

Stabilité conformationnelle des protéines : Comprendre et prédire la structure 3D

Conformational stability of proteins: understanding and predicting the 3D structure

Claudine MAYER
Université Paris Diderot
Institut Pasteur



Unité de Microbiologie Structurale

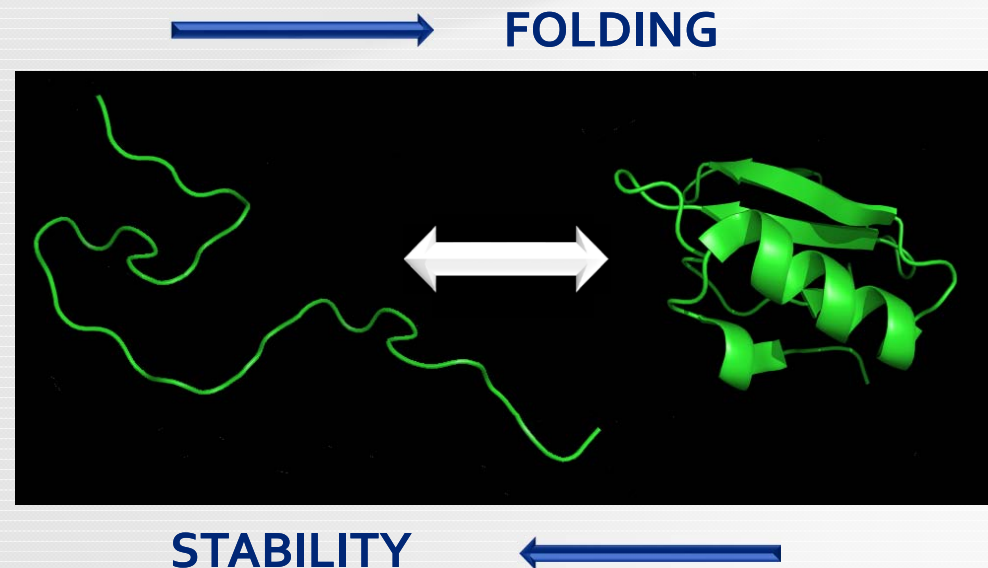
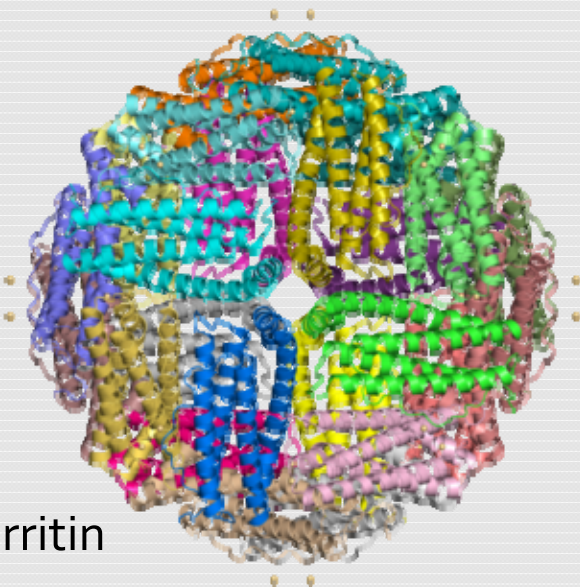
A definition of protein stability ?

The ability of a protein to retain its structural conformation or its activity when subjected to physical or chemical manipulations

Structural conformation = 3D structure, quaternary structure (oligomerisation state), interaction with its partners (in case of macromolecular assemblies)

The word is used in different ways by different people. For example, a physical biochemist and a biotechnologist may each mean something different when they speak of stability.

Quaternary structure



A definition of protein stability ?

The conformational stability is the free energy difference between the native folded state and unfolded state under defined conditions

Defined conditions = ambient or physiological conditions

Physical biochemist

Thermodynamic stability

Easy to study for reversible systems

The larger and more positive ΔG , the more stable is the protein to denaturation

Biotechnologist

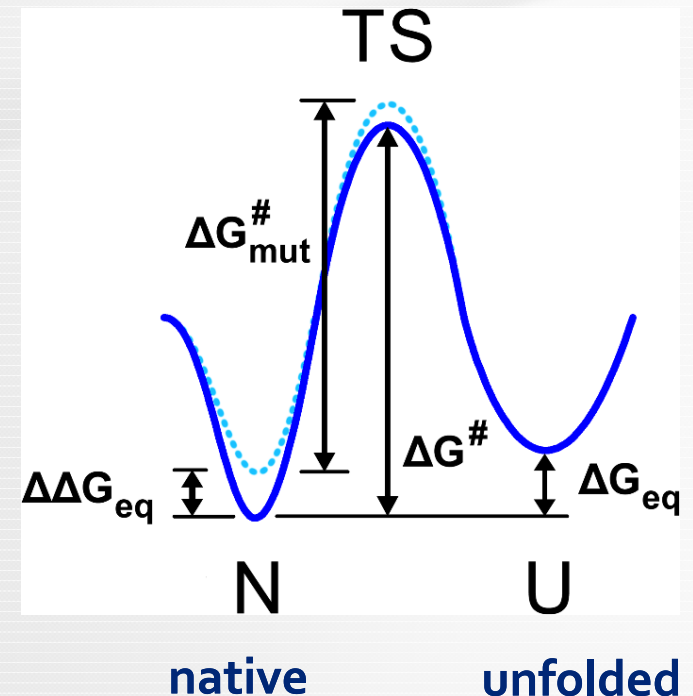
Kinetic stability

Determination of the rate of unfolding

Factors affecting stability are the relative free energies of the folded (N) and the transition state (TS)

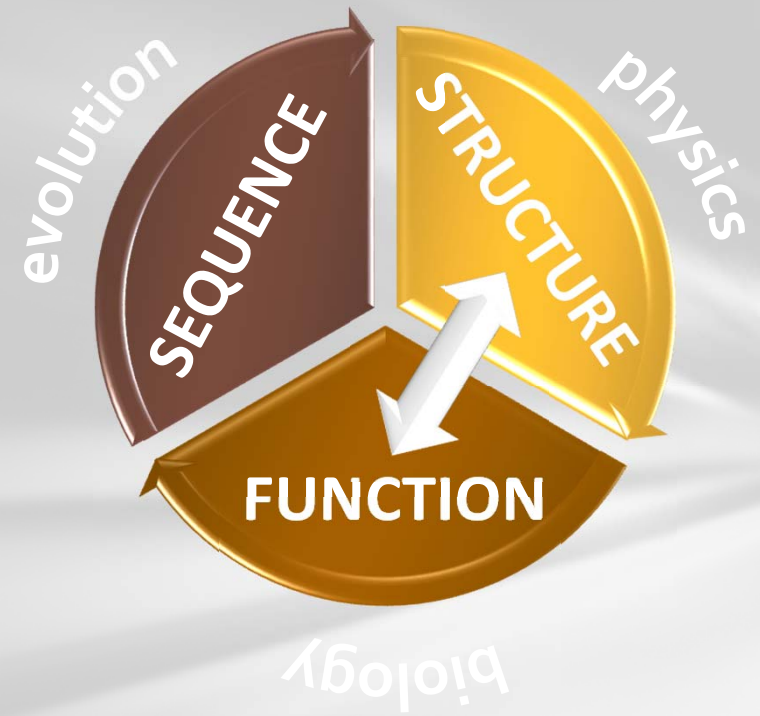
$$\Delta G = G_U - G_N$$

transition state

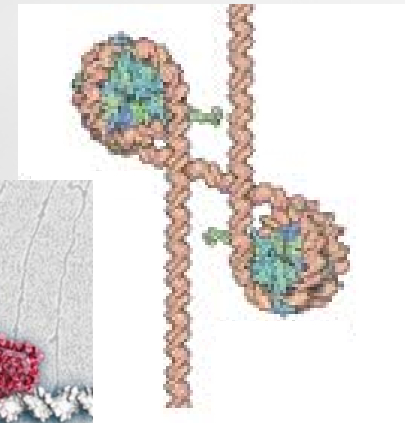
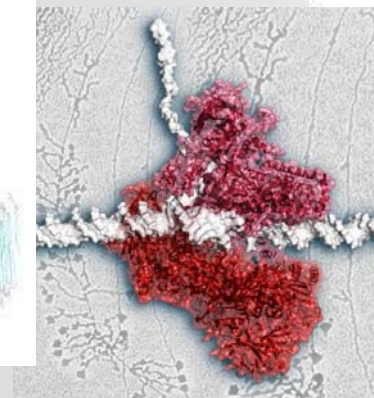
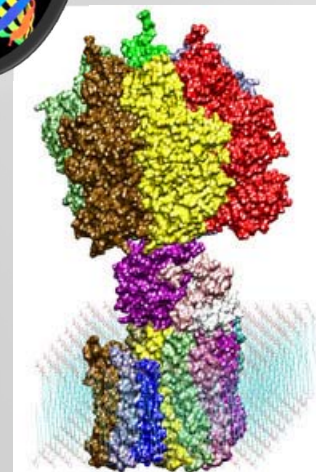
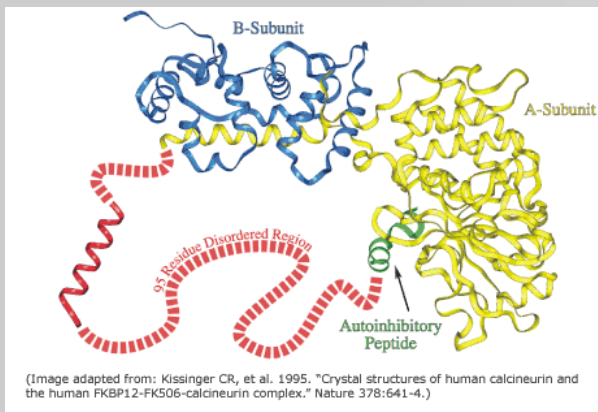


Structure-function relationships

The 3D structure of a protein defines not only its size and its shape but also its function

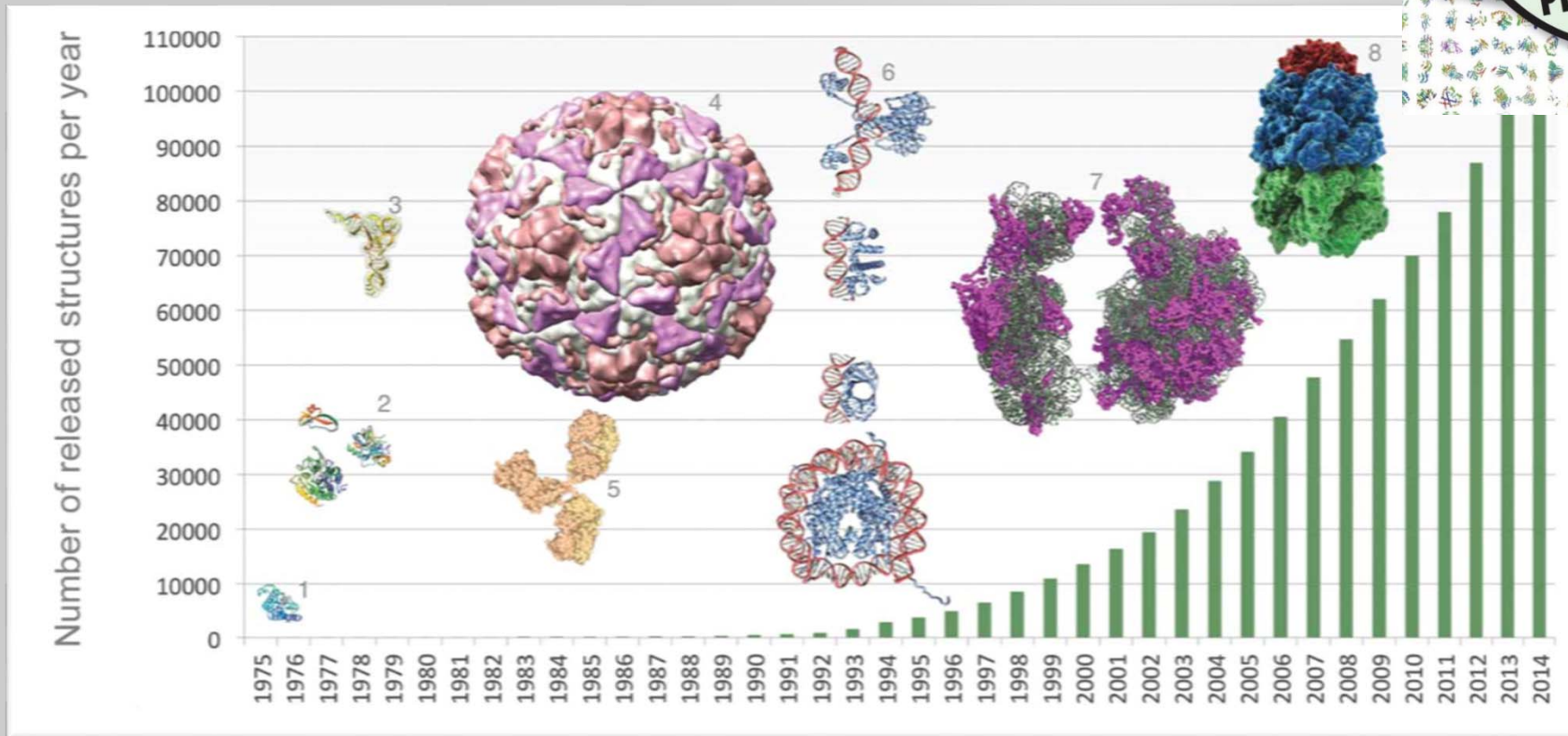


Exception
IDPs & IDRs



Available structural data in the PDB

111 956 structures in the PDB : most of the protein can be structurally studied e.g. the structure exists or it can be modelled



What is a structure ?

