Part II

Protein Sequence Analysis

Seq1:

meepqsdlsi elplsqetfs pksakralpt ntsssppkk meepqsdlsi elplsqetfs

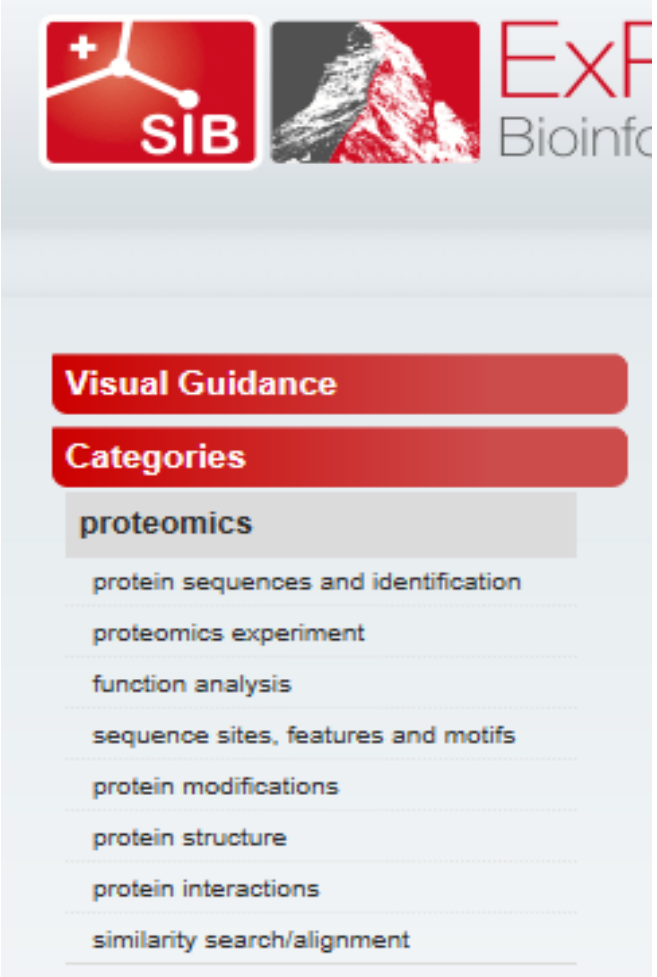
pksakralpt ntsssppkk meepqsdlsi elplsqetfs pksakralpt pksakralpt

想…想象…

你想知道啥？

Sequence analysis of proteins

- Back-translation
- Molecular weights, pIs
- Amino acid composition
- Hydropathy profile



1. Back—translation

- Protein -> DNA
- Use for cloning protein of interest where it may be present in low amount.
- Beware of codon bias and degeneracy of codons.

UUU-Phe	UCU-Ser	UAU-Tyr	UGU-Cys
UUC-Phe	UCC-Ser	UAC-Tyr	UGC-Cys
UUA-Leu	UCA-Ser	UAA-Stop	UGA-Stop
UUG-Leu	UCG-Ser	UAG-Stop	UGG-Trp
CUU-Leu	CCU-Pro	CAU-His	CGU-Arg
CUC-Leu	CCC-Pro	CAC-His	CGC-Arg
CUA-Leu	CCA-Pro	CAA-Gln	CGA-Arg
CUG-Leu	CCG-Pro	CAG-Gln	CGG-Arg
AUU-Ile	ACU-Thr	AAU-Asn	AGU-Ser
AUC-Ile	ACC-Thr	AAC-Asn	AGC-Ser
AUA-Ile	ACA-Thr	AAA-Lys	AGA-Arg
AUG-Met	ACG-Thr	AAG-Lys	AGG-Arg
GUU-Val	GCU-Ala	GAU-Asp	GGU-Gly
GUC-Val	GCC-Ala	GAC-Asp	GGC-Gly
GUA-Val	GCA-Ala	GAA-Glu	GGA-Gly
GUG-Val	GCG-Ala	GAG-Glu	GGG-Gly

Biased codon usage

Amino acid	Codon	Bacteria	Yeast	Fruit Fly	Human
Leu	UUA				
	UUG		Preferred		
	CUU				
	CUC				
	CUA				
	CUG	Preferred		Preferred	Preferred
Val	GUU	Preferred	Preferred		
	GUC				
	GUA				
	GUG			Preferred	Preferred

Back-translation Tool

Sequence Manipulation Suite:

Reverse Translate

Reverse Translate accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned. Use Reverse Translate when designing PCR primers to anneal to an unsequenced coding sequence from a related species.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 20,000,000 characters.

```
>sample sequence
ACDEFGHIKLMNPQRSTVWY*
```

Submit | Clear | Reset

Enter the codon table you wish to use (in GCG format). The default codon usage table was generated using all the E. coli coding sequences in GenBank. It was obtained from the Codon Usage Database.

AmAcid	Codon	Number	/1000	Fraction	..
Gly	GGG	50527.00	11.12	0.15	
Gly	GGA	39036.00	8.59	0.12	
Gly	GGT	114185.00	25.14	0.34	
Gly	GGC	130043.00	28.63	0.39	

*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Mon Nov 6 02:56:29 2017
Valid XHTML 1.0; Valid CSS

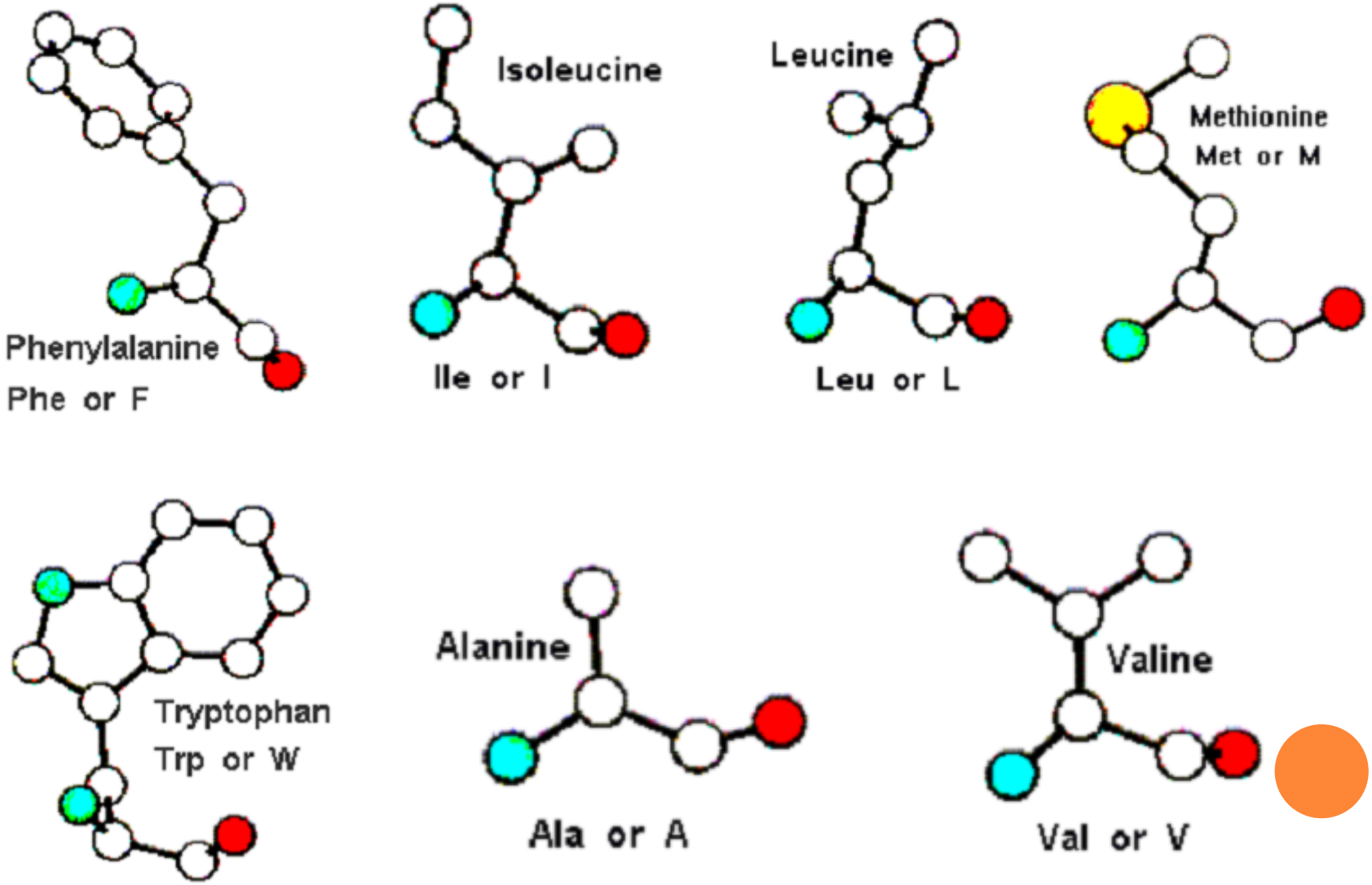
2. Molecular weights, pIs

- Aid in designing of purification experiments e.g. SDS-PAGE, IEF, 2-Dimensional Gel, Column chromatography etc.

