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

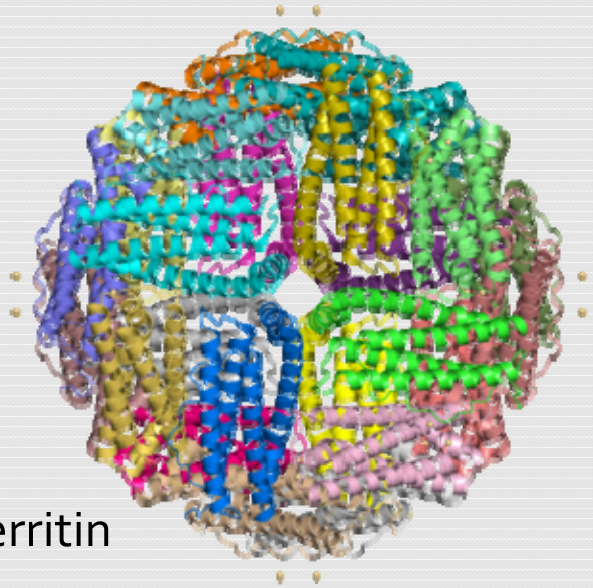
A definition of protein stability ?

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

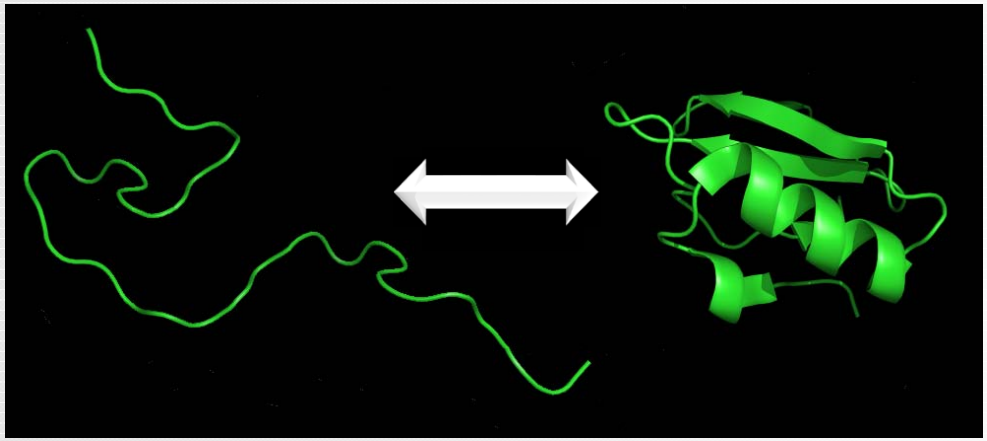
Structural conformation = 3D structure, quaternary structure (oligomerisation state), interaction with its partners (in case of macromolecular assemblies)

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



ferritin



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

$$\Delta G = G_U - G_N$$

Physical biochemist

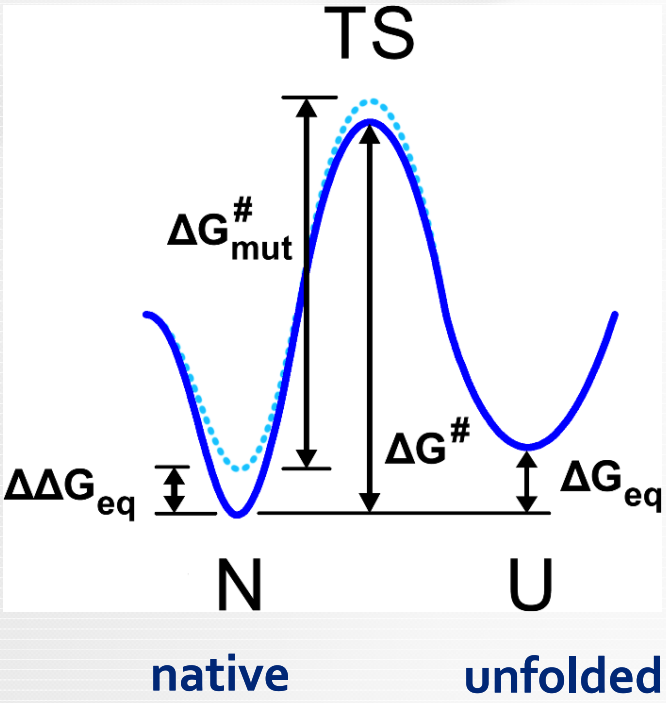
Thermodynamic stability

Easy to study for reversible systems
The larger and more positive ΔG, the more stable is the protein to denaturation

Biotechnologist

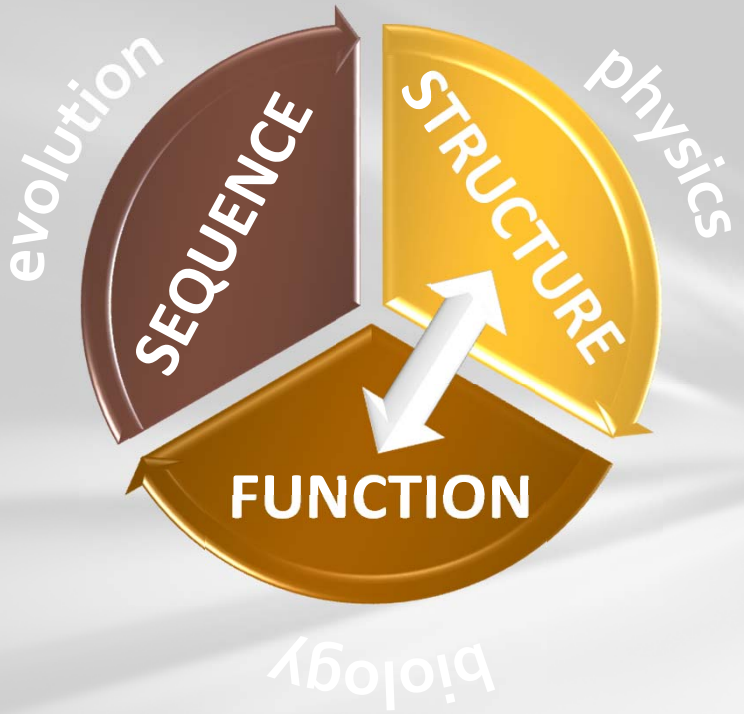
Kinetic stability

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

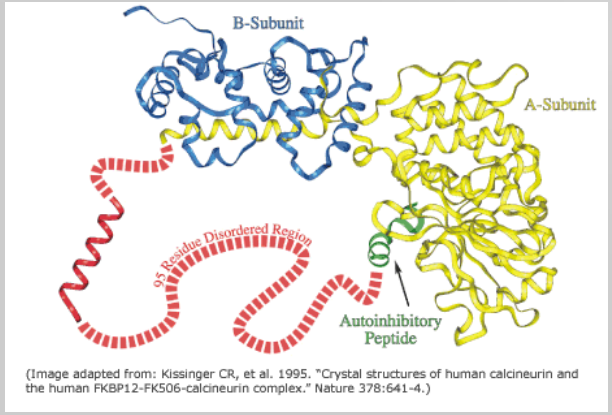


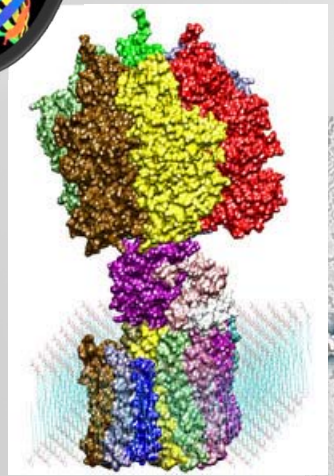
Structure-function relationships

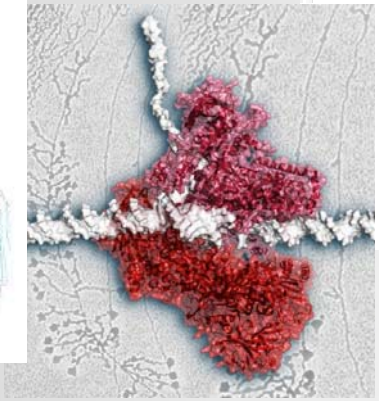
The 3D structure of a protein defines not only its size and its shape but also its function

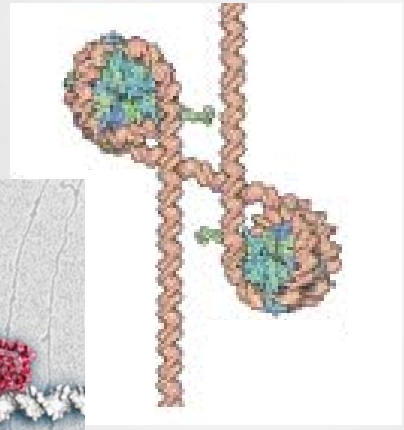


Exception
IDPs & IDRs





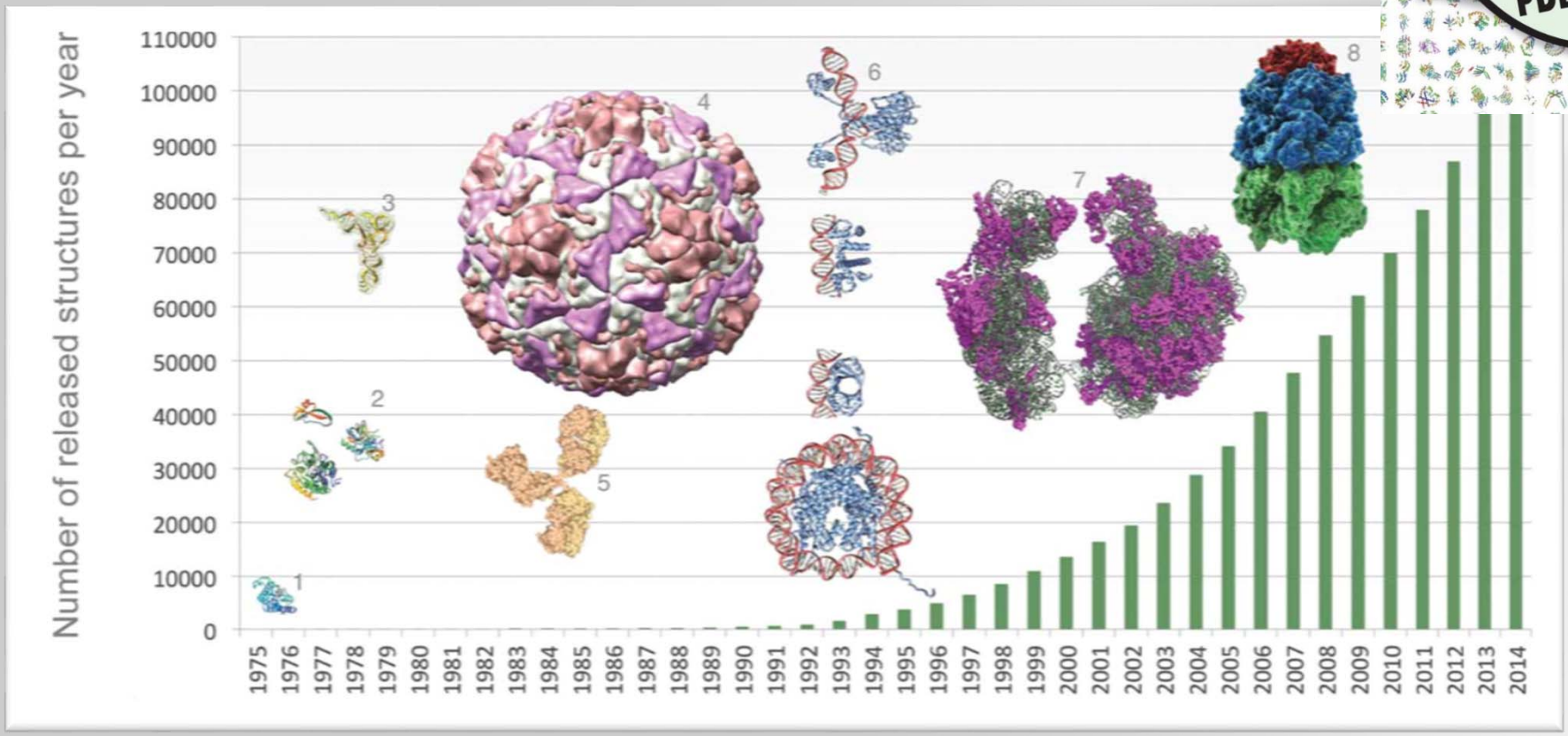




Available structural data in the PDB

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

RCSB PDB
PROTEIN DATA BANK



What is a structure ?

