

# Prediction of cell penetrating peptides and their uptake efficiency using random forest-based feature selections

Peng Liu<sup>1</sup>, Yijie Ding<sup>1</sup>, Ying Rong<sup>2</sup>, and Dong Chen<sup>3</sup>

<sup>1</sup>University of Electronic Science and Technology of China

<sup>2</sup>Beidahuang Industry Group General Hospital

<sup>3</sup>Quzhou University

December 8, 2021

## Abstract

Cell penetrating peptides (CPPs) are short peptides that can carry biomolecules of varying sizes across the cell membrane into the cytoplasm. Correctly identifying CPPs is the basis for studying their functions and mechanisms. Here, we propose a novel CPP predictor that is able to predict CPPs and their uptake efficiency. In our method, five feature descriptors are applied to encode the sequence and compose a hybrid feature vector. Afterward, the wrapper + random forest algorithm is employed, which combines feature selection with the prediction process to find features that are crucial for identifying CPPs. The jackknife cross validation result shows that our predictor is comparable to state-of-the-art CPP predictors, and our method reduces the feature dimension, which improves computational efficiency and avoids overfitting, allowing our predictor to be adopted to identify large-scale CPP data.

## Hosted file

Manuscript.docx available at <https://authorea.com/users/450318/articles/548642-prediction-of-cell-penetrating-peptides-and-their-uptake-efficiency-using-random-forest-based-feature-selections>



