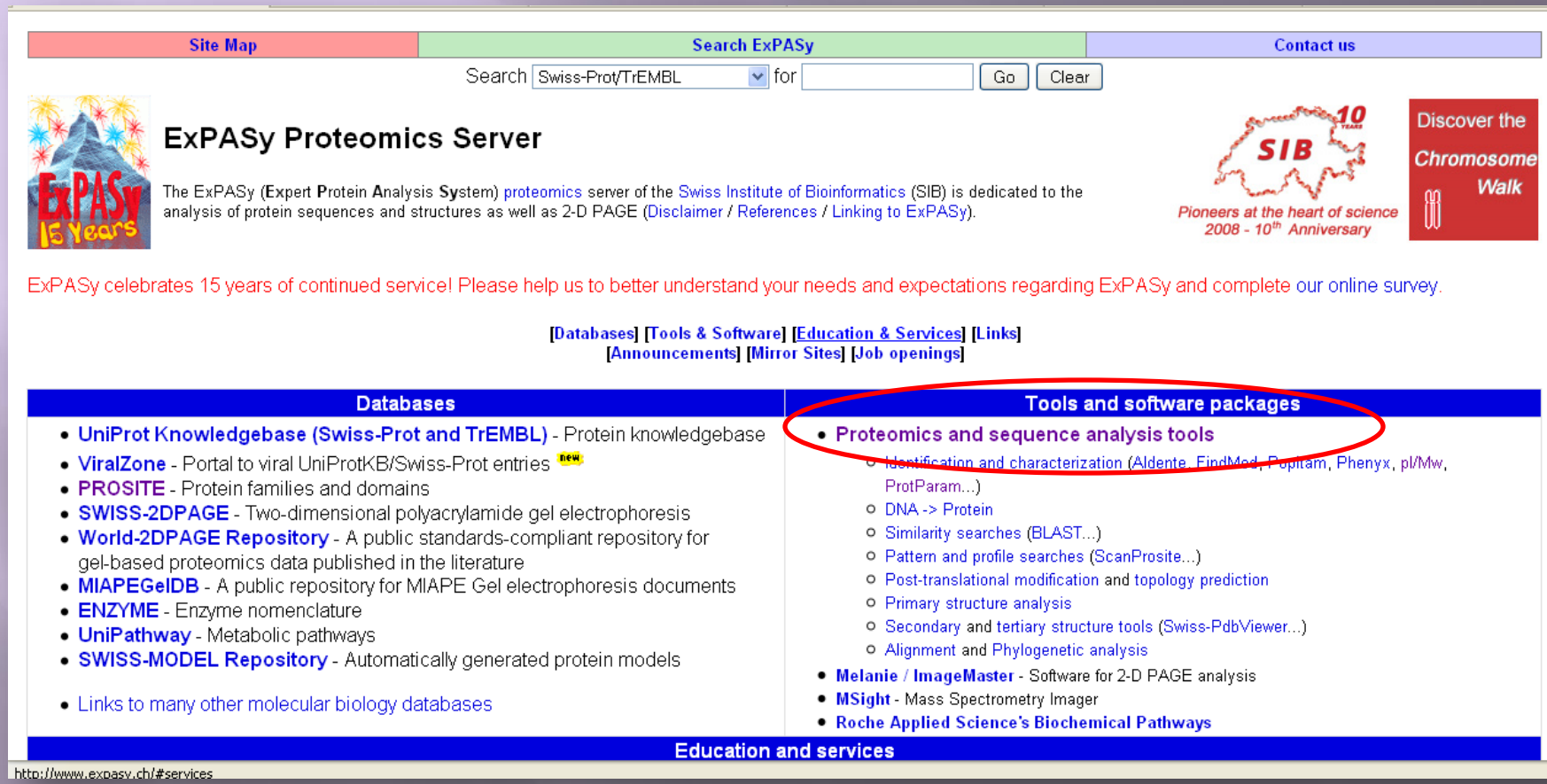
Red BIO-BIO-MOL

Curso de Posgrado: Vistas panorámicas desde el gen hasta la cristalización de una enzima para conocer su función

Análisis de una secuencia proteica: Empleo de bases de datos

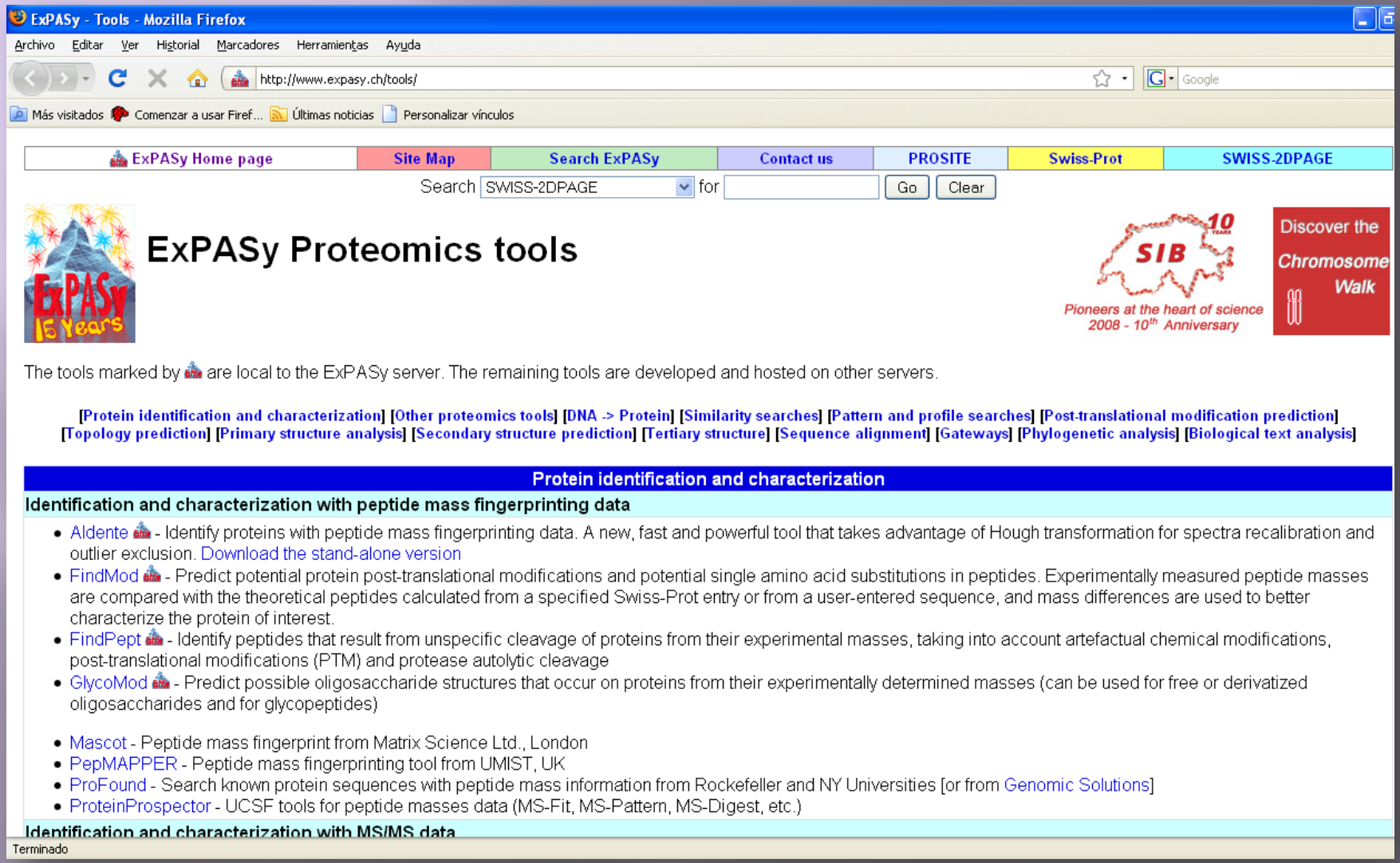
Dra. Paola Beassoni
UNRC
pbeassoni@exa.unrc.edu.ar