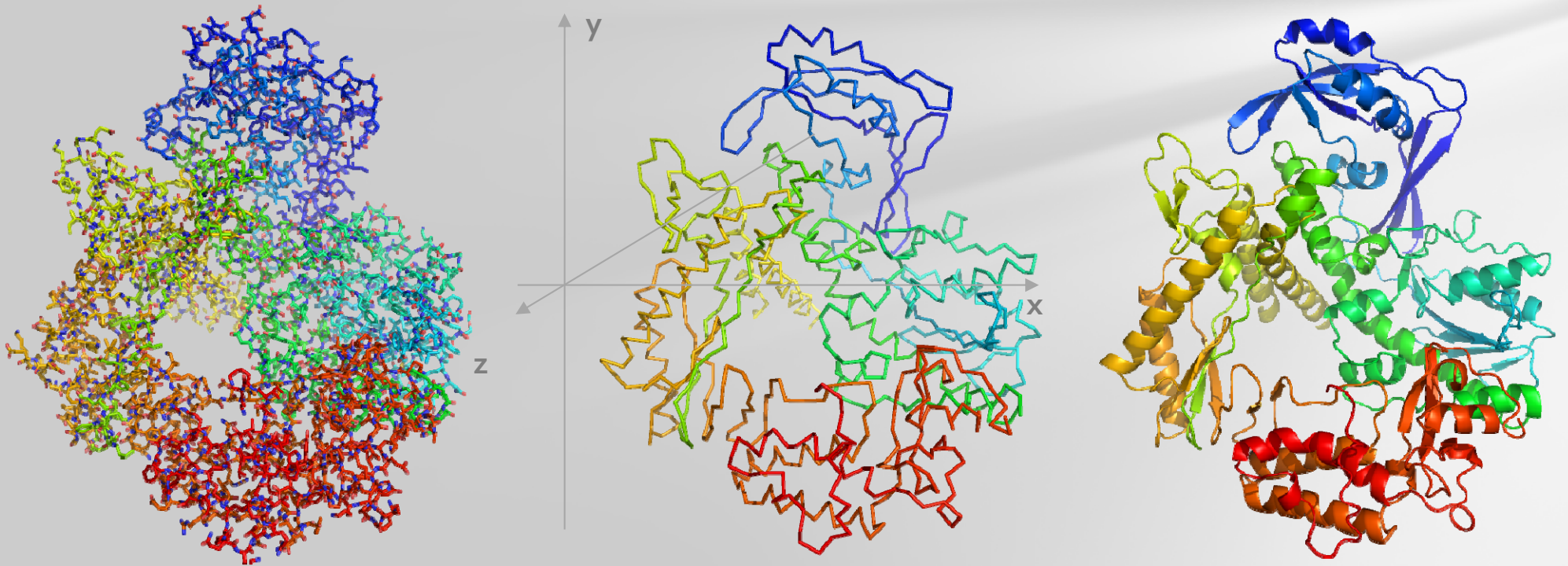
Structure prediction

Stability prediction

What is a structure ?

A 3-dimensional description of all atoms

Coordinates (x,y,z) of all atoms of the aminoacids that compose the protein
The quality of the structure depends on the resolution of the experimental data provided by the method used to obtain it (X-ray, NMR, cryo-EM)

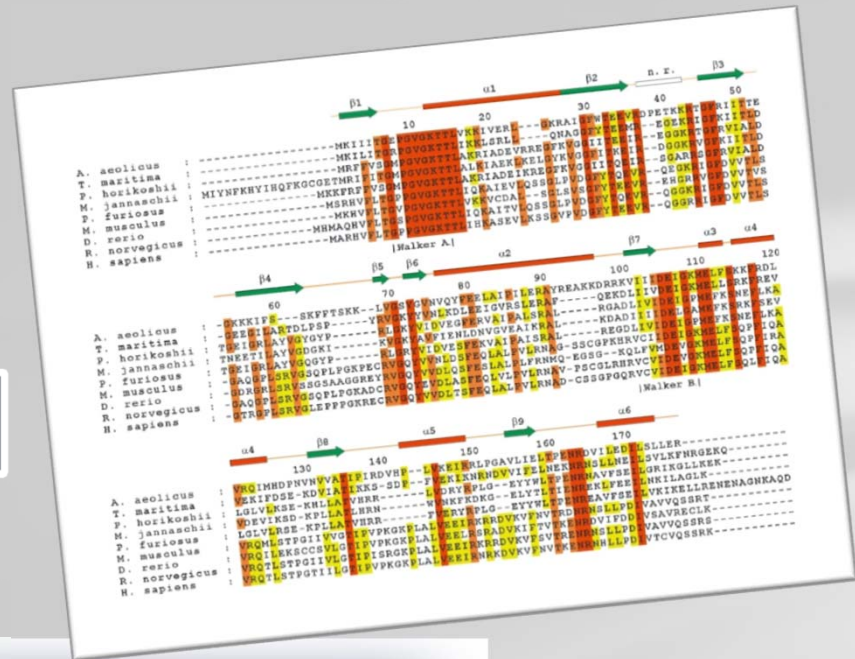
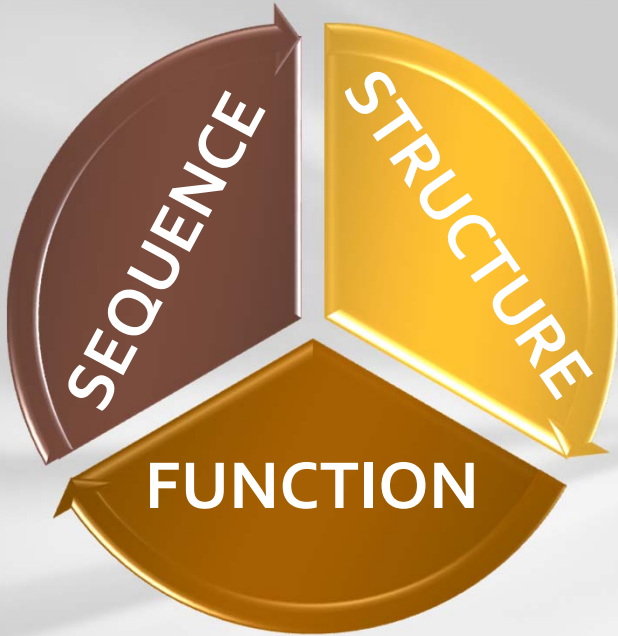


Sequence determines structure

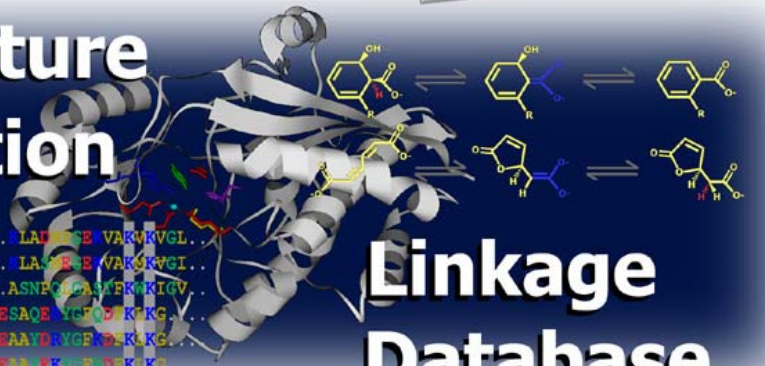
The aminoacid sequence of a protein determines its three-dimensional structure

Proteins that share sequence identity > 25 % are structurally similar

Homology modeling



**Structure
Function**



Y AAPLC...NGDPDDLIL...LAE...E VAK...KVL...
 YHTAPLC...YGDPELYA...LAE...E VAK...KVI...
 PMSLSYQ...LLPTGEAALDYL...ASN...GAS...FRK...KIV...
 HE...EALTPPEAIV...LAESAQE...YF...K...KG...
 RNK...EALTPESVVALAEAA...YDRYGF...DK...KG...
 RHE...EAM...PEAVR...LAEAA...AK...E...DFK...K...
 RHQ...KAMNSEAVVRLAEASQDRYGF...DK...K...KG...

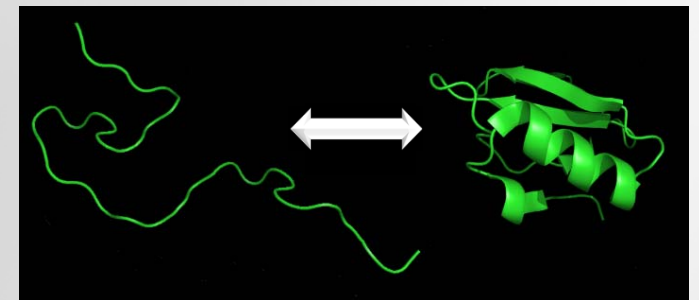
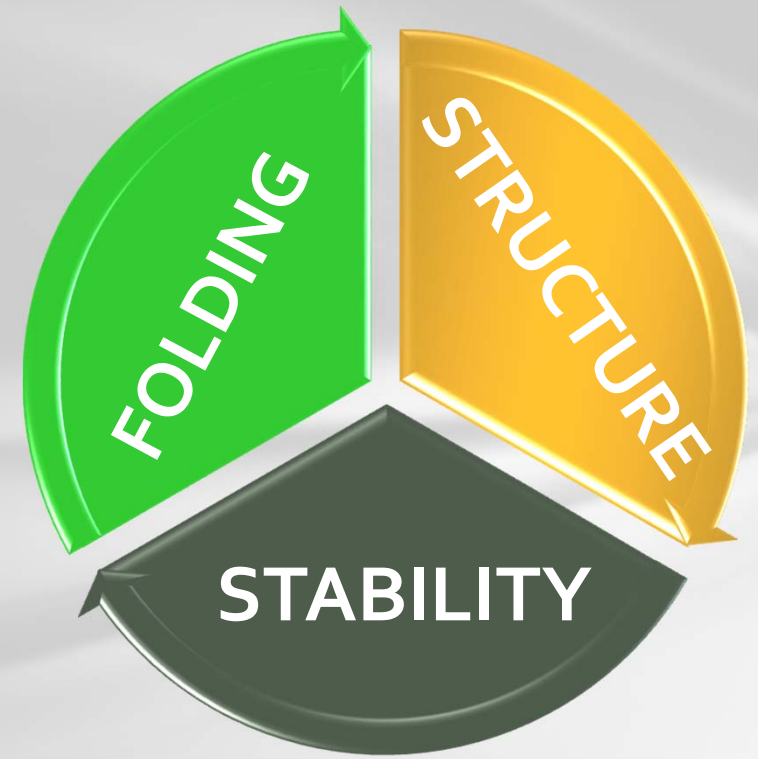
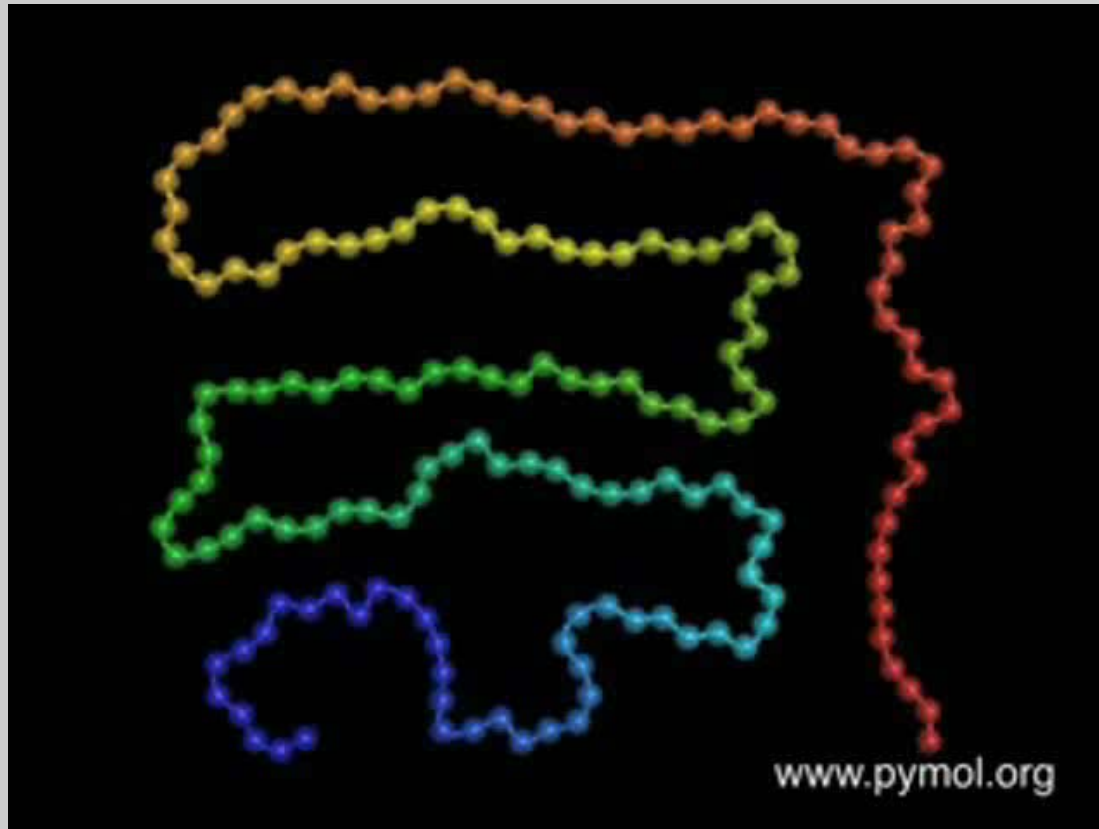
**Linkage
Database**