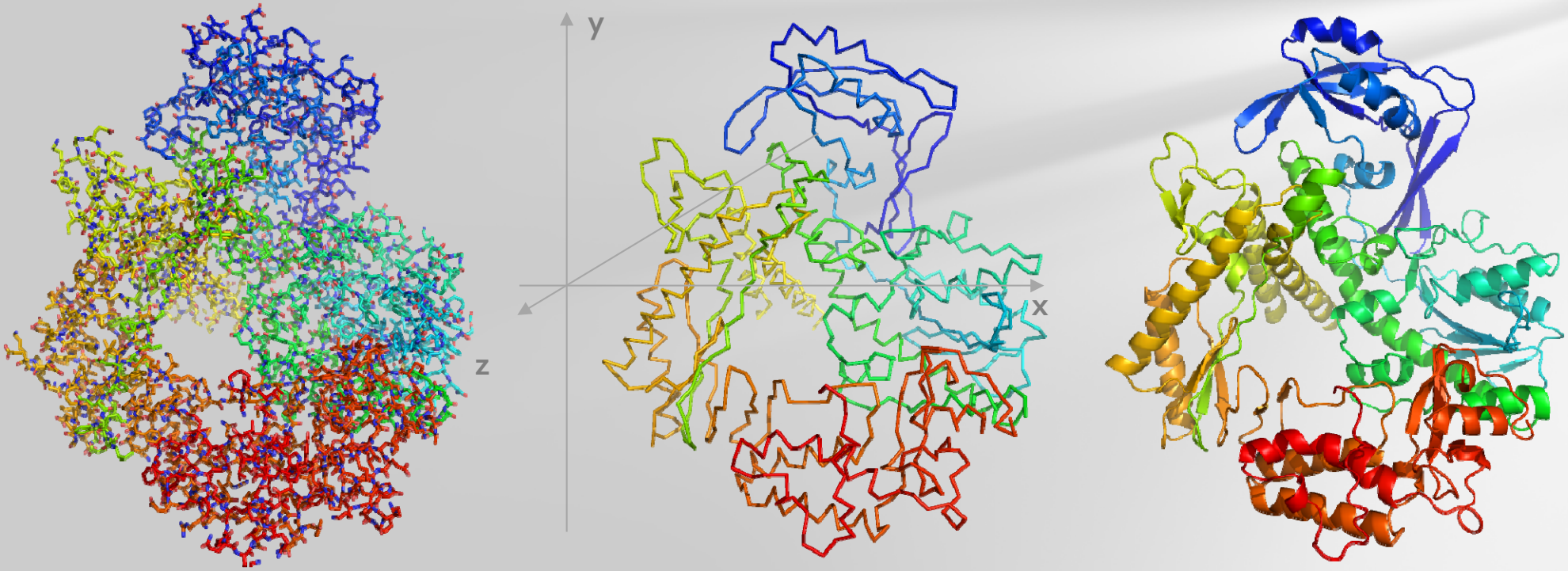
Structure prediction

Stability prediction

What is a structure ?

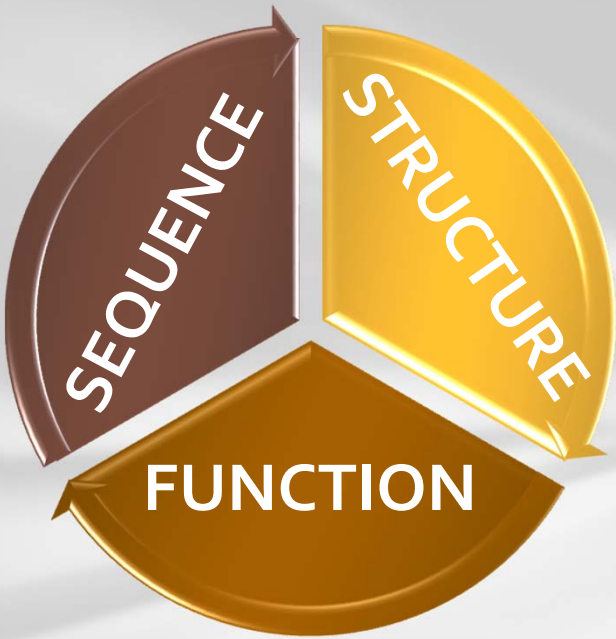
A 3-dimensional description of all atoms

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)

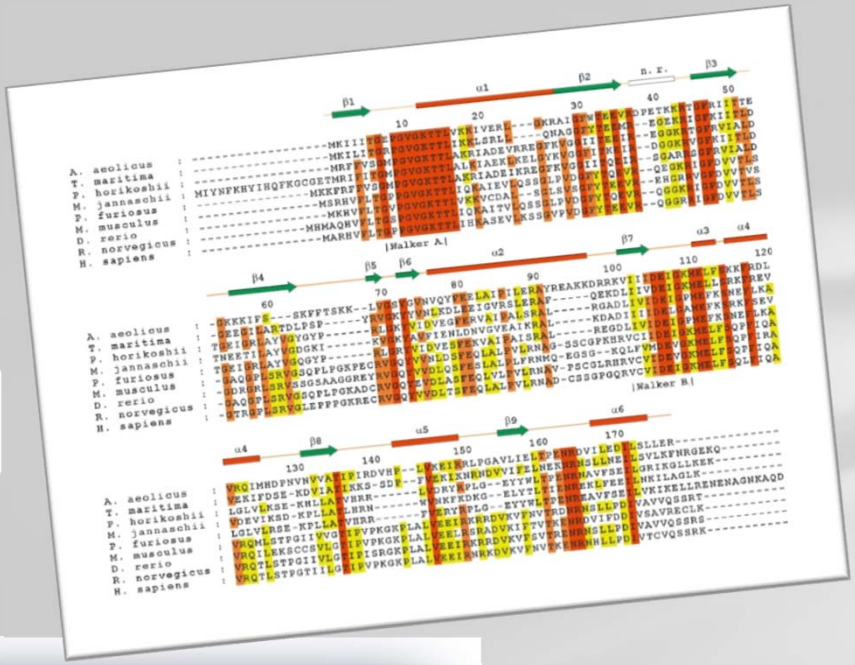


Sequence determines structure

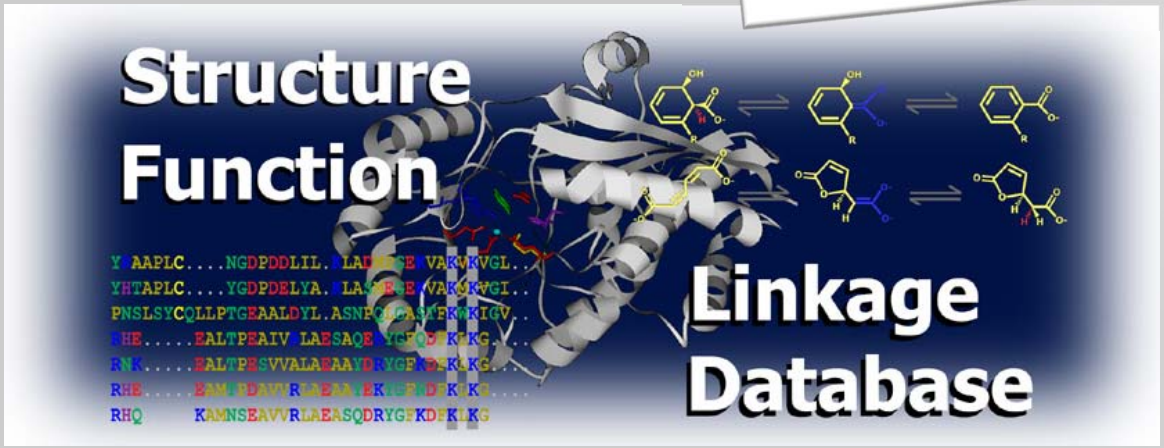
The aminoacid sequence of a protein determines its three-dimensional structure



Proteins that share sequence identity > 25 % are structurally similar



Homology modeling

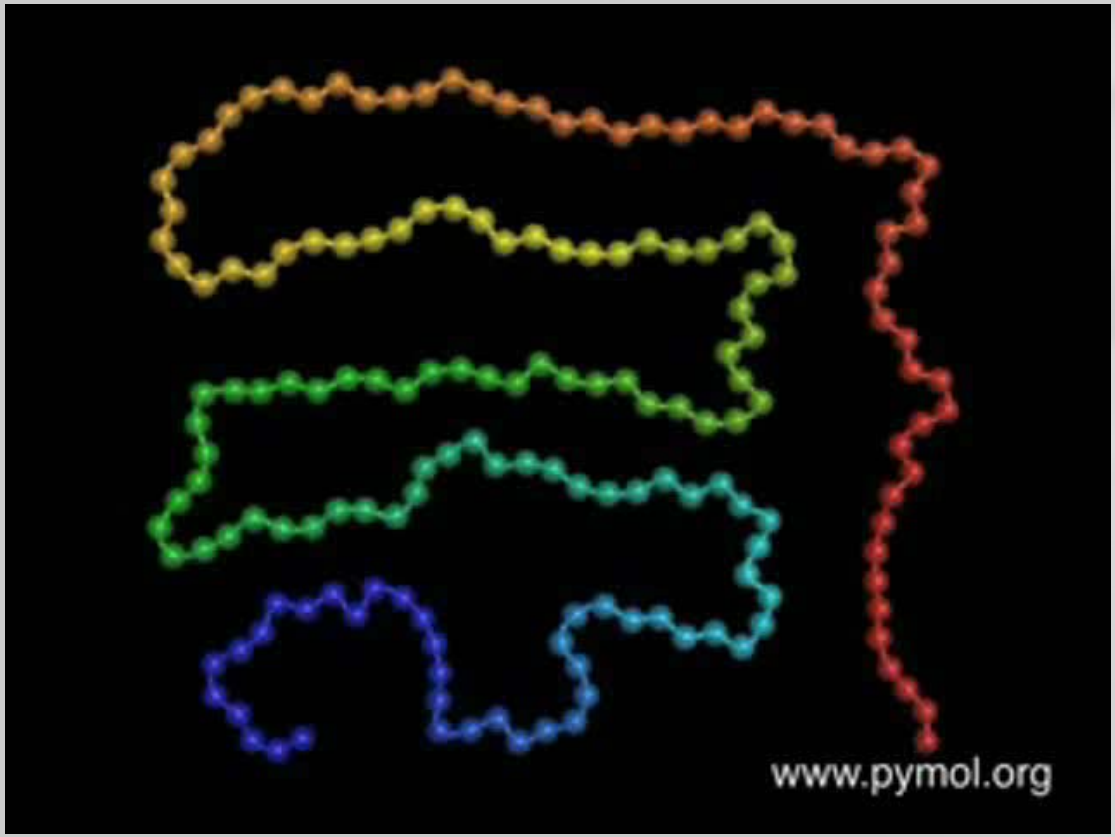


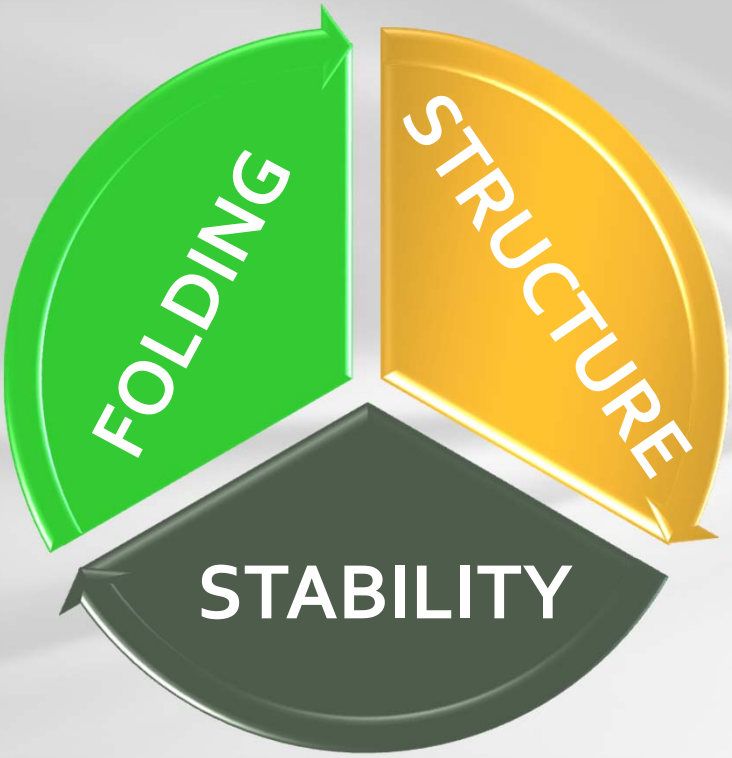
… And structure is highly correlated to function

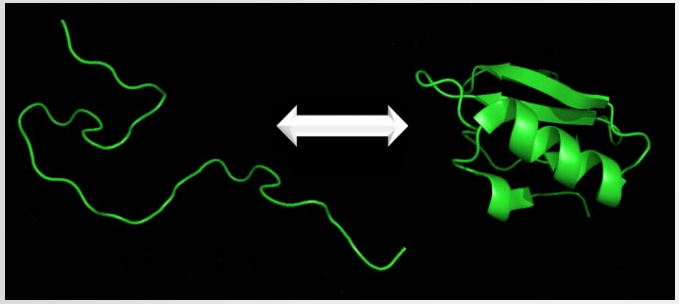
http://sfld.rbvi.ucsf.edu/django/

Mechanism of protein folding

What are the mechanisms that govern protein folding ?