Servidor ExPASy

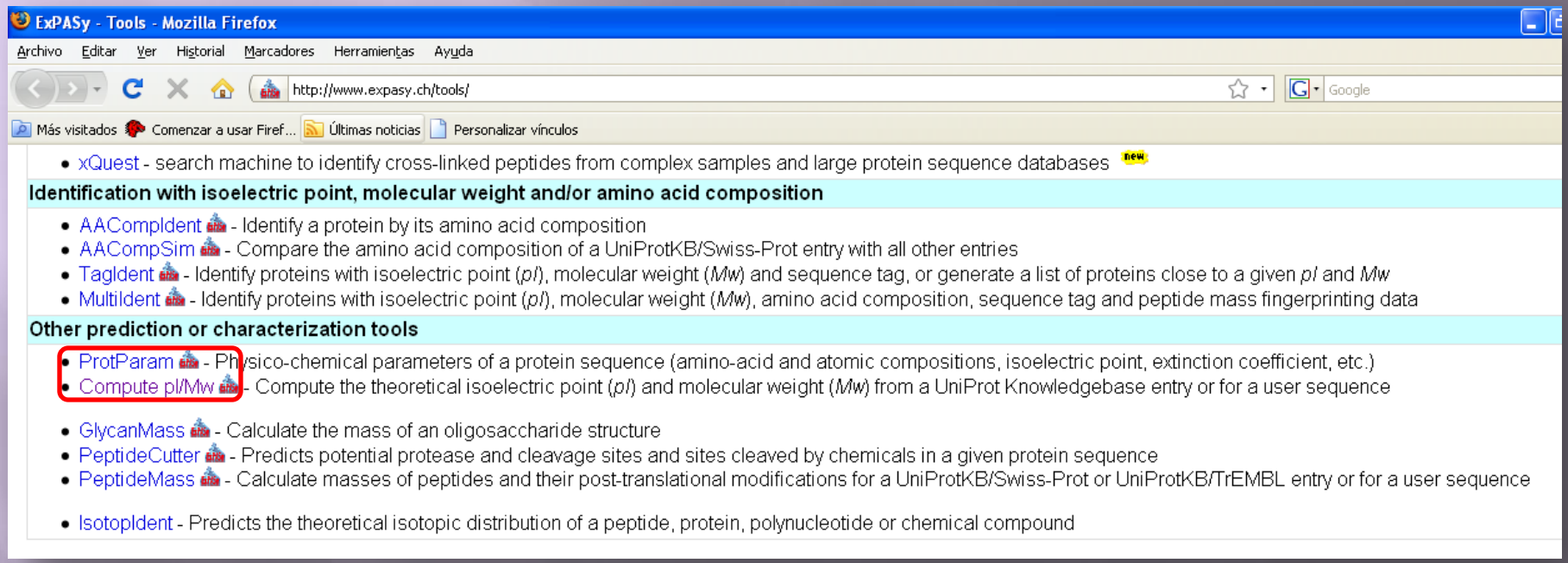


Servidor ExPASy

Identificación de Proteínas



Herramientas Básicas:
Calculo de Mw/pI composición relativa etc



Ejemplo ProtParam

Number of amino acids: 506

Molecular weight: 56419.3

Theoretical pI: 6.25

Amino acid composition: CSV format

Ala (A)	51	10.1%
Arg (R)	45	8.9%
Asn (N)	12	2.4%
Asp (D)	24	4.7%
Cys (C)	4	0.8%
Gln (Q)	22	4.3%
Glu (E)	46	9.1%
Gly (G)	45	8.9%
His (H)	18	3.6%
Ile (I)	27	5.3%
Leu (L)	62	12.3%
Lys (K)	18	3.6%
Met (M)	10	2.0%
Phe (F)	16	3.2%
Pro (P)	16	3.2%
Ser (S)	30	5.9%
Thr (T)	12	2.4%
Trp (W)	6	1.2%
Tyr (Y)	11	2.2%
Val (V)	31	6.1%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

Ext. coefficient 49640
Abs 0.1% (=1 g/l) 0.880, assuming ALL Cys residues appear as half cystines

Ext. coefficient 49390
Abs 0.1% (=1 g/l) 0.875, assuming NO Cys residues appear as half cystines

