

# Statistical prediction of protein structural, localization and functional properties by the analysis of its fragment mass distributions after proteolytic cleavage

Bogachev M., Kayumov A., Markelov O., Bunde A.  
*Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia*

---

## Abstract

Structural, localization and functional properties of unknown proteins are often being predicted from their primary polypeptide chains using sequence alignment with already characterized proteins and consequent molecular modeling. Here we suggest an approach to predict various structural and structure-associated properties of proteins directly from the mass distributions of their proteolytic cleavage fragments. For amino-acid-specific cleavages, the distributions of fragment masses are determined by the distributions of inter-amino-acid intervals in the protein, that in turn apparently reflect its structural and structure-related features. Large-scale computer simulations revealed that for transmembrane proteins, either  $\alpha$ -helical or  $\beta$ -barrel secondary structure could be predicted with about 90% accuracy after thermolysin cleavage. Moreover, 3/4 intrinsically disordered proteins could be correctly distinguished from proteins with fixed three-dimensional structure belonging to all four SCOP structural classes by combining 3-4 different cleavages. Additionally, in some cases the protein cellular localization (cytosolic or membrane-associated) and its host organism (Firmicute or Proteobacteria) could be predicted with around 80% accuracy. In contrast to cytosolic proteins, for membrane-associated proteins exhibiting specific structural conformations, their monotopic or transmembrane localization and functional group (ATP-binding, transporters, sensors and so on) could be also predicted with high accuracy and particular robustness against missing cleavages.

<http://dx.doi.org/10.1038/srep22286>

---