



Les protéines intrinsèquement désordonnées

Véronique Receveur-Bréchet

IMR – CNRS, Marseille

7^e rencontres autour de la plateforme Bioinformatique
Gen'Ouest



Structural Biology

↗ structures in the pdb (structural genomics, etc....)



More complex systems

- ✓ Membran Proteins
- ✓ Macromolecular Complexes
- ✓ Flexible multi-domain Proteins
- ✓ **Intrinsically disordered proteins (IDPs)**

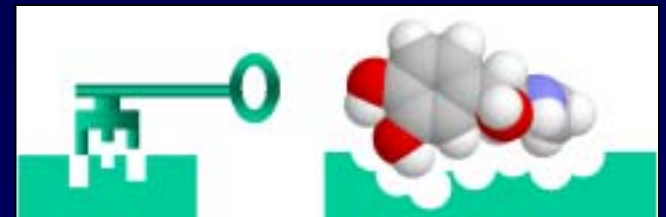
The Protein structure-function Paradigm

1 function ↔ 1 definite structure

Emil FISCHER (Nobel Prize in Chemistry, 1902) :



Lock and Key
Model

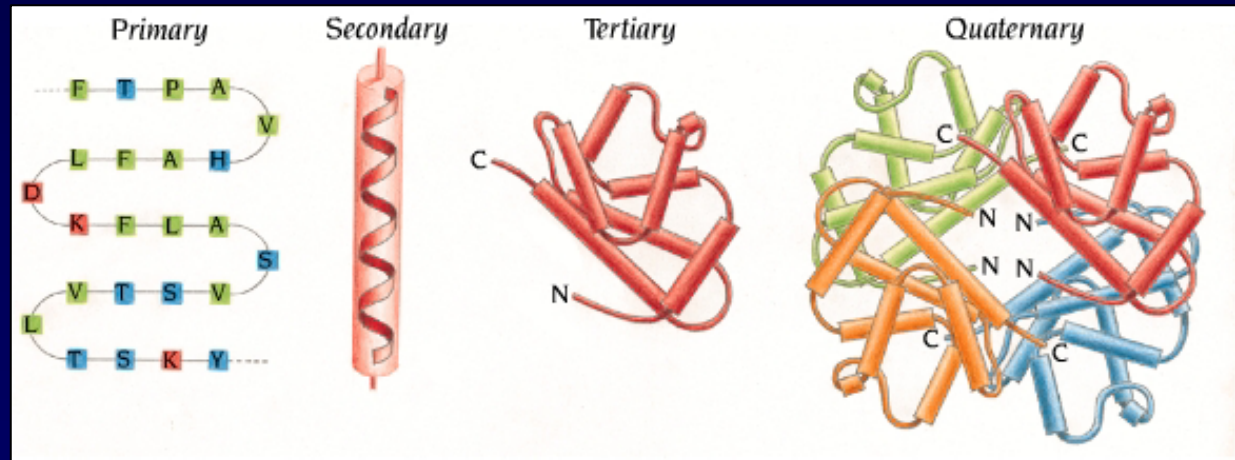


- Structure-function relationship studies
- Unstructured protein = denatured protein
- Drug Design

The Protein Folding paradigm

1 sequence \longleftrightarrow 1 definite structure

Christian ANFINSEN, Nobel Prize in Chemistry 1972



- Protein folding studies
- prediction of secondary structures
- 3D-modelling

The end of a Paradigm (1)

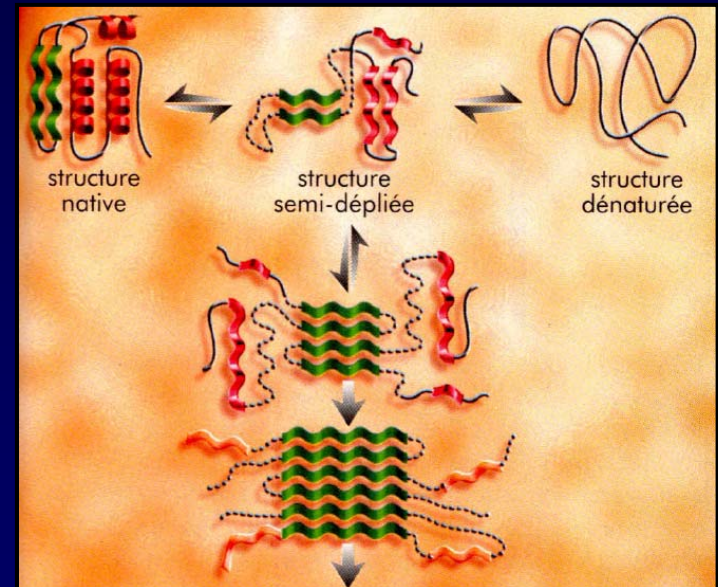
Amyloids

~~1 sequence~~

~~1 definite structure~~

Alternative amyloidogenic
misfolding

Conformational diseases



➔ The ability to form amyloids is a
generic property of proteins

The end of a Paradigm (2)

Intrinsically disordered proteins

~~1 function ↔ 1 definite structure~~

Proteins may be "unstructured" and functional

Article No. jmbi.1999.3110 available online at <http://www.idealibrary.com> on IDEAL® *J. Mol. Biol.* (1999) **293** 321–331