... And structure is highly correlated to function

<http://sfld.rbvi.ucsf.edu/django/>

Mechanism of protein folding

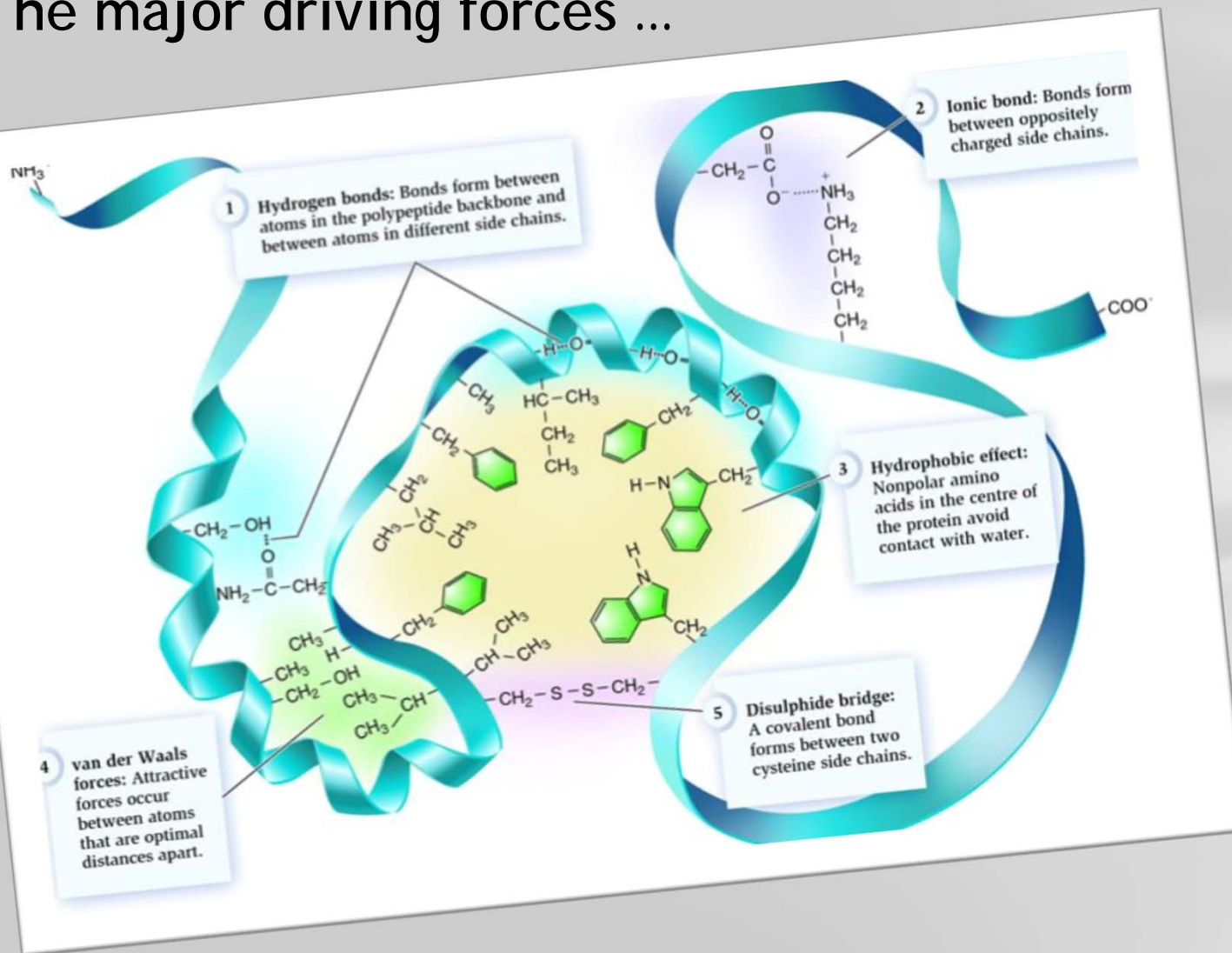
What are the mechanisms that govern protein folding ?



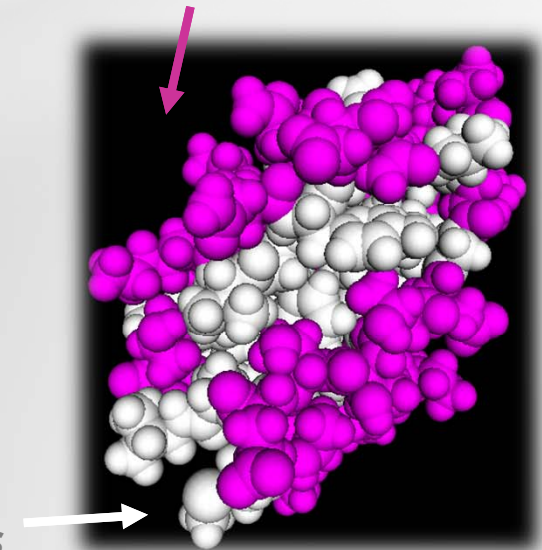
Factors determining protein folding

The major driving forces ...

- Hydrophobic effect
- H-bonds
- Conformational entropy
- Ionic interactions...



Charged and polar side chains are situated on the solvent-exposed surface where they interact with surrounding water molecules

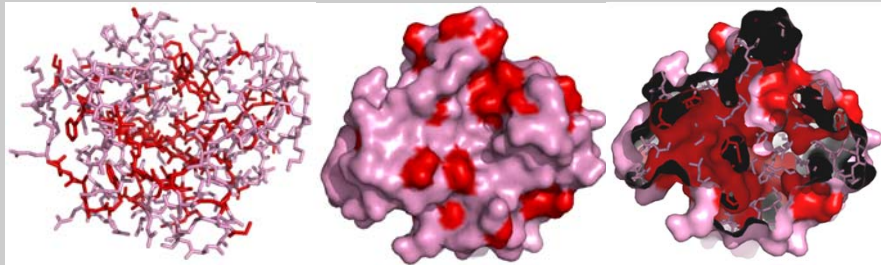


*Hydrophobic core in which side chains are buried from water
Minimizing the number of hydrophobic side chains exposed to water is the principal driving force behind the folding process*

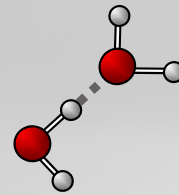
The Major Factors Affecting Protein Stability

... are the forces that contribute to protein stability

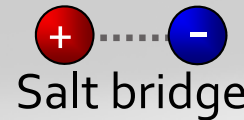
Hydrophobic interactions



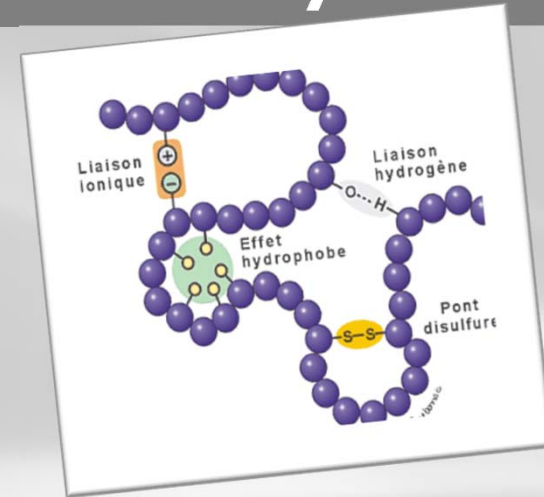
Non covalent bonds



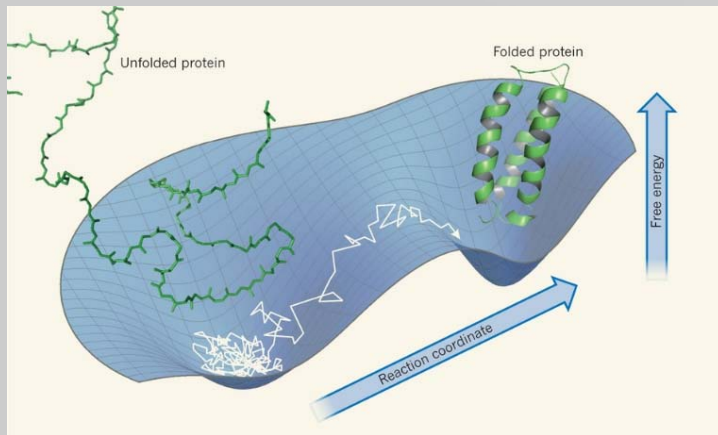
H-bond



Salt bridge



Conformational entropy



Sum of these interactions gives rise to the **final stability** of a protein

Difficult to predict protein stability from structure with computational methods !

Destabilising	
Conformational Entropy	-177
Groups Buried	-81
Groups Buried	-28
Stabilising	-286
g	
Ionisation	+4
e Bonds	+7
Hydrophobic Groups Buried	+94
Hydrogen Bonding	+166
Total Stabilising	+271
G (estimate)	-15
G (measured)	+9

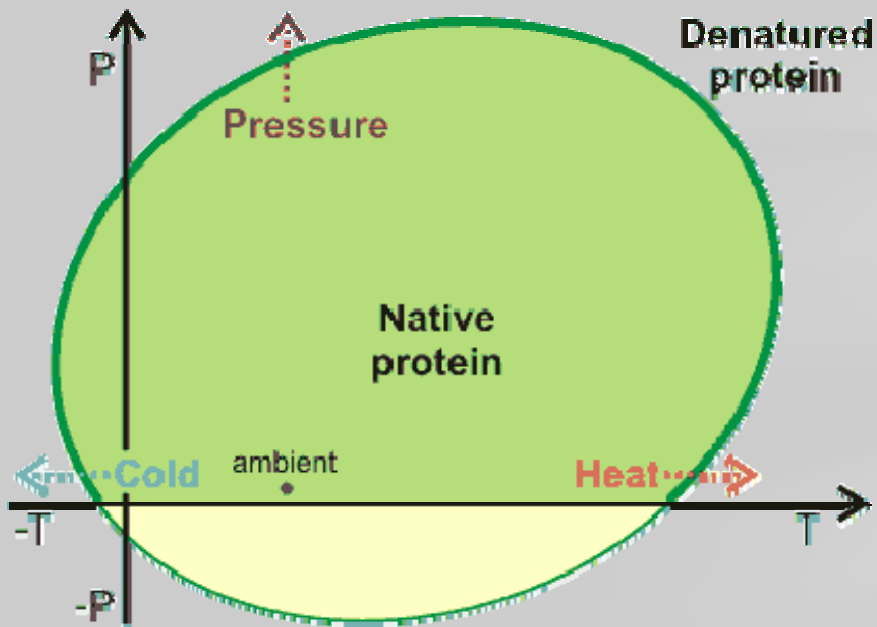
Free Energy (kcal/mol)

Rnase T1

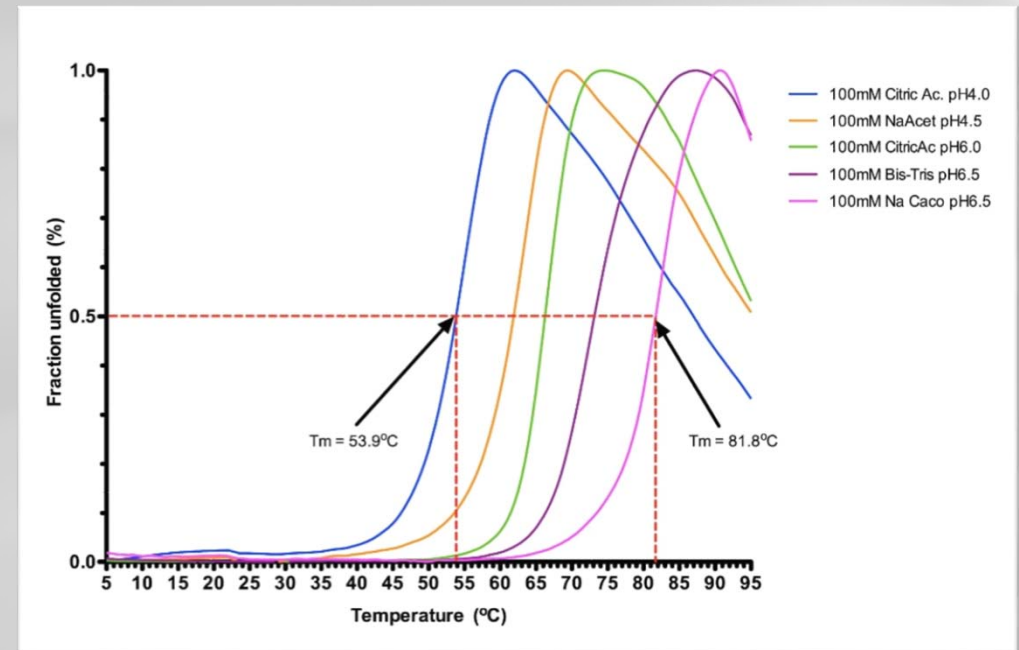
Effect of environment on protein stability

Every parameter that affect these interactions modify protein stability : temperature, pressure, ionic force, pH, ...

