Classification of the short-chain dehydrogenase/reductase family

Yvonne Kallberg, Johan Nilsson & Bengt Persson

Department of Medical Biochemistry and Biophysics,
Stockholm Bioinformatics Center,
Karolinska Institutet
Stockholm, Sweden
yvonne.kallberg@mbb.ki.se

The short-chain dehydrogenase/reductase (SDR)[1] family of proteins constitutes a widespread group of enzymes with residue identities typically 15-30% in pair-wise comparisons. Many different enzyme activities are represented. Alcohols, aromatic compunds, sugars and steriods are examples of substrates. A current estimate gives around 1500 known protein sequences divided into two main groups: the classical SDRs with a typical length of 250 residues, and the extended SDRs with an 100-residue C-terminal extension.

We have classified this family by applying methods for detecting remote protein homologies, comparing SAM-T99[2], PSI-BLAST[3], and our own knowledge-based approach. We have also made a subfamily classification, implementing a method for phylogenetic inference [4] and compared this to our own hierarchical approach.

The low residue identity between some of the protein sequences and the large number of family members provides a true challenge for current sequence comparison techniques of protein homology search and subfamily classification.

References

1. Jörnvall H, et. al. (1995) Short-chain Dehydrogenases/Reductases (SDR), *Biochemistry*, **34**:6003-6013
2. Karplus K, Barrett C & Hughey R (1998) Hidden Markov Models for Detecting Remote Protein Homologies, *Bioinformatics*, **14**:846-856
3. Jones DT (1999) Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices, *J. Mol. Biol.*, **292**:195-202
4. Sjölander K (1997) A Baysian-Information Theoretic Method for Evolutionary Inference in Proteins, Ph.D. Dissertation, University of California Santa Cruz