JMB



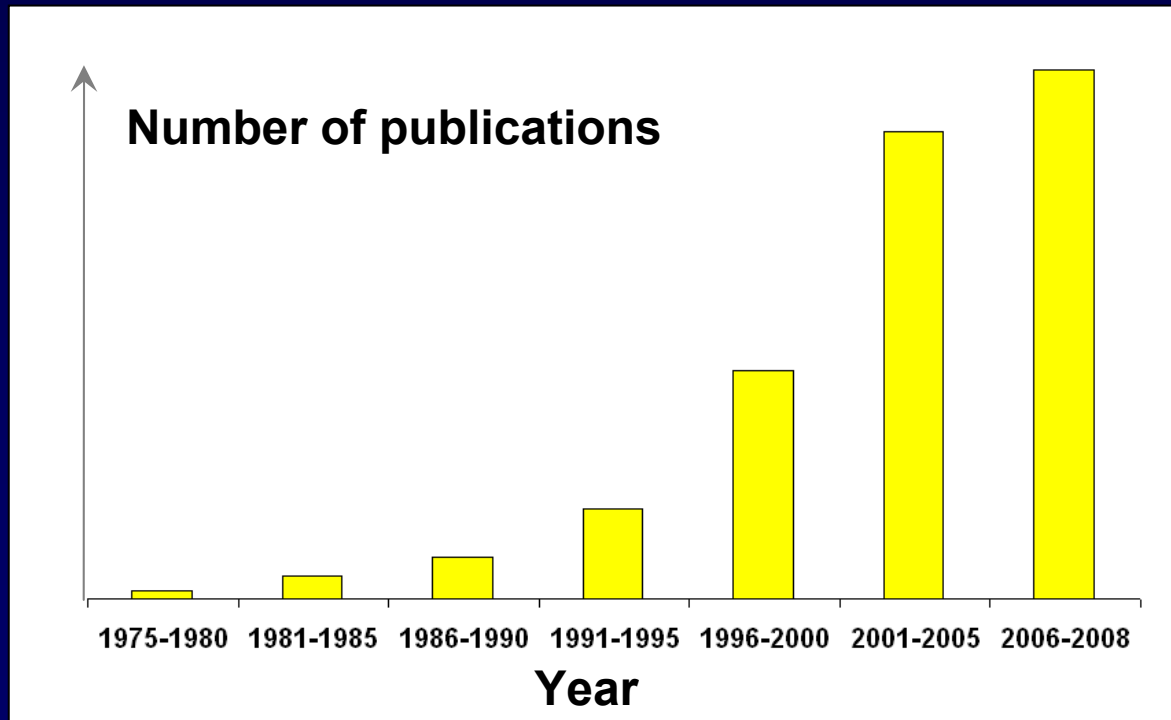
Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm

Peter E. Wright* and H. Jane Dyson*

Intrinsic disorder : A culture shock !

A quite novel concept: why?

- ascribed to “errors” and artefacts
- hard to conceive
- poorly informative



Abundance of disorder

Disorder is encoded by the aa sequence
(polar, charged aa ++, hydrophobic aa -)



Predictors of disorder



Data bank analyses



Abundance of disorder

Complete genomes

- Eubacteria : 4-5%
- Archaea : 2%
- Eucarya : 33% !

Virus ?

- ▶ 3% (0.8%) of *E. coli* proteins
- ▶ 31% (19%) of *S. cerevisiae* proteins
- ▶ 34% (19%) of *Arabidopsis thaliana* proteins
- ▶ 35% (22%) of *Homo sapiens* proteins
- ▶ 36% (22%) of *Drosophila melanogaster* proteins
- ▶ 0.5% (0.1%) pdb

Contain long
disordered regions
> 30 residues
(>50 residues)

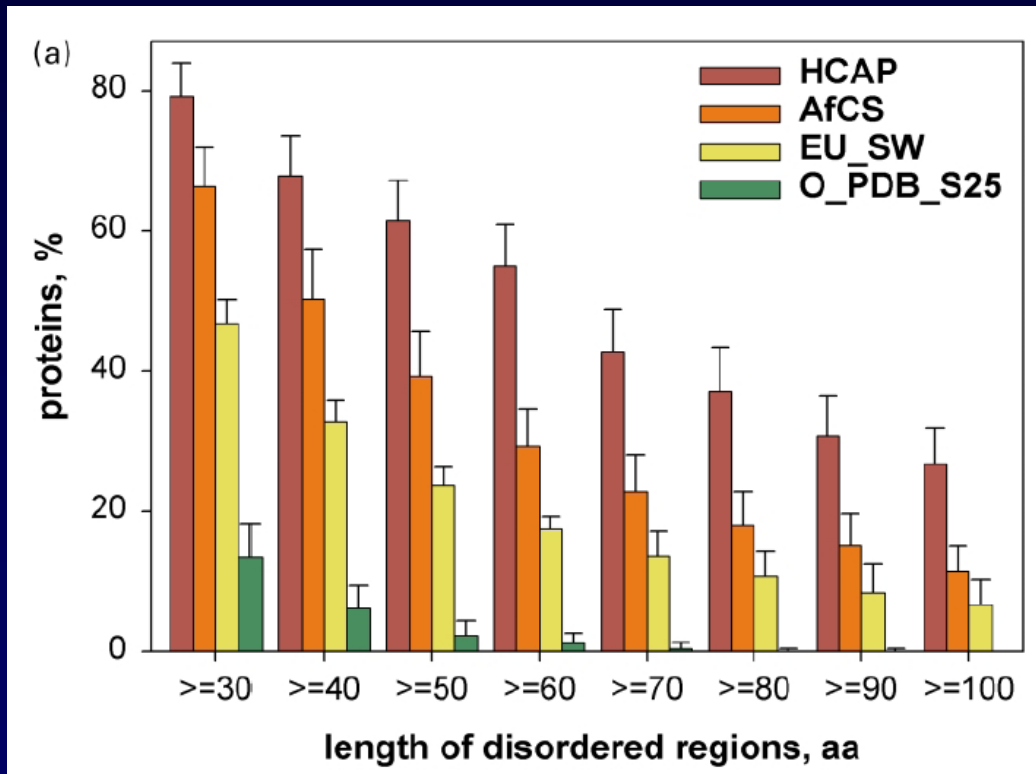
Which proteins are intrinsically disordered ?

- ▶ Multi-domain proteins : spacers, entropic springs...
- ▶ Proteins with multiple partners : Protein-protein interactions, DNA/RNA-binding, ligand binding : **Hubs**
- ▶ Proteins located in the nucleus



Signalling, cell cycle, regulation processes of transcription, translation, etc...

