

Attribute Information:

1. mcg: McGeoch's method for signal sequence recognition.
2. gvh: von Heijne's method for signal sequence recognition.
3. alm: Score of the ALOM membrane spanning region prediction program.
4. mit: Score of discriminant analysis of the amino acid content of the N-terminal region (20 residues long) of mitochondrial and non-mitochondrial proteins.
5. erl: Presence of "HDEL" substring (thought to act as a signal for retention in the endoplasmic reticulum lumen). Binary attribute.
6. pox: Peroxisomal targeting signal in the C-terminus.
7. vac: Score of discriminant analysis of the amino acid content of vacuolar and extracellular proteins.
8. nuc: Score of discriminant analysis of nuclear localization signals of nuclear and non-nuclear proteins.
9. class: Presence or absence of protein {positive, negative}.

Further information:

Following are the articles that have used this dataset:

1. "Expert Sytem for Predicting Protein Localization Sites in Gram-Negative Bacteria", Kenta Nakai & Minoru Kanehisa, PROTEINS: Structure, Function, and Genetics 11:95-110, 1991.
2. "A Knowledge Base for Predicting Protein Localization Sites in Eukaryotic Cells", Kenta Nakai & Minoru Kanehisa, Genomics 14:897-911, 1992.