Pressure vs temperature



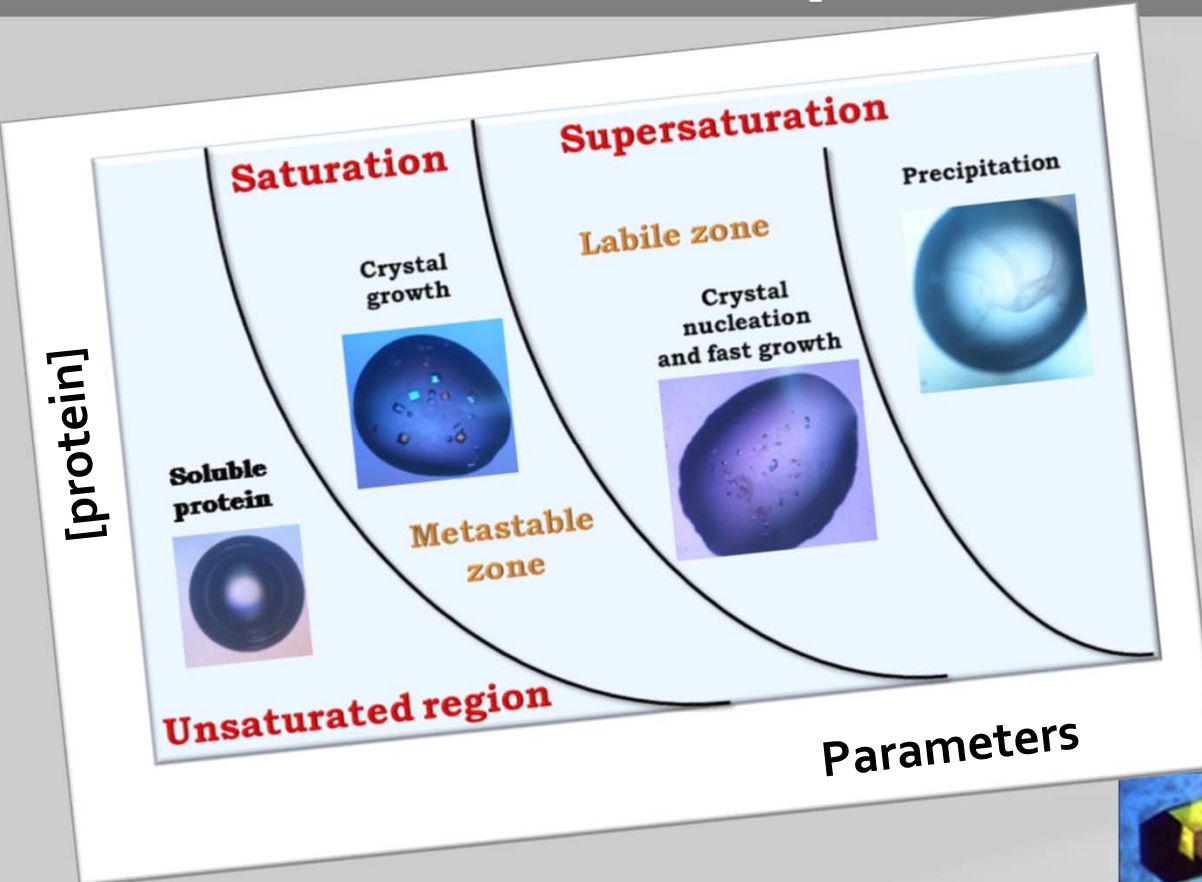
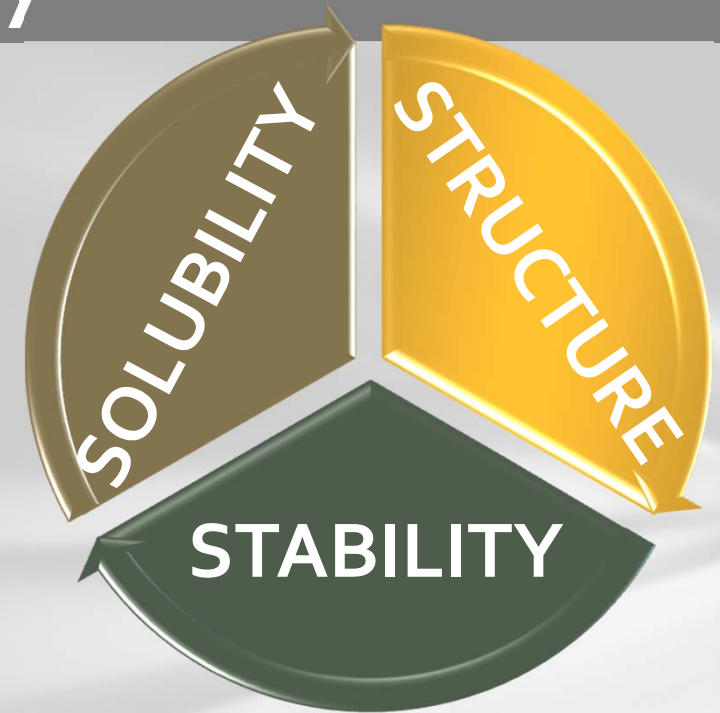
Temperature vs pH



Typical parameters

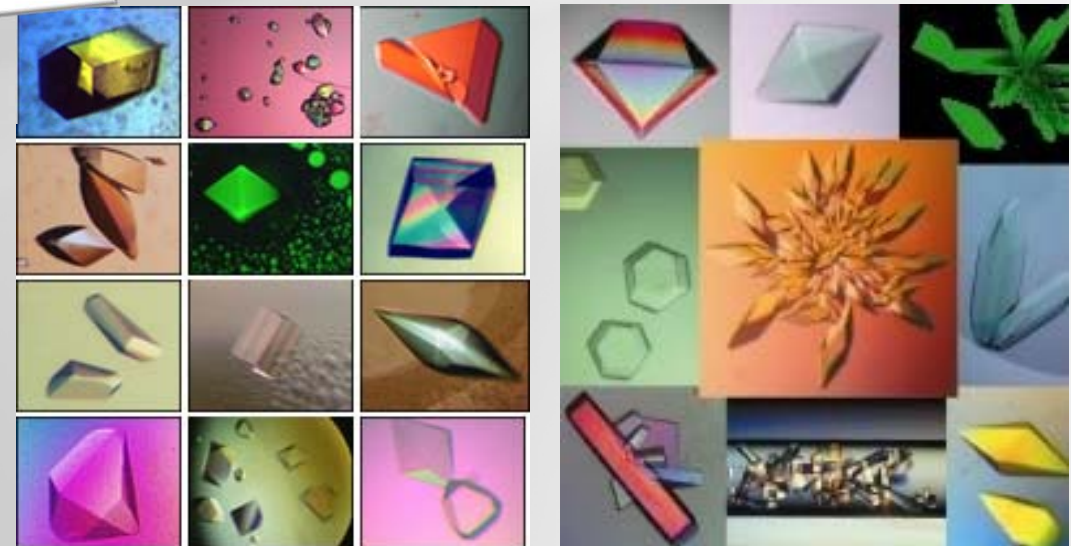
- External: external conditions and buffer solutions
- Internal: aa modifications, aa substitutions

Structure, Stability and solubility



For a crystallographer

- Stability in terms of solubility
- Objective is to obtain protein in solid phase (pack 10^{15} molecule in a crystal) without denature it



What is a structure ?

Structure prediction

Stability prediction

Predictions of order/disorder

Algorithms based on the analysis of the sequence
Information associated to solubility often directly related to folding state, aggregation or denaturation and secondary structure

Several servers and programs

<http://www.disprot.org/predictors.php>

Protein Disorder Predictors

DISPROT

<http://www.ist.temple.edu/disprot/Predictors.html>

DisEMBL

<http://dis.embl.de/>

MEDOR

<http://www.vazymolo.org/MeDor/>

GLOBPLOT₂

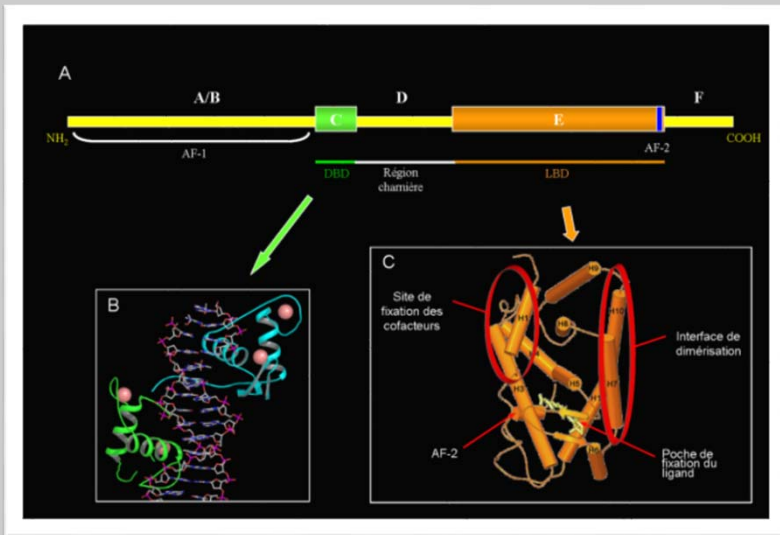
<http://globplot.embl.de/>

FoldIndex

<http://bip.weizmann.ac.il/fldbin/findex/>

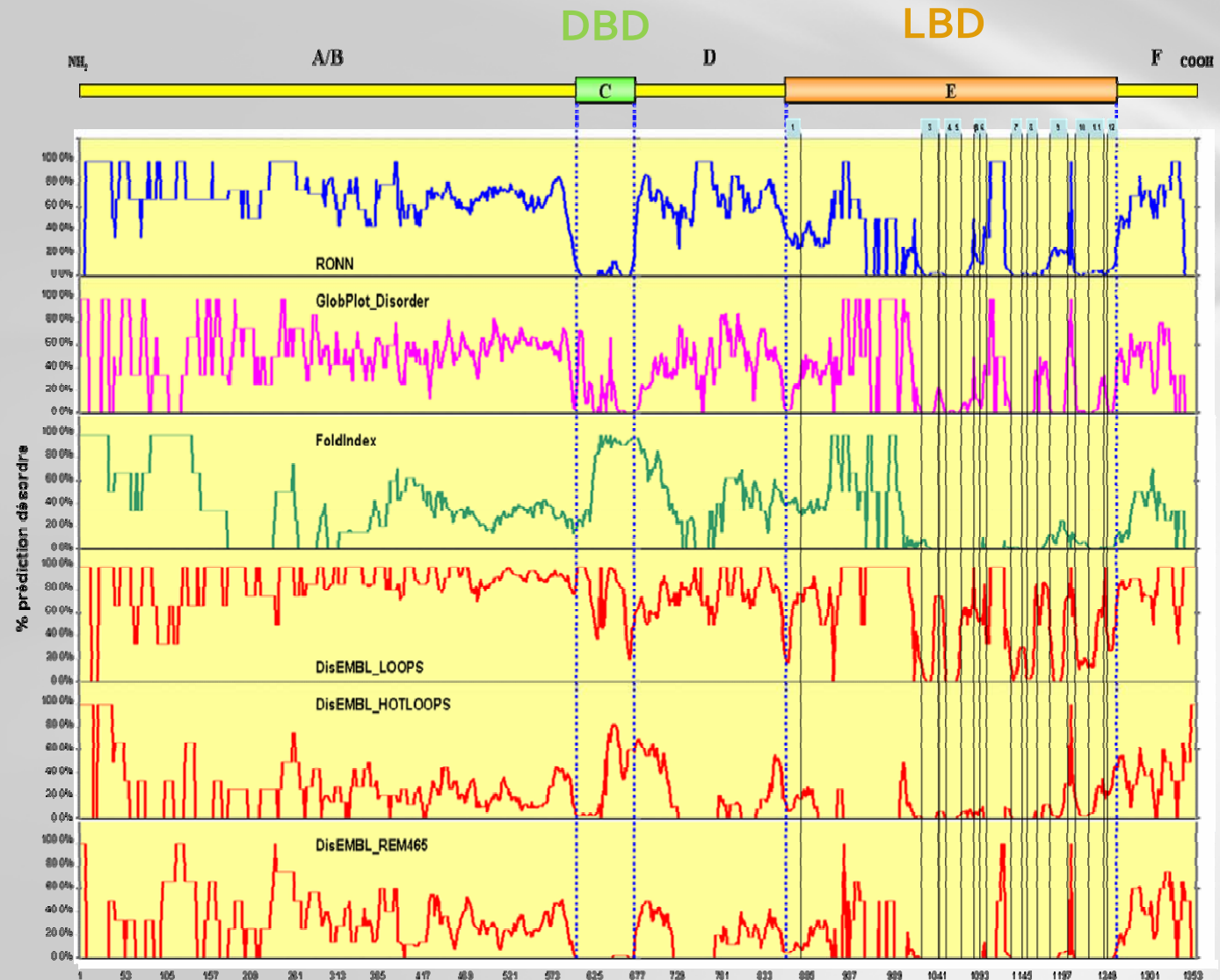
Predictions of order/disorder

Exemple : human nuclear receptors



Predictions of disorder using different programs and a multiple alignment of 48 human nuclear receptors

Transcription factors sensing hydrophobic ligands (steroids, thyroid hormones, ...) regulating gene expression