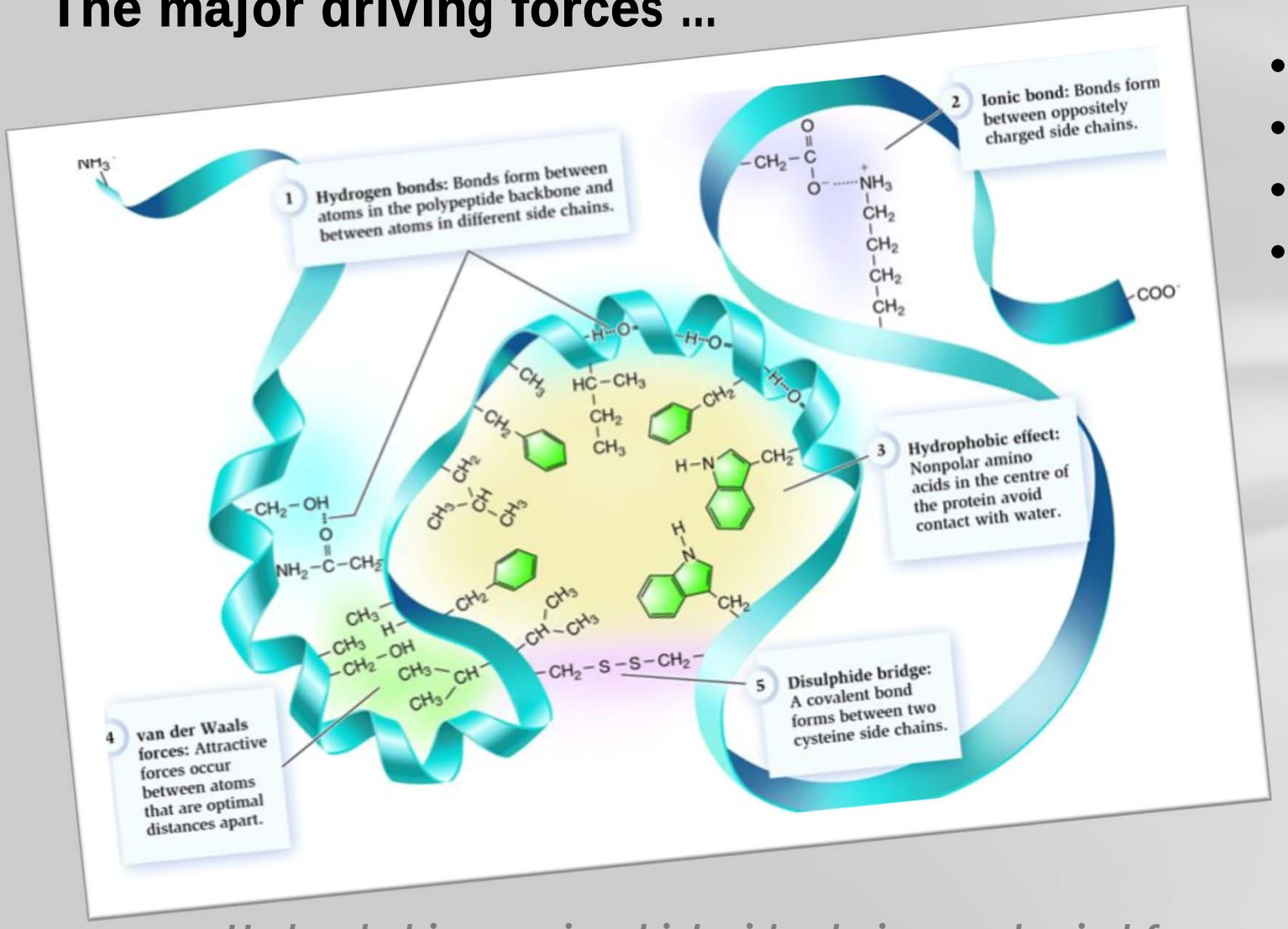






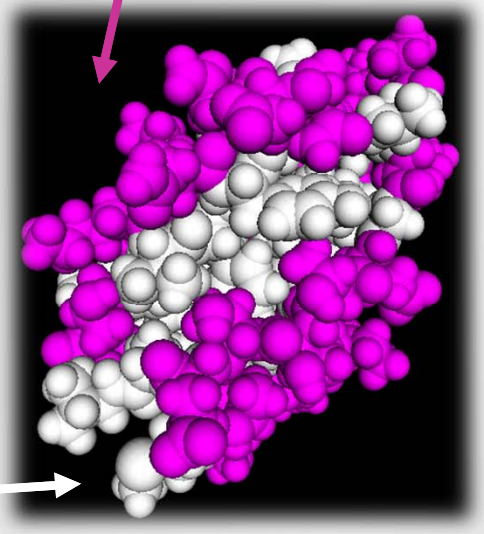
Factors determining protein folding

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

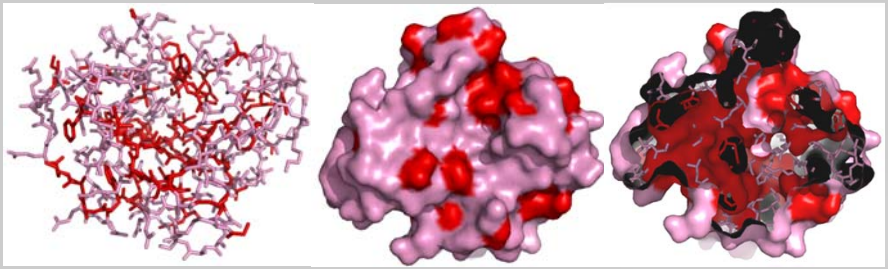


Hydrophobic core in which side chains are buried from water
Minimizing the number of hydrophobic side chains exposed to water is the principal driving force behind the folding process

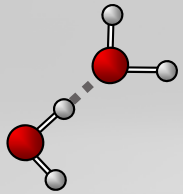
The Major Factors Affecting Protein Stability

… are the forces that contribute to protein stability

Hydrophobic interactions

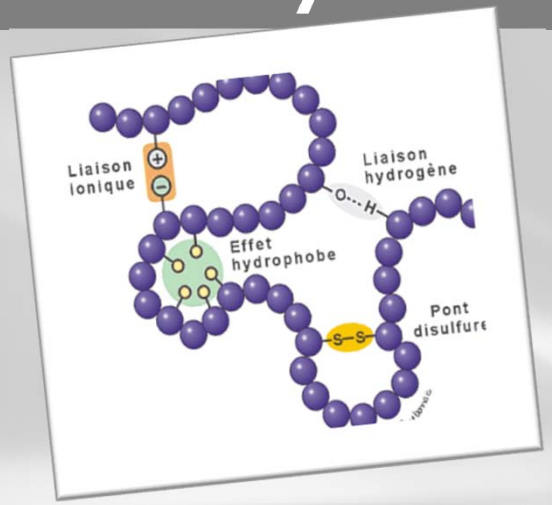


Non covalent bonds

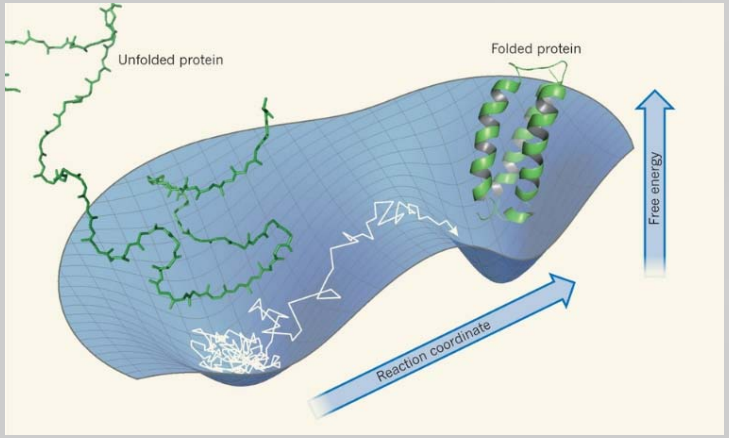


H-bond

Salt bridge



Conformational entropy



Difficult to predict protein stability from structure with computational methods !

Destabilising	
Conformational Entropy	-177
roups Buried	-81
ups Buried	-28
tabilising	-286
g	
lonisation	+4
Bonds	+7
bic Groups Buried	+94
Hydrogen Bonding	+166
Total Stabilising	+271
G (estimate)	-15
G (measured)	+9

Free Energy (kcal/mol)

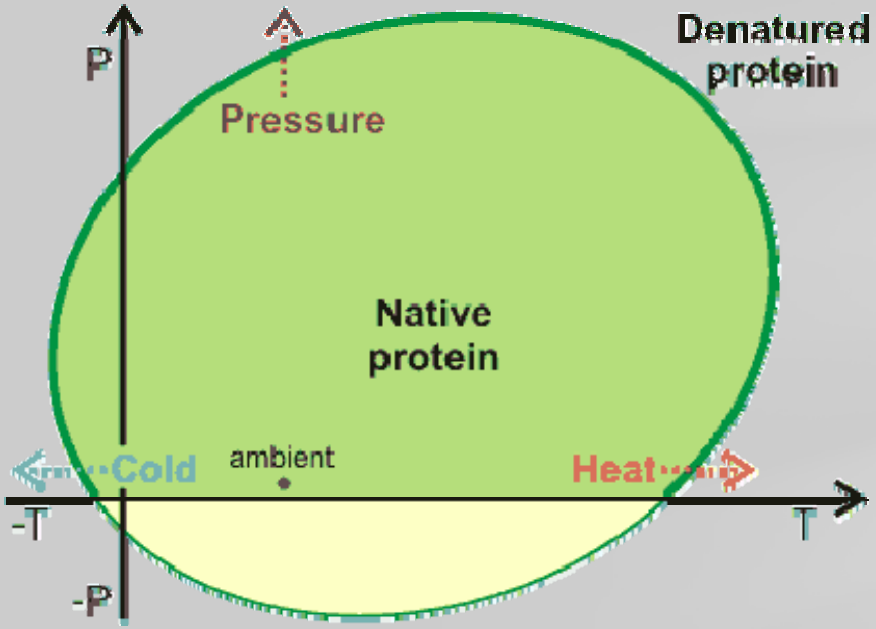
Rnase T1

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

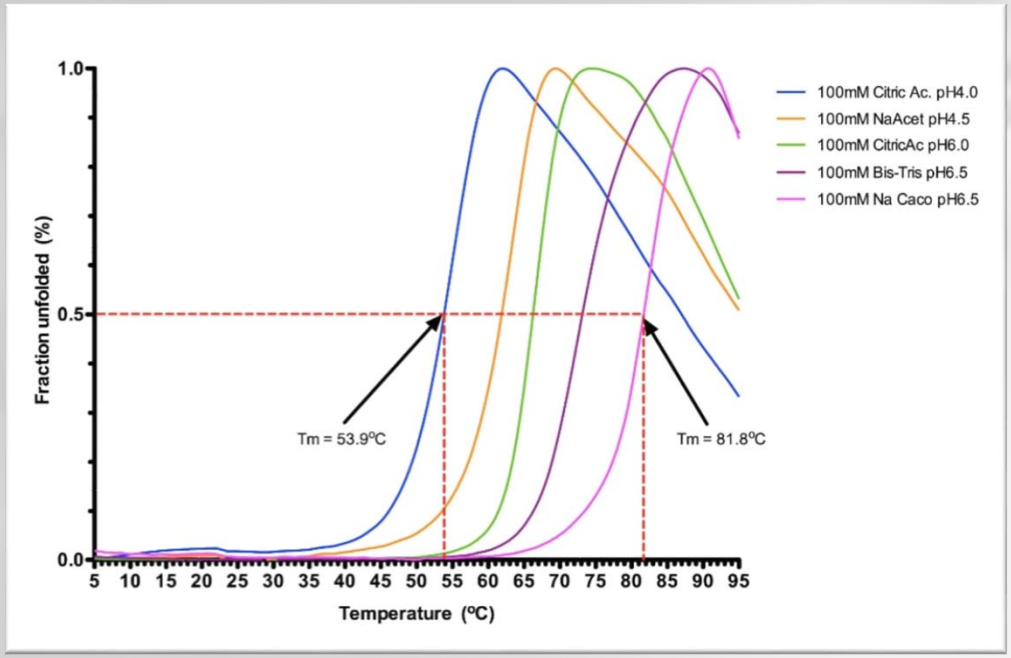
Effect of environment on protein stability