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

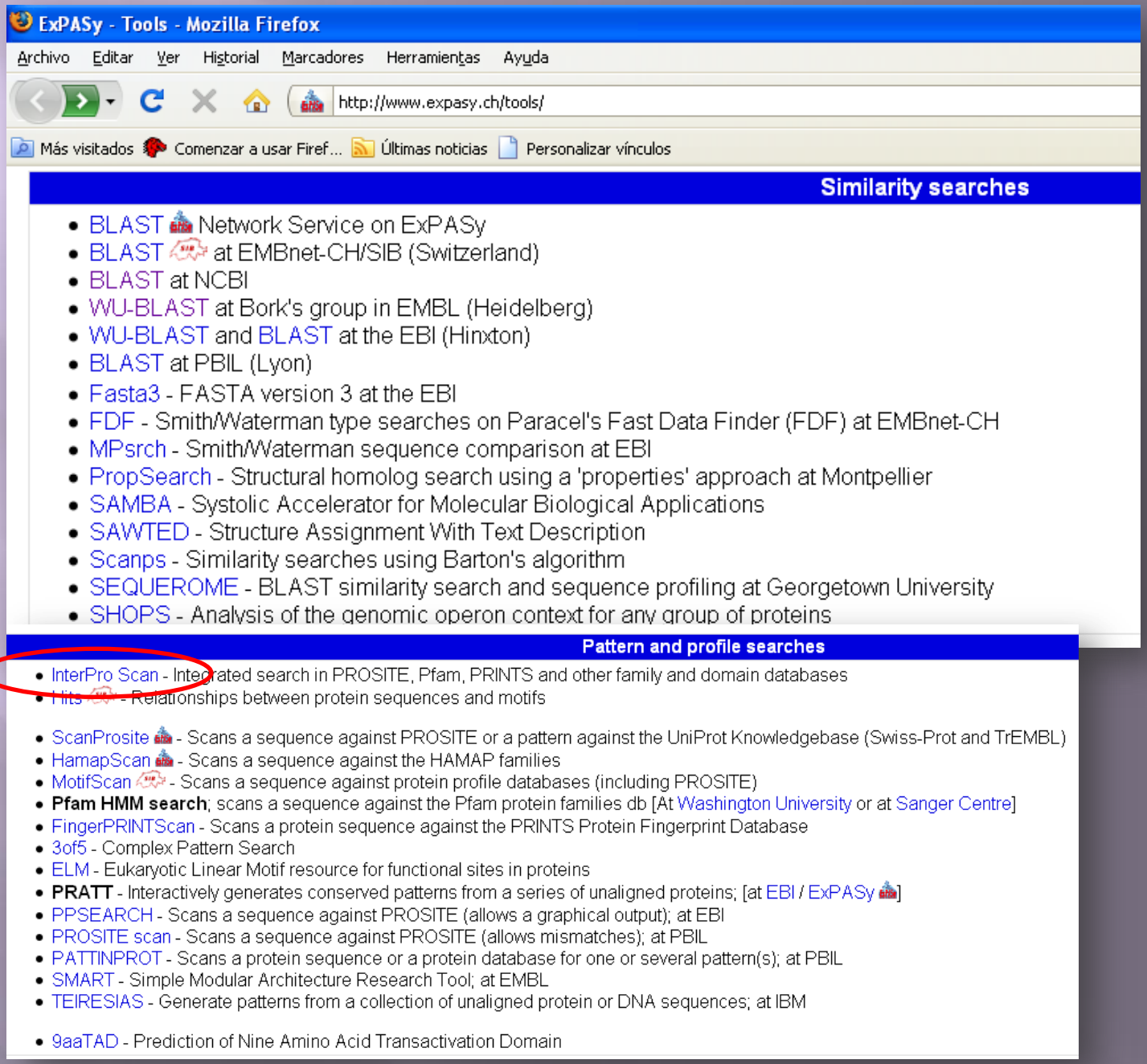
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

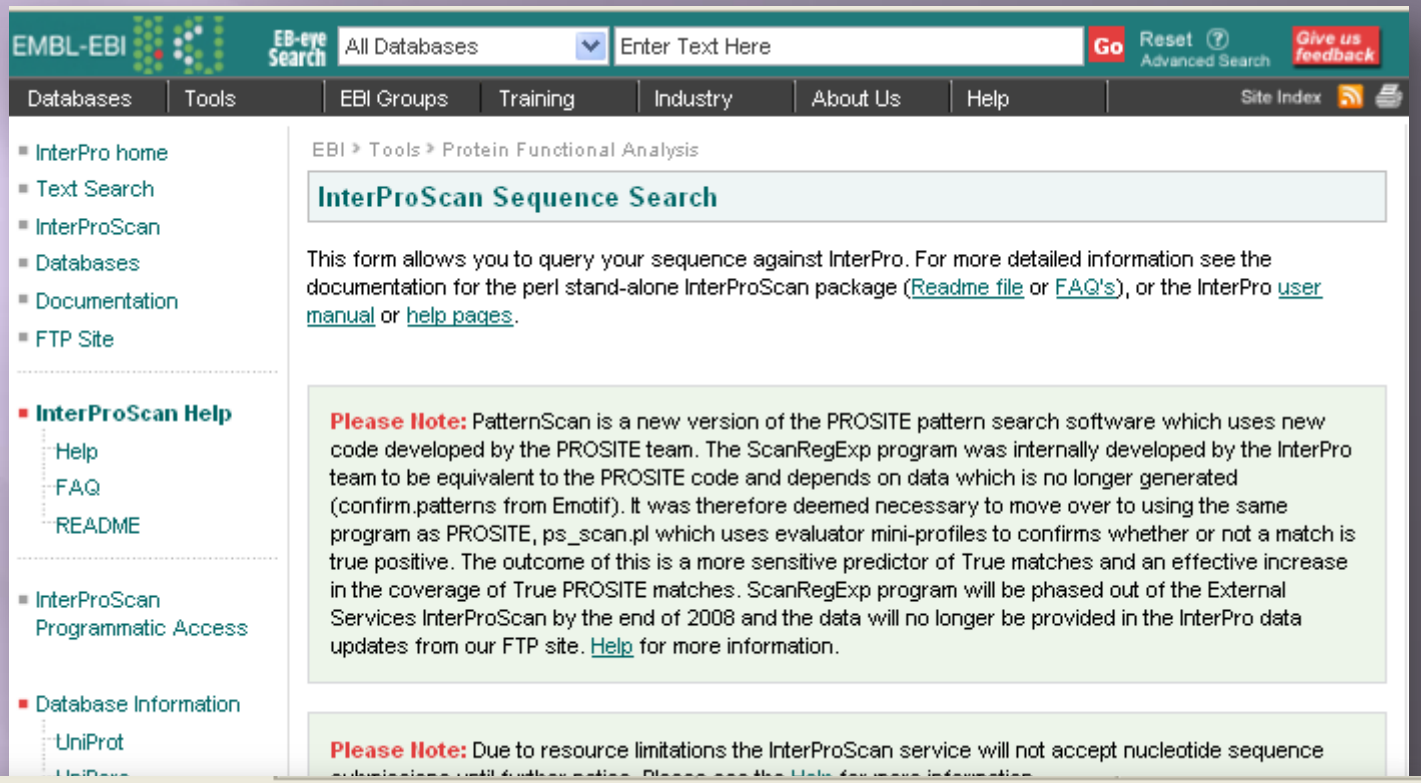
Instability index:

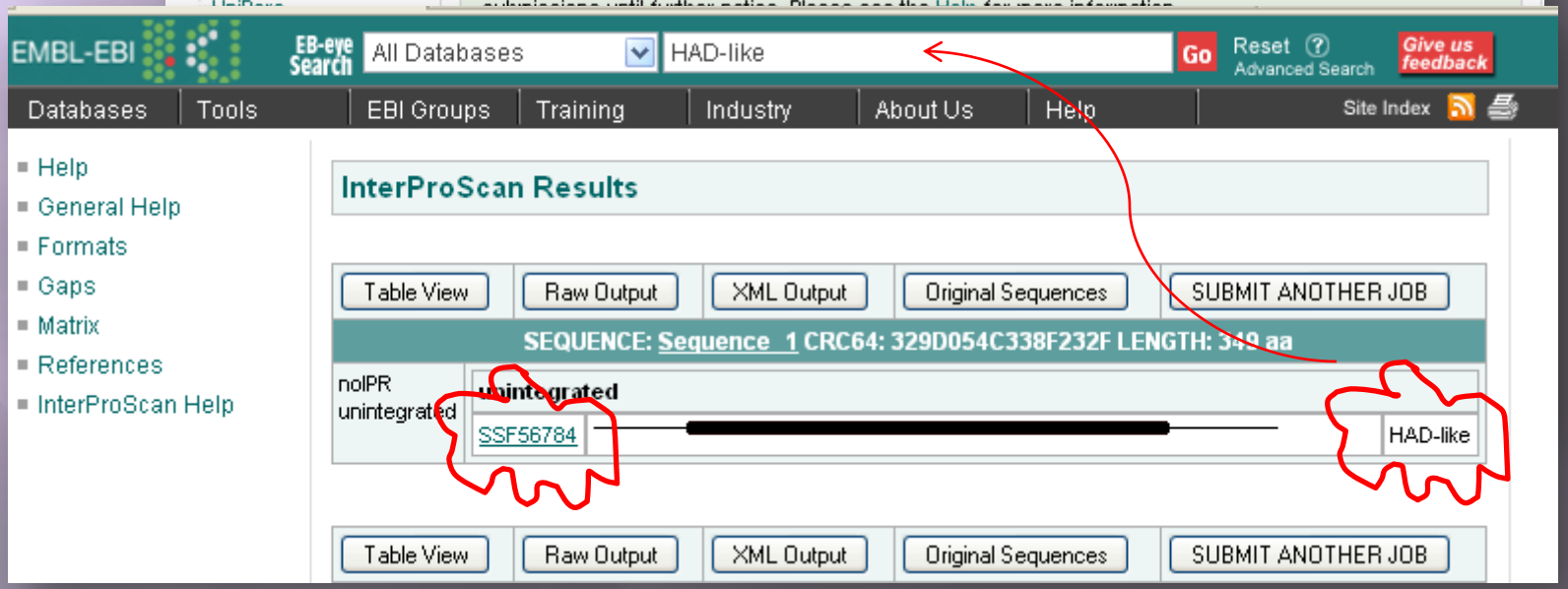
The instability index (II) is computed to be 45.39
This classifies the protein as unstable.