IDPs and Cancer



HCAP : Human Cancer associated Proteins

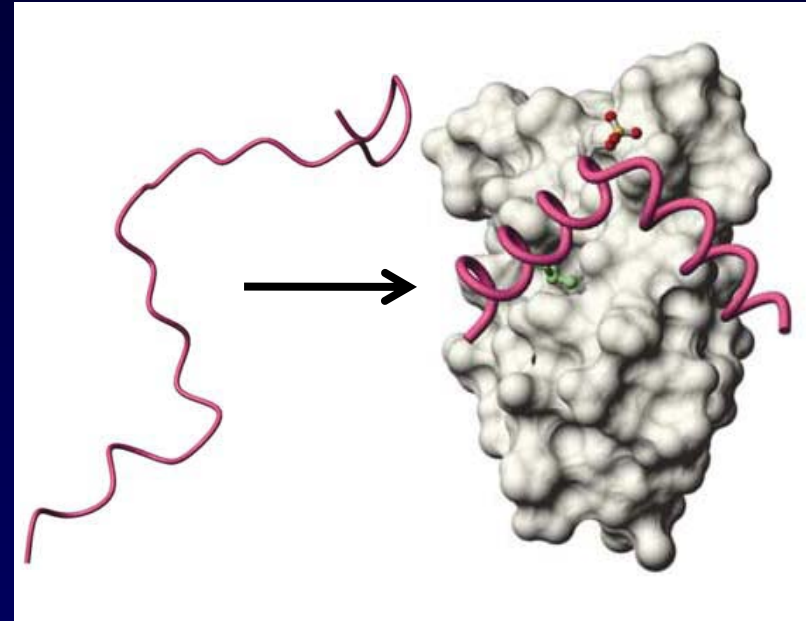
AfCS : signaling Proteins

EU-SW : Eukaryotic proteins (SwissProt)

p53, p21, p27, p57, c-myc, Mdm2, c-fos, Bcl-2, etc...

Functional benefits of disorder

- ▶ Plasticity
- ▶ Interaction with multiple partners
- ▶ Induced folding upon binding
- ▶ Large molecular interface
- ▶ Genome compaction



A unique molecular recognition mechanism !