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

Pressure vs temperature



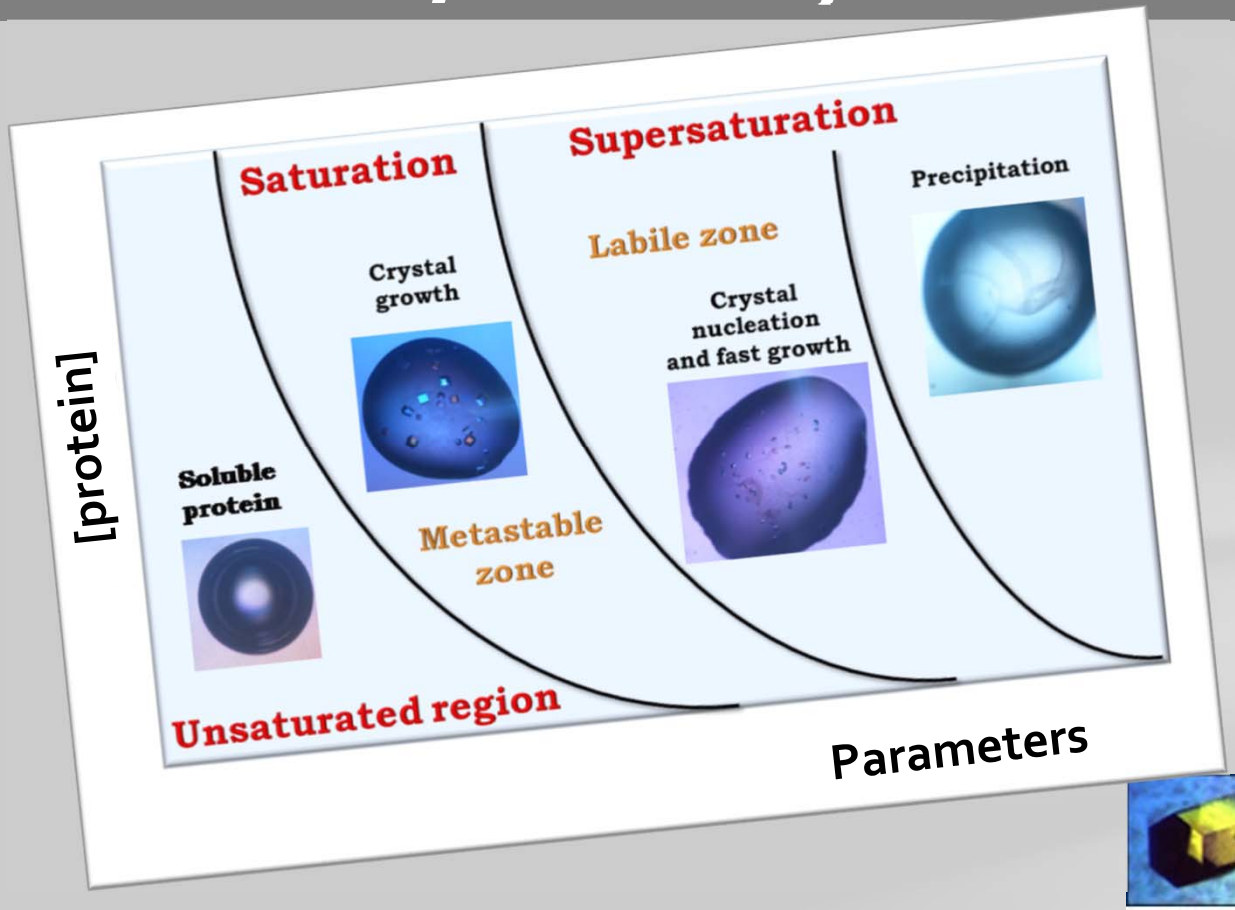
Temperature vs pH

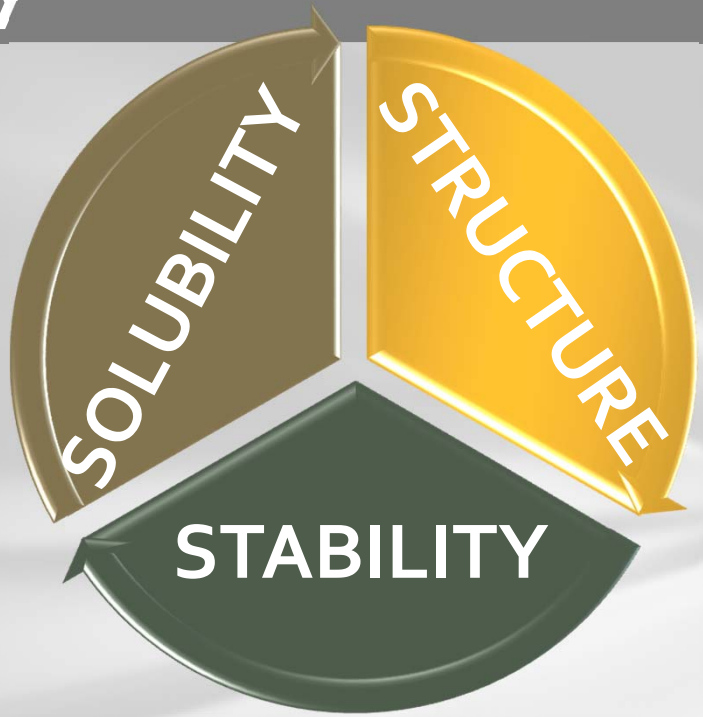


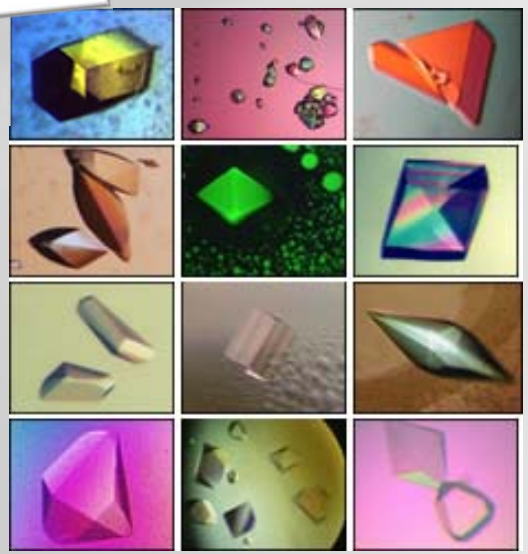
Typical parameters

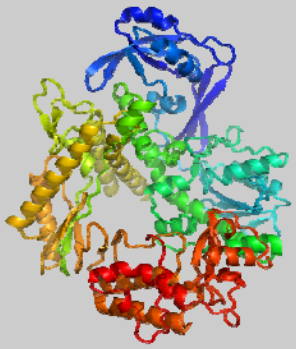
- **External: external conditions and buffer solutions**
- **Internal: aa modifications, aa substitutions**

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, agregation or denaturation and secondary structure

Several servers and programs

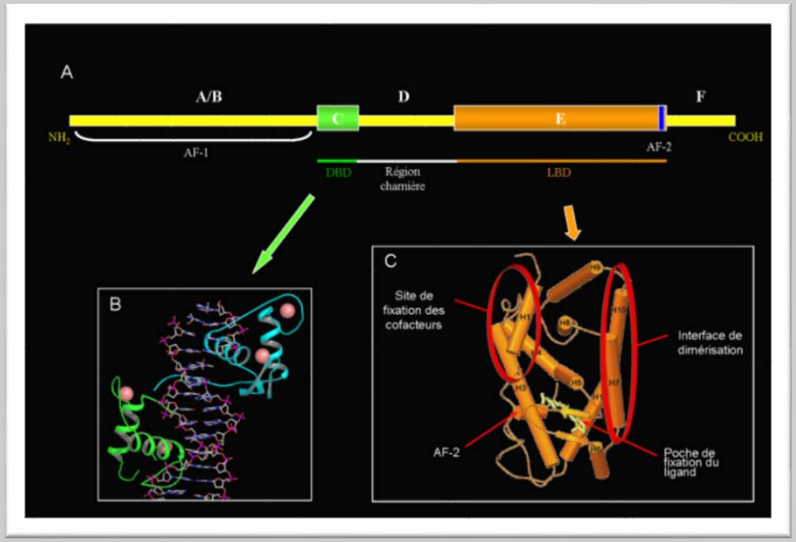
http://www.disprot.org/predictors.php

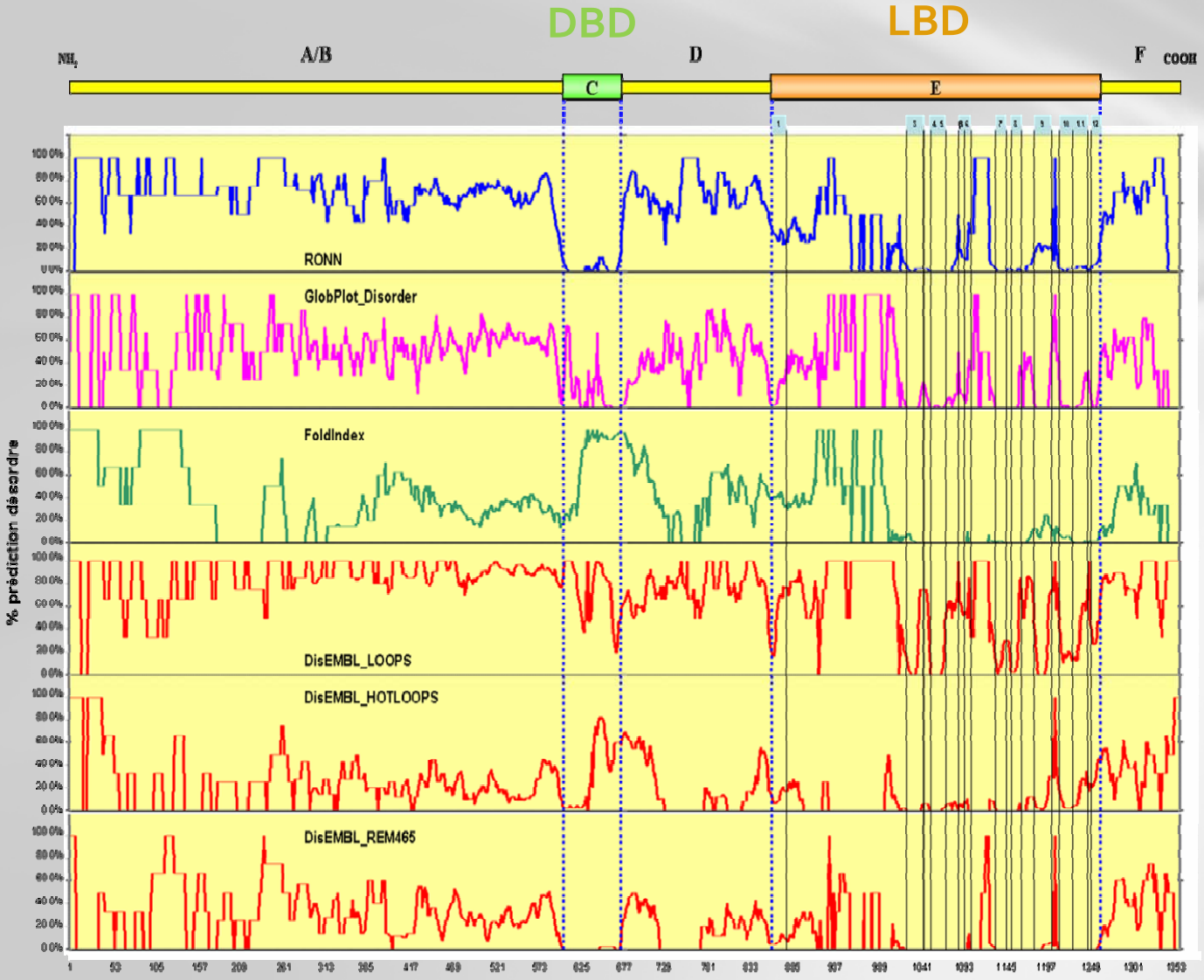
Protein Disorder Predictors

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





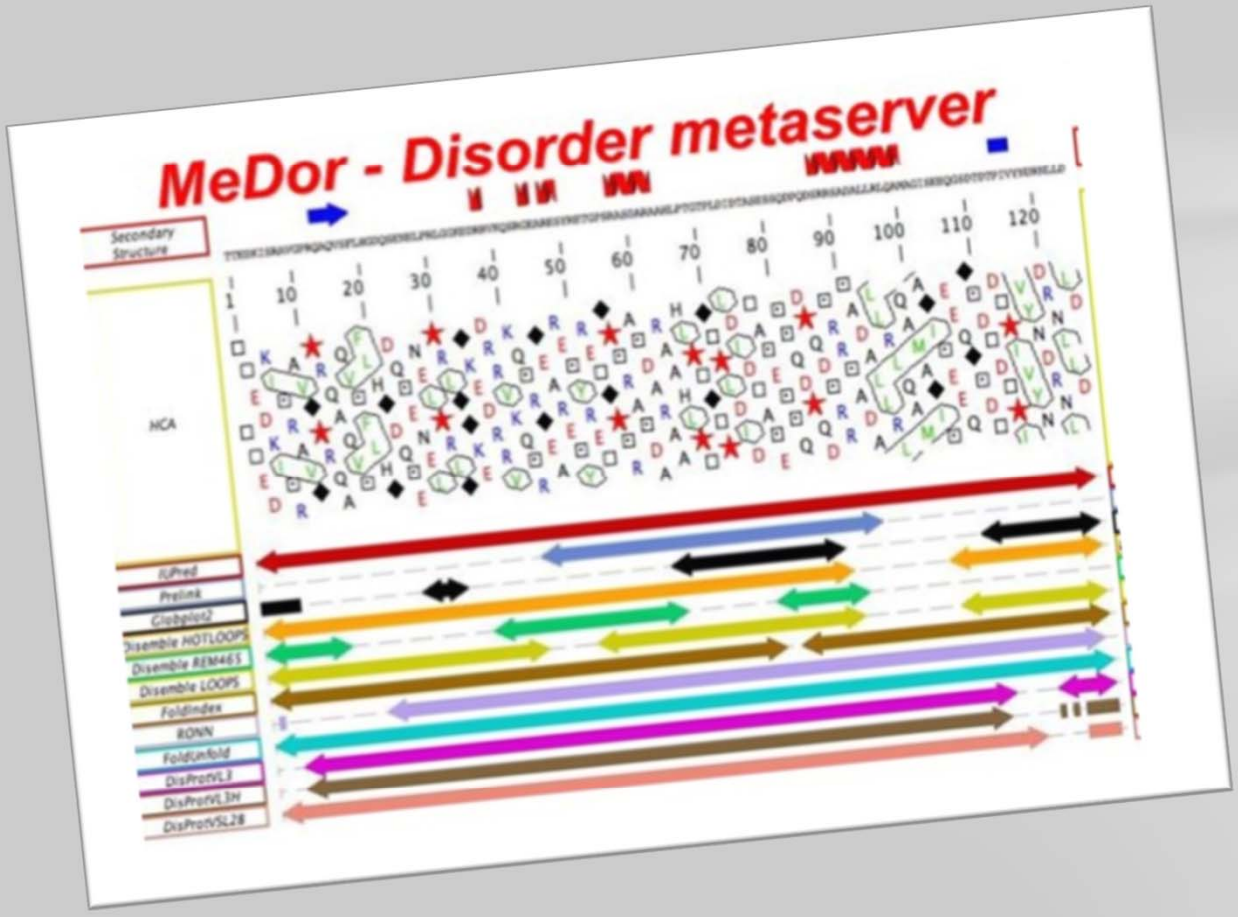
Predictions of disorder using different programs and a multiple alignment of 48 human nuclear receptors

