

Function prediction by automatized protein annotation

Piroska Dévay* and Frederic Bornancin

***IK@N, ISS and Novartis Institutes for BioMedical Research, Vienna, Austria**

Keywords: protein annotation, function prediction, homology search, protein interaction

The protein function prediction pipeline (FPP) aids the prediction of function of have been yet not annotated proteins. A submitted cDNA sequence is translated in six reading frames and these peptides obtained serve as templates for multiple database, domain and motif searches. Based on the annotation of homologue sequences or significant domain or motif hits the query protein can be annotated as well.

From a partial sequence that has shown homology to the N-terminal region of ceramide kinases we annotated this sequence as a ceramide kinase-like protein and cloned the full sequence (ref.). The CERK-like protein shows strong homology to a Drosophila protein with known interactions. Using FPP we can annotate these proteins and search for human relatives. Our findings suggest that the CERK-like protein interacts with a number of transcription factors and might have an important role in the regulation of cell growth.

The workflow of the program includes:

- a) search for transmembrane domains, signal peptide and motifs characteristic for extracellular or nuclear localization
- b) search in full sequence databases
- c) search for conserved motifs and patterns

DNA is translated in each reading frame. Results for each protein are summarized in a table and a graphical overview showing the location of TM regions, domain and pattern hits is provided. Additional information (alignments, scores, database entries) can be found via hyperlinks.

Using this program we could annotate a partial sequence as ceramide kinase-like and subsequently clone the full protein coding region.

The CERK-like protein is homologous to the Drosophila NP_649530 protein (Fig.1). The NP_649530 protein has known interactions (Fig.2). We have found human proteins related to 3 of the interacting proteins (NP_523814, NP_524951, NP_524434 and NP_648033): NP_004463, XP_370580, NP_006553 and Q9BU97 respectively. The first 3 were annotated as transcription factors by the FPP and the last contains SAPS domain.

Conclusion:

On the basis of our findings we hypothesize that the CERK-like protein plays a role in the regulation of cell cycle. Indeed, immunostaining shows that the CERK-like protein is localized in the nucleus and it is especially abundant in nucleoli.



Figure 1 : CERK-like protein vs. the Drosophila NP_649530 protein The black area corresponds to the diacylglycerol kinase catalytic domain, and the gray C-terminal region contains motifs that occur only in CERK-like proteins.



Figure 2 : Interactions of the Drosophila NP_649530 protein.

Reference:

Bornancin F, Mechtcheriakova D, Stora S, Graf C, Wlachos A, Devay P, Urtz N, Baumruker T, Billich A. . 2005. Characterization of a ceramide kinase-like protein. *Biochim Biophys Acta*;1687(1-3):31-43.