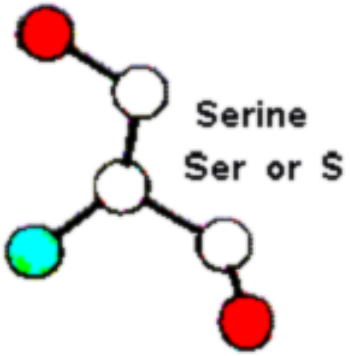
3. Amino Acid Composition

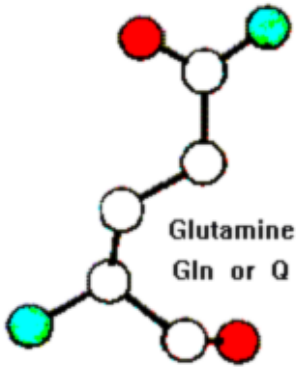
- Determine the percentages of amino acid residues present in a protein molecule.
- Uses:
  - determine the lifestyles of organisms: high percentages of Glu (-) and both Lys and Arginine (+) in hyperthermophiles vs. mesophiles -> absent (Tekaia *et al.*, 2002).
  - predict structural class (Luo *et al.*, 2002).

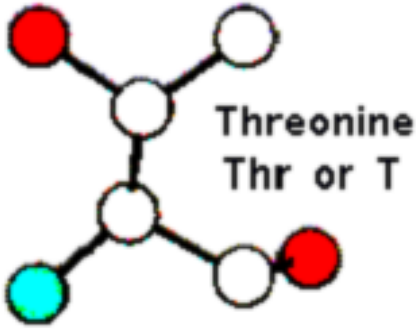
Nonpolar amino acids (FILMWAV)

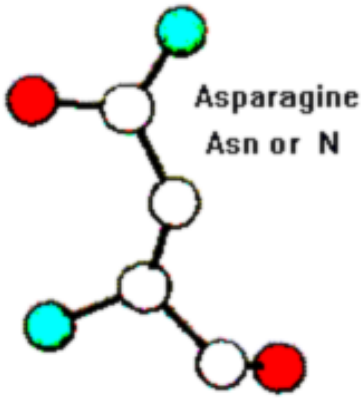


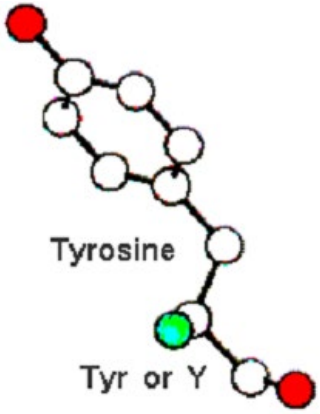
Polar uncharged (S⁻Q⁺T⁻N⁺Y⁻)



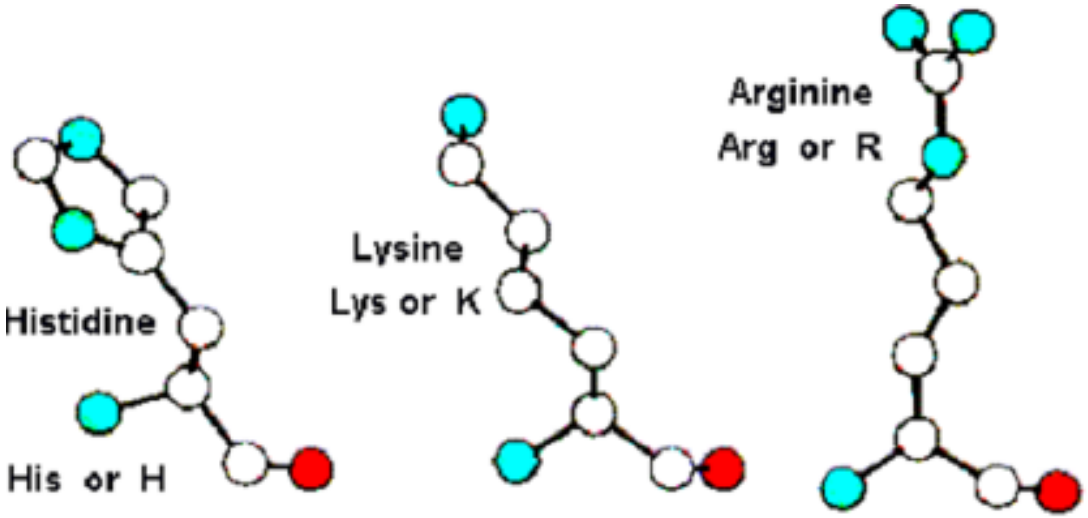


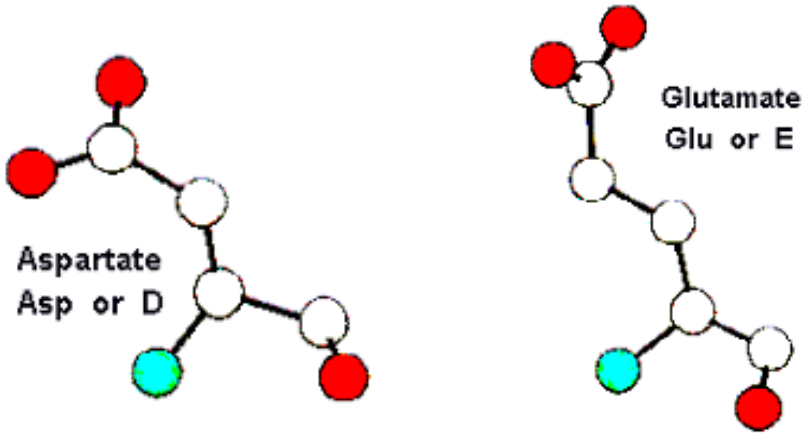




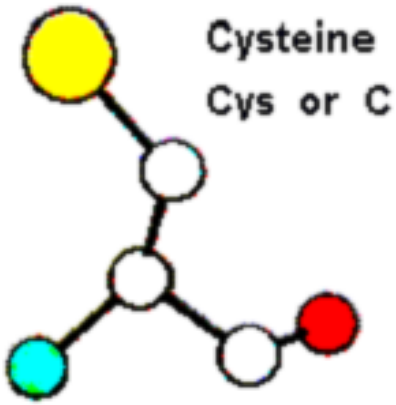


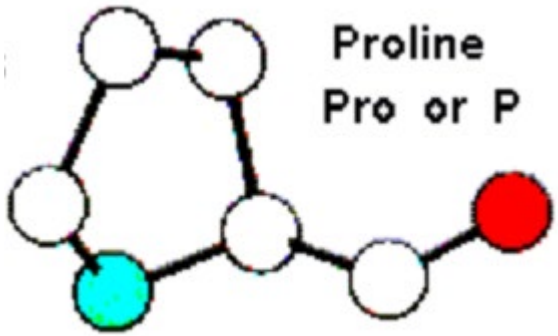
Polar charged (KHERD)

