

Highly accessed

Open access

## Developing a powerful *In Silico* tool for the discovery of novel caspase-3 substrates: a preliminary screening of the human proteome

Muneef Ayyash, Hashem Tamimi and Yaqoub Ashhab\*

\* Corresponding author: Yaqoub Ashhab [yashhab@ppu.edu](mailto:yashhab@ppu.edu)  
For all author emails, please [log on](#).

*BMC Bioinformatics* 2012, **13**:14 doi:10.1186/1471-2105-13-14  
Published: 23 January 2012

### Abstract

#### Background

Caspases are a family of cysteinyl proteases that regulate apoptosis and other biological processes. Caspase-3 is considered the central executioner member of this family with a wide range of substrates. Identification of caspase-3 cellular targets is crucial to gain further insights into the cellular mechanisms that have been implicated in various diseases including: cancer, neurodegenerative, and immunodeficiency diseases. To date, over 200 caspase-3 substrates have been identified experimentally. However, many are still awaiting discovery.

#### Results

Here, we describe a powerful bioinformatics tool that can predict the presence of caspase-3 cleavage sites in a given protein sequence using a Position-Specific Scoring Matrix (PSSM) approach. The present tool, which we call CAT3, was built using 227 confirmed caspase-3 substrates that were carefully extracted from the literature. Assessing prediction accuracy using 10 fold cross validation, our method shows AUC (area under the ROC curve) of 0.94, sensitivity of 88.83%, and specificity of 89.50%. The ability of CAT3 in predicting the precise cleavage site was demonstrated in comparison to existing state-of-the-art tools. In contrast to other tools which were trained on cleavage sites of various caspases as well as other similar proteases, CAT3 showed a significant decrease in the false positive rate. This cost effective and powerful feature makes CAT3 an ideal tool for high-throughput screening to identify novel caspase-3 substrates.

The developed tool, CAT3, was used to screen 13,066 human proteins with assigned gene ontology terms. The analyses revealed the presence of many potential caspase-3 substrates that are not yet described. The majority of these proteins are involved in signal transduction, regulation of cell adhesion, cytoskeleton organization, integrity of the nucleus, and development of nerve cells.

#### Conclusions

CAT3 is a powerful tool that is a clear improvement over existing similar tools, especially in reducing the false positive rate. Human proteome screening, using CAT3, indicate the presence of a large number of possible caspase-3 substrates that exceed the anticipated figure. In addition to their involvement in various expected functions such as cytoskeleton organization, nuclear integrity and adhesion, a large number of the predicted substrates are remarkably associated with the development of nerve tissues.

**Keywords:** Apoptosis; Caspase-3; Caspase substrates; Cleavage site prediction; Position-Specific Scoring Matrix (PSSM); Human proteome; Bioinformatic tool; Pattern recognition