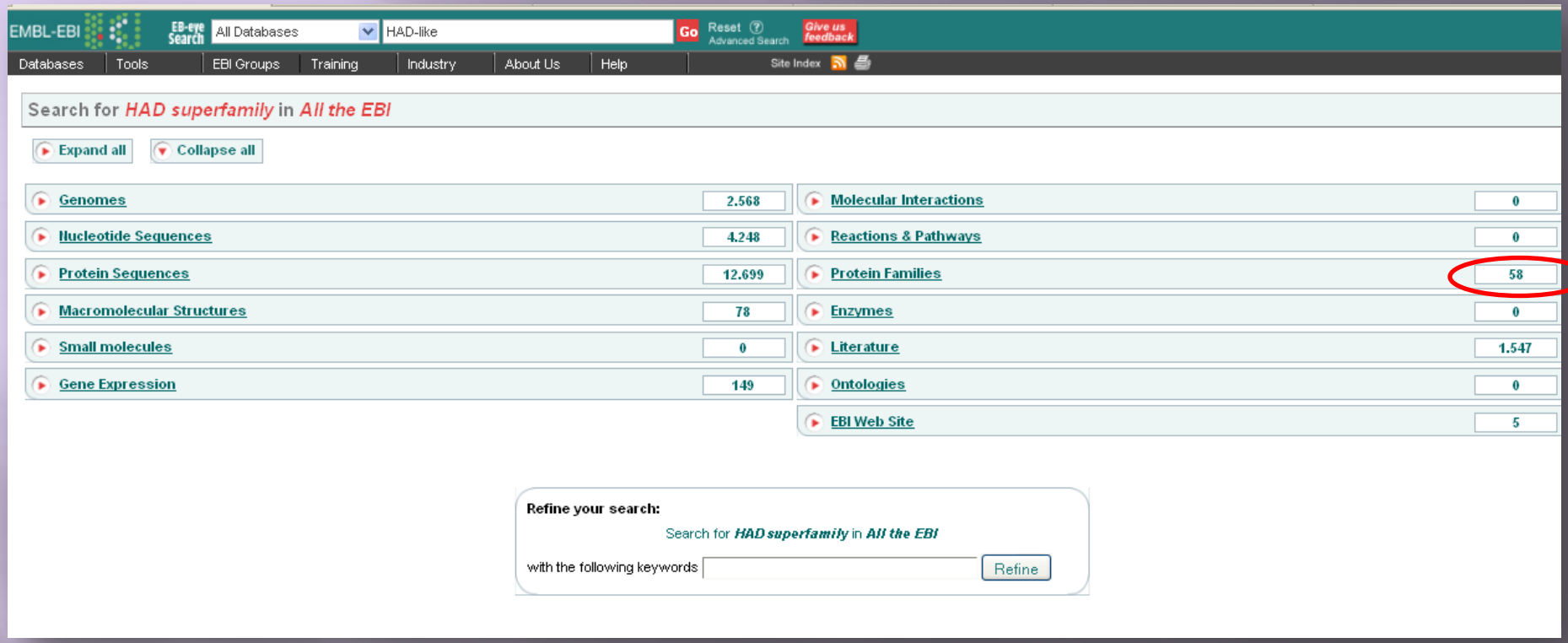
Aliphatic index: 96.44

Grand average of hydropathicity (GRAVY): -0.271


ExPASy - Tools - Mozilla Firefox
Archivo Editar Ver Historial Marcadores Herramientas Ayuda
http://www.expasy.ch/tools/
Más visitados Comenzar a usar Firef... Últimas noticias Personalizar vínculos
Similarity searches
BLAST Network Service on ExPASy
BLAST at EMBnet-CH/SIB (Switzerland)
BLAST at NCBI
WU-BLAST at Bork's group in EMBL (Heidelberg)
WU-BLAST and BLAST at the EBI (Hinxton)
BLAST at PBIL (Lyon)
Fasta3 - FASTA version 3 at the EBI
FDF - Smith/Waterman type searches on Paracel's Fast Data Finder (FDF) at EMBnet-CH
MPsrch - Smith/Waterman sequence comparison at EBI
PropSearch - Structural homolog search using a 'properties' approach at Montpellier
SAMBA - Systolic Accelerator for Molecular Biological Applications
SAWTED - Structure Assignment With Text Description
Scanps - Similarity searches using Barton's algorithm
SEQUEROME - BLAST similarity search and sequence profiling at Georgetown University
SHOPS - Analysis of the genomic operon context for any group of proteins
Pattern and profile searches
InterPro Scan - Integrated search in PROSITE, Pfam, PRINTS and other family and domain databases
Hits - Relationships between protein sequences and motifs
ScanProsite - Scans a sequence against PROSITE or a pattern against the UniProt Knowledgebase (Swiss-Prot and TrEMBL)
HamapScan - Scans a sequence against the HAMAP families
MotifScan - Scans a sequence against protein profile databases (including PROSITE)
Pfam HMM search; scans a sequence against the Pfam protein families db [At Washington University or at Sanger Centre]
FingerPRINTScan - Scans a protein sequence against the PRINTS Protein Fingerprint Database
3of5 - Complex Pattern Search
ELM - Eukaryotic Linear Motif resource for functional sites in proteins
PRATT - Interactively generates conserved patterns from a series of unaligned proteins; [at EBI / ExPASy]
PPSEARCH - Scans a sequence against PROSITE (allows a graphical output); at EBI
PROSITE scan - Scans a sequence against PROSITE (allows mismatches); at PBIL
PATTINPROT - Scans a protein sequence or a protein database for one or several pattern(s); at PBIL
SMART - Simple Modular Architecture Research Tool; at EMBL
TEIRESIAS - Generate patterns from a collection of unaligned protein or DNA sequences; at IBM
9aaTAD - Prediction of Nine Amino Acid Transactivation Domain


EMBL-EBI
EB-eye Search
All Databases
Enter Text Here
Go
Reset
Advanced Search
Give us feedback
Databases
Tools
EBI Groups
Training
Industry
About Us
Help
Site Index
InterPro home
Text Search
InterProScan
Databases
Documentation
FTP Site
InterProScan Help
Help
FAQ
README
InterProScan Programmatic Access
Database Information
UniProt
EBI › Tools › Protein Functional Analysis
InterProScan Sequence Search
This form allows you to query your sequence against InterPro. For more detailed information see the documentation for the perl stand-alone InterProScan package (Readme file or FAQ's), or the InterPro user manual or help pages.
Please Note: PatternScan is a new version of the PROSITE pattern search software which uses new code developed by the PROSITE team. The ScanRegExp program was internally developed by the InterPro team to be equivalent to the PROSITE code and depends on data which is no longer generated (confirm.patterns from Emotif). It was therefore deemed necessary to move over to using the same program as PROSITE, ps_scan.pl which uses evaluator mini-profiles to confirms whether or not a match is true positive. The outcome of this is a more sensitive predictor of True matches and an effective increase in the coverage of True PROSITE matches. ScanRegExp program will be phased out of the External Services InterProScan by the end of 2008 and the data will no longer be provided in the InterPro data updates from our FTP site. Help for more information.
Please Note: Due to resource limitations the InterProScan service will not accept nucleotide sequence


EMBL-EBI
EB-eye Search
All Databases
HAD-like
Go
Reset
Advanced Search
Give us feedback
Databases
Tools
EBI Groups
Training
Industry
About Us
Help
Site Index
Help
General Help
Formats
Gaps
Matrix
References
InterProScan Help
InterProScan Results
Table View
Raw Output
XML Output
Original Sequences
SUBMIT ANOTHER JOB
SEQUENCE: Sequence_1 CRC64: 329D054C338F232F LENGTH: 349 aa
nolPR unintegrated
unintegrated
SSF56784
HAD-like
Table View
Raw Output
XML Output
Original Sequences
SUBMIT ANOTHER JOB


EMBL-EBI
EB-eye Search
All Databases
HAD-like
Go
Reset
Advanced Search
Give us feedback
Databases
Tools
EBI Groups
Training
Industry
About Us
Help
Site Index
Search for HAD superfamily in All the EBI
Expand all
Collapse all
Genomes 2.568
Nucleotide Sequences 4.248
Protein Sequences 12.699
Macromolecular Structures 78
Small molecules 0
Gene Expression 149
Molecular Interactions 0
Reactions & Pathways 0
Protein Families 58
Enzymes 0
Literature 1.547
Ontologies 0
EBI Web Site 5
Refine your search:
Search for HAD superfamily in All the EBI
with the following keywords
Refine

Children IPR006328 Haloacid dehalogenase, type II

Contains IPR005834 Haloacid dehalogenase-like hydrolase

GO Term annotation

Function GO:0019120 hydrolase activity, acting on acid halide bonds, in C-halide compounds

InterPro annotation

Abstract

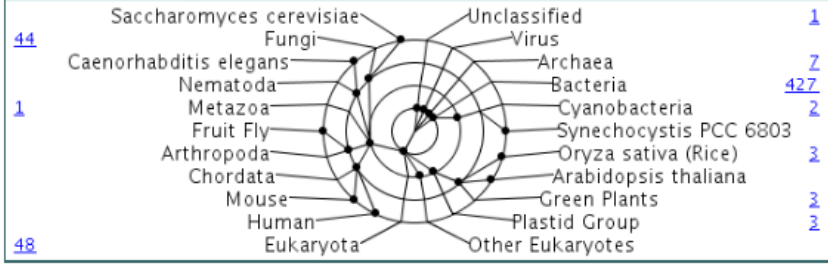
This group of sequences represent part of one structural subfamily of the Haloacid Dehalogenase (HAD) superfamily of aspartate-nucleophile hydrolases. The superfamily is defined by the presence of three short catalytic motifs [1]. The subfamilies are defined [2] based on the location and the observed or predicted fold of a so-called capping domain [3], or the absence of such a domain. Subfamily I consists of sequences in which the capping domain is found in between the first and second catalytic motifs. Subfamily II consists of sequences in which the capping domain is found between the second and third motifs. Subfamily III sequences have no capping domain in either of these positions. The Subfamily IA and IB capping domains are predicted by PSI-PRED to consist of an alpha helical bundle. Subfamily I encompasses such a wide region of sequence space (the sequences are highly divergent) that modelling it with a single HMM is impossible, resulting in an overly broad description which allows in many unrelated sequences. Subfamily IA and IB are separated based on an apparent phylogenetic bifurcation. Subfamily IA is still too broad to model, but cannot be further subdivided into large chunks based on phylogenetic trees. Of the three motifs defining the HAD superfamily, the third has three variant forms [2]: (1) hhhhsDxxx(x)D, (2) hhhhssxxx(x)D and (3) hhhhDDxxx(x)s where s refers to a small amino acid and h to a hydrophobic one. All three of these variants are found in subfamily IA.