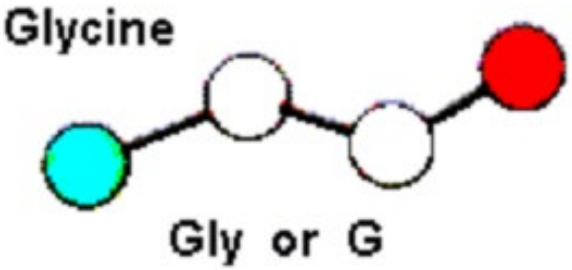




Unique Properties







Protein functions from specific residues

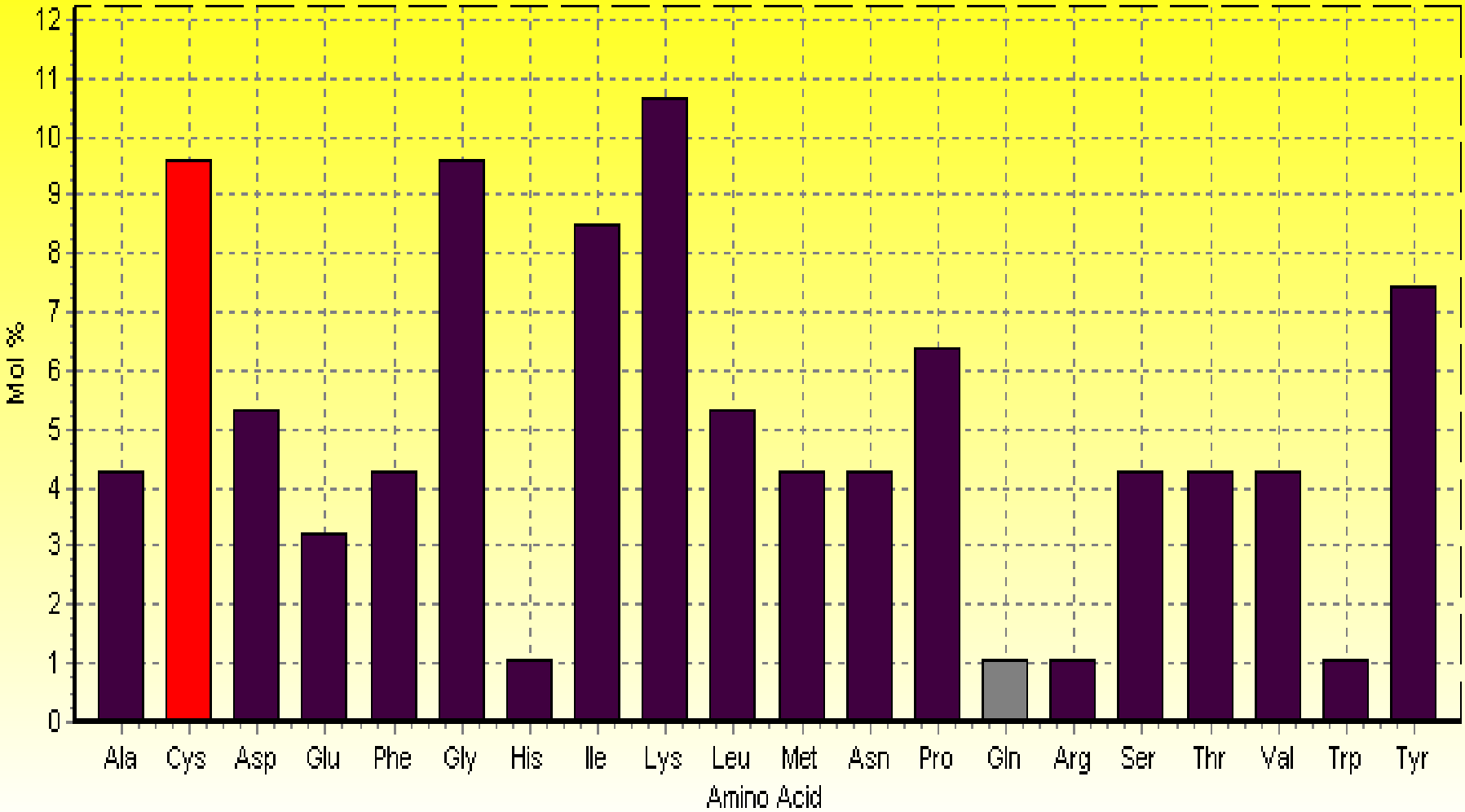
- C Disulphide-rich, zinc fingers
- DE Acidic proteins (unknown)
- G Collagens
- H Histidine-rich glycoprotein
- KR Nuclear proteins, nuclear localisation
- P Collagen, filaments
- ST Mucins (high molecular weight glycosylated proteins that form a major part of a protective biofilm on the surface of epithelial cells)

Protein functions from specific residues

- Polar (C, D, E, H, K, N, Q, R, S, T) – active sites
- Aromatic (F, H, W, Y) – protein ligand-binding sites
- Zn^{2+}-coordinates (C, D, E, H, N, Q) – active site, zinc finger
- Ca^{2+}-coordinates (D, E, N, Q) – ligand-binding site
- Mg/Mn-coordinates (D, E, N, S, R, T) – Mg^{2+} or Mn^{2+} catalysis, ligand binding
- Phosphate-binding (H, K, R, S, T) – phosphate and sulfate binding

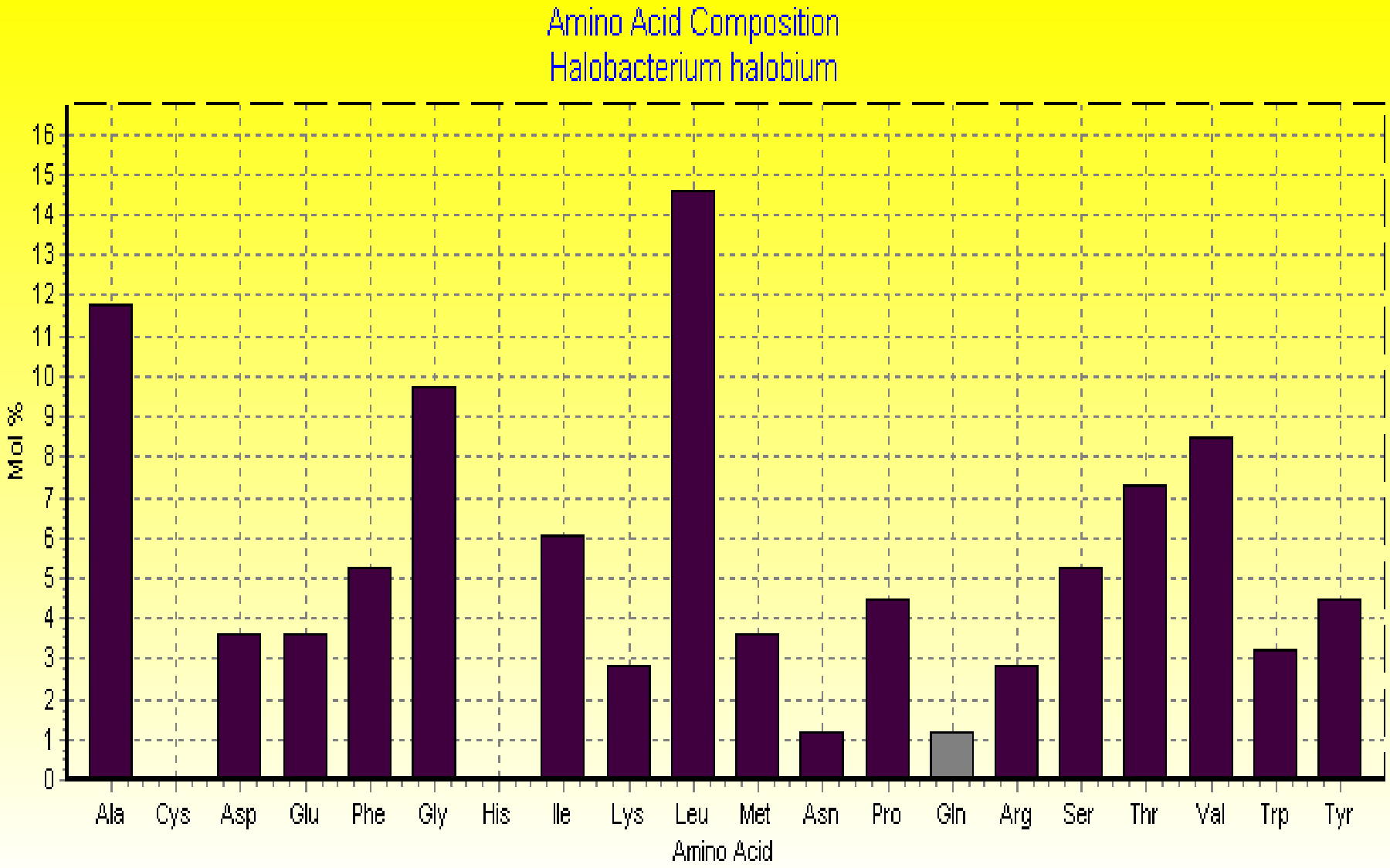
Amino Acid Composition: Hottentotta judaica

Protein: Hottentotta judaica
Length: 94 amino acids
Molecular weight: 10510.93 Daltons