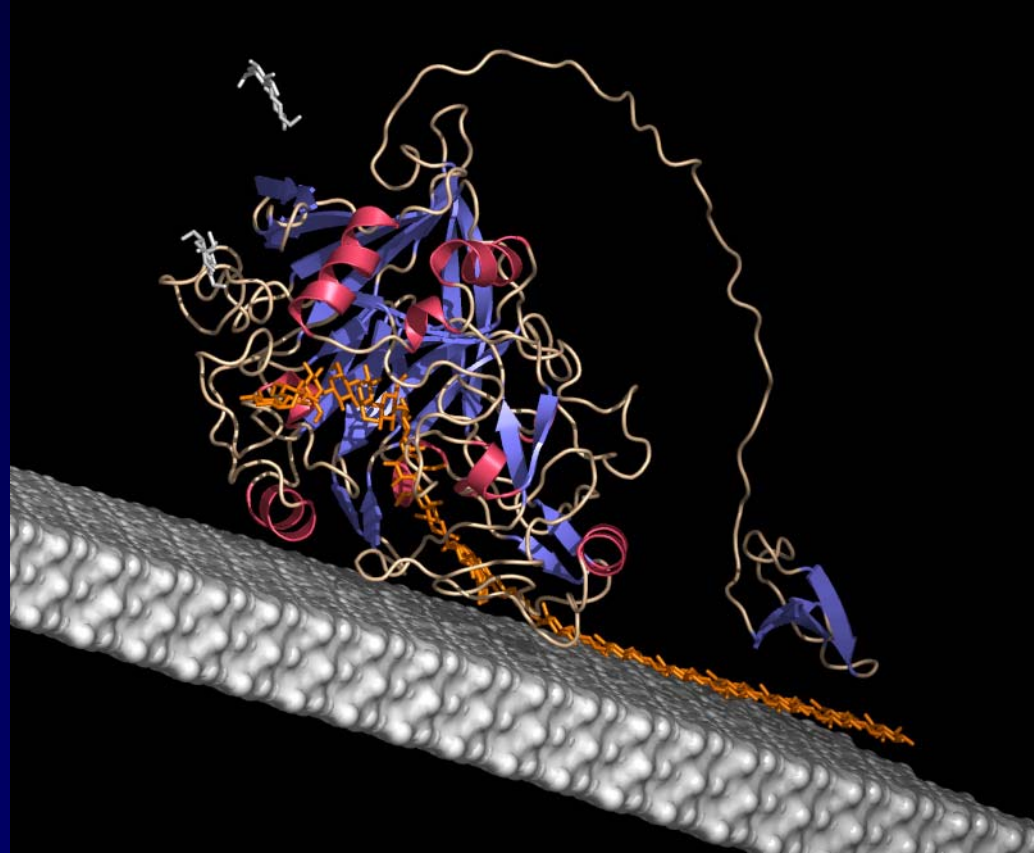
Linkers and Entropic springs

Cellulases

Catalytic domain

Cellulose binding domain

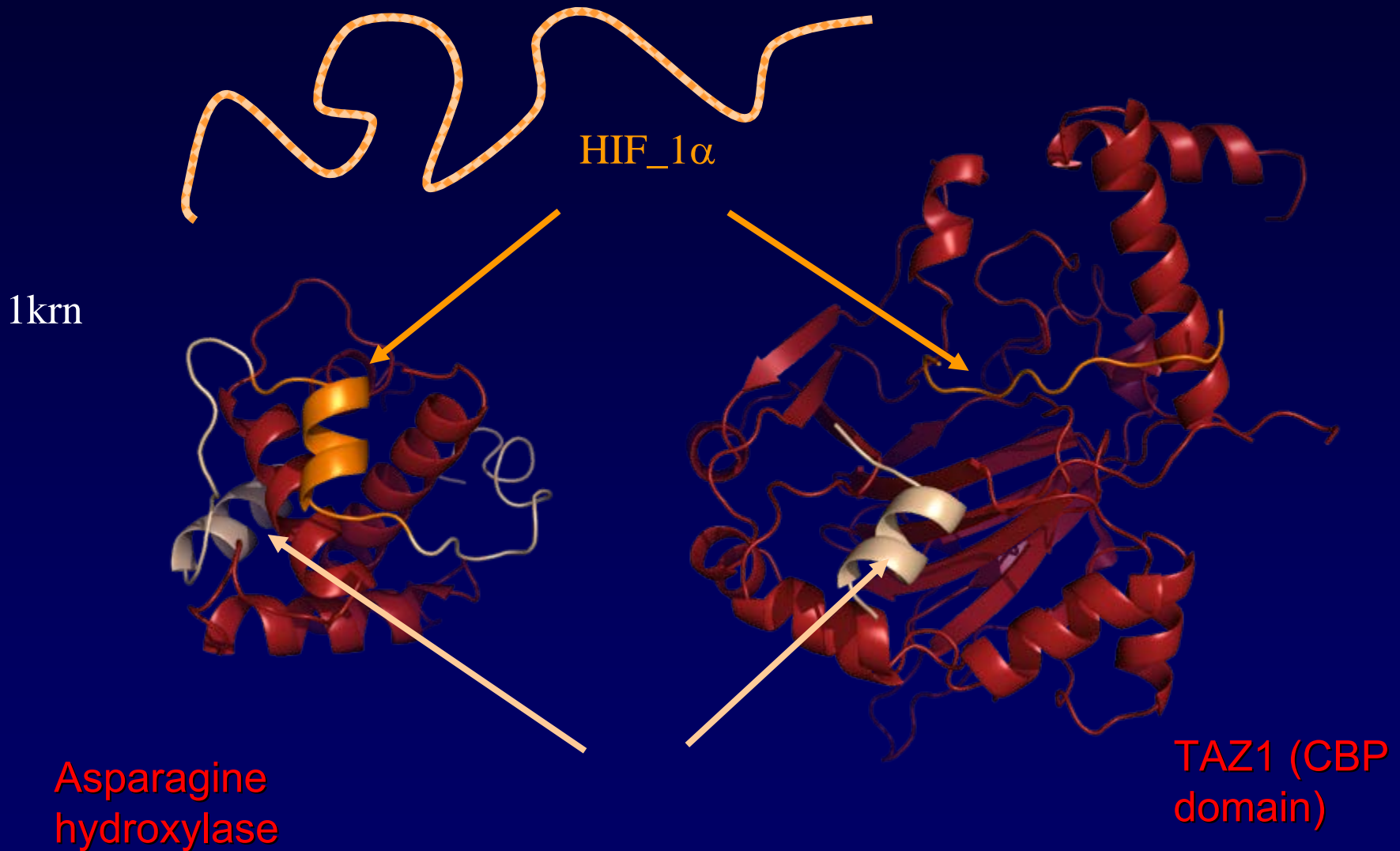
Crystalline cellulose



Courtesy of C. Divne, Upsala University, Sweden

Receveur et al (2003) *JBC* 277: 40887
Von Ossowski et al (2005) *JMB* 88:2823
Violot et al (2005) *JMB* 348:1211

Induced folding



Dyson & Wright (2005)
Mol. Cell. Biol.

Intrinsic disordered proteins

- **No definite 3D structure**
- **No stable secondary and tertiary structure**
(at physiological pH and salinity, in absence of partner or ligand)
- **Disorder required for function**

Sequence specificities of disordered proteins

- **Biased Composition** (rich in polar and charged residues)
- **Low complexity**
- **High variability among related proteins** (weak evolution constraints)
- **Low content in predicted secondary structure**

Prediction of disorder

PONDR <http://www.pondr.com>

Charge/hydropathy method <http://www.pondr.com>

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

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

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

Hydrophobic cluster analysis (HCA)

<http://smi.snv.jussieu.fr/hca/hca-seq.html>

DISOPRED <http://bioinfo.cs.ucl.ac.uk/disopred>

NORSp <http://cubic.bioc.columbia.edu/services/NORSp>

IUPRED <http://iupred.enzim.hu>

RONN <http://www.strubi.ox.ac.uk/RONN>

Métaserveur : MEDOR (AFMB)

DisProt: data base of Disordered Proteins (ca 200 proteins)
<http://www.disprot.org>

Utilisation de prédicteurs reposant sur des paramètres différents

Philosophies différentes :

- prédicteurs *entraînés* sur datasets (Pondr, RONN, etc.)
- prédicteurs *à priori* (IUPred, charge/hydrophobie)

Paramètres physico-chimiques différents

Hydrophobicité, arômaticité, contenu en Cys/Trp, etc.

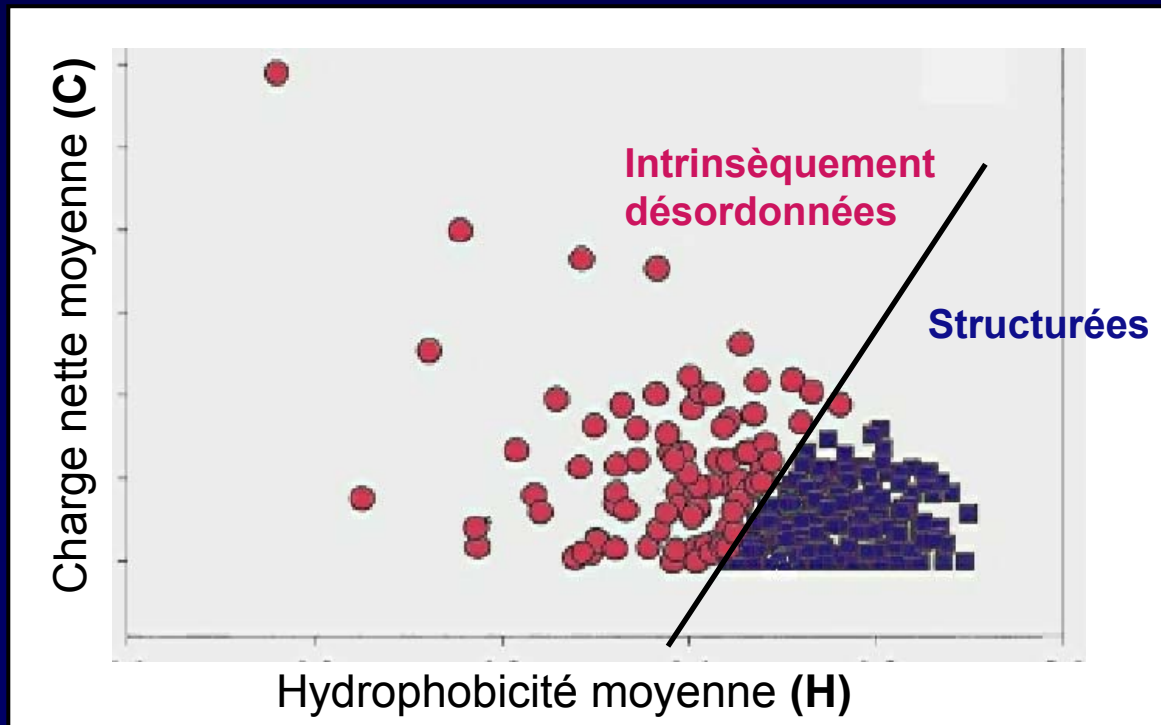
Concepts différents

Analyse de la complexité des séquences (SEG)

→ Indispensable d'utiliser des prédicteurs différents (et complémentaires !)