Structural links

CATH: 1.10.164.10.2 , 3.40.50.1000.3
SCOP: c.108.1.1
PDB - click here

Database links Enzyme: EC:3.8.1.2

Taxonomic coverage



Overlapping InterPro entries

IPR006388	Numbers of overlapping proteins	Average numbers of overlapping amino acids
IPR006328 % Overlap: 99	81 403 1	N/A
IPR005834 % Overlap: 94	26 458 16734	N/A
IPR005833 % Overlap: 77	110 374 3009	N/A

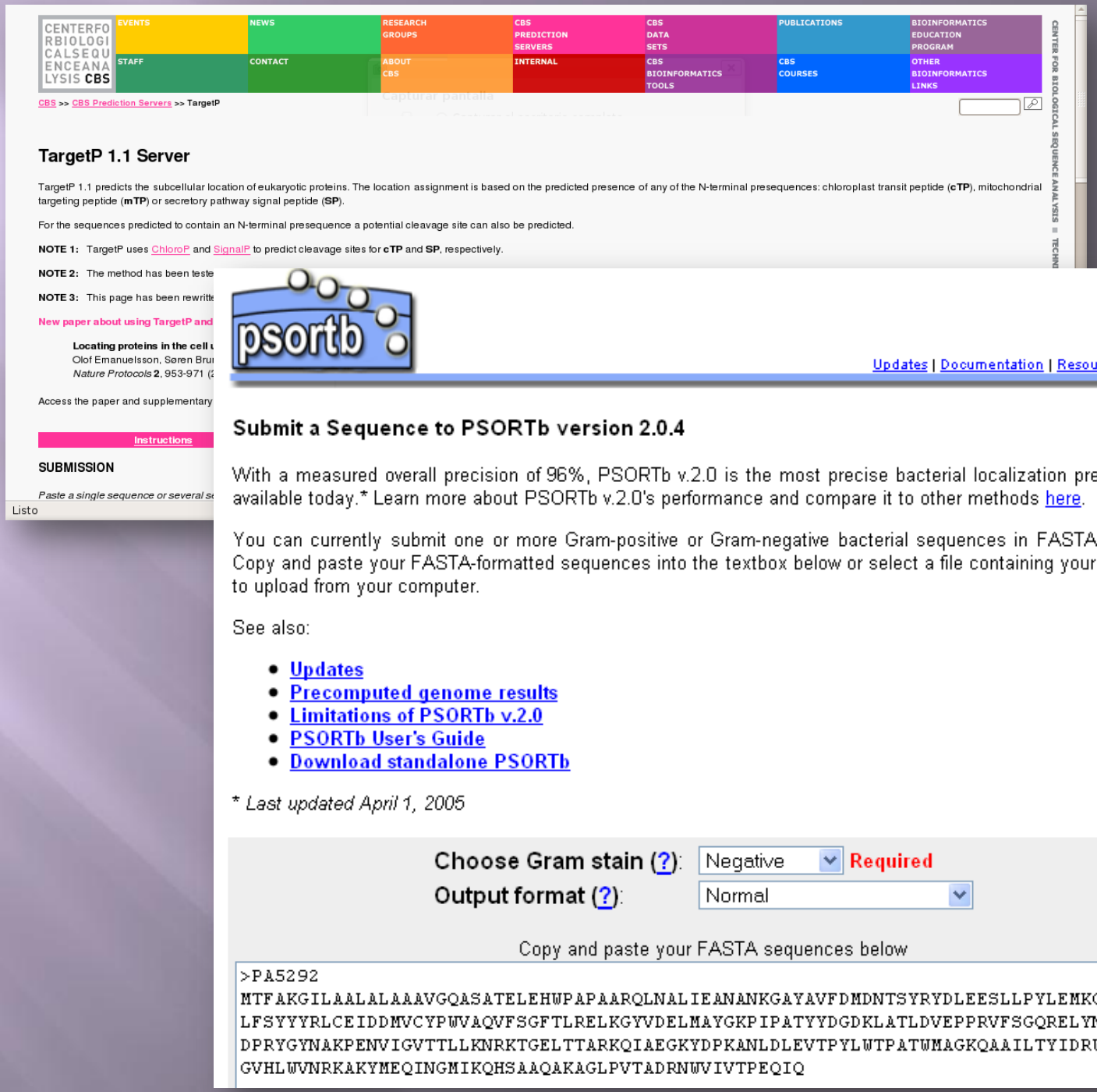
Terminado

Predicción de topología
Localización subcelular

Topology prediction

- NetNES Leucine-rich nuclear export signals (NES) in eukaryotic proteins
- PSORT - Prediction of protein subcellular localization
- SecretomeP - Non-classical and leaderless secretion of proteins
- TargetP - Prediction of subcellular location
- TatP - Twin-arginine signal peptides

- DAS - Prediction of transmembrane regions in prokaryotes using the Dense Alignment Surface method (Stockholm University)
- HMMTOP - Prediction of transmembrane helices and topology of proteins (Hungarian Academy of Sciences)
- PredictProtein - Prediction of transmembrane helix location and topology (Columbia University)
- SOSUI - Prediction of transmembrane regions (Nagoya University, Japan)
- TMAP - Transmembrane detection based on multiple sequence alignment (Karolinska Institut; Sweden)
- TMHMM - Prediction of transmembrane helices in proteins (CBS; Denmark)
- TMpred - Prediction of transmembrane regions and protein orientation (EMBnet-CH)
- TopPred - Topology prediction of membrane proteins (France)



TargetP 1.1 Server

TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (**cTP**), mitochondrial targeting peptide (**mTP**) or secretory pathway signal peptide (**SP**).

For the sequences predicted to contain an N-terminal presequence a potential cleavage site can also be predicted.

NOTE 1: TargetP uses ChloroP and SignalP to predict cleavage sites for **cTP** and **SP**, respectively.

