



BioDec

New Software for the Postgenomics

www.BioDec.com

WHY BIOINFORMATICS

It is now established that the three-dimensional structure of protein is unequivocally determined by its aminoacidic sequence, and that its function is strongly dependent on its spatial structure. The recent successes of genomics and proteomics generated an enormous amount of data, measurable in million of sequences both for the genomes and for the proteomes encoded within the genome themselves. However only for a few thousands proteins their three-dimensional structure is known.

Protein sequence analysis is the discipline that tries to fill the knowledge gap between protein sequence and protein function. It is branch of Bioinformatics and its goal is to derive from the biological sequence the maximum amount of information possible, when the protein structure is unknown, and to understand function and the interactions between different proteins, whenever their structure is known. The most important techniques in sequence analysis, namely *sequence comparison*, *pattern recogni-*

tion and data mining, originates from computer science. However, despite their origin, in order to successfully use these techniques to the life sciences, you need also biological and chemical know-how.

WHAT IS BIODEC

BioDec (<http://www.biodec.com>) is a SME specialized in *bioinformatics*, with special emphasis on protein sequence analysis and "*knowledge management*" software. BioDec started as a spin-off of an academic research group and of Information Technology professionals.

The mission of the company is to provide high quality products and services for bioinformatics. Our main products are integrated tools (BIODECODING MACHINES) for decoding and annotating information from biological sequences, with special emphasis on trans-membrane protein detection and classification. When applicable, we build our products as web-based application servers, in order to ease the fruition of the tools by

non-specialized users. We also strive to use technological standards and *state-of-the-art algorithms*, and we try to achieve this goal by an ongoing interaction with a world-class academic research group in bioinformatics (Prof. Rita Casadio, University of Bologna).

OUR SOLUTIONS

I. Bioinformation Management System

The *Bioinformation Management System* (BMS) is a web application where all the data about your bioinformatic research can be stored. This means that biosequences, annotations, protein structural information, files, documents, and so on will be managed by a single, rational and unifying multipurpose system. Then the users will be able to create new virtual experiments, store results, search the annotations and the biosequences, share their data among different users groups: all through a simple unified interface. The idea behind this system is to provide researchers with a sort of virtual laboratory diary, where all the experiments of *in silico* bioinformatic research are managed.

The BMS comes in different flavors, depending on which tools do you want to use, but all releases share a common interface based on the PLONE (<http://plone.org>) technology.

Trample 2.0

The Trans-Membrane Protein Labelling Environment is our application server, built on top of our BMS technology, dedicated to the detection and annotation of putative *trans-membrane protein* sequences. It contains tools for the detection and an-

notation of beta-barrels and all-alpha helical trans-membrane proteins, signal peptides, coiled-coils and disordered regions together with a fast fold-recognition algorithm to find suitable templates for homology modelling.

ZenDock and ZenPatches

A BMS built around a predictor of the most probable *protein-protein interaction* sites on a protein surface, that, for every aminoacid on the protein surface, allows its classification as an interaction candidate or not, and estimate the prediction reliability. ZenPatches is a supplement of the PDB (Protein Data Bank) database, where all the structures have been analyzed and annotated using ZenDock. It's very convenient if you do not need to analyze your own crystal structures.

Tangram (Fold recognition system)

Tangram fold recognition code detects the sequence *remote homologies*, based on profile-profile alignments validated through entropy-filtering, as described in our Proteins paper [Proteins 54 (2004) 351-360]. The detected remote homology are suitable as templates for model building or for functional annotation of the target sequence.

II. BioDecoders

BioDecoders are state-of-the-art pipelines for the annotation of proteome biosequences: we have tools for prokaryotic and eukaryotic organisms. A specialized tool has been built for *G-Protein Coupled Receptor proteins* (GPCRs). The annotation results are published in multiple presentation formats (the Proteome Atlases) such as hyper-linked PDF ebooks for

reading and printing, HTML pages for web publishing, XML for data interchange and are available as annotated databases.

Angler (Proteome Annotation System)

Angler consists of two modules, one suited for Prokaryotes (Angler-PRO) and the other for Eukaryotes (Angler-EU). They are based on a hierarchical combination of different BioDec Tools. Angler-PRO is our fully-automated pipeline for the annotation of gram-negative bacterial proteomes, focussed on the detection of *outer-membrane and membrane-attached proteins*. Angler results are available either as browseable electronic "proteome atlases" or as relational databases. Angler-EU shares with Angler-PRO the same philosophy. However it is based on a different pipeline, with remarkably different end classes. This is due to the several different organelles typical of Eukaryotic cell, requiring subtle and specific calibration of the classification algorithms.

Cribrum (Proteome Annotation System)

This is a specific tool that searches for *G-Protein Coupled Receptor* proteins (GPCR) sequences in proteomes. Cribrum classifies the putative GPCR sequences in probable family or subfamilies at different tree-depth, according to the level of knowledge and availability of the different GPCR subfamily members. These are derived from a newly consistent annotation of the loops and transmembrane segments for the GPCRs found in public (or custom) databases and reclassified according to a BioDec procedure.

OUR SERVICES

BioDec offers a range of products and services to support researchers in the field of genomics and proteomics.

Thanks to our academic and business management experience, we can offer cutting-edge bioinformatic consulting as well as the outsourcing of bioinformatics and conventional IT research projects. We can also:

- develop and engineer *custom software applications* on your specifications,
- *integrate* those tools in our framework,
- generate non-disclosed *annotation of genomes*.

Contacts:

BioDec S.r.l.
Via Calzavecchio, 20/2
I-40033 Casalecchio di Reno (BO) – Italy
Phone: (+39)-051-0548263
Fax: (+39)-051-7459582
E-mail: info@biodec.com
WWW: <http://www.biodec.com>

BioDec software distributor (Italy):

S.-In. Soluzioni Informatiche
Via G. Salvemini, 9
I-36100 Vicenza – Italy
Phone: (+39)-0444-240341
Fax: (+39)-0444-533954
E-mail: info@s-in.it
WWW: <http://www.s-in.it>