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

Prediction of order/disorder

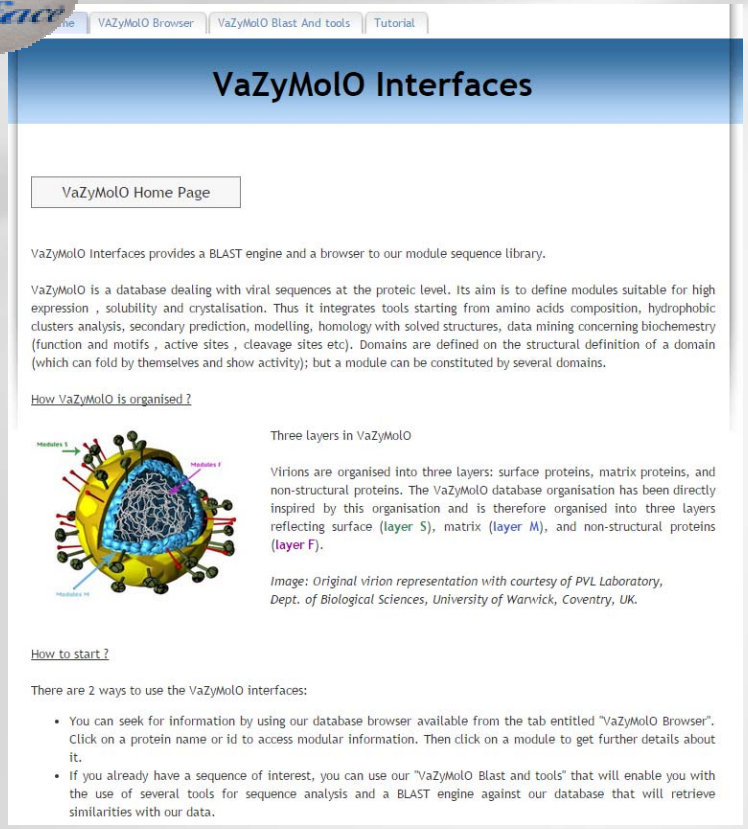
MEDOR **http://www.vazymolo.org/MeDor/**

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



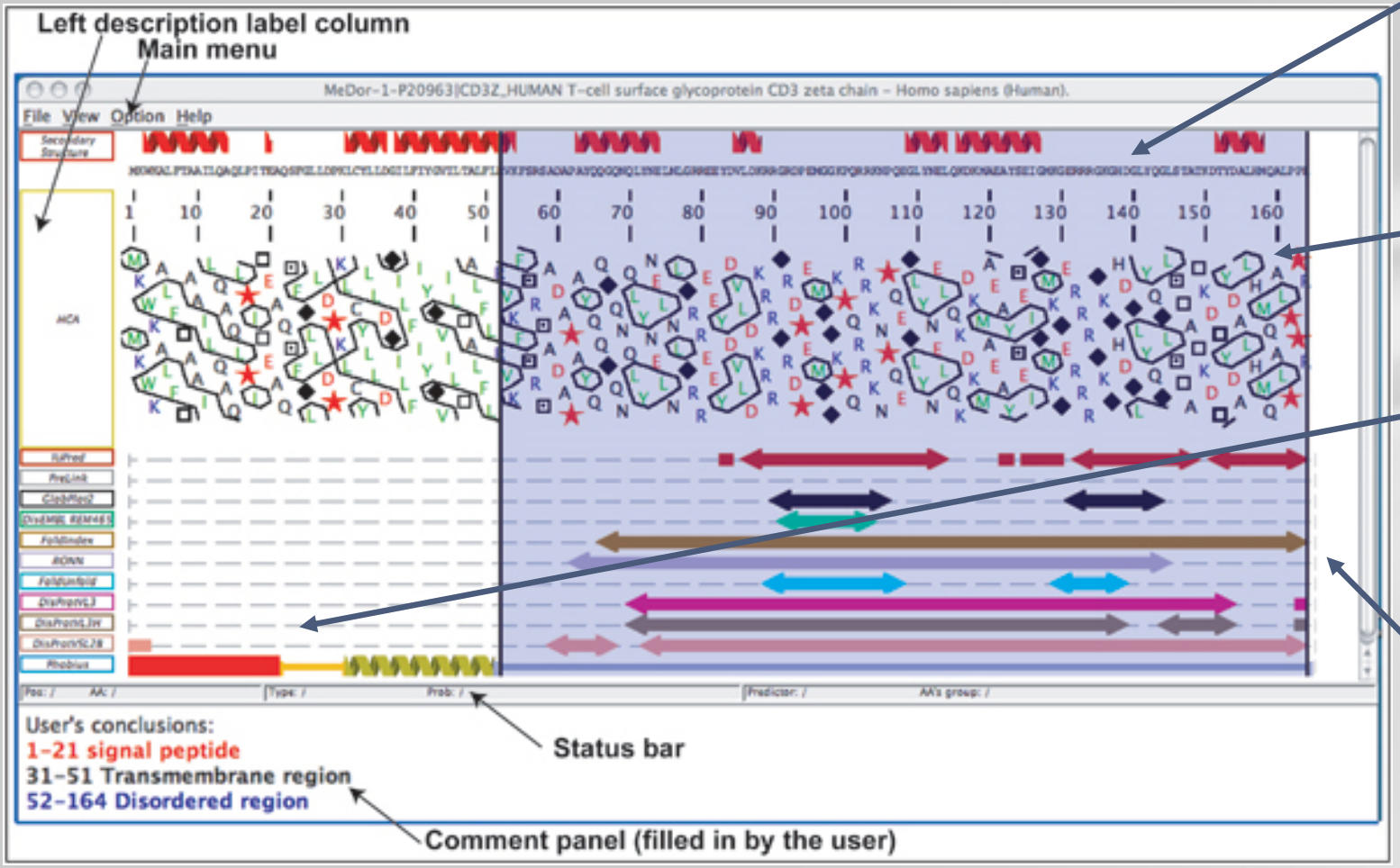
Database of viral proteins

Its aim is to define modules suitable for high expression, solubility and crystallization



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.

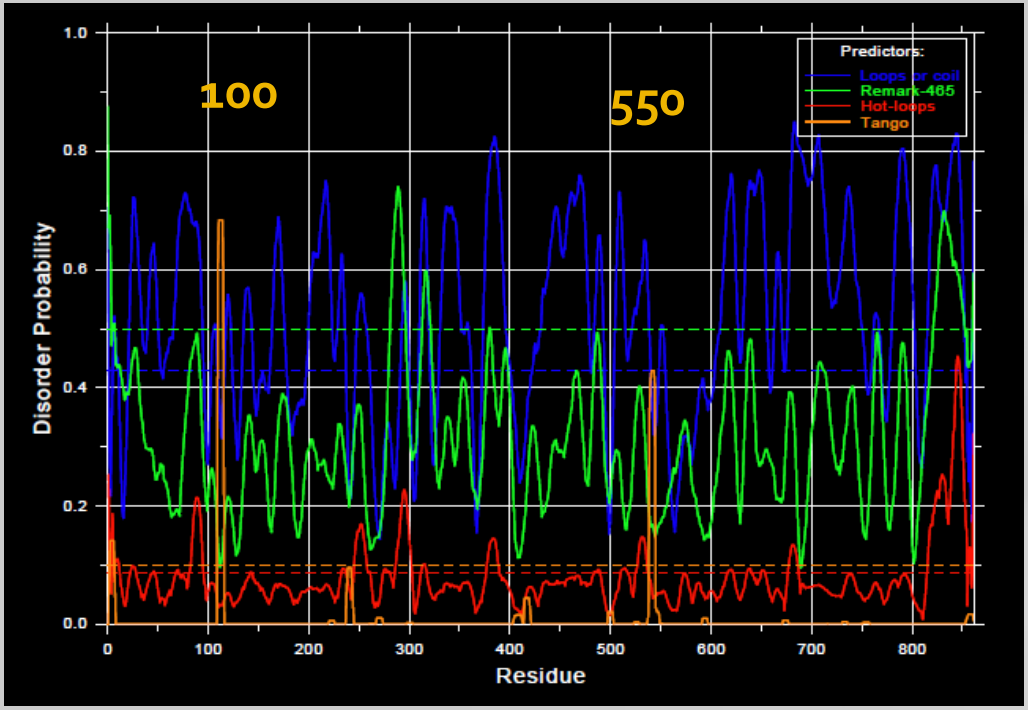
DisProt entry DP00200 human T cell glycoprotein CD3 Z chain (P20963)

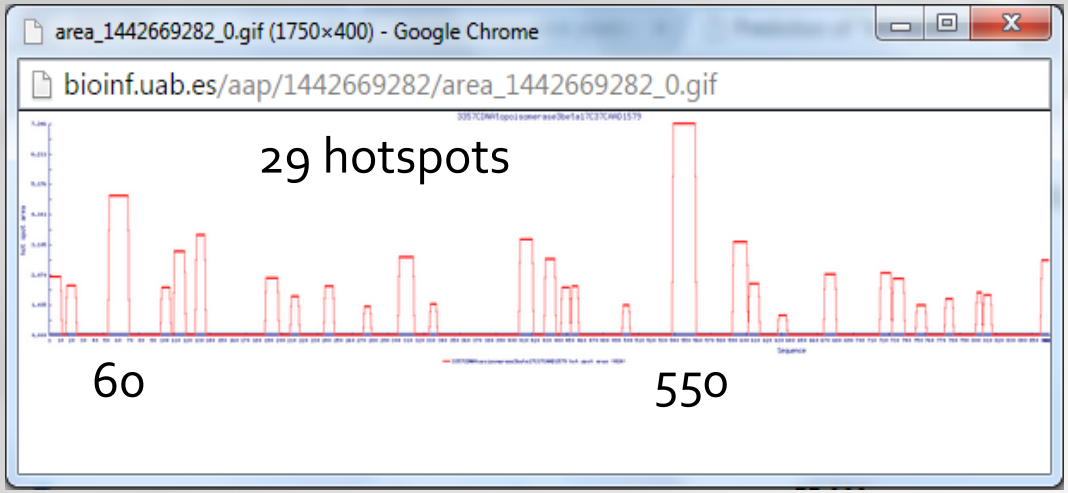
Prediction of aggregation

TANGO **http://dis.embl.de/**

AGGRESCAN **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

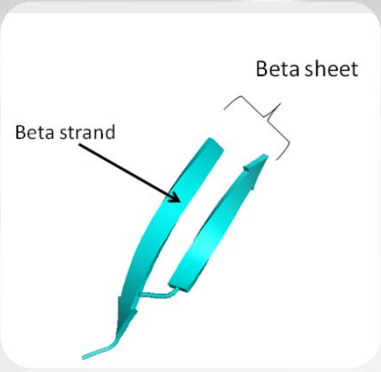
Secondary structure prediction

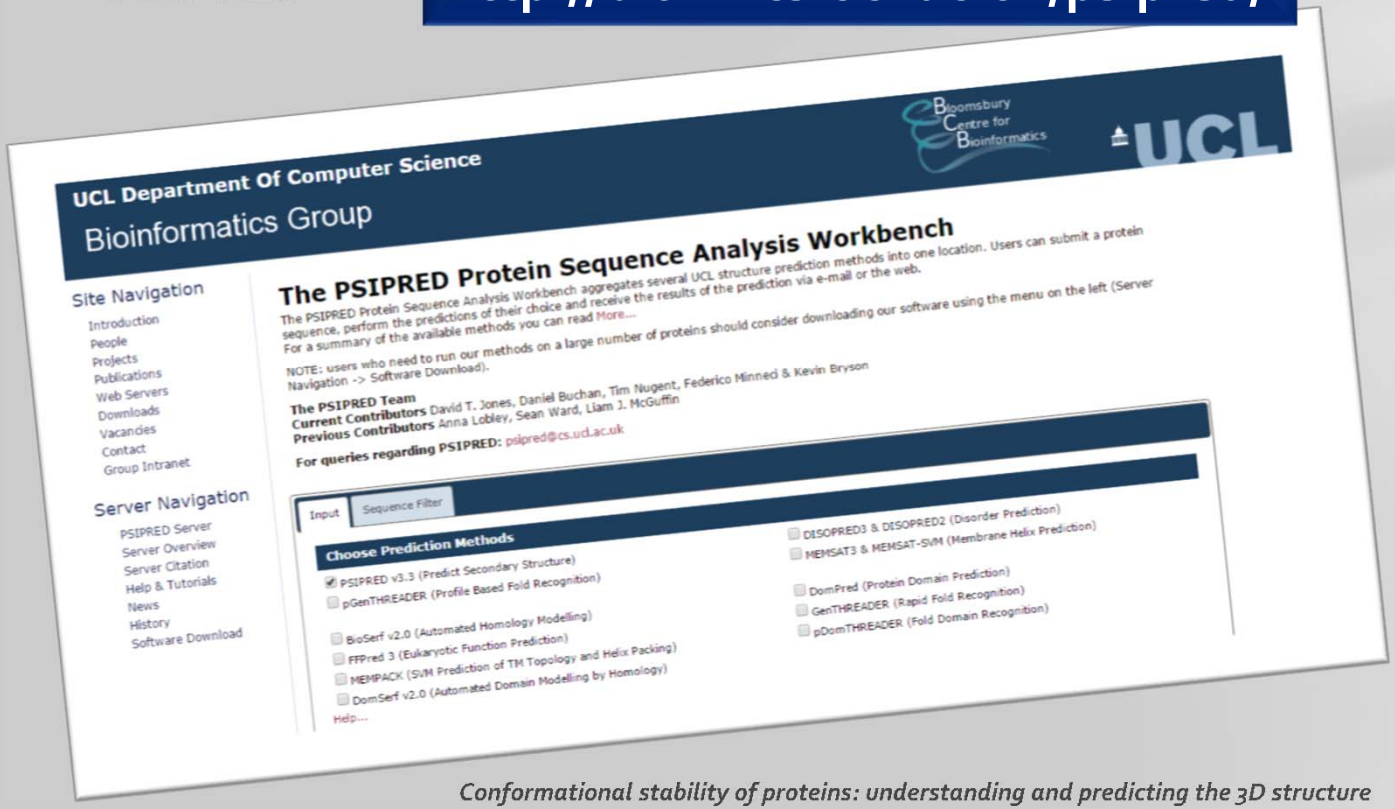
Based on the propensity of each aminoacid to form a secondary structure (helix and strand)