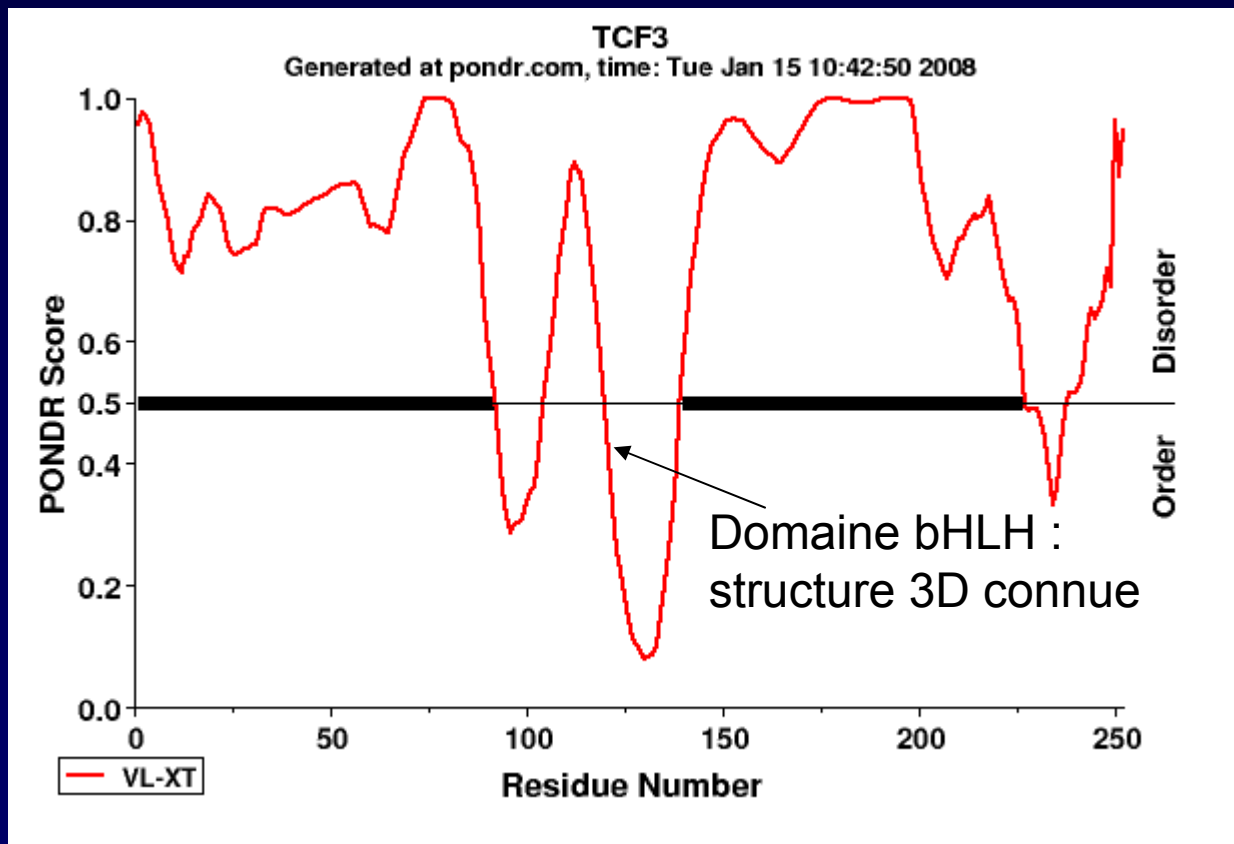
Analyse globale de la séquence

- Résidus promoteurs d'ordre : WFY VL I M
- Résidus promoteurs de désordre : ST P DEQ KR
- Basse complexité de séquence (motifs répétés)
- Hydrophobicité / Hydrophilie



Prédicteurs entraînés PONDR (Predictor Of Naturally Disordered Regions)