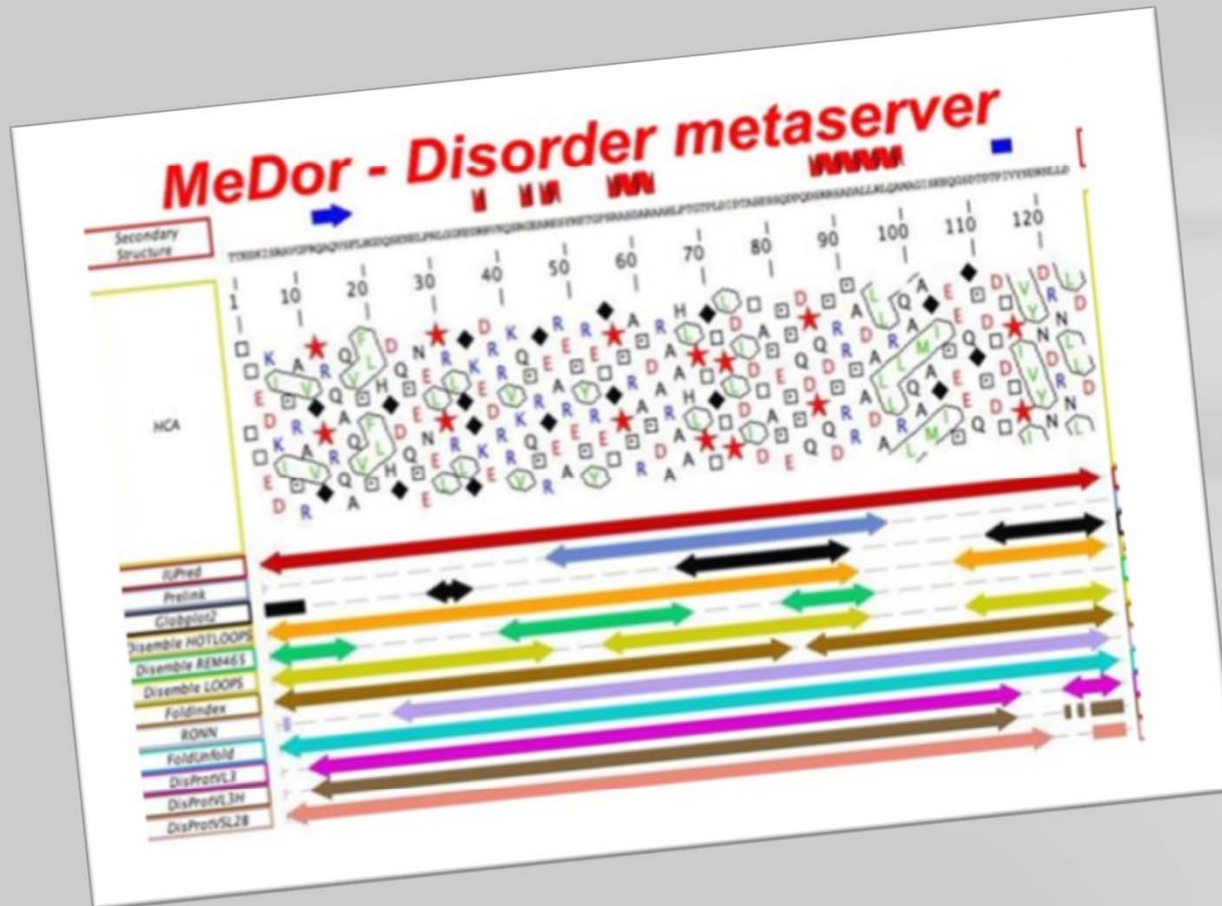
Prediction of order/disorder

MEDOR

<http://www.vazymolo.org/MeDor/>

Prediction of regions sensitive to defolding, of potential interacting partners, ...

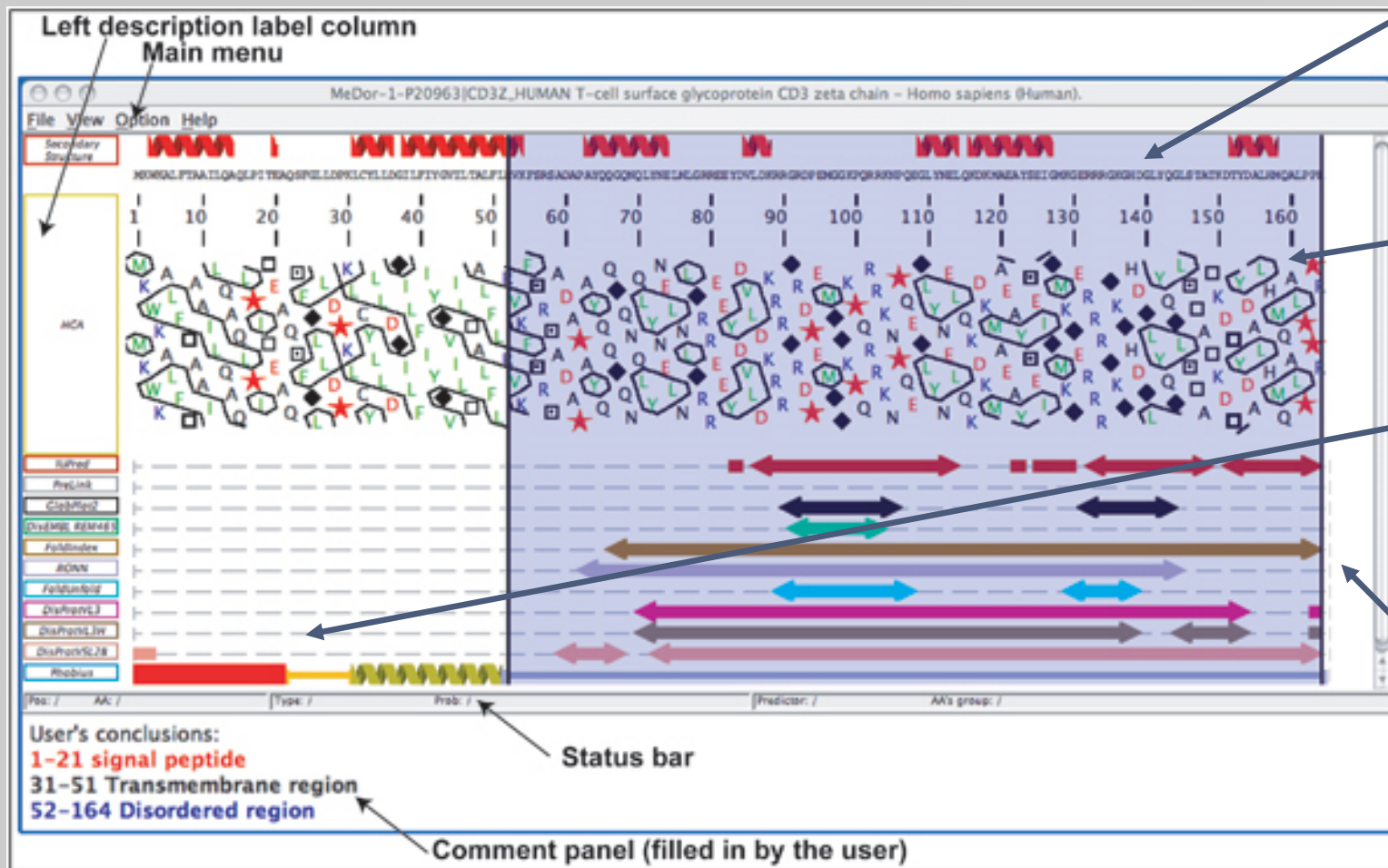
Database of viral proteins
Its aim is to define modules suitable for high expression, solubility and crystallization



The image shows a screenshot of the VaZyMoLO Interfaces website. At the top, there are navigation tabs: "Home", "VaZyMoLO Browser", "VaZyMoLO Blast And tools", and "Tutorial". The main heading is "VaZyMoLO Interfaces". Below this, there is a "VaZyMoLO Home Page" button. The text describes the database's purpose: "VaZyMoLO Interfaces provides a BLAST engine and a browser to our module sequence library. VaZyMoLO is a database dealing with viral sequences at the proteic level. Its aim is to define modules suitable for high expression, solubility and crystallization. Thus it integrates tools starting from amino acids composition, hydrophobic clusters analysis, secondary prediction, modelling, homology with solved structures, data mining concerning biochemistry (function and motifs, active sites, cleavage sites etc). Domains are defined on the structural definition of a domain (which can fold by themselves and show activity); but a module can be constituted by several domains." Below this, there is a section titled "How VaZyMoLO is organised?" which includes a diagram of a virus particle and the text: "Three layers in VaZyMoLO. Virions are organised into three layers: surface proteins, matrix proteins, and non-structural proteins. The VaZyMoLO database organisation has been directly inspired by this organisation and is therefore organised into three layers reflecting surface (layer S), matrix (layer M), and non-structural proteins (layer F)." The diagram shows a virus particle with three layers: "Modules S" (surface proteins), "Modules M" (matrix proteins), and "Modules F" (non-structural proteins). Below this, there is a section titled "How to start?" which lists two ways to use the VaZyMoLO interfaces: 1. You can seek for information by using our database browser available from the tab entitled "VaZyMoLO Browser". Click on a protein name or id to access modular information. Then click on a module to get further details about it. 2. If you already have a sequence of interest, you can use our "VaZyMoLO Blast and tools" that will enable you with the use of several tools for sequence analysis and a BLAST engine against our database that will retrieve similarities with our data.

Prediction of order/disorder

Example of a MeDor output



The sequence is represented below the predicted secondary structure elements (β -strands are represented by blue arrows, and α -helices are drawn in red)

HCA plot

Peptide signals and TM domains predicted by Phobius are highlighted as red bars and yellow helices

Predicted disordered regions are represented by bidirectional arrows of different colors as a function of predictors.

DisProt entry DP00200 human T cell glycoprotein CD3 Z chain (P20963)

Prediction of aggregation

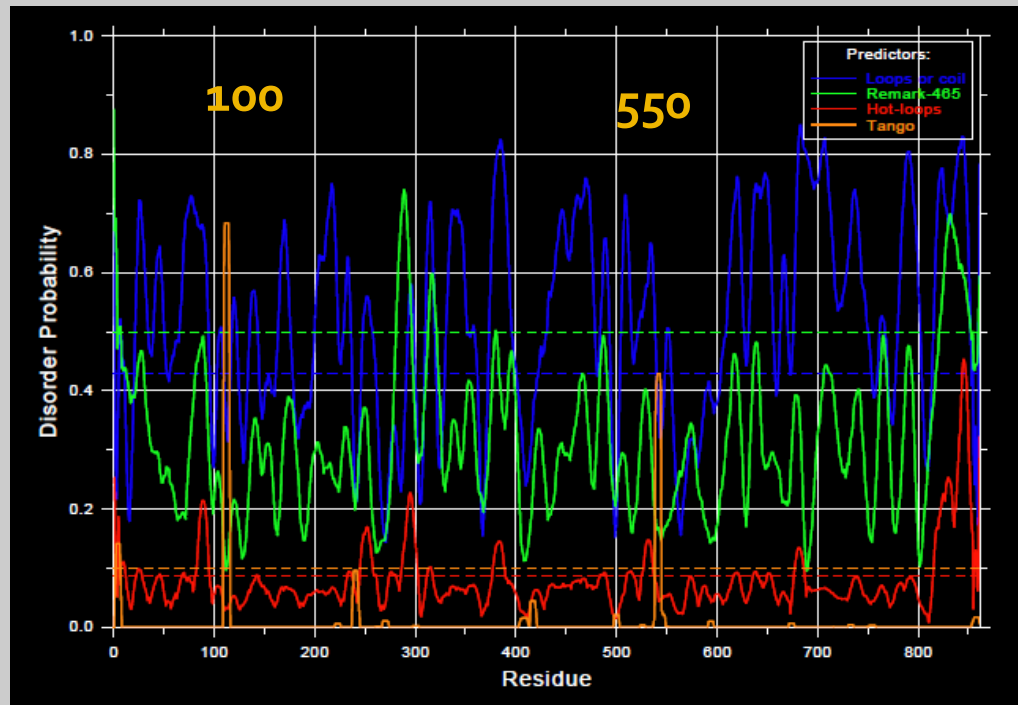
TANGO

<http://dis.embl.de/>

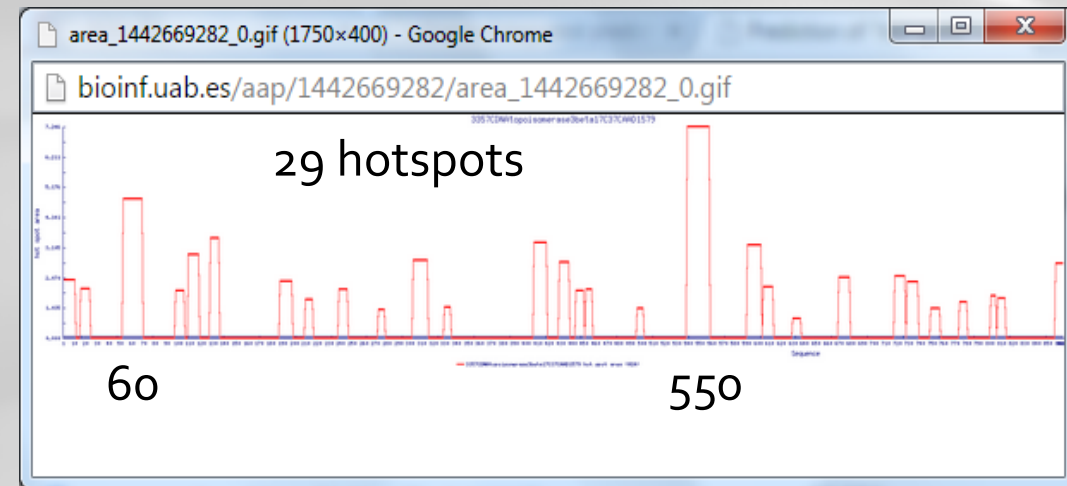
AGGRESKAN

<http://bioinf.uab.es/aggrescan/>

Prediction of domains with propensity to aggregate



Based on simple physico-chemical principles of secondary structure formation extended by the assumption that the core regions of an aggregate are fully buried



Based on an aggregation-propensity scale for natural amino acids derived from *in vivo* experiments and on the assumption that short and specific sequence stretches modulate protein aggregation

Secondary structure prediction

Based on the propensity of each aminoacid to form a secondary structure (helix and strand)

SOPMA

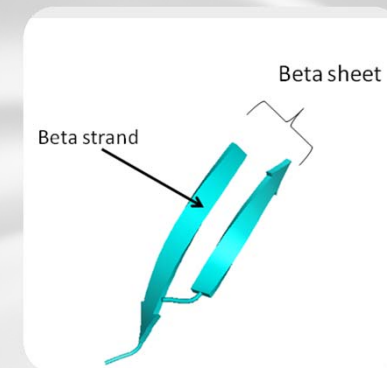
<https://npsa-prabi.ibcp.fr/>

Jpred 4

<http://www.compbio.dundee.ac.uk/jpred/>

PSIPRED

<http://bioinf.cs.ucl.ac.uk/psipred/>



UCL Department Of Computer Science
Bioinformatics Group

The PSIPRED Protein Sequence Analysis Workbench

The PSIPRED Protein Sequence Analysis Workbench aggregates several UCL structure prediction methods into one location. Users can submit a protein sequence, perform the predictions of their choice and receive the results of the prediction via e-mail or the web. For a summary of the available methods you can read [More...](#)

NOTE: users who need to run our methods on a large number of proteins should consider downloading our software using the menu on the left (Server Navigation -> Software Download).