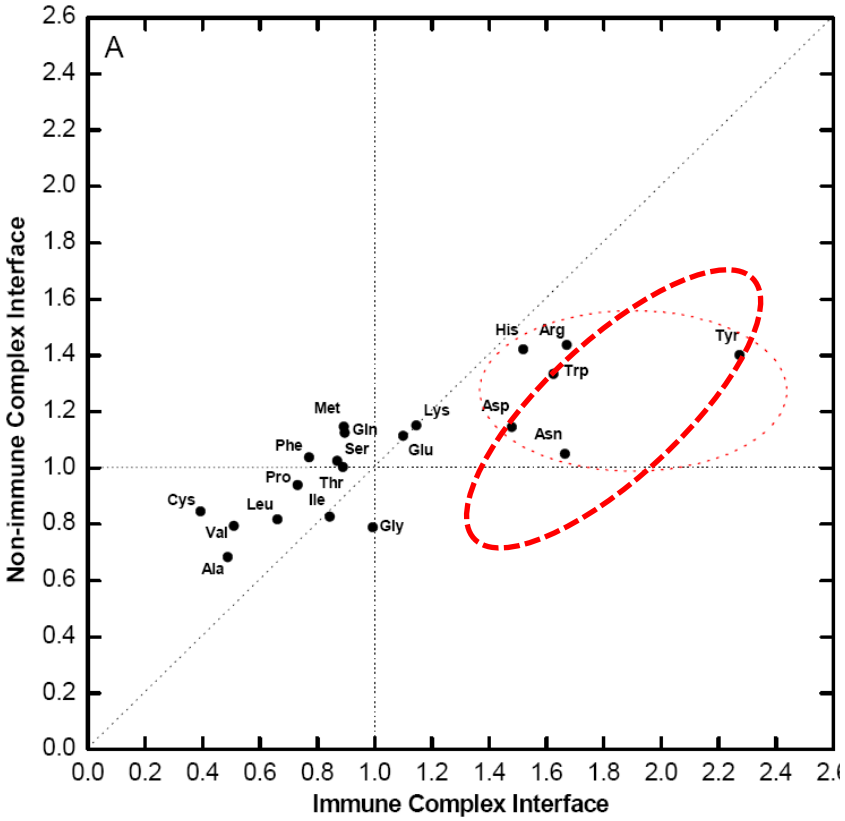
Amino Acid Composition: Halobacterium halobium

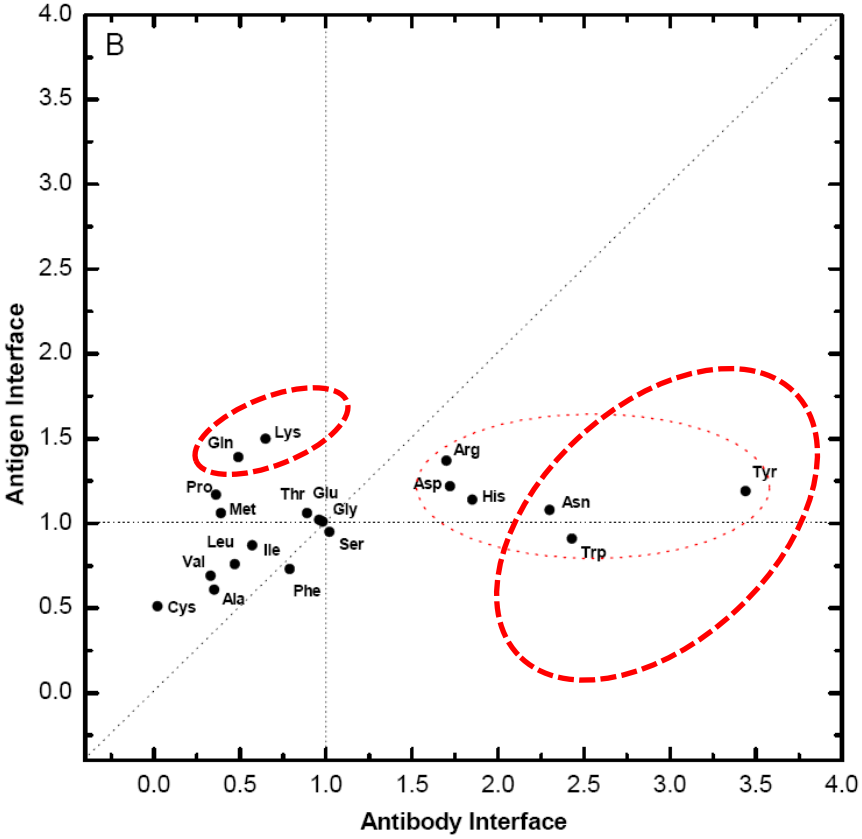
Protein: Halobacterium halobium
Length: 247 amino acids
Molecular weight: 26723.79 Daltons