- Long Disordered Regions (LDR) > 30 résidus

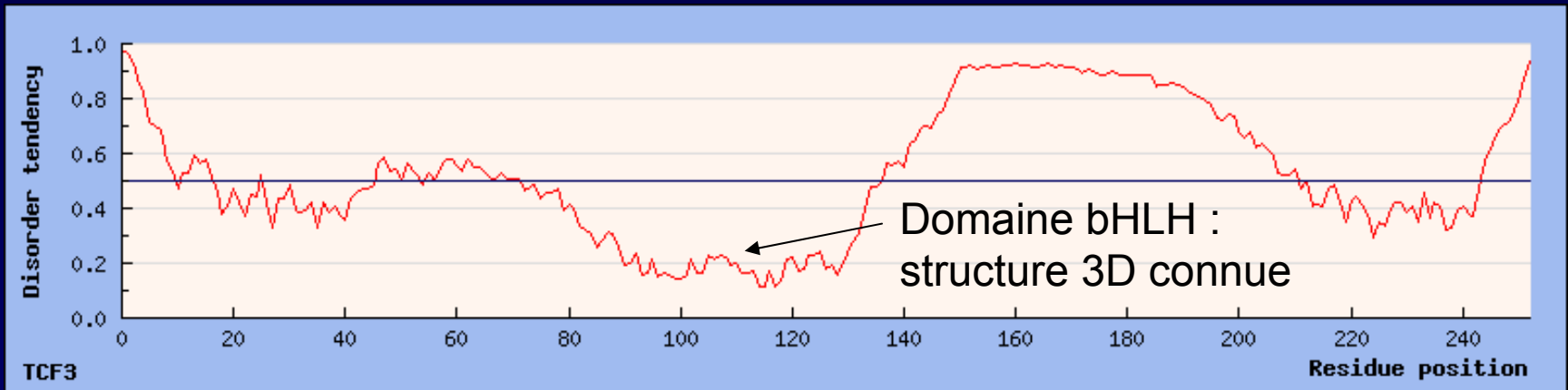


Prédicteurs ab initio

IUPRED

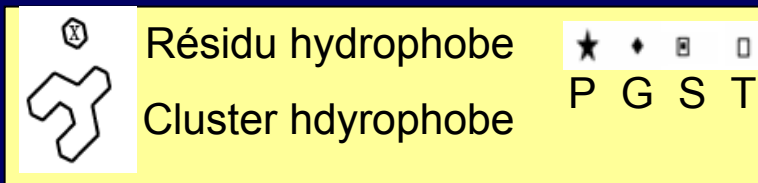
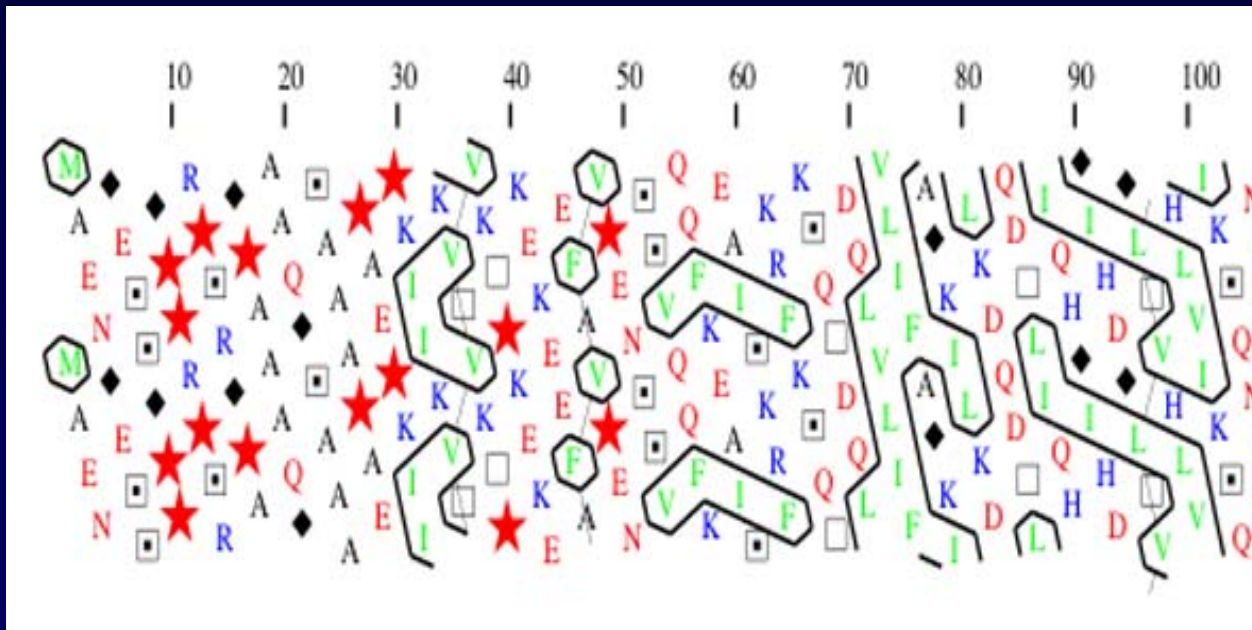
(Prediction of Intrinsically Unstructured Proteins)

- Fenêtre de 20 résidus
- Énergie d'interaction entre résidus proches dans la séquence

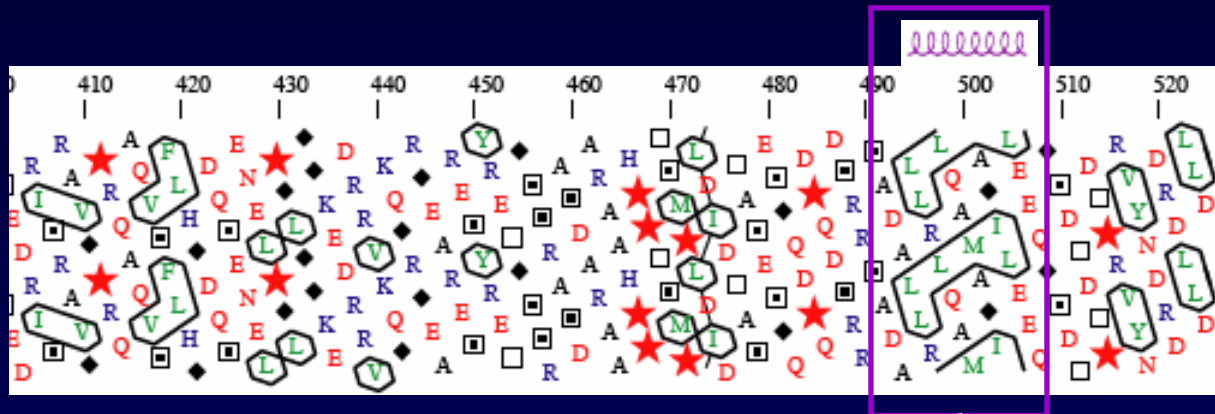


```
MELEQREGTMAAVGFEEFSAPPGSELALPPLFGGHILESELETEVEFVSGGLGGSGLRERDEEEEAARGRRR  
QRELNRRKYQALGRRCREIEQVNERVLNRLHQVQRITRRLQQERRFLMRVLDYGGDDYRASQFTIVLEDEGSQGTDA  
PTPGNAENEPPEKETLSPPRRTPAPPEPGSPAPGEGPSGRKRRRVPRDGRRAGNALTPELAPVQIKVEEDFGFEADEA  
LDSSWVSRGPKLLPYPTLASPASD
```

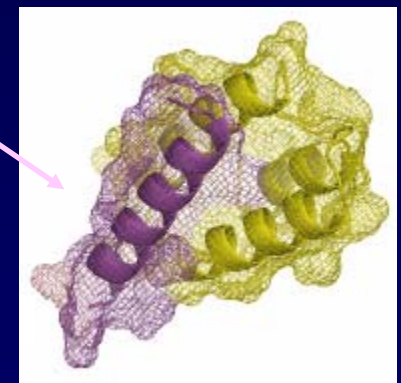
Analyse fine Plots HCA (Hydrophobic Cluster Analysis)



Prédiction du repliement induit



Cluster hydrophobe associé à une α -hélice au milieu d'une région désordonnée



→ repliement en α -hélice de cette région induit par la liaison à son partenaire

Pour en savoir plus...