The PSIPRED Team
Current Contributors David T. Jones, Daniel Buchan, Tim Nugent, Federico Minned & Kevin Bryson
Previous Contributors Anna Lobley, Sean Ward, Liam J. McGuffin
For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

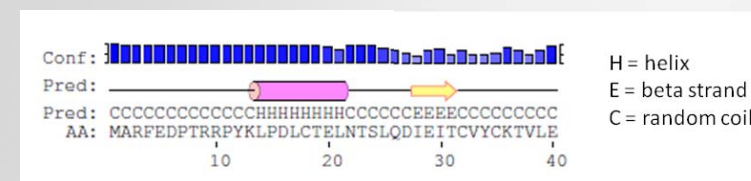
Site Navigation
Introduction
People
Projects
Publications
Web Servers
Downloads
Vacancies
Contact
Group Intranet

Server Navigation
PSIPRED Server
Server Overview
Server Citation
Help & Tutorials
News
History
Software Download

Choose Prediction Methods

- PSIPRED v3.3 (Predict Secondary Structure)
- pGenTHREADER (Profile Based Fold Recognition)
- BioSerf v2.0 (Automated Homology Modelling)
- FFPred 3 (Eukaryotic Function Prediction)
- MEMPACK (SVM Prediction of TM Topology and Helix Packing)
- DomSerf v2.0 (Automated Domain Modelling by Homology)
- DISOPRED3 & DISOPRED2 (Disorder Prediction)
- MEMSAT3 & MEMSAT-SVM (Membrane Helix Prediction)
- DomPred (Protein Domain Prediction)
- GenTHREADER (Rapid Fold Recognition)
- pDomTHREADER (Fold Domain Recognition)

Prediction all along the sequence with a confidence index



Secondary structure prediction

Example of a PSIPRED output

UCL Department Of Computer Science
Bioinformatics Group

Blomsbury Centre for Bioinformatics UCL

Sequence analysis results for job: kkk
ID: ac4b98fa-5948-11e5-9f10-00163e110593

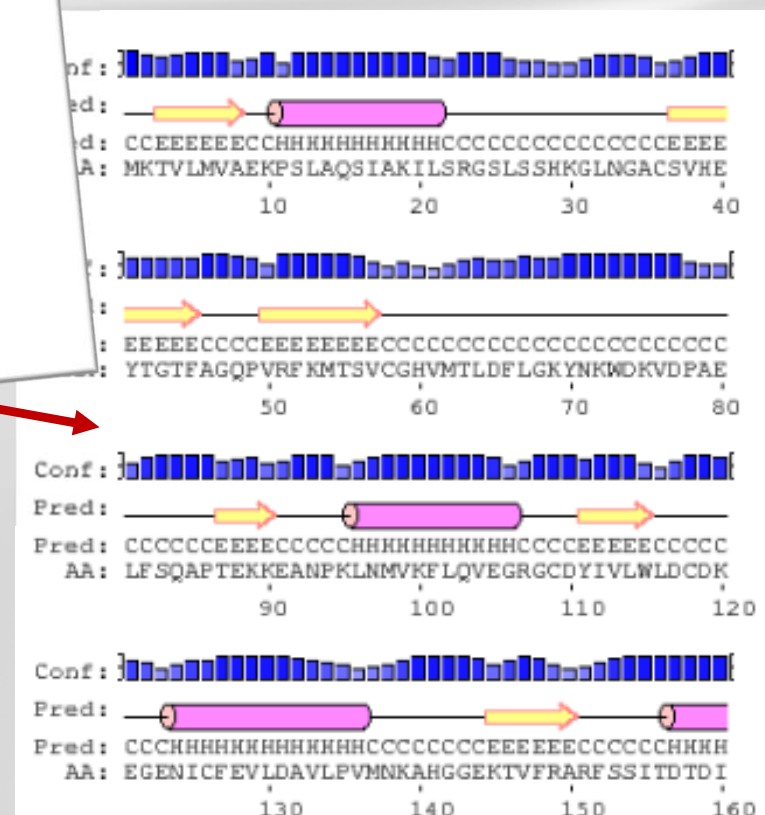
Summary PSIPRED Downloads

Please click on the thumbnail(s) to see the full version of the diagram in PNG format - or see the 'Downloads' tab for more, high-quality options.

Site Navigation
Introduction
People
Projects
Publications
Web Servers
Downloads
Vacancies
Contact
Group Intranet

Server Navigation
PSIPRED Server
Server Overview
Server Citation
Help & Tutorials
News
History
Software Download

human Topo III α

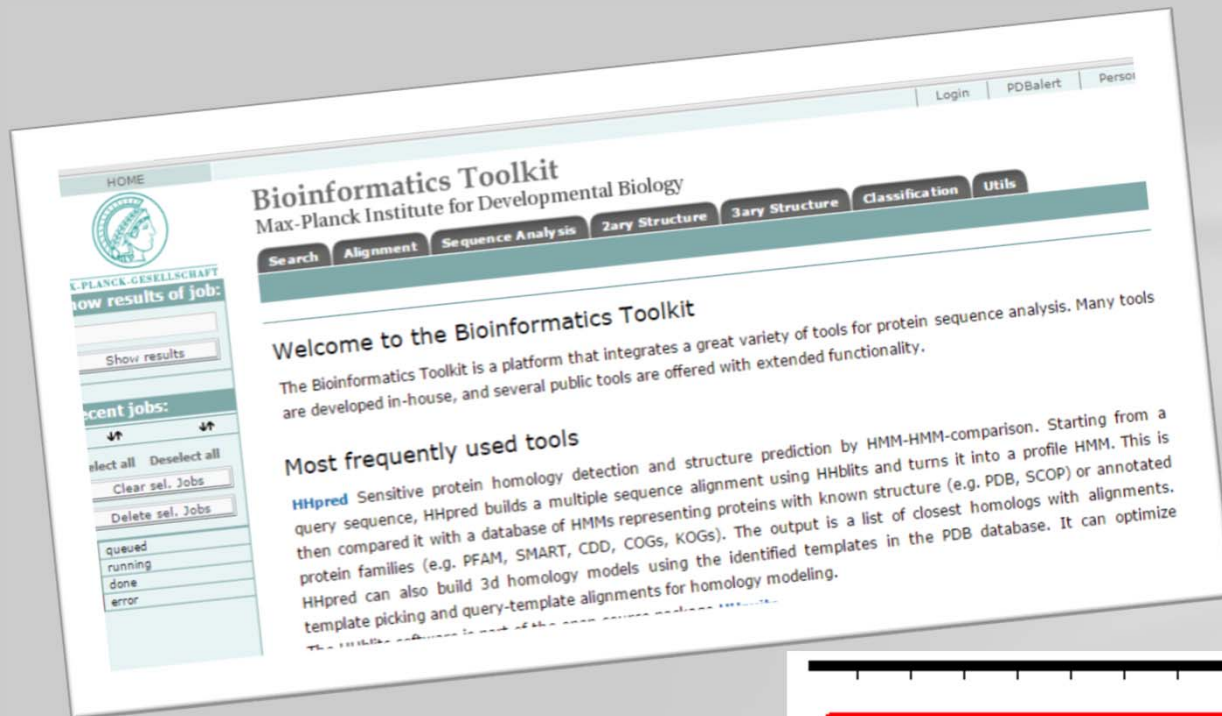


From sequence to structure prediction

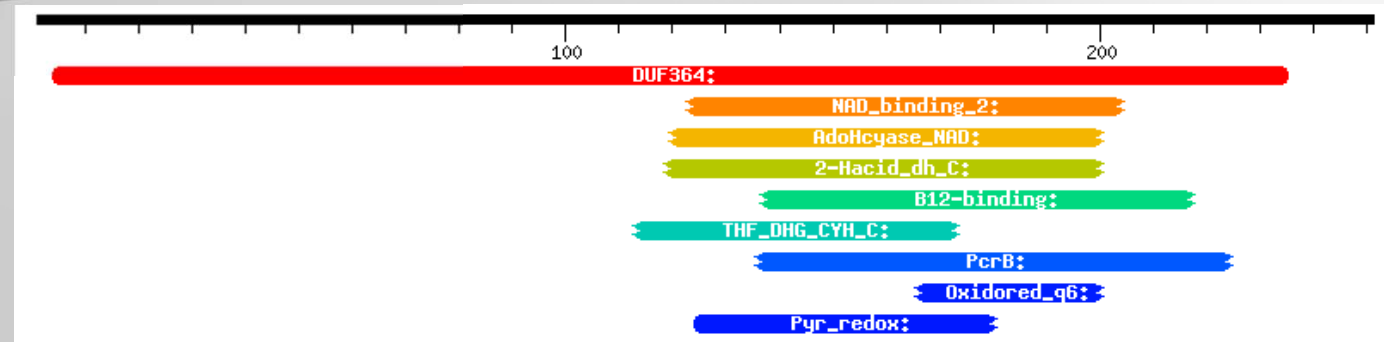
Homology detection and structure prediction by HMM-HMM comparison

HHpred

<http://toolkit.tuebingen.mpg.de/hhpred>



HHpred is often used for remote homology detection and homology-based function prediction. It runs with the free, open-source software package **HH-suite** for fast sequence searching, protein threading and remote homology detection.

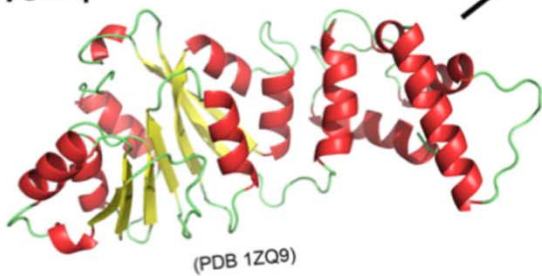


Homology modeling

Target sequence

HLLKNPGLDKIIYAAKIKSSDIVLEIGCGTGNLTVKLLPLAKKV
ITIDIDSRMISEVKKRCLYEGYNNLEVEYEGDAIKTVFPKFDVCTA
NIPYKISSPLIFKLIHRPLFKCAVLMFQKEFAERMLANVGDSDNY
SRLTINVKLFCKVTKVCNVNRSFPKVDVSVIVKLIKESFSL
TNFDEWDLNLRICFSRKRKTLHAIFKRNAVLMLEHNYKNWCTLN
KQVPVNFPPKCYCLDVLEHLMCEKRSINLDENDFLKLLLEFNKK
GIRHF
(T0295 from CASP7)

Template structure



```
12Q9 QHILKNPLIINSIIDKAAALRPTDVLVLEVGPGTGNMTVKLLEKAKKVACELDPRLVAELH 60
T0295 -HLLKNPGLDKIIYAAKIKSSDIVLEIGCGTGNLTVKLLPLAKKVITIDIDSRMISEVK 59
*:*** **:* * :*:***:*** **:*:***:*** **:*:***:***

12Q9 KRVOGTPVASKLQVLVGDVLTDLPPFDTCVANLFPYQISSPFVKLLHRPFRCALMF 120
T0295 KRCLYEGYN-NLEVEYEGDAIKTVFPKFDVCTANIPYKISSPLIFKLIHRPLFKCAVLMF 118
** :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: *

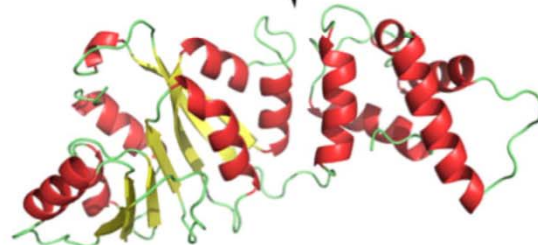
12Q9 QREFALRLVAKPGDKLYCRLSINTQLLARVDHLMKVGKNNFRPPKVESSVRIEPKNPP 180
T0295 QKEFAERMLANVGDSDNYSRLTINVKLFCKVTKVCNVNRSFPKVDVSVIVKLIKES 178
*:** * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: *

12Q9 PPINFQEWDLVRITFVRKNKTLSPAAPKSSAVQQLLEKNYRHCNSVHNIIPEDFSIADK 240
T0295 FLTNFDEWDLNLRICFSRKRKTLHAIFKRNAVLMLEHNYKNWCTLN-KQVPVNFPPK 237
*:** * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: *

12Q9 IQQILTSTGFSDKRRARMSDIDDIFIRLLHGFNAEGIHFS 278
T0295 CLDVLEHLMCEKRSINLDENDFLKLLLEFNKKGIRHF 275
:* .:***: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: * :*: *
```