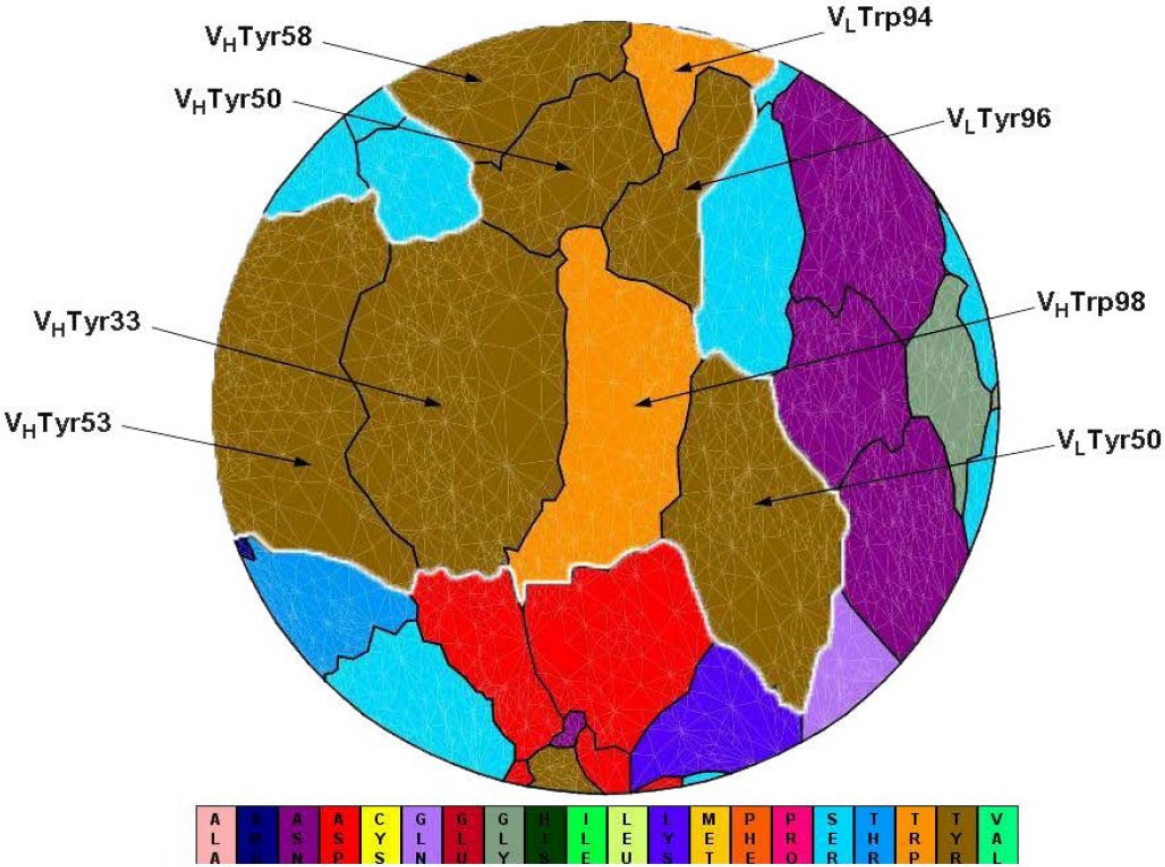


作用界面抗体侧氨基酸组成规律

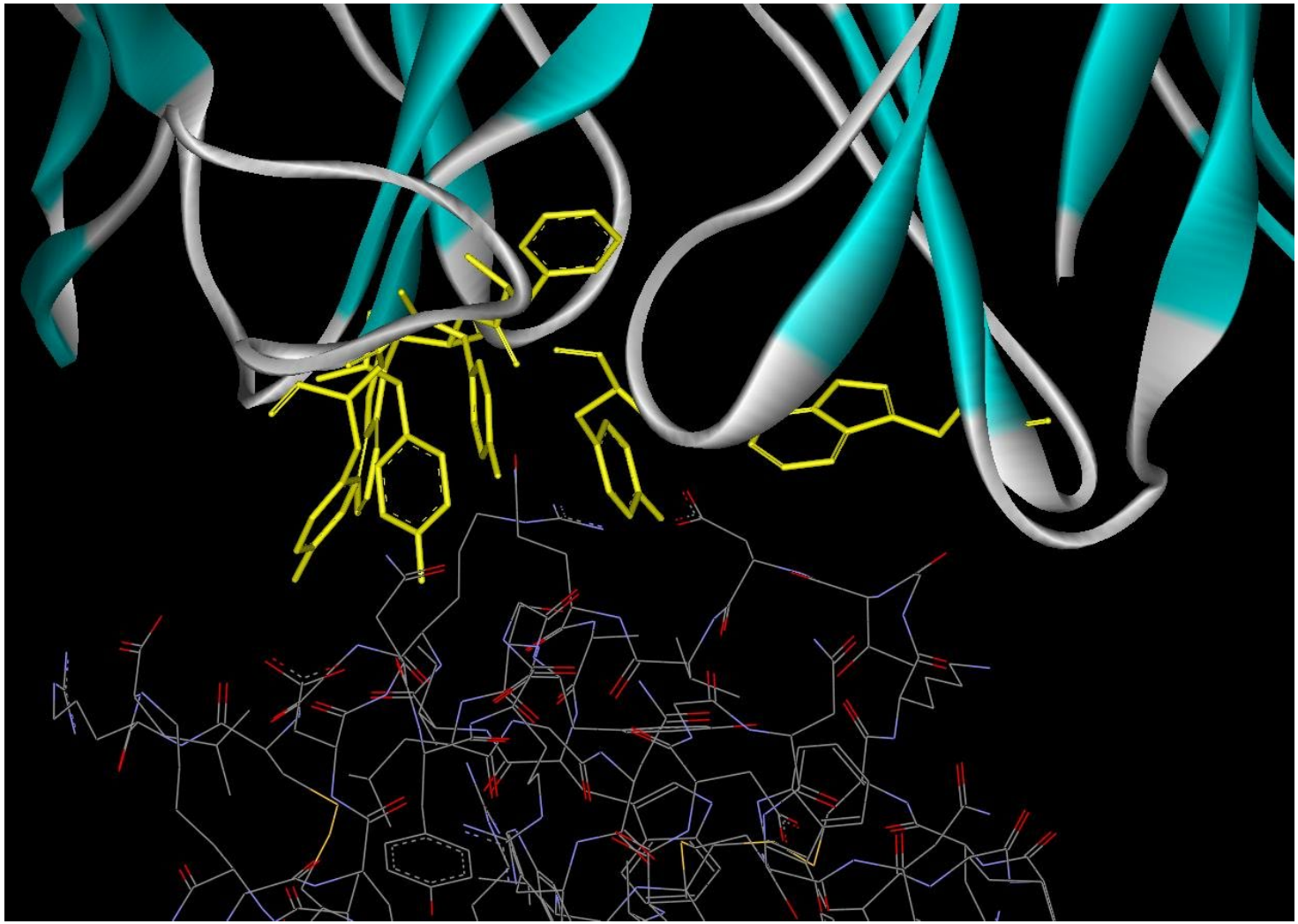
2009.0



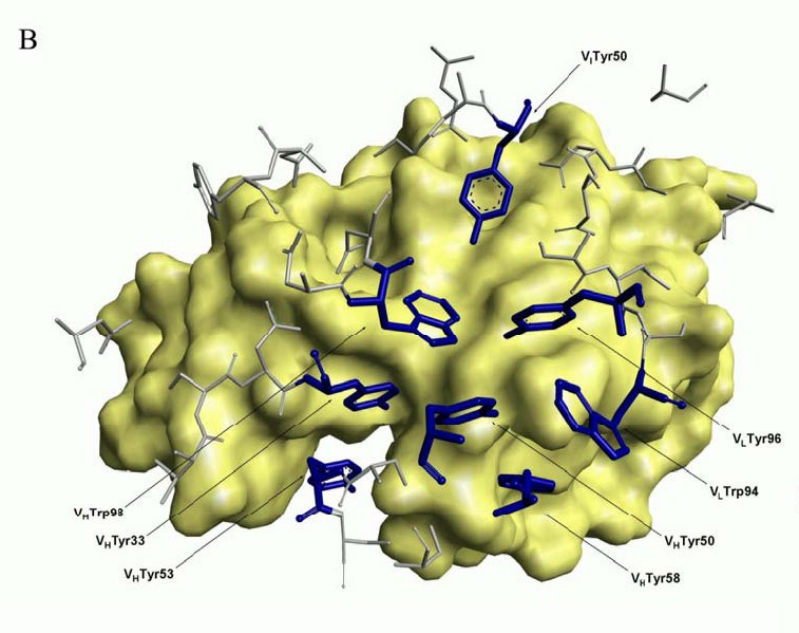


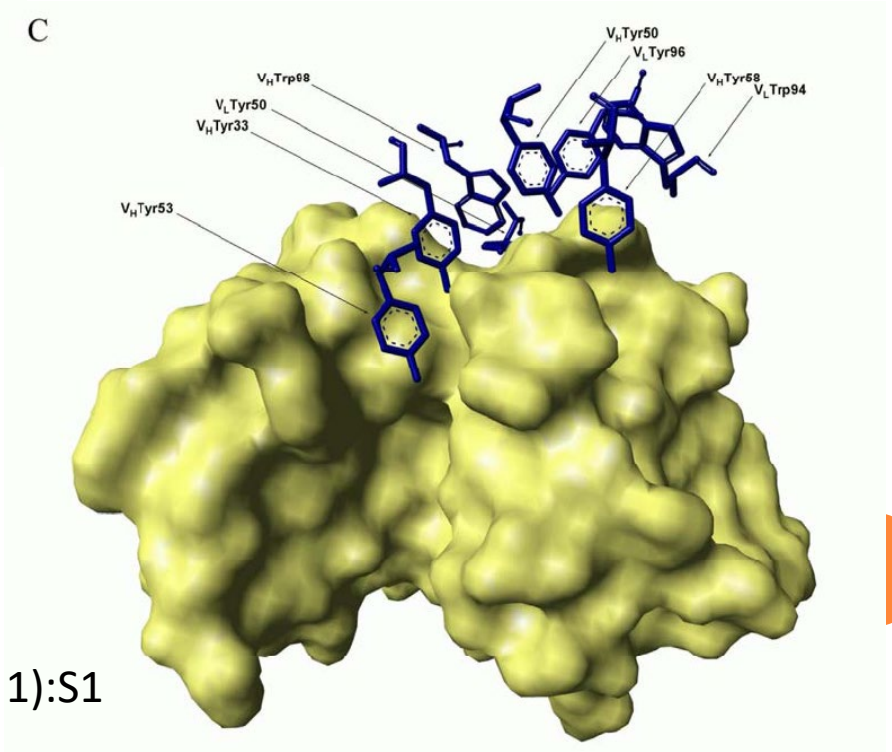

V_HTyr58
V_LTrp94
V_HTyr50
V_LTyr96
V_HTyr33
V_HTrp98
V_HTyr53
V_LTyr50
ALA
ASN
ASP
CYS
GLN
GLU
GLY
ILE
LEU
MET
PHE
PRO
SER
THR
TRP
TYR
VAL

1BVK



Ab-Ag界面上富集的芳香族氨基酸空间分布特征





immunome research, 2010 , 6(Suppl 1):S1

AACompIdent tool

AACompIdent is a tool which allows the identification of a protein from its amino acid composition [references]. It searches the Swiss-Prot and / or TrEMBL databases for proteins, whose amino acid compositions are closest to the amino acid composition given.

Documentation is available.

Few amino acid analysis techniques produce composition results for all amino acids. We currently have indexed Swiss-Prot and TrEMBL for the following constellations. Please choose one of them:

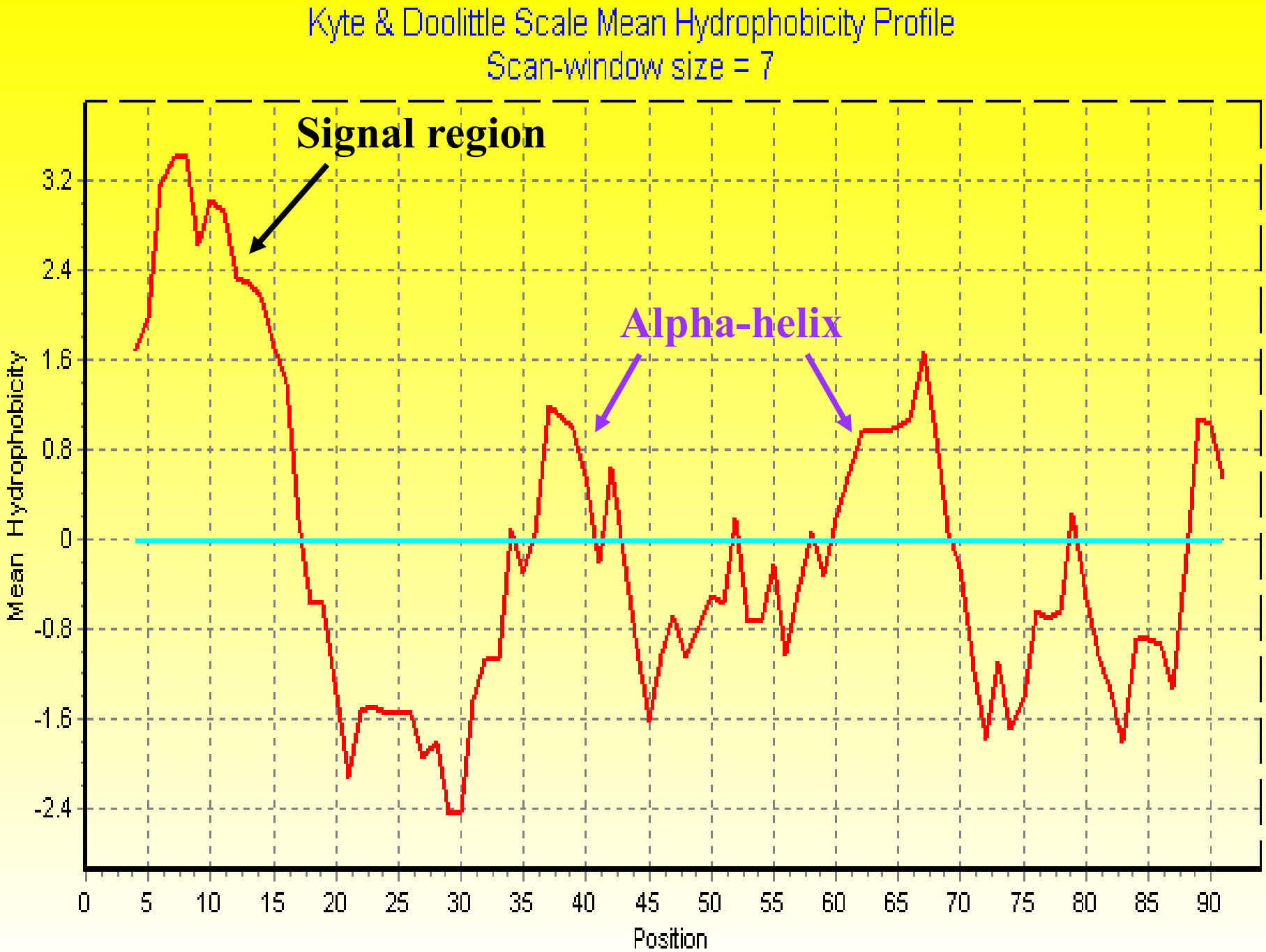
1. Constellation 0: **ALL amino acids**: Ala, Ile, Pro, Val, Arg, Leu, Ser, Thr, Gly, Met, His, Phe, Tyr, Lys, Asp, Asn, Gln, Glu, Cys and Trp.
2. Constellation 1: Ala, Ile, Pro, Val, Arg, Leu, Ser, Asx, Thr, Glx, Gly, Met, His, Phe and Tyr.
   (Asp+Asn=Asx; Gln+Glu=Glx; Lys, Cys and Trp are not considered).
3. Constellation 2: Ala, Ile, Pro, Val, Arg, Leu, Ser, Asx, Lys, Thr, Glx, Gly, Met, His, Phe and Tyr.
   (Asp+Asn=Asx; Gln+Glu=Glx; Cys and Trp are not considered).
4. Constellation 5: Ala, Ile, Pro, Val, Arg, Leu, Ser, Asx, Lys, Thr, Glx, Gly, Met, His, Phe, Tyr and Cys.
   (Asp+Asn=Asx; Gln+Glu=Glx; Trp is not considered).

Hydropathy Profiles

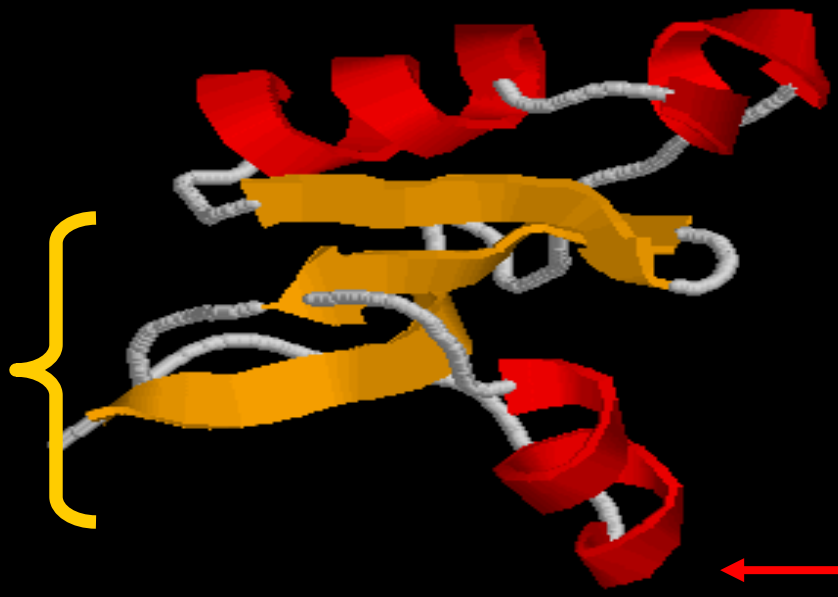
- Hydropathy – describe the hydrophobicity and hydrophilicity of a protein sequence.
- A graph in which hydropathy values are calculated within a sliding window and plotted for each residue in a protein sequence.

A sliding window

M	K	F	F	L	M	C	L	I	I	F	P	I	M	G	V	L	G


Kyte & Doolittle Scale Mean Hydrophobicity Profile
Scan-window size = 7
Signal region
Alpha-helix
Mean Hydrophobicity
3.2
2.4
1.6
0.8
0
-0.8
-1.6
-2.4
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
Position

MKFFLMCLIIFPIMGVLGKKNGYPLDRNGKTTECSGVNAIAPHYCNSECT
KVYYAKSGYCCWGACYCFGLEDDKPIGPMKDITKKYCDVQIIPS
(Signal region:1-18 Toxin region:19-94
Disulphide)