PROTEINS: Structure, Function, and Bioinformatics 65:1–14 (2006)

REVIEW

A Practical Overview of Protein Disorder Prediction Methods

François Ferron,^{1,2} Sonia Longhi,^{1*} Bruno Canard,¹ and David Karlin³

¹*Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS et Universités Aix-Marseille I et II, Marseille, France*

²*Boston Biomedical Research Institute, Watertown, Massachusetts 02472*

³*Ecole de l'ADN, INMED, Marseille, France*

Current Protein and Peptide Science, 2007, 8, 000-000

Prediction of Protein Disorder at the Domain Level

Zsuzsanna Dosztányi, Márk Sándor, Peter Tompa and István Simon*

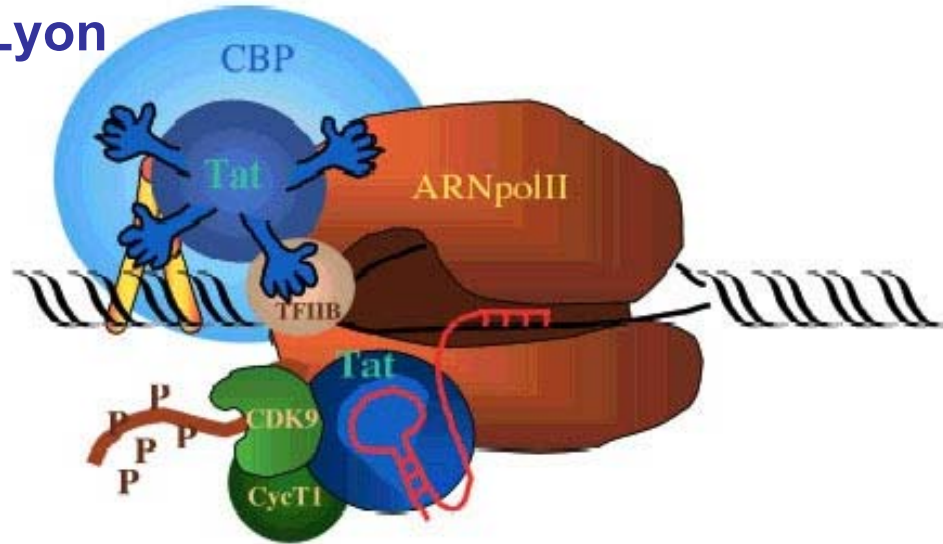
FINAL

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, Budapest, Hungary, 1518 Budapest, PO Box 7, Hungary

HIV-1 - Tat function



Coll. P.Gouet, IBCP, Lyon



- ✓ Transactivator of transcription
- ✓ Interaction with TAR via cellular Proteins
- ✓ Stimulation of the production of HIV virions
- ✓ Decrease of the antiviral immunity

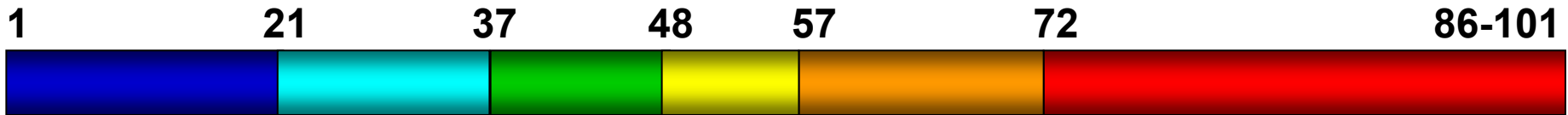
Excellent Target for antiviral therapy and Drug design

HIV-1 Tat Sequence



Exon 1

Exon 2



Acidic and Pro rich domain

Cys rich domain

Core domain

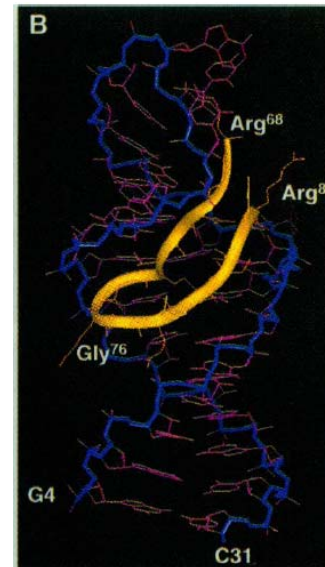
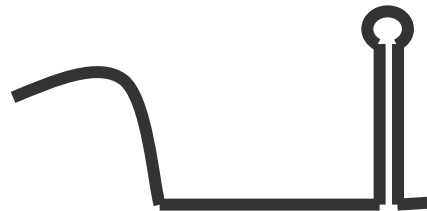
Basic domain

Gln rich domain

C-terminal domain

Interaction with Cyclin T1

TAR binding



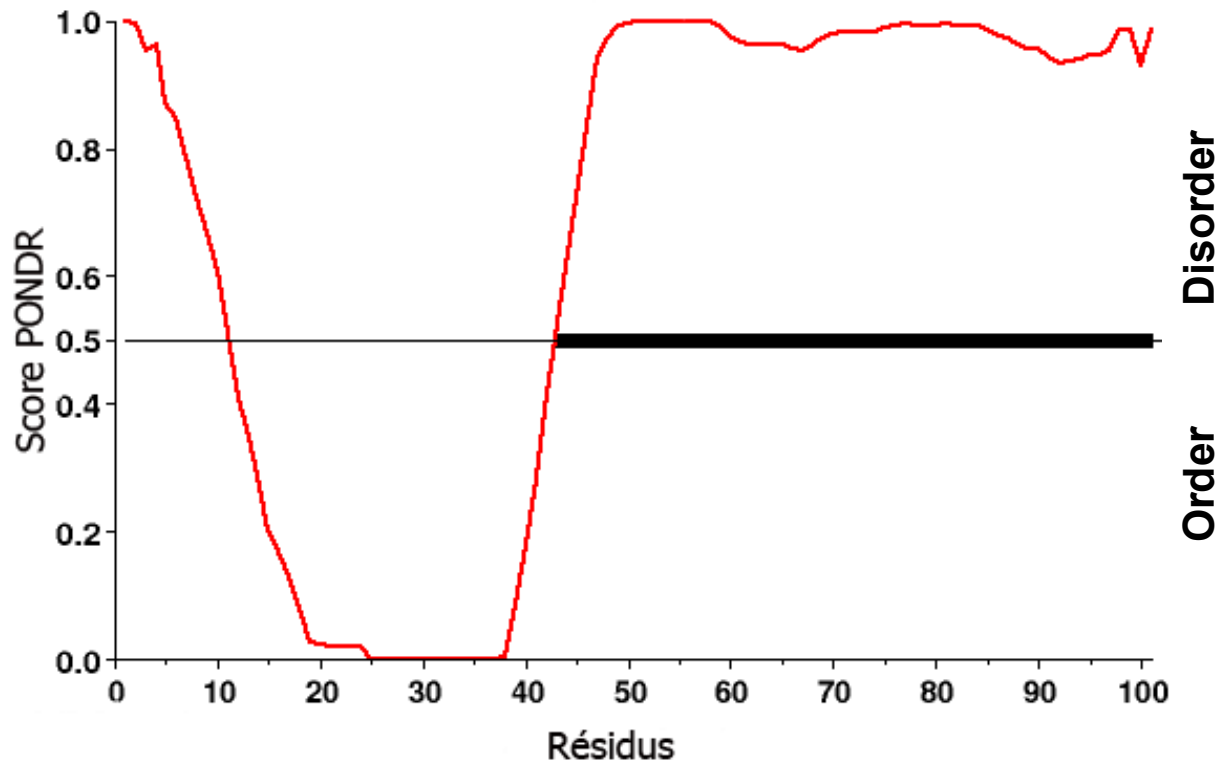
Cell adhesion

Puglisi et al (1995)

