Accepted Manuscript

Prediction of cell-penetrating peptides with feature selection techniques

Hua Tang, Zhen-Dong Su, Huan-Huan Wei, Wei Chen, Hao Lin

Please cite this article as: H. Tang, Z.-D. Su, H.-H. Wei, W. Chen, H. Lin, Prediction of cell-penetrating peptides with feature selection techniques, *Biochemical and Biophysical Research Communications* (2016), doi: 10.1016/j.bbrc.2016.06.035.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Prediction of cell-penetrating peptides with feature selection techniques

Hua Tang[1,*], Zhen-Dong Su[3], Huan-Huan Wei[3], Wei Chen[2,3], Hao Lin[3]*

[1] Department of Pathophysiology, Southwest Medical University, Luzhou 646000, China;

[2] Department of Physics, School of Sciences, Center for Genomics and Computational Biology, North China University of Science and Technology, Tangshan 063009, China;

[3] Key Laboratory for NeuroInformation of Ministry of Education, Center of Bioinformatics and Center for Information in Biomedicine, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China.

Email:

Hua Tang: tanghua771211@aliyun.com

Hao Lin: hlin@uestc.edu.cn

* Corresponding authors

Abstract: Cell-penetrating peptides are a group of peptides which can transport different types of cargo molecules such as drugs across plasma membrane and have been applied in the treatment of various diseases. Thus, the accurate prediction of cell-penetrating peptides with bioinformatics methods will accelerate the development of drug delivery systems. The study aims to develop a powerful model to accurately identify cell-penetrating peptides. At first, the peptides were translated into a set of vectors with the same dimension by using dipeptide compositions. Secondly, the Analysis of Variance-based technique was used to reduce the dimension of the vector and explore the optimized features. Finally, the support vector machine was utilized to discriminate cell-penetrating peptides from non-cell-penetrating peptides. The five-fold cross-validated results showed that our proposed method could achieve an overall prediction accuracy of 83.6%. The results indicated that our model could provide more precise predictions for new peptides. Based on the proposed model, we constructed a free webserver called C2Pred (http://lin.uestc.edu.cn/server/C2Pred).