( $S_n$ ), specificity ( $S_p$ ), overall predictive accuracy ( $Q_A$ ) and Matthew's correlation coefficients ( $MCC$ ) are defined by

$$S_n = \frac{TP}{TP + FN} \quad (1)$$

$$S_p = \frac{TN}{TN + FP} \quad (2)$$

$$Q_A = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

where  $TP$  denotes the numbers of the correctly recognized positives,  $FN$  denotes the numbers of the positives recognized as negatives,  $FP$  denotes the numbers of the negatives recognized as positives,  $TN$  denotes the numbers of correctly recognized negatives.

**Jackknife cross-validation test.** In statistical prediction, the following three cross-validation methods are often used to examine a predictor for its effectiveness in practical application: independent dataset test, sub-sampling test, and jackknife cross-validation test<sup>12</sup>, among the three cross-validation methods, the jackknife test is deemed the most objective, and has been used to examine the performance of various predictors<sup>13–16</sup>. Hence the jackknife test was used to evaluate the performance of our method. During the process of jackknife test, each protein is singled out in turn for testing and the remaining proteins are merged for training.

## Discussion

Hajisharifi *et al.*<sup>10</sup> used SVM-based classification on the base of their PseAAC parameter and the local alignment kernel as string kernel method in conjunction with 5-fold cross-validation test for the same benchmark dataset (see Equation (5)). For the purpose of comparing the predictive capability of our method, the predicted results of Hajisharifi's and Chen's method are enumerated in Table 2 for the same dataset. Compared results show that the performance of our method is more superior to that of Hajisharifi's method. The  $s_n$ ,  $s_p$ ,  $Q_A$  and  $MCC$  of our method are about 5.40%, 3.92%, 4.49% and 0.095 higher than the predictive results of Hajisharifi's method with 5-fold cross-validation test, respectively. Although only the  $s_n$  obtained in our method is higher than that of Chen's method with 5-fold cross-validation test, the features of anticancer peptides are obtained more comprehensive in our method. At least our method can play a complimentary role to the existing methods in this area.

The predictive result indicates that the combined parameter AAC + RAAC + acACS is effective to the prediction of anticancer peptides proteins. From the discussion above, it can be seen that our method has advantage of more comprehensive features and higher predictive success rates. The combined parameter AAC + RAAC + acACS successfully enhance the prediction quality for the anticancer peptides. This method may have broad application in protein and DNA motif identification.

## Materials and Method

**Datasets.** The Benchmark dataset was generated by Hajisharifi *et al.*<sup>10</sup> and can be expressed as

$$S = S_{anticancer} \cup S_{non-anticancer} \quad (5)$$

where  $S_{anticancer}$  consists of 138 anticancer peptides and  $S_{non-anticancer}$  consists of 206 non-anticancer peptides. The anticancer peptides derived from the antimicrobial peptides database (ADP2) and 206 non-anticancer peptides were selected from Universal Protein Resource<sup>17</sup>.

In order to estimate the effectiveness of the new prediction method, an independent dataset was employed and can be expressed as

$$S' = S'_{\text{anticancer}} \cup S'_{\text{non-anticancer}} \quad (6)$$

where  $S'_{\text{anticancer}}$  consists of 150 anticancer peptides and  $S'_{\text{non-anticancer}}$  consists of 150 non-anticancer peptides. The samples in  $S'_{\text{anticancer}}$  and  $S'_{\text{non-anticancer}}$  were fetched from the dataset used by Chen *et al.*<sup>11</sup> and none of the sequences in  $S'$  was the same with the sequences in  $S$ .

**Support Vector Machine (SVM).** The support vector machine (SVM) is a widely used classification method developed based on the statistical learning theory<sup>18–25</sup>. The SVM is particularly attractive to biological sequence analysis due to its ability to handle noise, large dataset and large input spaces. The SVM model is a representation of the examples as points in space, mapped by a kernel function so that the examples are divided by a clear gap that is widely enough. The new examples are mapped into the same space and predicted according to which side of the gap they fall on. The radial basis kernel function (RBF) was used to obtain the best classification hyperplane. The regularization parameter  $C$  and the kernel width parameter  $\gamma$  were tuned via the grid search method. For a brief formulation of SVM and how it works, see the papers<sup>26,27</sup>. In this paper, the LibSVM algorithm<sup>28</sup> has been used to predict the anticancer peptide, which can be downloaded from <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