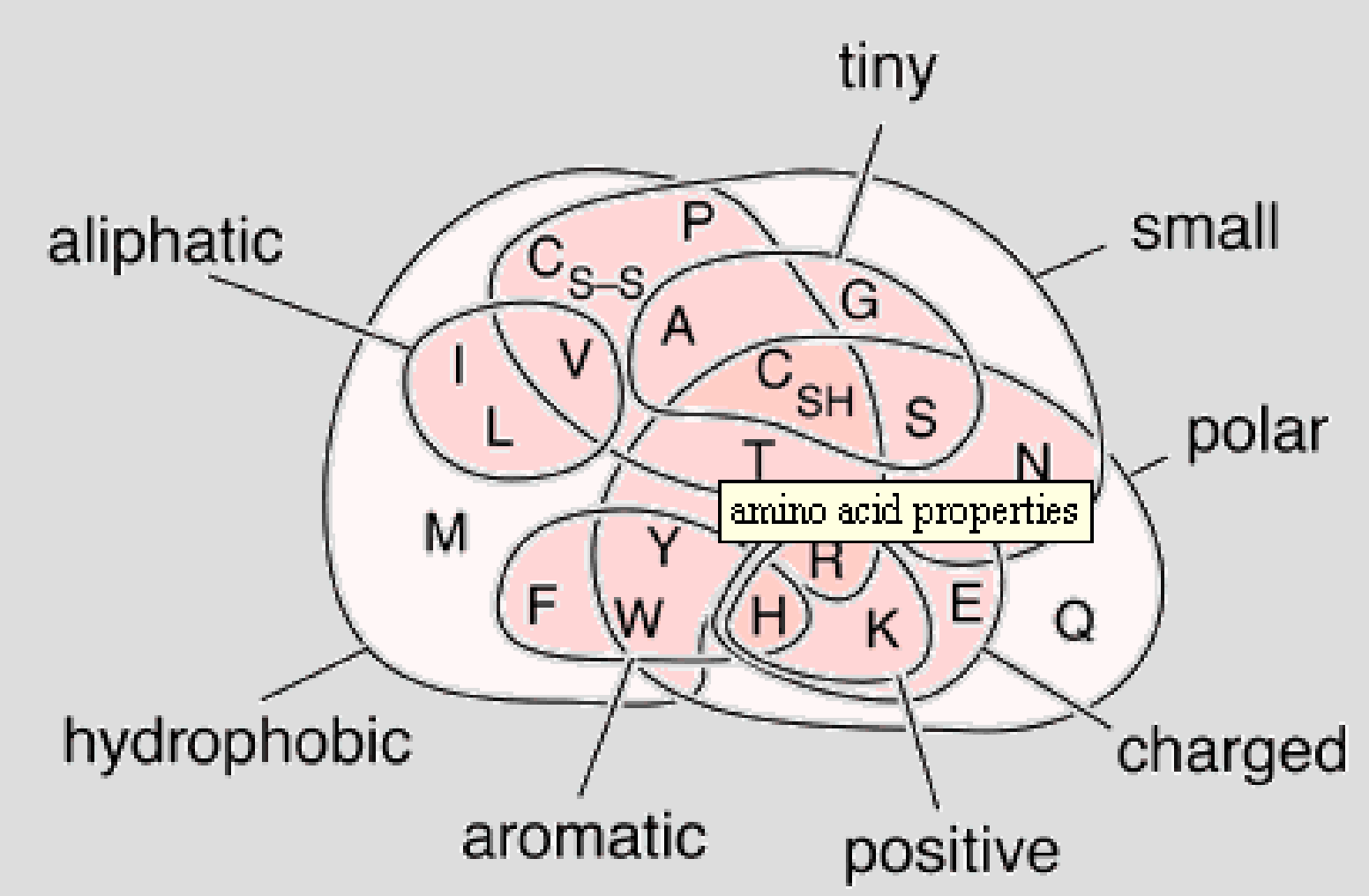


A SCHEMATIC REPRESENTATION OF A 3-D STRUCTURE OF A SCORPION TOXIN

Hydropathy Profiles

- Hydropathy scale – each amino acid is assigned a value reflecting its relative hydrophobicity and hydrophilicity.
- 2 broad classes of scales:
  - Environmental characteristics of protein residues.
  - Experimental measurements of amino acid physiochemical properties.

Venn Diagram of the 20 amino acid physiochemical properties



Hydropathy Profiles

- Basic ranking: internal {FILMV}, external {DEHKNQR}, ambivalent {ACGPSTWY}

Scale Name	Residue ranking
Kyte	IVLFCMAGTSWYPHDNEQKR
Eisenberg	IFVLWMAGCYPTSHENQDKR
Cornette	FILVMHYCWAQRTGESPNKD

Hydropathy Profiles

- Detect possible transmembrane domains (consecutive 20-25 runs of hydrophobic amino acids).
- Hydrophobic protein cores

protein sequnce analysis too | ExPASy: SIB Bioinformatics R

https://www.expasy.org

最常访问 火狐官方站点 新手上路 常用网址 京东商城

ExPASy
Bioinformatics Resource Portal

Home About Contact

Query all databases | search | help

Visual Guidance

Categories

- proteomics
- genomics
- structure analysis
- systems biology
- evolutionary biology
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- medicinal chemistry
- glycomics

Resources A..Z

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

Featuring today

Compute pI/MW

Compute isoelectric point and molecular weight from protein sequence
[details]

FindPept

Identify peptides resulting from unspecific protein cleavage from their experimental masses
[details]

FindMod

Predict potential protein post-translational modifications and find potential single amino

Popular resources

- UniProtKB
- SWISS-MODEL
- STRING
- PROSITE

Latest News

Protein Spotlight: A way in - 2020-03-26

As children in Scotland, back home from school and when the weather was dry, we would fling our schoolbags into the hall and grab a few golf clubs, a ball and a tee. There was no need for a change of clothes or shoes, ...More.

New neXtProt application release 2020-02-28 - 2020-02-28

Modifications to SPARQL queries. For

ProtScale

ProtScale [Reference / Documentation] allows you to compute and represent the profile produced by any amino acid scale on a selected protein.

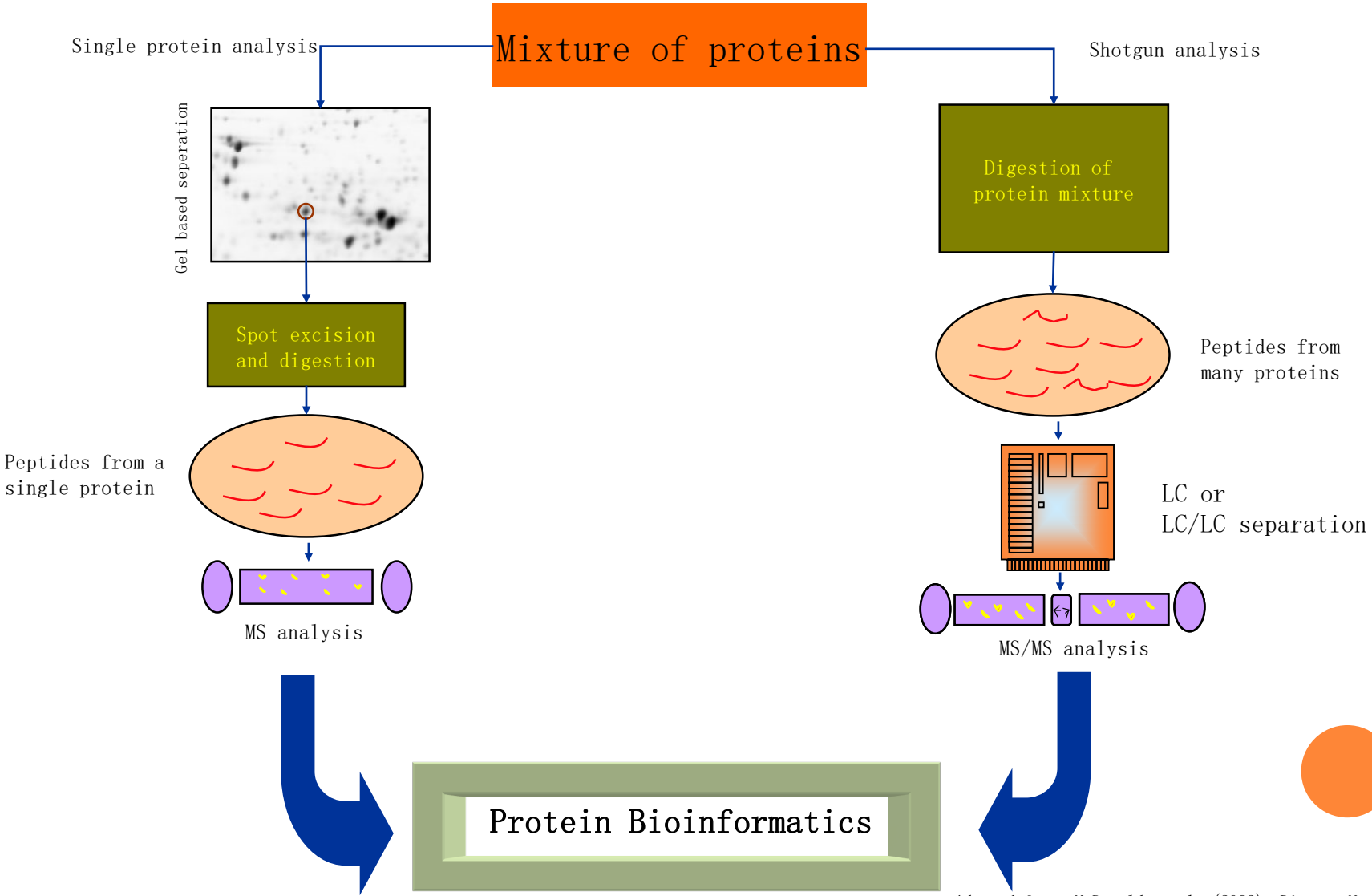
An **amino acid scale** is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides 57 predefined scales entered from the litera

- ○ Molecular weight
- ○ Bulkiness
- ○ Polarity / Grantham
- ○ Recognition factors
- ○ Hphob. OMH / Sweet et al.
- ● Hphob. / Kyte & Doolittle
- ○ Hphob. / Abraham & Leo
- ○ Hphob. / Bull & Breese
- ○ Hphob. / Guy
- ○ Hphob. / Miyazawa et al.
- ○ Hphob. / Roseman
- ○ Hphob. / Wolfenden et al.
- ○ Hphob. HPLC / Wilson & al
- ○ Hphob. HPLC pH3.4 / Cowan
- ○ Hphob. / Rf mobility
- ○ HPLC / TFA retention
- ○ HPLC / retention pH 2.1
- ○ % buried residues
- ○ Hphob. / Chothia
- ○ Ratio hetero end/side
- ○ Average flexibility
- ○ beta-sheet / Chou & Fasman
- ○ alpha-helix / Deleage & Roux
- ○ beta-turn / Deleage & Roux
- ○ alpha-helix / Levitt
- ○ beta-turn / Levitt
- ○ Antiparallel beta-strand
- ○ Number of codon(s)
- ○ Polarity / Zimmerman
- ○ Refractivity
- ○ Hphob. / Eisenberg et al.
- ○ Hphob. / Hopp & Woods
- ○ Hphob. / Manavalan et al.
- ○ Hphob. / Black
- ○ Hphob. / Fauchere et al.
- ○ Hphob. / Janin
- ○ Hphob. / Rao & Argos
- ○ Hphob. / Tanford
- ○ Hphob. / Welling & al
- ○ Hphob. HPLC / Parker & al
- ○ Hphob. HPLC pH7.5 / Cowan
- ○ HPLC / HFBA retention
- ○ Transmembrane tendency
- ○ HPLC / retention pH 7.4
- ○ % accessible residues
- ○ Hphob. / Rose & al
- ○ Average area buried
- ○ alpha-helix / Chou & Fasman
- ○ beta-turn / Chou & Fasman
- ○ beta-sheet / Deleage & Roux
- ○ Coil / Deleage & Roux
- ○ beta-sheet / Levitt
- ○ Total beta-strand
- ○ Parallel beta-strand