NOTE 2: The method has been teste

NOTE 3: This page has been rewritte

New paper about using TargetP and

Locating proteins in the cell
Olof Emanuelsson, Søren Bru
Nature Protocols **2**, 953-971 (2

Access the paper and supplementary

Instructions

SUBMISSION

Paste a single sequence or several s

Listo

psortb

Submit a Sequence to PSORTb version 2.0.4

With a measured overall precision of 96%, PSORTb v.2.0 is the most precise bacterial localization prediction tool available today.* Learn more about PSORTb v.2.0's performance and compare it to other methods here.

You can currently submit one or more Gram-positive or Gram-negative bacterial sequences in FASTA format (?). Copy and paste your FASTA-formatted sequences into the textbox below or select a file containing your sequences to upload from your computer.

See also:

- Updates
- Precomputed genome results
- Limitations of PSORTb v.2.0
- PSORTb User's Guide
- Download standalone PSORTb

* *Last updated April 1, 2005*

Choose Gram stain (?): Negative **Required**

Output format (?): Normal

Copy and paste your FASTA sequences below

```
>PA5292
MTFAKGILAALALAAAVGQASATELEHWPAPAARQLNALIEANANKGAYAVFDMDNTSYRYDLEESLLPYLEMKGVLTRDRL
LFSYYYRLCEIDDMVCYPWVAQVFSGFTLRELKGYVDELMAYGKPIPATYYDGDKLATLDVEPPRVFSGQRELYNKLMENGI
DPRYGYNAKPENVIGVTTLLKNRKTGELTTARKQIAEGKYDPKANLDLEVTPYLWTPATWMAGKQAAILTYIDRWKRPILVA
GVHLWVNRKAKYMEQINGMIKQHSAAQAKAGLPVTADRNWVIVTPEQIQ
```

psortb

Submit Sequences | Docume

PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: PA5292
  Analysis Report:
    CMSVM-              Unknown                    [No details]
    CytoSVM-            Unknown                    [No details]
    ECSVM-              Unknown                    [No details]
    HMMTOP-             Unknown                    [1 internal helix found]
    Motif-              Unknown                    [No motifs found]
    OMPMotif-           Unknown                    [No motifs found]
    OMSVM-              Unknown                    [No details]
    PPSVM-              Unknown                    [No details]
    Profile-            Unknown                    [No matches to profiles found]
    SCL-BLAST-          Unknown                    [No matches against database]
    SCL-BLASTe-         Unknown                    [No matches against database]
    Signal-             Non-Cytoplasmic            [Signal peptide detected]
  Localization Scores:
    Cytoplasmic             0.00
    CytoplasmicMembrane     2.50
    Periplasmic             2.50
    OuterMembrane           2.50
    Extracellular           2.50
  Final Prediction:
    Unknown
```

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

TargetP 1.1 Server - prediction results

Technical University of Denmark

```
### targetp v1.1 prediction results ###################################
Number of query sequences:  1
Cleavage site predictions included.
Using NON-PLANT networks.