Sequence alignment

exploit the 3D similarity between
a known template structure
and the target sequence
to build models



The quality of the homology model is dependent on the quality of the sequence alignment and template structure

Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure

3D protein structure is evolutionarily more conserved than would be expected on the basis of sequence conservation alone

The **sequence alignment** and **template structure** are then used to produce a structural model of the target

Homology modeling

- I-TASSER** is the best server for protein structure prediction according to the 2006-2012 **CASP** experiments
- RaptorX** excels at aligning hard targets according to the 2010 **CASP9** experiments
RaptorX generates the significantly better alignments for the hardest 50 **CASP9** template-based modeling targets than other servers
- MODELLER** is a popular software tool for producing homology models by satisfaction of spatial restraints using methodology derived from NMR data processing
The **ModWeb** comparative protein structure modeling web-server uses primarily **MODELLER** for automatic comparative modeling
- SWISS-MODEL** provides an automated web server for protein structure homology modeling
- Robetta** widely used servers for protein structure prediction
- SPARKSx** is one of the top performing servers in the **CASP** focused on the remote fold recognition
- PEP-FOLD** is a *de novo* approach aimed at predicting peptide structures from amino acid sequences, based on a HMM structural alphabet
- QUARK** is an algorithm developed for *ab initio* protein structure modeling

Homology modeling

Phyre2

is amongst the top performing server in the CASP international blind trials of structure prediction in homology modelling and remote fold recognition, and are designed with an **emphasis on ease of use for non-experts**

The screenshot shows the Phyre2 web interface. At the top, there's a 'Final Model' section with a 3D protein structure and a 'Confidence Summary' bar chart. Below that, there's a 'Sequence analysis' section with links for 'Secondary structure and disorder prediction', 'Domain analysis', and 'Detailed template information'. The main part of the interface is a table of template alignments.

Template	Alignment Coverage	3D Model	Confidence	% I.T.	Template Information
1 c2h2AA	Alignment		100.0	40	PDB header: c2h2AA Chain: A: PDB Molecular topology: 3 alpha; PDBTitle: crystal structure of the human topoisomerase II alpha-mt2 complex with bound calcium ion
2 c2h2AA	Alignment		100.0	22	PDB header: c2h2AA Chain: B: PDB Molecular topology: 2; PDBTitle: structure of a c. coli topoisomerase II complex with an 8-2 base single stranded oligonucleotide. Enzyme in green/2 alpha 1,3
3 c2h2AA	Alignment		100.0	22	PDB header: c2h2AA Chain: A: PDB Molecular topology: 1 DNA topoisomerase Superfamily: topoisomerase type I DNA topoisomerase Family: topoisomerase type I DNA topoisomerase
4 c2h2AA	Alignment		100.0	24	PDB header: c2h2AA Chain: A: PDB Molecular topology: 1 PDBTitle: structure of full length topoisomerase I from Thermotoga maritima (2 monomeric crystal form)
5 c2h2AA	Alignment		100.0	27	PDB header: c2h2AA Chain: A: PDB Molecular topology: 1 DNA topoisomerase Superfamily: topoisomerase type I DNA topoisomerase Family: topoisomerase type I DNA topoisomerase
6 c2h2AA	Alignment		100.0	22	PDB header: c2h2AA Chain: A: PDB Molecular topology: gyrase; PDBTitle: thermotoga maritima reverse gyrase, trimeric form

At the bottom, there's a 'Binding site prediction' section and a note: 'The final model of your protein (97% modelled at >90% confidence) has been submitted to the 3DLigandSite server to predict potential binding sites. Results will appear [here](#) when complete.'

5 templates were selected to model your protein based on heuristics to maximise confidence, percentage identity and alignment coverage. Below is a table indicating where your sequence was covered by each template, colour-coded by the confidence of the match to that template overall.

29 residues were modelled by ab initio. Please note: ab initio modelling is **highly** unreliable.

Template	Confidence
1 c2h2AA	98%
2 c2h2AA	99%
3 c2h2AA	100%
4 c2h2AA	98%
5 c2h2AA	96%
101 c2h2AA	98%
101 c2h2AA	99%
101 c2h2AA	100%
101 c2h2AA	98%
101 c2h2AA	96%
201 c2h2AA	98%
201 c2h2AA	99%
201 c2h2AA	100%
201 c2h2AA	98%
201 c2h2AA	96%
301 c2h2AA	98%
301 c2h2AA	99%
301 c2h2AA	100%
301 c2h2AA	98%
301 c2h2AA	96%
401 c2h2AA	98%
401 c2h2AA	99%
401 c2h2AA	100%
401 c2h2AA	98%
401 c2h2AA	96%
501 c2h2AA	98%
501 c2h2AA	99%
501 c2h2AA	100%
501 c2h2AA	98%
501 c2h2AA	96%
601 c2h2AA	98%
601 c2h2AA	99%
601 c2h2AA	100%
601 c2h2AA	98%
601 c2h2AA	96%
701 c2h2AA	98%
701 c2h2AA	99%
701 c2h2AA	100%
701 c2h2AA	98%
701 c2h2AA	96%
801 c2h2AA	98%
801 c2h2AA	99%
801 c2h2AA	100%
801 c2h2AA	98%
801 c2h2AA	96%

Phyre is for **non-commercial** use only
Commercial users please contact [Michael Sternberg](#)

Please cite: The Phyre2 web portal for protein modeling, prediction and analysis.
Kelley LA et al. *Nature Protocols* 10, 845-858 (2015) [DOI] [Citation link]


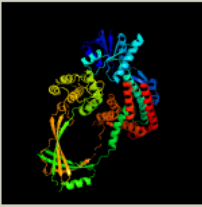
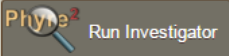
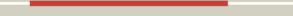
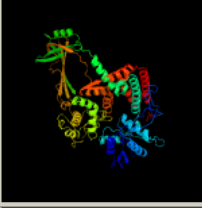
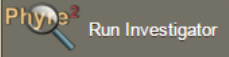
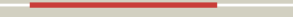
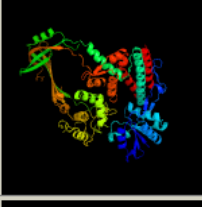
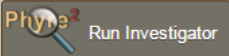


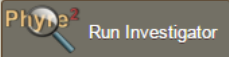


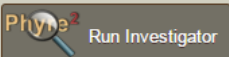
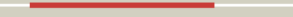

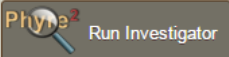
If you use the binding site predictions from 3DLigandSite, please also cite:
3DLigandSite: predicting ligand-binding sites using similar structures.
Wass MN, Kelley LA and Sternberg MJ *Nucleic Acids Research* 38, W469-73 (2010) [PubMed]

© Structural Bioinformatics Group
Imperial College London
Laurence Kelley, Michael Sternberg
Gladimir
Teresa and Conditions

Component software
Template detection: [I-TASSER](#) 1.2.1
Secondary structure prediction: [DSSP](#) 2.2.3
Disorder prediction: [Disorder](#) 2.6
Transmembrane prediction: [Thomson](#) 2.0.0
Multi-template modelling and ab initio: [EvoPro](#) 1.0

Homology modeling

Phyre2

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c4chtA	 Alignment		100.0	40	PDB header: cell cycle Chain: A; PDB Molecule: dna topoisomerase 3-alpha; PDBTitle: crystal structure of the human topoisomerase iii alpha-rmi12 complex with bound calcium ion 
2	c2o59B	 Alignment		100.0	22	PDB header: isomerase/dna Chain: B; PDB Molecule: dna topoisomerase 3; PDBTitle: structure of e. coli topoisomerase iii in complex with an 8-2 base single stranded oligonucleotide. frozen in glycerol3 ph 8.0 
3	d1i7da	 Alignment		100.0	22	Fold: Prokaryotic type I DNA topoisomerase Superfamily: Prokaryotic type I DNA topoisomerase Family: Prokaryotic type I DNA topoisomerase 
4	c2gajA	 Alignment		100.0	24	PDB header: isomerase Chain: A; PDB Molecule: dna topoisomerase i; PDBTitle: structure of full length topoisomerase i from thermotoga maritima in2 monoclinic crystal form 
5	d1mw9x	 Alignment		100.0	27	Fold: Prokaryotic type I DNA topoisomerase Superfamily: Prokaryotic type I DNA topoisomerase Family: Prokaryotic type I DNA topoisomerase 
6	c4ddvA	 Alignment		100.0	22	PDB header: hydrolase Chain: A; PDB Molecule: reverse gyrase; PDBTitle: thermotoga maritima reverse gyrase, triclinic form 

...s of
... and are

What is a structure ?

Structure prediction

Stability prediction

Structural analyses

Calculations or estimation of structural parameters that contributes to protein stability

PROPKA

estimation of the pKa values of ionisable aa

DALI

comparison of structural homologues, prediction of function

VADAR

structure validation server that allows to calculate volumes, accessible surfaces, contact surface, ...

MarkUs

analysis and comparison of the structural and functional properties of proteins

HotSpot Wizard

a tool for automatic identification of hot spot sites for engineering of substrate specificity, activity or enantioselectivity of enzymes

FoldX

a protein design algorithm that uses an empirical force field. It can determine the energetic effect of point mutations as well as the interaction energy of protein complexes (including Protein-DNA)

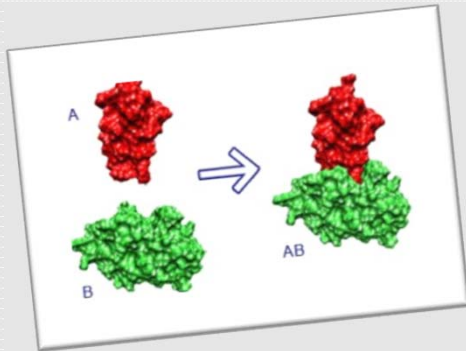
Stability: prediction of free energy changes between alternative structures

Prediction of protein-protein interactions

“Here one should remember that any protein fails to execute its function unless it interacts with other biomolecules”

Ito et al. (2001) *Proc. Natl. Acad. Sci. USA* 98, 4569

A comprehensive two-hybrid analysis to explore the yeast protein interactome



Webservers

Rosettadock

Prediction of interaction from 2 structural models

Patchdock

Docking based on surface complementarity , easy to use

Firedock

docking protein-protein, easy to use

ClusPro

tops the competition in the latest rounds of CAPRI experiment

Zdock

rigid-body search of docking orientations

Stand-alone

HADDOCK

Docking with possibility to implement experimental data (mutagenesis, cross-linking, NMR chemical shift, ...)

HEX

protein-protein docking, webserver also exists

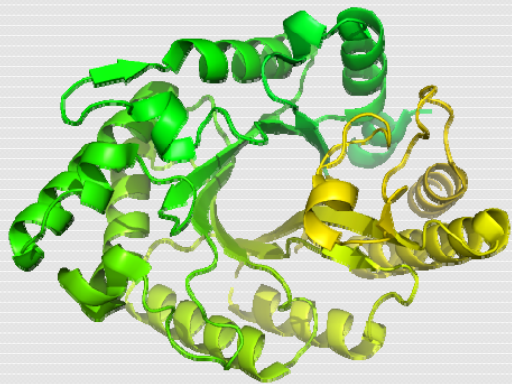
Protein thermostability

Effect of the temperature on protein stability

A powerful method is the comparison of mesophilic and thermostable homologous proteins

- ➔ Presence of **extra hydrogen bonds** and **salt bridges** in thermostable proteins
the protein structure is more resistant to unfolding
- ➔ Other factors are **compactness** of protein structure, **oligomerization** and **interaction strength** between subunits

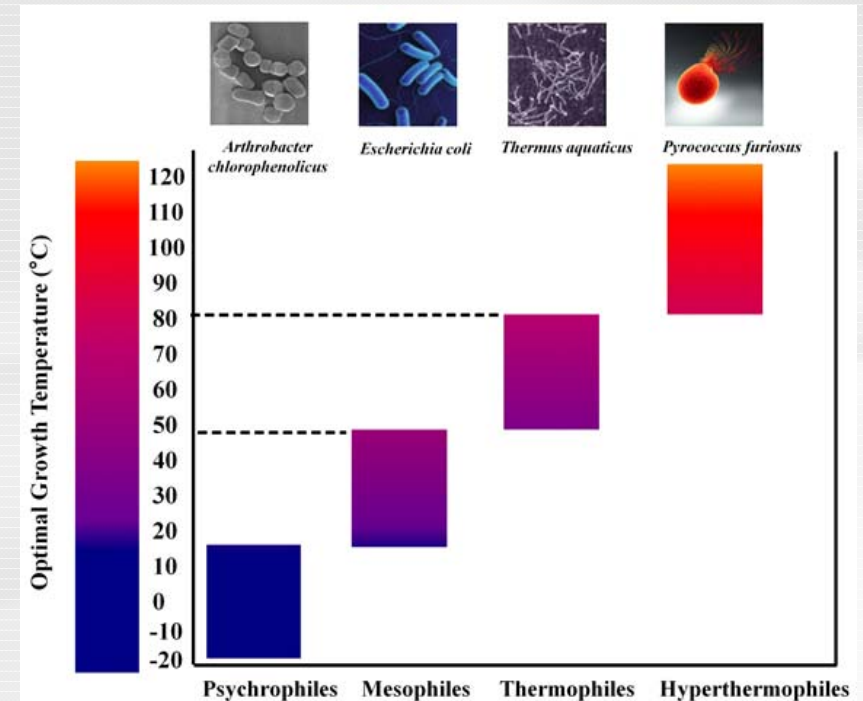
How to increase the thermostability of target proteins ?