Primary structure analysis

- ProtParam - Physico-chemical parameters of a protein sequence (amino-acid and atomic compositions, isoelectric point, extinction coefficient, etc.)
- Compute pI/Mw - Compute the theoretical isoelectric point (*pI*) and molecular weight (*Mw*) from a UniProt Knowledgebase entry or for a user sequence
- ScanSite pI/Mw - Compute the theoretical *pI* and *Mw*, and multiple phosphorylation states
- MW, pI, Titration curve - Computes *pI*, composition and allows to see a titration curve
- Scratch Protein Predictor
- HeliQuest - A web server to screen sequences with specific alpha-helical properties
- Radar - De novo repeat detection in protein sequences
- REP - Searches a protein sequence for repeats
- REPRO - De novo repeat detection in protein sequences
- T-REKS - De novo detection and alignment of repeats in protein sequences
- TRUST - De novo repeat detection in protein sequences
- XSTREAM - De novo tandem repeat detection and architecture modeling in protein sequences
- SAPS - Statistical analysis of protein sequences at EMBnet-CH [Also available at EBI]
- Coils - Prediction of coiled coil regions in proteins (Lupas's method) at EMBnet-CH [Also available at PBIL]
- Paircoil - Prediction of coiled coil regions in proteins (Berger's method)
- Paircoil2 - Prediction of the parallel coiled coil fold from sequence using pairwise residue probabilities with the Paircoil algorithm.
- Multicoil - Prediction of two- and three-stranded coiled coils
- 2ZIP - Prediction of Leucine Zippers
- ePESTfind - Identification of PEST regions
- HLA_Bind - Prediction of MHC type I (HLA) peptide binding
- PEPVAC - Prediction of supertypic MHC binders
- RANKPEP - Prediction of peptide MHC binding
- SYFPEITHI - Prediction of MHC type I and II peptide binding
- ProtScale - Amino acid scale representation (Hydrophobicity, other conformational parameters, etc.)
- Drawhca - Draw an HCA (Hydrophobic Cluster Analysis) plot of a protein sequence

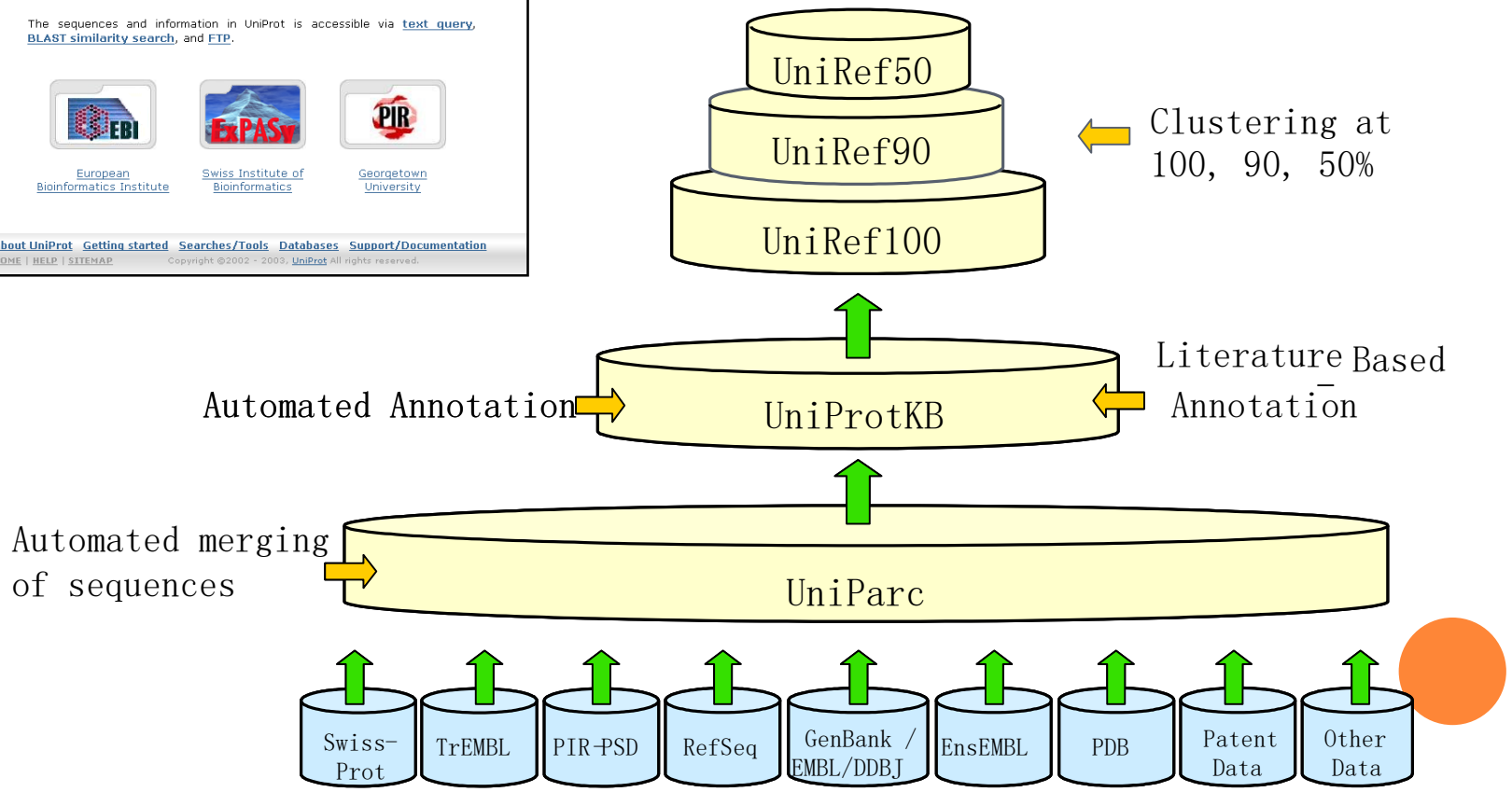
Single protein and shotgun analysis



Adapted from: McDonald et al. (2002). Disease Markers 18:99-

Universal Protein Resource

http://www.uniprot.org/



protein sequnce analysis too | ExPASy: SIB Bioinformatics R

https://www.expasy.org

最常访问 火狐官方站点 新手上路 常用网址 京东商城

ExPASy
Bioinformatics Resource Portal

Home About Contact

Query all databases | search | help

Visual Guidance

Categories

- proteomics
- genomics
- structure analysis
- systems biology
- evolutionary biology
- population genetics
- transcriptomics
- biophysics
- imaging
- IT infrastructure
- medicinal chemistry
- glycomics

Resources A..Z

ExPASy is the **SIB Bioinformatics Resource Portal** which provides access to scientific databases and software tools (i.e., *resources*) in different areas of life sciences including proteomics, genomics, phylogeny, systems biology, population genetics, transcriptomics etc. (see Categories in the left menu). On this portal you find resources from many different SIB groups as well as external institutions.

Featuring today

Compute pI/MW

Compute isoelectric point and molecular weight from protein sequence
[details]

FindPept

Identify peptides resulting from unspecific protein cleavage from their experimental masses
[details]

FindMod

Predict potential protein post-translational modifications and find potential single amino

Popular resources

- UniProtKB
- SWISS-MODEL
- STRING
- PROSITE

Latest News

Protein Spotlight: A way in - 2020-03-26

As children in Scotland, back home from school and when the weather was dry, we would fling our schoolbags into the hall and grab a few golf clubs, a ball and a tee. There was no need for a change of clothes or shoes, ...More.

New neXtProt application release 2020-02-28 - 2020-02-28

Modifications to SPARQL queries. For

Try Swiss Institute of Bioinformatics

- https://www.expasy.org/
- http://www.uniprot.org
- How to calculate theoretical MW/PI from protein sequence?