Name                    Len            mTP    SP    other   Loc   RC   TPlen
-------------------------------------------------------------------------------
Sequence                349          0.038  0.940  0.040    S     1     22
-------------------------------------------------------------------------------
cutoff                               0.000  0.000  0.000
```

Explain the output. Go back.

Post-translational modification prediction

- ChloroP - Prediction of chloroplast transit peptides
- LipoP - Prediction of lipoproteins and signal peptides in Gram negative bacteria
- MITOPROT - Prediction of mitochondrial targeting sequences
- PATS - Prediction of apicoplast targeted sequences
- PlasMit - Prediction of mitochondrial transit peptides in Plasmodium falciparum
- Predotar - Prediction of mitochondrial and plastid targeting sequences
- PTS1 - Prediction of peroxisomal targeting signal 1 containing proteins
- SignalP - Prediction of signal peptide cleavage sites

- DictyOGlyc - Prediction of GlcNAc O-glycosylation sites in Dictyostelium
- NetCGlyc - C-mannosylation sites in mammalian proteins
- NetOGlyc - Prediction of O-GalNAc (mucin type) glycosylation sites in mammalian proteins
- NetGlycate - Glycation of epsilon amino groups of lysines in mammalian proteins
- NetNGlyc - Prediction of N-glycosylation sites in human proteins
- OGPET - Prediction of O-GalNAc (mucin-type) glycosylation sites in eukaryotic (non-protozoan) proteins
- YinOYang - O-beta-GlcNAc attachment sites in eukaryotic protein sequences

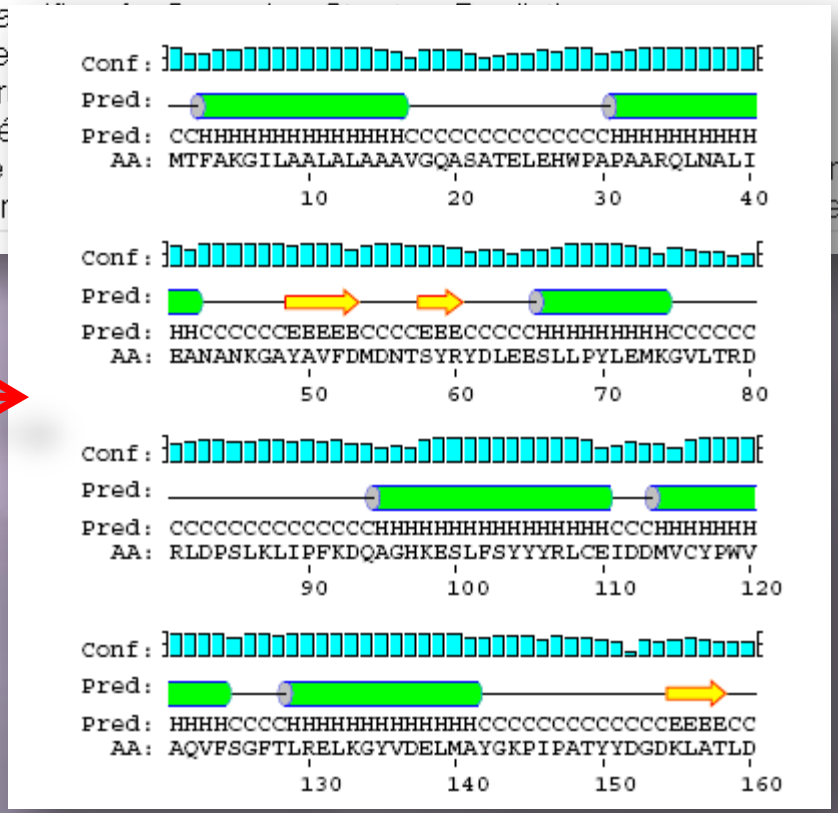
- big-PI Predictor - GPI Modification Site Prediction
- DGPI - Prediction of GPI-anchor and cleavage sites (Mirror site)
- GPI-SOM - Identification of GPI-anchor signals by a Kohonen Self Organizing Map
- Myristoylator - Prediction of N-terminal myristoylation by neural networks
- NMT - Prediction of N-terminal N-myristoylation
- CSS-Palm - Palmitoylation site prediction with CSS
- PrePS - Prenylation Prediction Suite

- NetAcet - Prediction of N-acetyltransferase A (NatA) substrates (in yeast and mammalian proteins)
- NetPhos - Prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins
- NetPhosK - Kinase specific phosphorylation sites in eukaryotic proteins
- NetPhosYeast - Serine and threonine phosphorylation sites in yeast proteins
- GPS - Prediction of kinase-specific phosphorylation sites for 408 human protein kinases in hierarchy new
- Sulfinator - Prediction of tyrosine sulfation sites

Secondary structure prediction

- AGADIR - An algorithm to predict the helical content of peptides
- APSSP - Advanced Protein Secondary Structure Prediction Server
- GOR - Garnier et al, 1996
- HNN - Hierarchical Neural Network method (Guermeur, 1997)
- HTMSRAP - Helical TransMembrane Segment Rotational Angle Prediction new
- Jpred - A consensus method for protein secondary structure prediction at University of Dundee
- JUFO - Protein secondary structure prediction from sequence (neural network)
- nnPredict - University of California at San Francisco (UCSF)
- Porter - University College Dublin
- PredictProtein - PHDsec, PHDacc, PHDhtm, PHDtopology, PHDthreader, MaxHom, EvalSec from Columbia University
- Prof - Cascaded Multiple Cla
- PSA - BioMolecular Enginee
- PSIpred - Various protein str
- SOPMA - Geourjon and Delé
- SSpro - Secondary structure
- DLP-SVM - Domain linker pr

niversity of California
echnology



Tertiary structure prediction

Comparative modeling

- SWISS-MODEL - An automated knowledge-based protein modelling server
- 3Djigsaw - Three-dimensional models for proteins based on homologues of known structure
- CPHmodels - Automated neural-network based protein modelling server
- ESyPred3D - Automated homology modeling program using neural networks
- Geno3d - Automatic modelling of protein three-dimensional structure
- SDSC1 - Protein Structure Homology Modeling Server

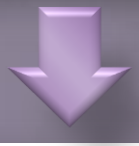
Threading

- 3D-PSSM - Protein fold recognition using 1D and 3D sequence profiles coupled with secondary structure information (Foldfit)
- Fugue - Sequence-structure homology recognition
- HHpred - Protein homology detection and structure prediction by HMM-HMM comparison
- Libellula - Neural network approach to evaluate fold recognition results
- LOOPP - Sequence to sequence, sequence to structure, and structure to structure alignment
- SAM-T02 - HMM-based Protein Structure Prediction
- Threader - Protein fold recognition

- ProSup - Protein structure superimposition
- SWEET - Constructing 3D models of saccharides from their sequences

Ab initio

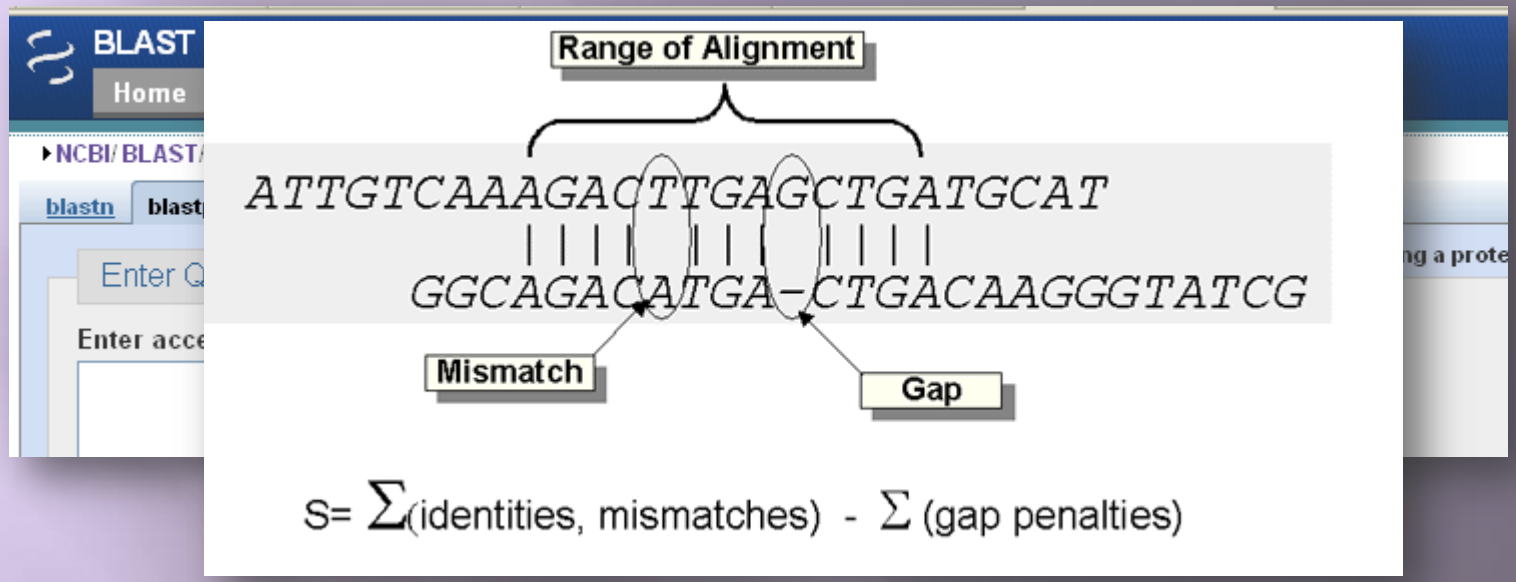
- HMMSTR/Rosetta - Prediction of protein structure from sequence



Assessing tertiary structure prediction

- Anolea - Atomic Non-Local Environment Assessment
- Biotech Validation Suite for Protein Structures
- EVA - EValuation of Automatic protein structure prediction
- LiveBench - Continuous Benchmarking of Structure Prediction Servers
- NQ-Flipper - validation and correction of asparagine and glutamine side-chain amide rotamers in protein structures solved by X-ra
- PROCHECK - Verification of the stereochemical quality of a protein structure
- ProSA-web - recognition of errors in 3D structures of proteins new
- What If - Protein structure analysis program for mutant prediction, structure verification, molecular graphics

Basic Local Alignment Search Tool



	A	C	D	E	F	G	H →
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	-1
G	0	-3	-1	-2	-3	6	-2
H	-2	-3	-1	0	-1	-2	8
↓							

BLOSUM 62