SOPMA https://npsa-prabi.ibcp.fr/

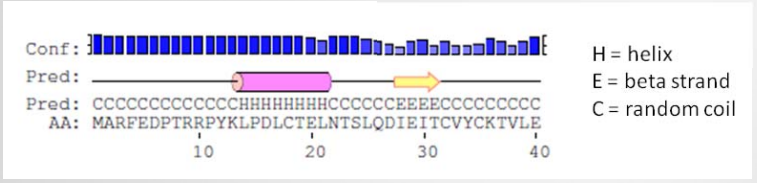
Jpred 4 http://www.compbio.dundee.ac.uk/jpred/

PSIPRED http://bioinf.cs.ucl.ac.uk/psipred/



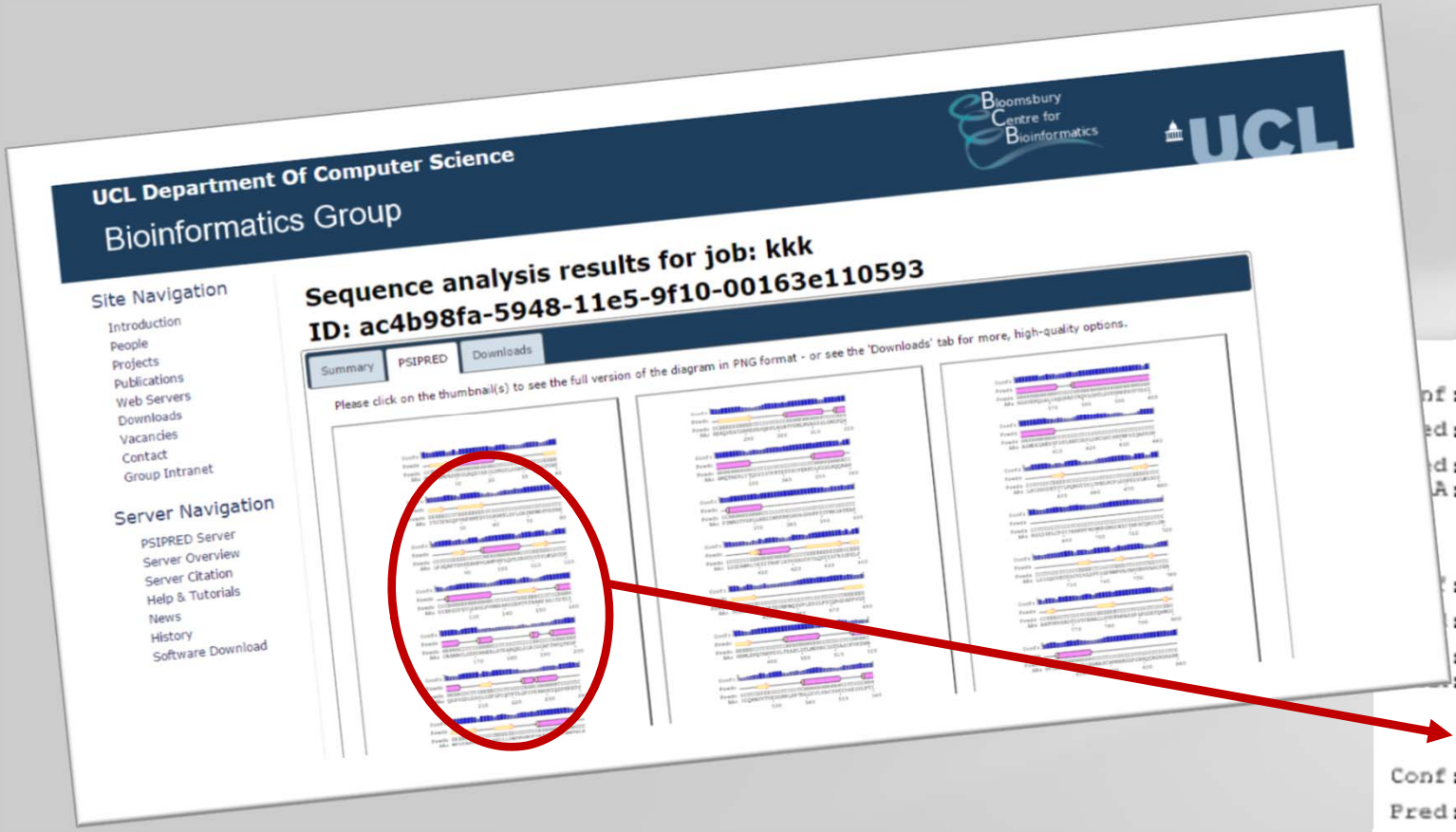


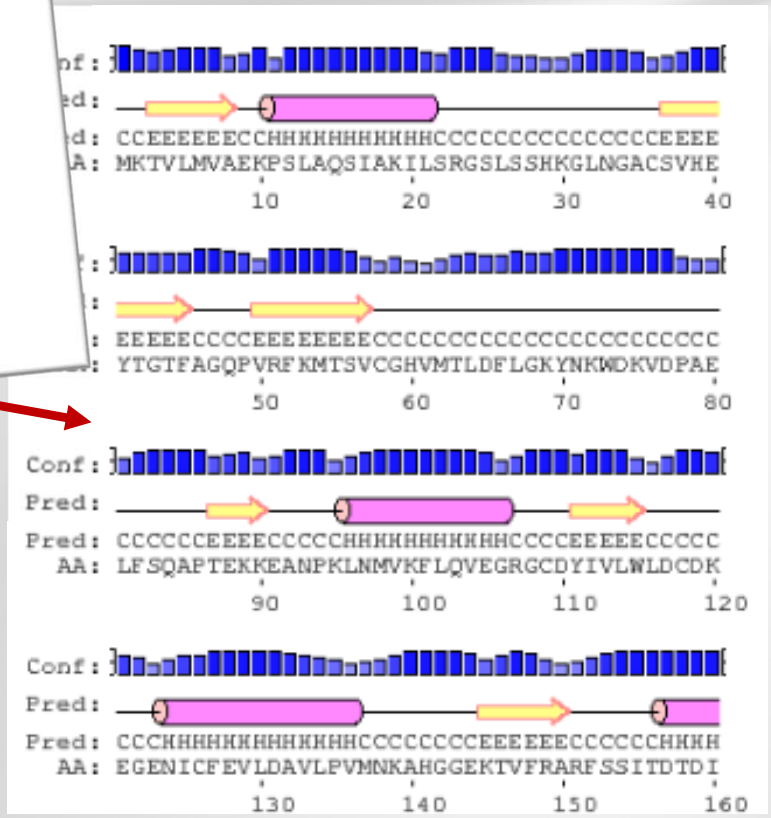
Prediction all along the sequence with a confidence index



Secondary structure prediction

Example of a PSIPRED output





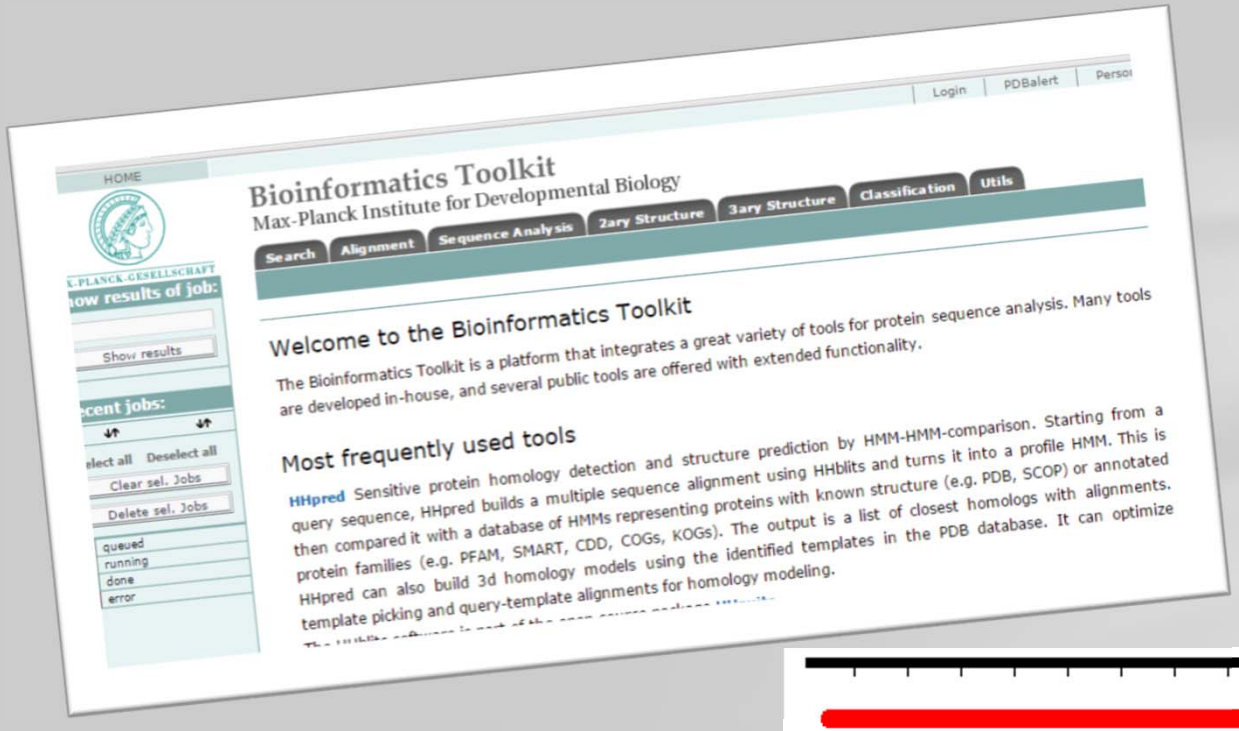
human Topo IIIα

From sequence to structure prediction

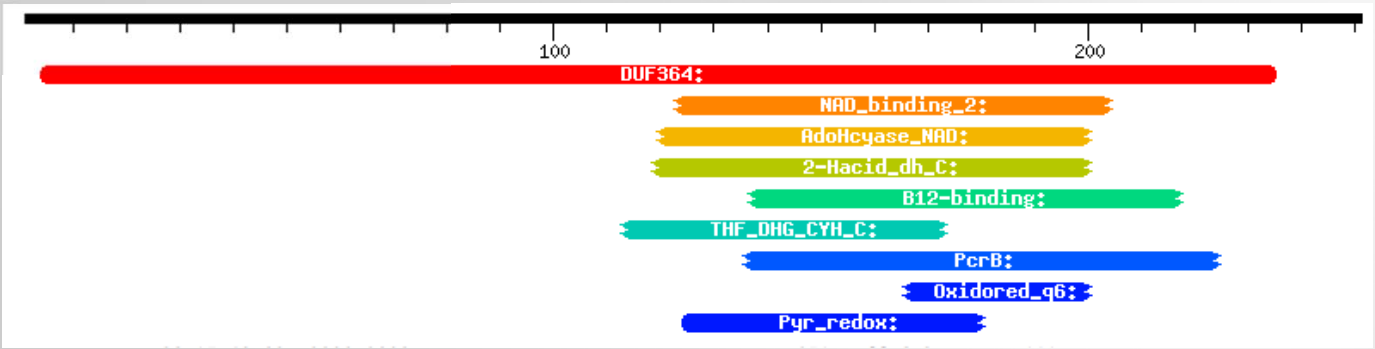
Homology detection and structure prediction by HMM-HMM comparison