**The local amino acids compositions (AAC).** The information parameters are very important for predicted algorithms. In a sequence-based predictor, the most important issue is the way in which to extract features from primary sequences of proteins<sup>29–32</sup>. The primary sequences of proteins are composed of 20 amino acids. The absolute occurrence frequencies of the 20 amino acids in protein are important features. Hence, the absolute occurrence frequencies of the 20 amino acids in protein sequence are considered as the information parameters of a protein and can be defined as

$$X = [x_1, x_2, \dots, x_p, \dots, x_{20}]^T \quad (7)$$

where  $x_i (i = 1, 2 \dots 20)$  is the absolute occurrence frequencies of the 20 native amino acids and calculated by

$$x_i = \frac{n_i}{L} \quad (8)$$

where  $n_i$  is the occurrence number of the 20 native amino acids of the protein;  $L$  is the length of the protein.

**Analysis of the amino acids compositions.** Using the dataset  $S$ , we analyzed the average amino acids compositions of anticancer peptides and non-anticancer peptides. The average amino acids compositions can be calculated as follow

$$y_i^m = \sum_{j=1}^{k_m} \frac{n_{i,j}^m}{N_j^m} \quad (i = 1, 2 \dots 20; m = 1, 2) \quad (9)$$

where  $n_{i,j}^m$  is the number of  $i$ -th amino acids of  $j$ -th protein in  $m$ -th group,  $N_j^m$  denotes the total number of amino acids of  $j$ -th protein in  $m$ -th group,  $k_m$  denotes the number of samples in the  $m$ -th group (here ( $k_1 = 138$ ,  $k_2 = 206$ )). We calculated the average amino acids compositions of anticancer peptides and non-anticancer peptides by using of Equation (9). The calculate results indicated that the amino acids compositions of anticancer peptides and non-anticancer peptides were different. Hence the amino acids compositions were suitable as features to distinguish anticancer peptides and non-anticancer peptides. The different distribution of the amino acids compositions in anticancer peptides and non-anticancer peptides were shown in Fig. 3.

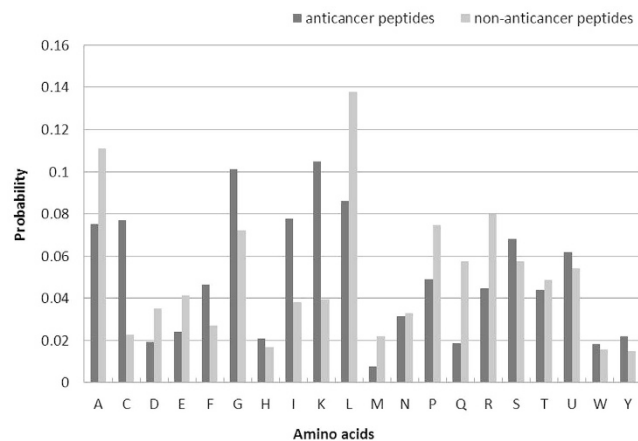
**Auto covariance of the average chemical shift (acACS) algorithm.** In a predictor, the most important issue is the way in which to extract features from primary sequences of protein. To achieve this, the acACS algorithm is proposed, which uses simple secondary structure information to represent the sample of a protein<sup>33</sup>. The average chemical shift of a protein has intrinsic correlation with the protein's secondary structure and the function of this protein. According to this point of view, there must be some relationship among the average chemical shift, protein structure and functions. So the acACS algorithm has been widely applied to predictions of protein attributes, such as predicting protein submitochondrial localization<sup>34</sup>, the subcellular locations of the mycobacterial proteins and DNA-binding proteins<sup>35</sup>, as well as for identifying acidic and alkaline enzymes<sup>36</sup> and discriminating between bioluminescent and nonbioluminescent proteins<sup>37</sup>. The acACS algorithm can be obtained from web server at <http://202.207.14.87:8032/fuwu/acacs/index.asp>.

For a protein  $P$ ,

$$P = [j_1, j_2 \dots j_i \dots j_L] \quad (10)$$

where  $L$  is the length of the protein sequence  $p$  and  $j$  represent the 20 native amino acids residues,  $p$  is then expressed as follow:

$$P_{\text{acACS}}^i = [\theta^i(0), \theta^i(1), \theta^i(2), \dots, \theta^i(\lambda)] \\ (i = {}^{15}\text{N}, {}^{13}\text{C}_\alpha, {}^1\text{H}_\alpha, {}^1\text{H}_\text{N}, 0 < \lambda < L) \quad (11)$$



**Figure 3.** The different distribution of the amino acids compositions in anticancer peptides and non-anticancer peptides.

where  $\theta^i(\lambda)$  is the correlation factor of the average chemical shift for  $j_i$  with the average chemical shift for  $j_{i+\lambda}$  along the protein sequence. The factor  $\lambda$  ( $0 < \lambda < L$ ) reflects the rank of correlation. The factor  $i$  can be the different composition of  $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$ ,  $^1\text{H}_\alpha$  and  $^1\text{H}_\text{N}$ . In order to obtain the best result, an appropriate number for factor  $\lambda$  and  $i$  should be determined according to the predicting results.

