Testset of Transmembrane Proteins

Steffen Möller, Michele Magrane, Rolf Apweiler
moeller@ebi.ac.uk, magrane@ebi.ac.uk, apweiler@ebi.ac.uk
EMBL Outstation, European Bioinformatics Instute, Hinxton, Cambridge, UK

Rita Casadio
casadio@kaiser.alma.unibo.it
Centro Interdipartimentale di Ricerche Biotecnologiche,
University of Bologna, Italy

Abstract: A new collection of transmembrane peptides was created and manually (re)annotated. This allows us to evaluate existing programs for the prediction or transmembrane segments and serves the scientific community as a better source for the development of new algorithms.

This testset is a superset of other frequently used datasets. It is always explicitly stated from which paper which part of the annotation was extracted, and also in which other testsets the sequence is found. The annotation not only includes the final estimation of transmembrane boundaries but also experimental evidence that led to the conclusion.

Different categories represent different levels of trust set into the annotation:

A - structure known
B - very good biochemical evidence
C - biochemical evidence
D - by similarity, predicted, not fully trusted