HHpred

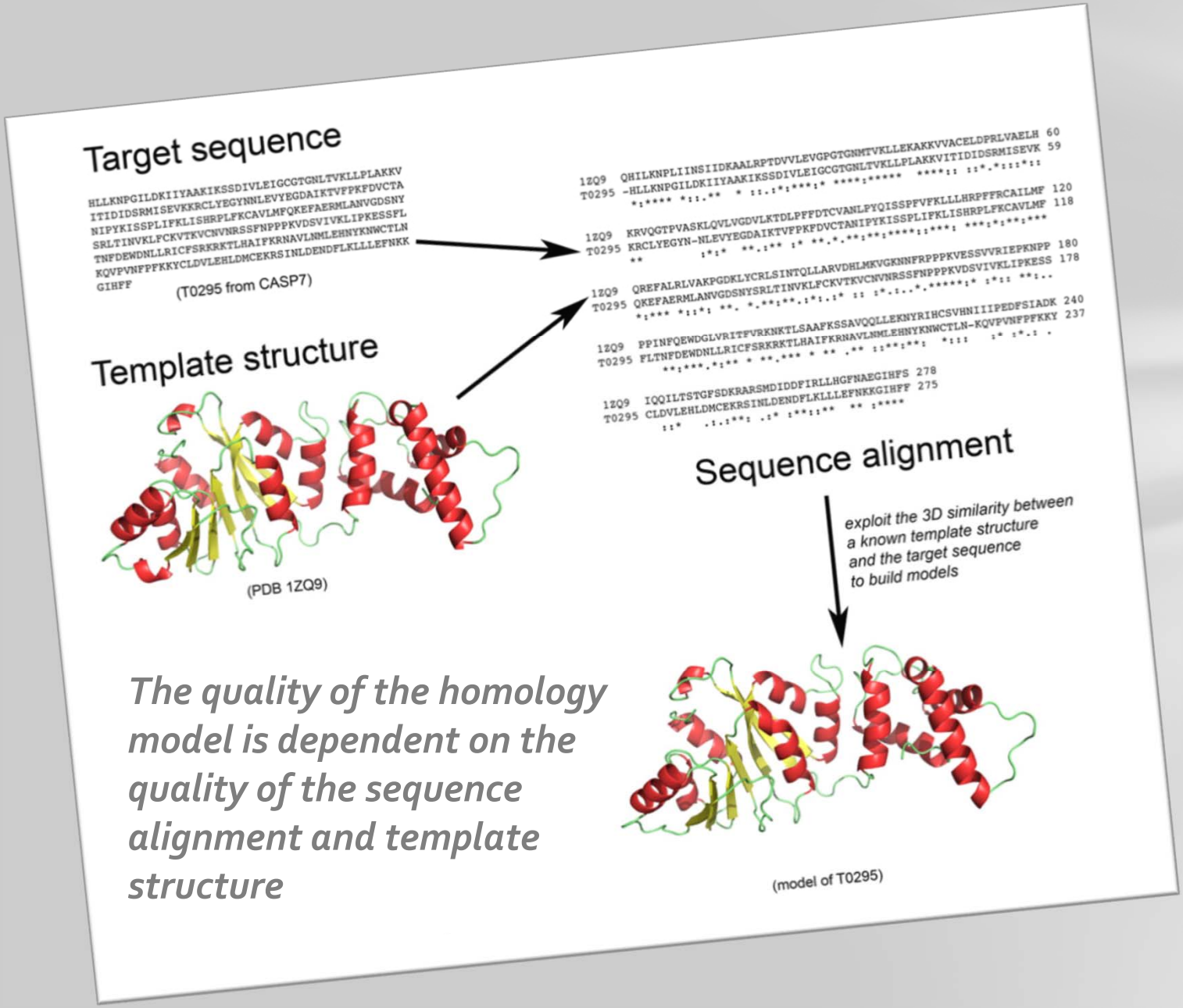
http://toolkit.tuebingen.mpg.de/hhpred



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



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

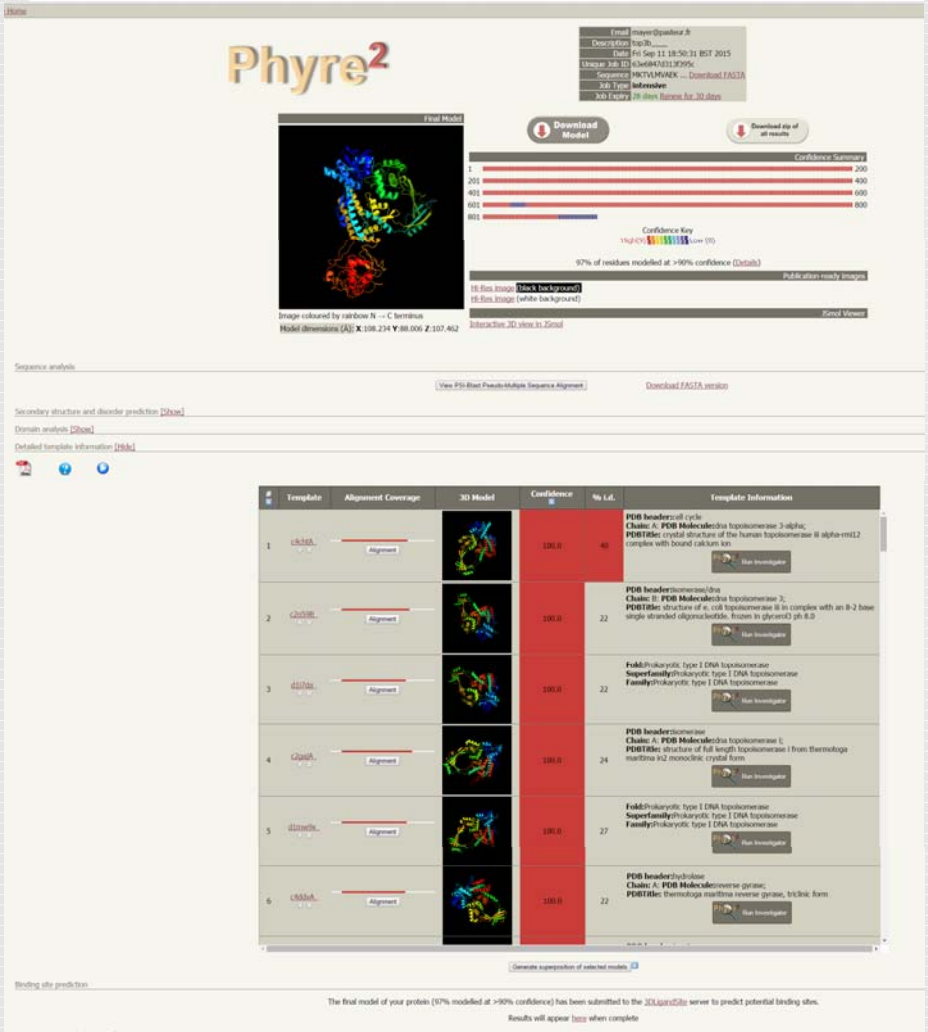
The sequence alignment and template structure are then used to produce a structural model of the target

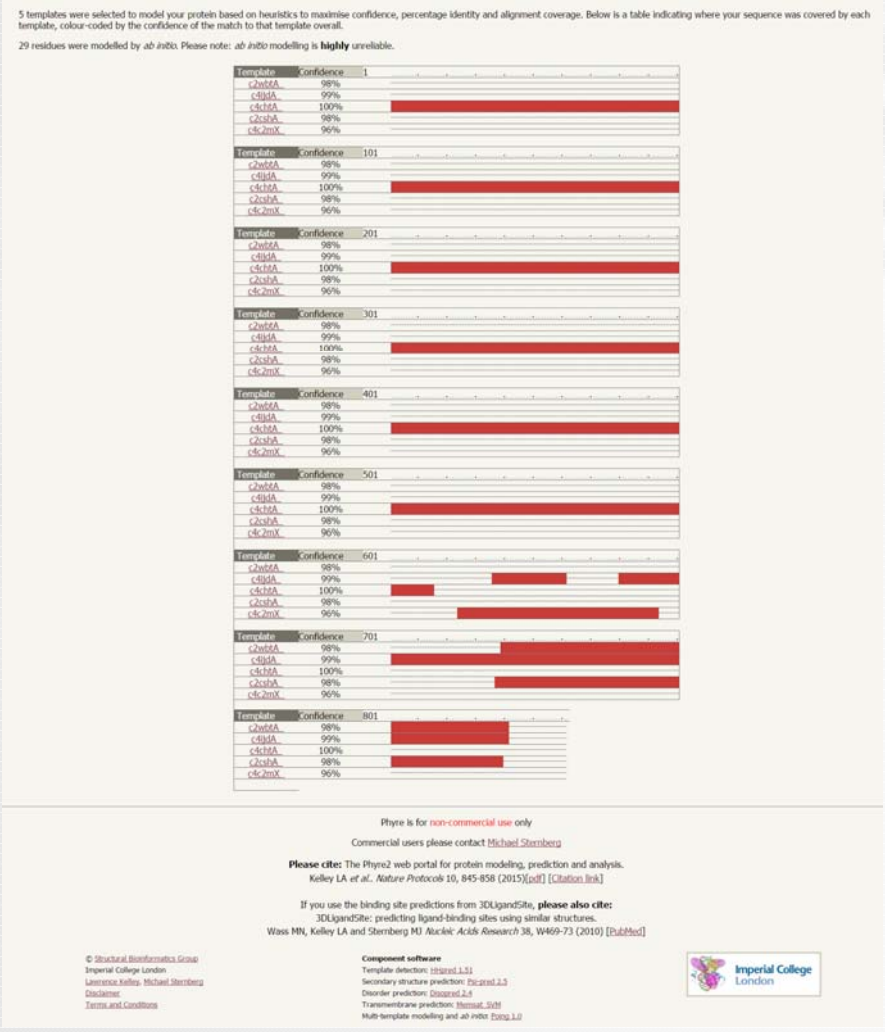
Homology modeling

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**



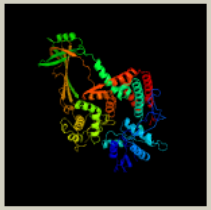
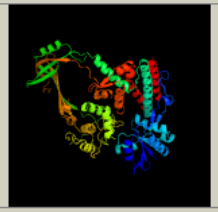


Homology modeling

Phyre

ls of

, and are

#	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 Run Investigator
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 Run Investigator
3	d1i7da	Alignment		100.0	22	**Fold:**Prokaryotic type I DNA topoisomerase **Superfamily:**Prokaryotic type I DNA topoisomerase **Family:**Prokaryotic type I DNA topoisomerase Run Investigator
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 Run Investigator
5	d1mw9x	Alignment		100.0	27	**Fold:**Prokaryotic type I DNA topoisomerase **Superfamily:**Prokaryotic type I DNA topoisomerase **Family:**Prokaryotic type I DNA topoisomerase Run Investigator
6	c4ddvA	Alignment		100.0	22	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**reverse gyrase; **PDBTitle:** thermotoga maritima reverse gyrase, triclinic form Run Investigator

What is a structure ?

Structure prediction

Stability prediction

Structural analyses

Calculations or estimation of structural parameters that contributes to protein stability

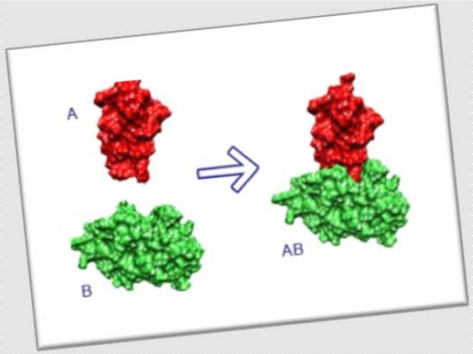
PROPKA	estimation of the pKa values of ionisable aa
DALI	comparison of structural homologues, prediction of function
VADAR	structure validation server that allows to calculate volumes, accessible surfaces, contact surface, …
MarkUs	analysis and comparison of the structural and functional properties of 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

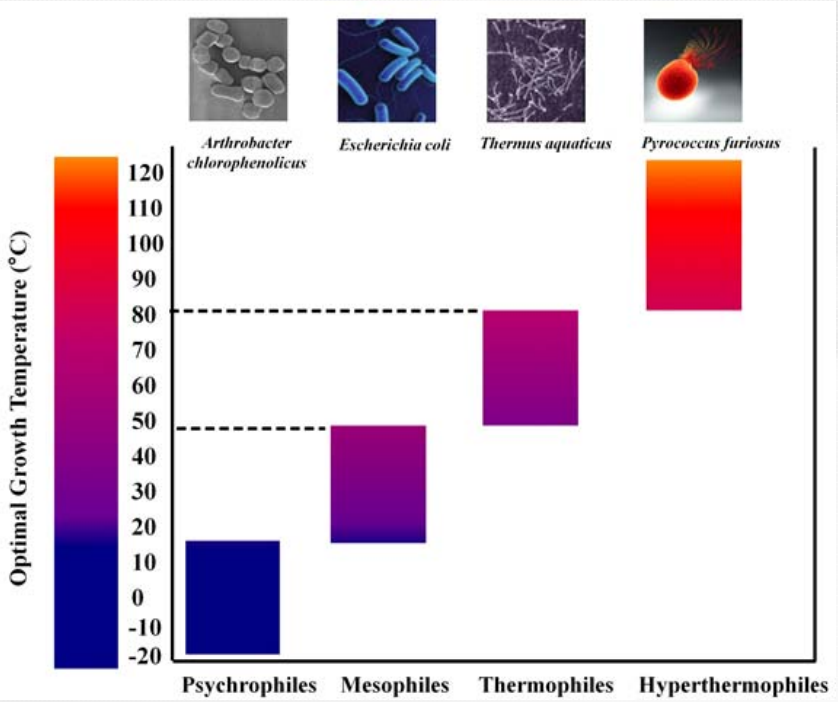
HADDOCK	Docking with possibility to implement experimental data (mutagenesis, cross-linking, NMR chemical shift, …)
HEX	protein-protein docking, webserver also exists