**The reduced amino acid composition (RAAC).** It was demonstrated that in the definition of global protein structure, the patterns of hydrophobic and hydrophilic residues have major significance. To obtain the hydrophathy characteristics, the amino acids were divided into groups using their individual hydrophathies according to the ranges of the hydrophathy scale. Therefore, a protein sequence with 20 amino acids can be represented by a sequence with 6 characters according to following schemes: Strongly hydrophilic or polar (R, D, E, N, Q, K, H), Strongly hydrophobic (L, I, V, A, M, F), Weakly hydrophilic or weakly hydrophobic (S, T, Y, W), Proline (P), glycine (G) and Cysteine (C)<sup>38,39</sup>. The dipeptide composition of the six characters were chosen and represented as follow:

$$Q = [q_1 \cdots q_i \cdots q_{36}]^T \quad (12)$$

where  $q_i$  ( $i = 1, 2, \dots, 36$ ) is the absolute occurrence frequencies of the 36 hydrophathy dipeptides and calculated by

$$q_i = \frac{n_i}{L - 1} \quad (13)$$

where  $n_i$  is the occurrence number of the 36 hydrophathy dipeptides of the protein,  $L$  is the length of the protein.

## References

- Jemal, A. *et al.* & Forman, D. Global cancer statistics. *CA- Cancer J. Clin.* **61**, 69–90 (2011).
- Boman, H. G. Nilsson, I. & Rasmuson, B. Inducible antibacterial defence system in *Drosophila*. *Nature* **237**, 232–235 (1972).
- Boman, H. G. Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* **254**, 197–215 (2003).
- Cho, J. H. Sung, B. H. & Kim, S. C. Buforins: histone H2A-derived antimicrobial peptides from toad stomach. *Biochim. Biophys. Acta* **1788**, 1564–1569 (2009).
- Shi, S. L. Wang, Y. Y. Liang, Y. & Li, Q. F. Effects of tachyplesin and n-sodium butyrate on proliferation and gene expression of human gastric adenocarcinoma cell line BGC-823. *World J. Gastroenterol.* **12**, 1694–1698 (2006).
- Chen, J. *et al.* Tachyplesin activates the classic complement pathway to kill tumor cells. *Cancer Res.* **65**, 4614–4622 (2005).
- Xiao, X. *et al.* iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal. Biochem.* **436**, 168–177 (2013).
- Chen, W. & Luo, L. F. Classification of antimicrobial peptide using diversity measure with quadratic discriminant analysis. *J. Microbiol. Methods* **78**, 94–96 (2009).
- Wang, P. *et al.* Prediction of antimicrobial peptides based on sequence alignment and feature selection method. *PLoS One* **6**, e18476 (2011).
- Hajisharifi, Z. *et al.* Predicting anticancer peptides with Chou's pseudo amino acid composition and investigating their mutagenicity via Ames test. *J. Theor. Biol.* **34**, 34–40 (2014).
- Chen, W. *et al.* iACP: a sequence-based tool for identifying anticancer peptides. *Oncotarget* **7**, 16895–16909 (2016).
- Chou, K. C. & Zhang, C. T. Review: Prediction of protein structural classes. *Crit. Rev. Biochem. Mol. Biol.* **30**, 275–349 (1995).
- Chen, W. Feng, P. M. Lin, H. & Chou, K. C. IRSpotPseDNC: identify recombination spots with pseudo dinucleotide composition. *Nucleic Acids Res.* **41**, e68 (2013).
- Chou, K. C. & Shen, H. B. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* **5**, e11335 (2010).
- Lin, W. Z. Fang, J. A. Xiao, X. & Chou, K. C. ILoc-animal: a multi-label learning classifier for predicting subcellular localization of animal proteins. *Mol. BioSyst.* **9**, 634–644 (2013).
- Liu, B. *et al.* Identification of real microRNA precursors with a pseudo structure status composition approach. *PLoS One* **10**, e0121501 (2015).
- Wang, G. Li, X. & Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.* **37**, D933–D937 (2009).

18. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn* **20**, 273–297 (1995).
19. Guo, S. H. *et al.* iNuc-PseKNC: a sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. *Bioinformatics* **30**, 1522–1529 (2014).
20. Liu, B. Fang, L. Wang, S. & Wang, X. Identification of microRNA precursor with the degenerate K-tuple or Kmer strategy. *J. Theor. Biol.* **385**, 153–159 (2015).
21. Liu, B. Fang, L. Long, R. & Lan, X. iEnhancer-2L: a two-layer predictor for identifying enhancers and their strength by pseudo k-tuple nucleotide composition. *Bioinformatics* **32**, 362–389 (2016).
22. Liu, B. *et al.* Combining evolutionary information extracted from frequency profiles with sequence-based kernels for protein remote homology detection. *Bioinformatics* **30**, 472–479 (2014).
23. Lin, H. *et al.* 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 (2014).
24. Cao, R. Wang, Z. Wang, Y. & Cheng, J. SMOQ: a tool for predicting the absolute residue-specific quality of a single protein model with support vector machines. *BMC Bioinform.* **15**, 120 (2014).
25. Cao, R. Wang, Z. & Cheng, J. Designing and evaluating the MULTICOM protein local and global model quality prediction methods in the CASP10 experiment. *BMC Struct. Biol.* **14**, 13 (2014).
26. Chou, K. C. & Cai, Y. D. Using functional domain composition and support vector machines for prediction of protein subcellular location. *J. Biol. Chem.* **277**, 45765–45769 (2002).
27. Cai, Y. D. & Zhou, G. P. Support vector machines for predicting membrane protein types by using functional domain composition. *Biophys. J.* **84**, 3257–3263 (2003).
28. Fan, R. E. Chen, P. H. & Lin, C. J. Working set selection using second order information for training support vector machines. *J. Mach. Learn. Res.* **6**, 1889–1918 (2005).
29. Li, F. M. & Li, Q. Z. Predicting protein subcellular localization using Chou's pseudo amino acid composition and improved hybrid approach. *Protein Pept. Lett.* **15**, 612–616 (2008).
30. Li, F. M. & Li, Q. Z. Using pseudo amino acid composition to predict protein subnuclear location with improved hybrid approach. *Amino acids* **34**, 119–125 (2008).
31. Lin, H. Deng, E. Z. Ding, H. & Chen, W. 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 (2014).
32. Lin, H. *et al.* Predicting cancerlectins by the optimal g-gap dipeptides. *Sci. Rep.* **5**, 16964 (2015).
33. Fan, G. L. *et al.* acACS: Improving the Prediction Accuracy of Protein Subcellular Location and Protein Classification by Incorporating the Average Chemical Shifts Composition. *Sci. World J.* **864135** (2014).
34. Fan, G. L. & Li, Q. Z. Predicting protein submitochondria locations by combining different descriptors into the general form of Chou's pseudo amino acid composition. *Amino Acids* **43**, 545–555 (2012).
35. Fan, G. L. & Li, Q. Z. 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 (2012).
36. Fan, G. L. Li, Q. Z. & Zuo, Y. C. Predicting acidic and alkaline enzymes by incorporating the average chemical shift and gene ontology informations into the general form of Chou's PseAAC. *Process Biochem.* **48**, 1048–1053 (2013).
37. Fan, G. L. & Li, Q. Z. 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 (2013).
38. Chen, Y. L. & Li, Q. Z. Prediction of the subcellular location of apoptosis proteins. *J. Theor. Biol.* **245**, 775–783 (2007).
39. Pánek, J. Eidhammer, I. & Aasland, R. A New Method for Identification of Protein (Sub) Families in a Set of Proteins Based on Hydropathy Distribution in Proteins. *Proteins: Struct. Funct. Bioinformatics* **58**, 923–934 (2005).

## Acknowledgements

The authors wish to thank the anonymous reviewers for their constructive comments, which were very helpful for strengthening the presentation of this study. This work was supported by National Natural Science Foundation of China (No. 31360206) and Project for Talent Development Foundation of Inner Mongolia Autonomous Region.

## Author Contributions

F.-M.L. conceived the selection of feature parameters. X.-Q.W. carried out the computation by SVM. F.-M.L. performed the results analysis and wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Li, F.-M. and Wang, X.-Q. Identifying anticancer peptides by using improved hybrid compositions. *Sci. Rep.* **6**, 33910; doi: 10.1038/srep33910 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016