mesophile



thermophile



Design stabilizing mutations

- mutations which truncate loops
- increase salt bridges or hydrogen bonds
- introduced disulfide bonds

Ligand binding can increase the stability

Effect of mutations on protein stability

Prediction of the effect of mutations on protein stability using structural knowledge

Design of new proteins

- Fundamental science : folding mechanisms
- Biotechnology : design of catalytically more efficient proteins or with longer half-life
- Molecular medicine : pathogenic missense mutations

Computational methods

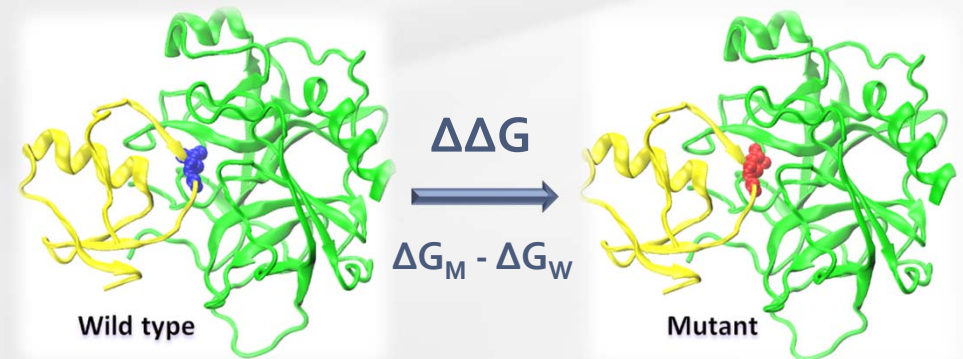
➔ Homology modeling

➔ Stability predictors

webservers or programs

Based either on atomic force field, statistical, empirical approaches or machine-learning methods or a combination of both

Use structure and/or sequence



Effect of mutations on protein stability

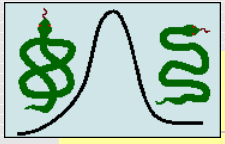
ProTherm

<http://www.abren.net/protherm>

The reference database for experimentally determined protein stability free energy or T_m changes by mutations

The screenshot shows the ProTherm website interface. At the top left is a logo featuring a green protein ribbon structure and a black curve representing a stability profile. Below the logo is a navigation bar with buttons for Home, ProTherm, ProNIT, and Biomolecules Gallery. The main heading reads "ProTherm Thermodynamic Database for Proteins and Mutants". A green banner below the heading states "Data updated: Feb. 22 2013" and includes a link to "Overview". The main content area contains a search box with a "Go" button and an "Advanced Search" button. Below the search box is a vertical menu with links: Overview, What's New, Statistics, Tutorial, More About ProTherm, Cross-References, Acknowledgement, and Members. The main text describes the database as a collection of numerical data on thermodynamic parameters like Gibbs free energy change, enthalpy change, heat capacity change, transition temperature, etc. for wild type and mutant proteins. It also mentions cross-linking with sequence databases (PIR and SWISS-PROT), structural databases (Protein Data Bank), functional databases (Protein Mutant Database), and literature databases (PubMed). A note at the bottom states: "Please note that this database is under constant development. There will be changes without prior notice. We welcome your comments and suggestions to improve this database." At the bottom right, there are navigation links: Home | ProTherm | ProNIT | Biomolecules Gallery.

Effect of mutations on protein stability



ProTherm Search

Please fill or choose necessary entries below, set display and sorting options.
 Explanations for the terms are [here](#)

Entry: - PDB Code: Start Clear

Protein: Source:

Mol-weight: To

Mutation: To Single Double Multiple Wild Type

Sec. Structure: Helix Sheet Turn Coil

Accessibility: Any Buried Partially Buried Exposed ASA: To %

Measure: Absorbance CD DSC Fluorescence NMR Others

Method: Thermal Denaturants Others

pH: To

dTm/Tm/T: dTm: To C

dH/dC/dG/dG_H2O: dH: To energy unit: kcal

ddG/ddG_H2O: ddG: To

State: 2 3 >3

Reversibility: Any

Keyword: OR

Author: OR

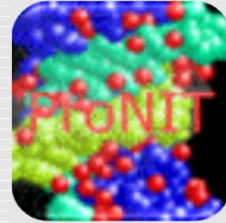
Year: Since Until

Display Option: Default Clear

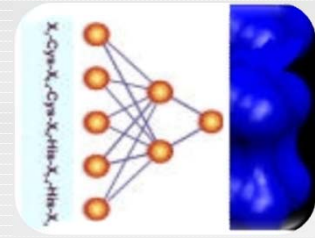
<input checked="" type="checkbox"/> ENTRY	<input checked="" type="checkbox"/> PROTEIN	<input checked="" type="checkbox"/> SOURCE	<input type="checkbox"/> AMINO LENGTH	<input type="checkbox"/> MOL-WEIGHT	<input type="checkbox"/> PIR
<input type="checkbox"/> E.C.NUMBER	<input type="checkbox"/> PMD.NO	<input checked="" type="checkbox"/> PDB_wild	<input checked="" type="checkbox"/> PDB_mutant	<input checked="" type="checkbox"/> MUTATION	<input checked="" type="checkbox"/> SEC-STR.
<input type="checkbox"/> ASA	<input type="checkbox"/> STATE	<input type="checkbox"/> dG_H2O	<input type="checkbox"/> ddG_H2O	<input checked="" type="checkbox"/> dG	<input checked="" type="checkbox"/> ddG
<input type="checkbox"/> T	<input checked="" type="checkbox"/> Tm	<input checked="" type="checkbox"/> dTm	<input type="checkbox"/> dHvH	<input type="checkbox"/> dHcal	<input type="checkbox"/> m
<input type="checkbox"/> Cm	<input type="checkbox"/> dCp	<input type="checkbox"/> pH	<input type="checkbox"/> BUFFER_NAME	<input type="checkbox"/> ION_NAME	<input type="checkbox"/> ADDITIVES
<input checked="" type="checkbox"/> MEASURE	<input type="checkbox"/> METHOD	<input type="checkbox"/> Reversibility	<input type="checkbox"/> ACTIVITY	<input type="checkbox"/> ACTIVITY_Km	<input type="checkbox"/> ACTIVITY_Kcat
<input type="checkbox"/> ACTIVITY_Kd	<input type="checkbox"/> KEY_WORDS	<input checked="" type="checkbox"/> REFERENCE	<input type="checkbox"/> AUTHOR		

Sorting By: OFF OFF OFF

Entries per page: 300



calculate protein-DNA interaction and DNA conformational energies



predicts the Real Values of Solvent Accessibility

Search Condition
 Mutation: Single
 Sec. Str.: Helix
 Method: Thermal
 Ph: 7 to 7

Entry	Protein	Source	PDB W	PDB M	Mut	dG	ddG	Tm	dTm	Measure	REFERENCE
2297	RNase T1	<i>Aspergillus oryzae</i>	1RN1	1RGC	Q25K	NULL	NULL	51.7	3.40	Fluores	J BIOL CHEM 264, 11621-11625 (1989)
2333	RNase T1	<i>Aspergillus oryzae</i>	1RN1	NULL	S17A	NULL	NULL	52.6	1.70	Fluores	BIOCHEM 31, 725-732 (1992) PMID: 1731929
14482	DsbA	<i>Escherichia coli</i>	1A23	NULL	H32Y	13.80	6.80	NULL	NULL	Fluores	PROTEIN SCI 6, 1893-1900 (1997)
14483	DsbA	<i>Escherichia coli</i>	1A23	NULL	H32L	12.30	5.30	NULL	NULL	Fluores	PROTEIN SCI 6, 1893-1900 (1997)
14484	DsbA	<i>Escherichia coli</i>	1A23	NULL	H32S	12.20	5.20	NULL	NULL	Fluores	PROTEIN SCI 6, 1893-1900 (1997)

Effect of mutations on protein stability

An unlimited number of webservers ...



AUTO-MUTE

AUTOMated server for predicting functional consequences of amino acid MUTations in protEins

A collection of programs for predicting functional changes to proteins upon single residue substitutions, developed by combining structure-based features with trained statistical learning models. For each type of...



MUpuro

Prediction of protein stability changes for single site mutations from sequences. Because MUpuro can accurately predict protein stability changes using primary sequence information only, it is applicable to many...



ENCoM

Elastic Network Contact Model

A coarse grained normal modes analysis method to evaluate thermostability of proteins. The ENCoM Server can be used by anyone to evaluate the effect of mutations on the stability of a structure. The server can also...



CUPSAT

Cologne University Protein Stability Analysis Tool

Assessment of protein stability upon point mutations. The prediction is based on the analysis of atom potentials and torsion angle distribution to assess the environment of the mutation site....



I-Mutant

A support vector machine based automatic prediction of protein stability changes upon mutations. I-Mutant predictions are based on the protein structure or, more...



SCide

A program to identify stabilization centers from known protein structures. These are residues involved in cooperative long-range contacts, which can be formed between various regions of a single polypeptide chain, or...



PoPMuSiC

A web server that predicts the thermodynamic stability changes caused by single site mutations in proteins, using a linear combination of statistical potentials whose coefficients depend on the solvent accessibility...



SDM

Site Directed Mutator

Predicts the effect that single point mutations have on protein stability. The method is based on knowledge of observed substitutions that have occurred in homologous proteins and which are encoded in...



<http://omictools.com/protein-stability-changes-c1478-p1.html>

Effect of mutations on protein stability

iStable

<http://predictor.nchu.edu.tw/istable/>

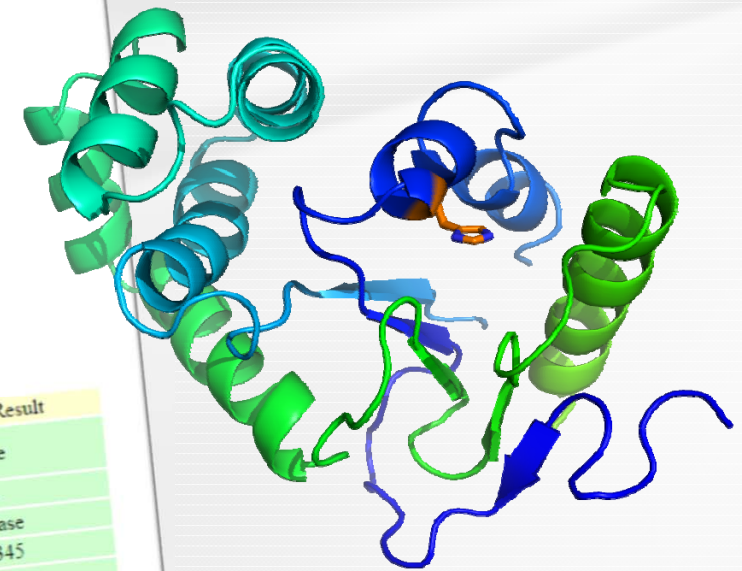
An integrated predictor constructed by using sequence information and prediction results from different element predictors. In the learning model, iStable adopted the support vector machine as an integrator, while not just choosing the majority answer given by element predictors

PDB ID: 1A23 Chain: A Wild-type: H Position: 32(32) Mutant: Y Temperature (°C, 0-100): 25 pH (0-14): 7

Protein sequence(no headers):
AQYEDGKQYTTLEKPVAGAPQVLEFFSFFCPCYQFEVLHISDNVKKLPEGVKHTKYHVFH
GGDLGKDLTQAWAVAMALGVEDKVTVPLFEGVQKQTIRASDIRDVFINAGIKGEEYDAAMIS
FVVKSLVAQQEKAADVQLRGVPAFVNGKYQLNPQGHDTSNMDVFVQQYADTVKYLSEK

Clear Submit

Predictor	i-Mutant2.0 PDB	i-Mutant2.0 SEQ	AUTO-MUTE SVM	Reference	MUpro	PoPMuSiC	CUPSAT	Meta Result
Result	7	31.9	5.55	helix	helix	0.99962109	null	32.44
Conf.			Increased	Increased	Increase	null	Increase	Increase
$\Delta\Delta G$	Increase	Increase	0.76	3.77		null	0.66	0.73345
RSA	6.50		1.51	buried				
SS			buried	buried				

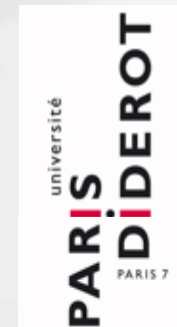


E. coli DsbA

Conclusion

- ➔ **3D structure and stability are strongly related**
- ➔ **Effect of post-translational modifications on protein stability**
- ➔ **Problem of the stability of splicing variants**
- ➔ **Is stability *in vitro* related to stability *in vivo* ?**

It's finished !



Pascal DHULSTER