Protein thermostability

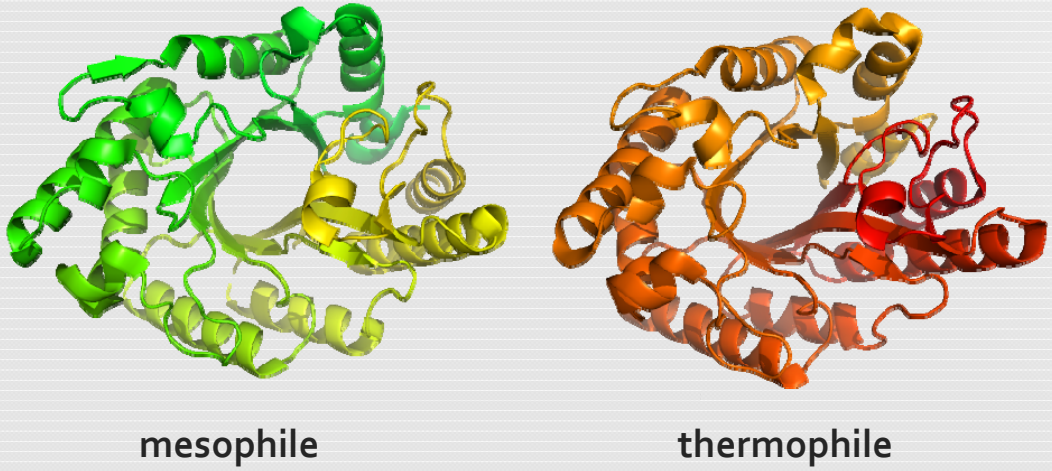
Effect of the temperature on protein stability

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

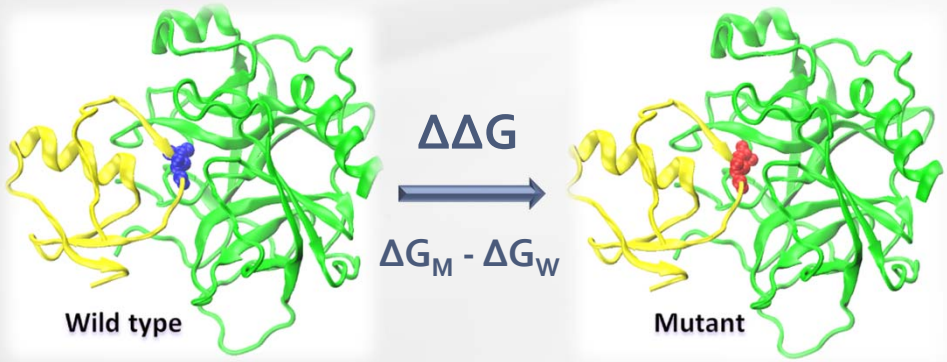
Effect of mutations on protein stability

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

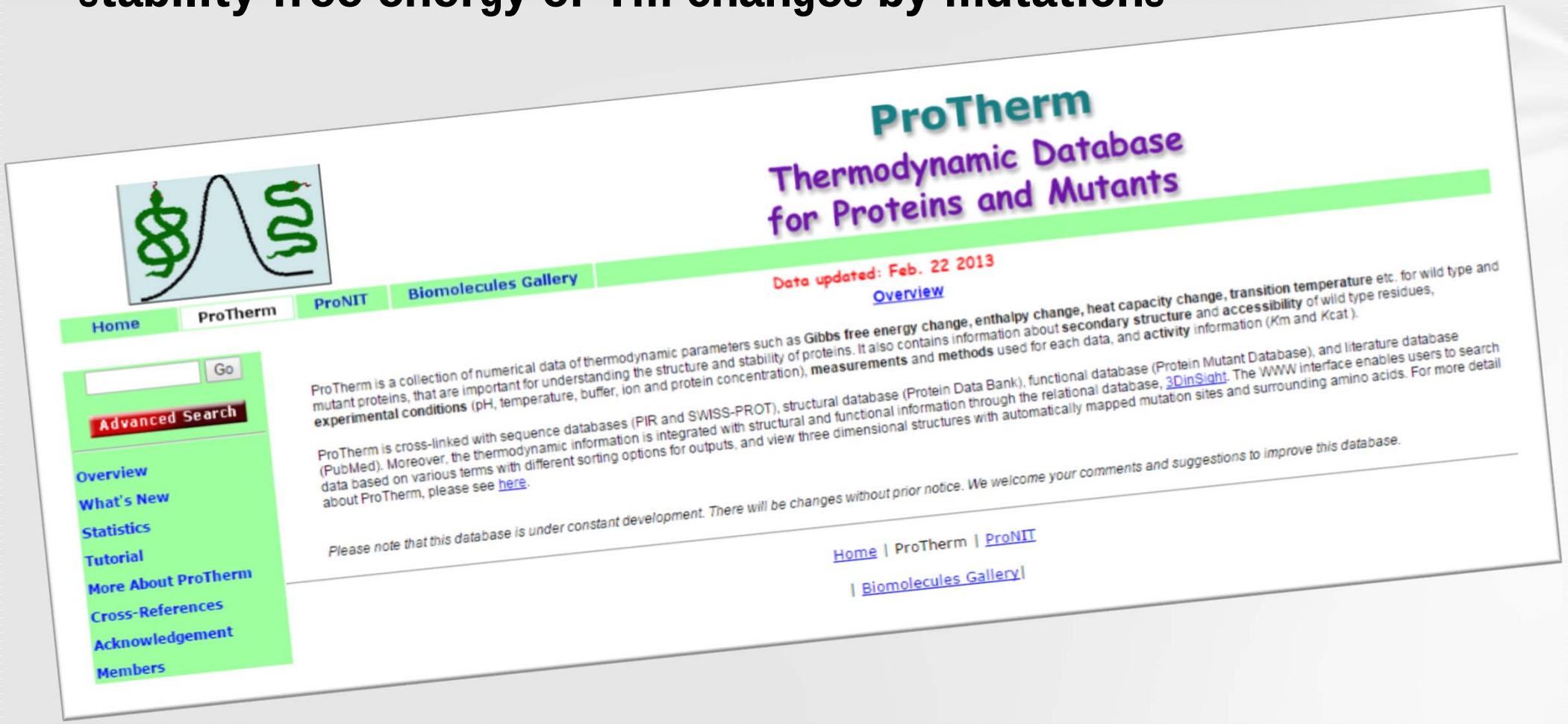
Based either on atomic force field, statistical, empirical approaches or machine-learning methods or a combination of both

Use structure and/or sequence

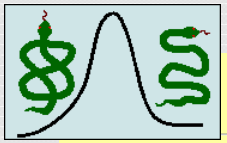
Effect of mutations on protein stability

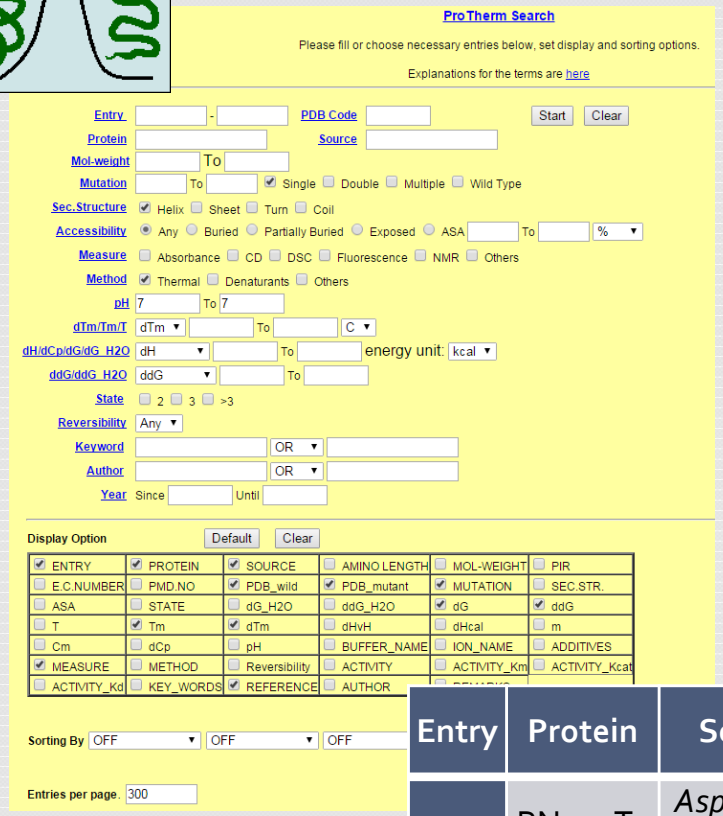
ProTherm **http://www.abren.net/protherm**

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

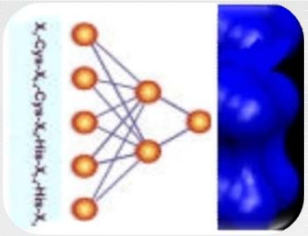


Effect of mutations on protein stability





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	*Aspergillus oryzae*	1RN1	1RGC	**Q25K**	NULL	NULL	51.7	3.40	Fluores	J BIOL CHEM 264, 11621-11625 (1989)
2333	RNase T1	*Aspergillus oryzae*	1RN1	NULL	**S17A**	NULL	NULL	52.6	1.70	Fluores	BIOCHEM 31, 725-732 (1992) PMID: 1731929
14482	DsbA	*Escherichia coli*	1A23	NULL	**H32Y**	13.80	6.80	NULL	NULL	Fluores	PROTEIN SCI 6, 1893-1900 (1997)
14483	DsbA	*Escherichia coli*	1A23	NULL	**H32L**	12.30	5.30	NULL	NULL	Fluores	PROTEIN SCI 6, 1893-1900 (1997)
14484	DsbA	*Escherichia coli*	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...

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...

I-Mutant

A support vector ma… …utomatic prediction of protein st… …mutations. I-Mutant predictions are …e protein structure or, more...

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...

MUpro

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

CUPSAT

Cologne University Pr…

A… …stability upon point mutations. The prediction …om potentials and torsion angle distribution to assess …environment of the mutation site....

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...

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

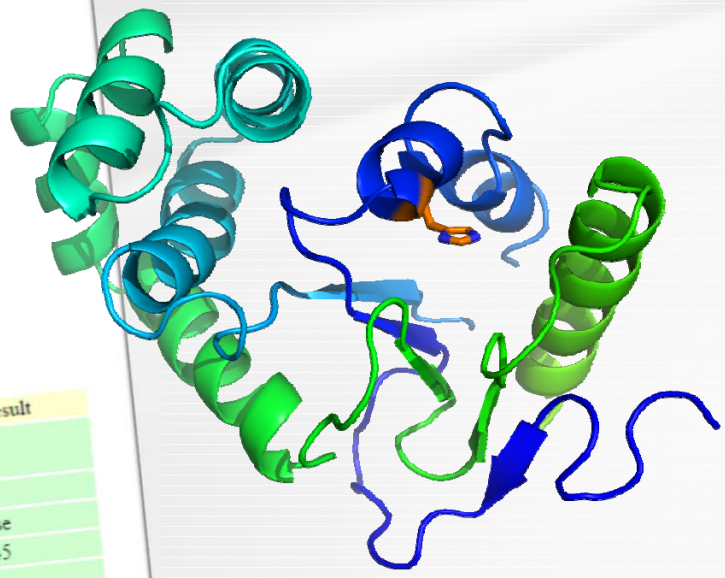
PDB ID: 1A23 Chain: A Wild-type: H Position: 32(32) Mutant: Y Temperature pH (°C, 0-100) (0-14): 25 7

Protein sequence(no headers)

AQYEDGKQYTTLEKPVAGAPQVLEFFSFFCPHCYQFEEVLHISDNVKKKLPEGVKMTKYHVNFM
GGDLGKDLTQAWAVAMALGVEDKVTVPLFEGVQKTQTIRSASDIRDVFINAGIKGEEYDAAWNS
FVVKSLVAQQEKAAADVQLRGVPAMFVNGKYQLNPQGMDTSNMDVFVQQYADTVKYLSEKK

Clear Submit

Predictor	i-Mutant2.0 PDB	i-Mutant2.0 SEQ	AUTO-MUTE SVM	Reference AUTO-MUTE RF	MUpro	PoPMuSiC	CUPSAT	Meta Result iStable
Result	7	31.9	5.55	helix	helix	0.99962109	null	32.44
Conf.			Increased	Increased	Increase			Increase
ΔΔG	Increase	Increase	0.76	0.95		null	Increase	0.73345
RSA	6.50		1.51	3.77		null	0.66	
SS			buried	buried				

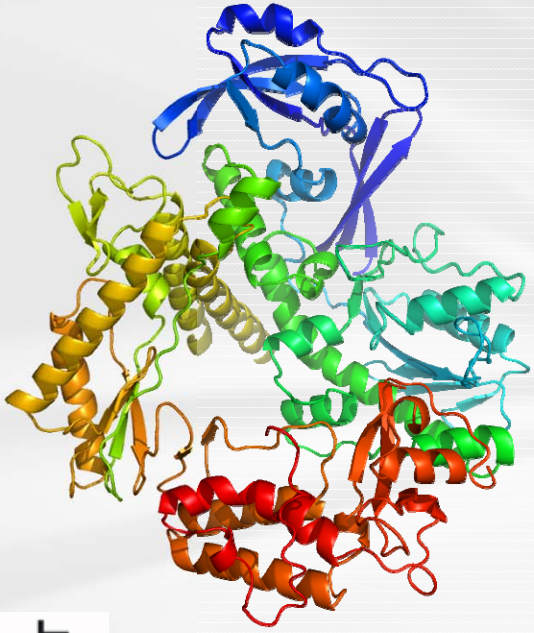


E. coli DsbA

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