Prediction of disordered regions

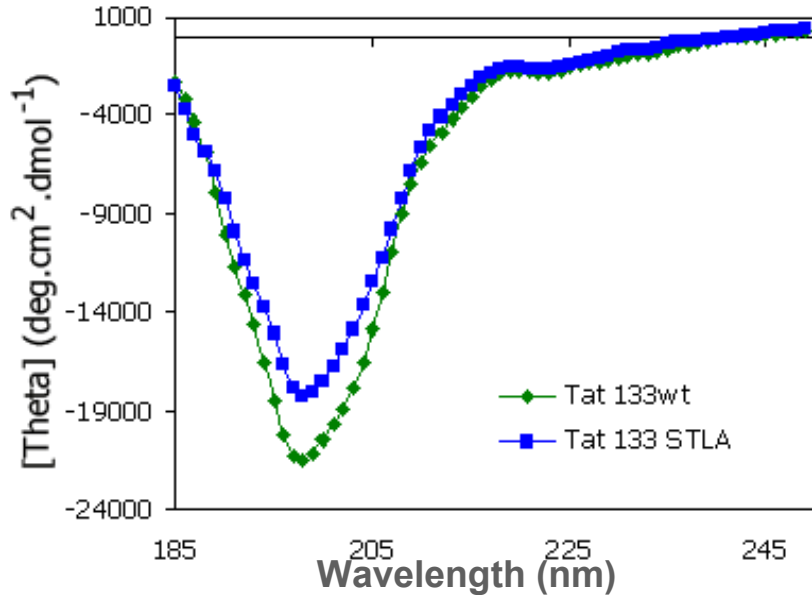


✓ DisEMBL (<http://dis.embl.de>), GlobPlot (<http://globplot.embl.de>) et PONDR® (<http://pondr.com>)



Tat is predicted as disordered in solution, except in the Cys-rich domain/core domain (aa 22 to 45) ?

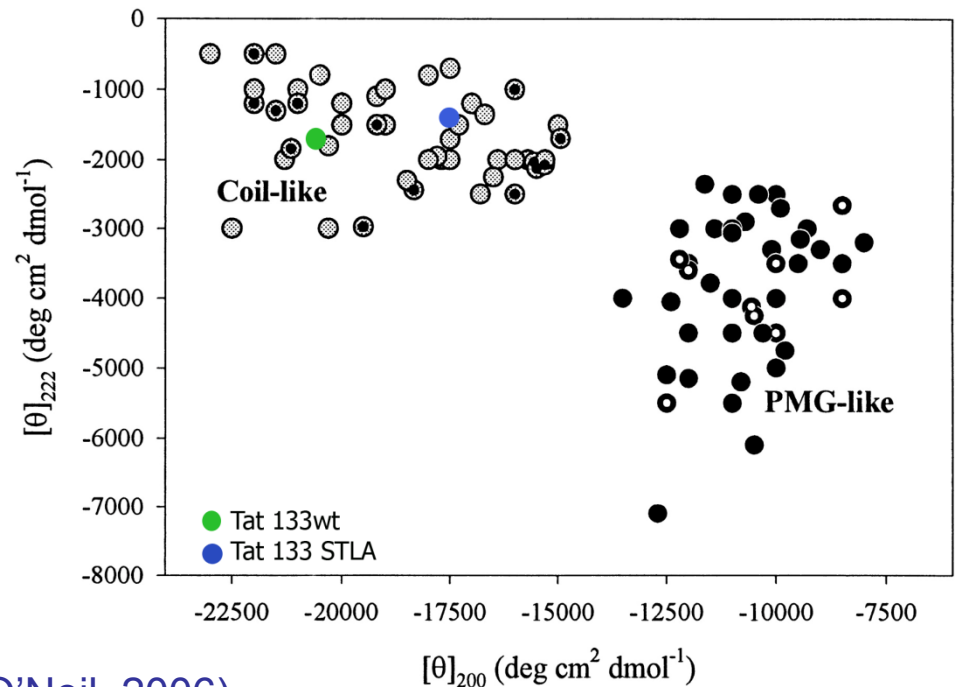
Far UV Circular Dichroism



**Tat lacks stable
Secondary Structure**

Tat is a Random coil

(Uversky, Proteins, 2002)



✓ NMR Structure and Dynamic (Shojania & O'Neil, 2006)

Small Angle X-ray Scattering



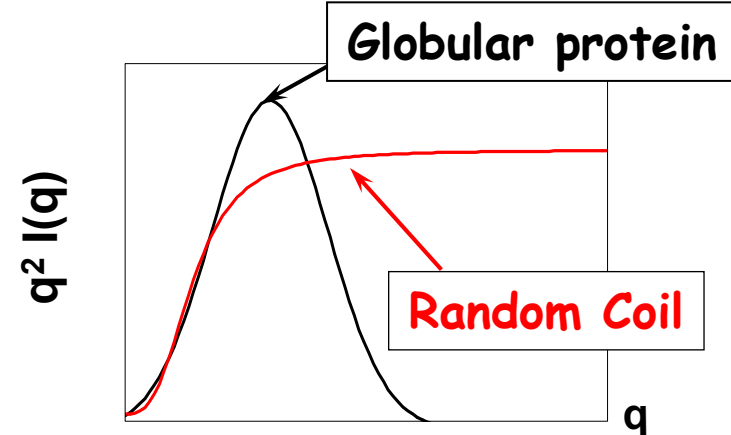
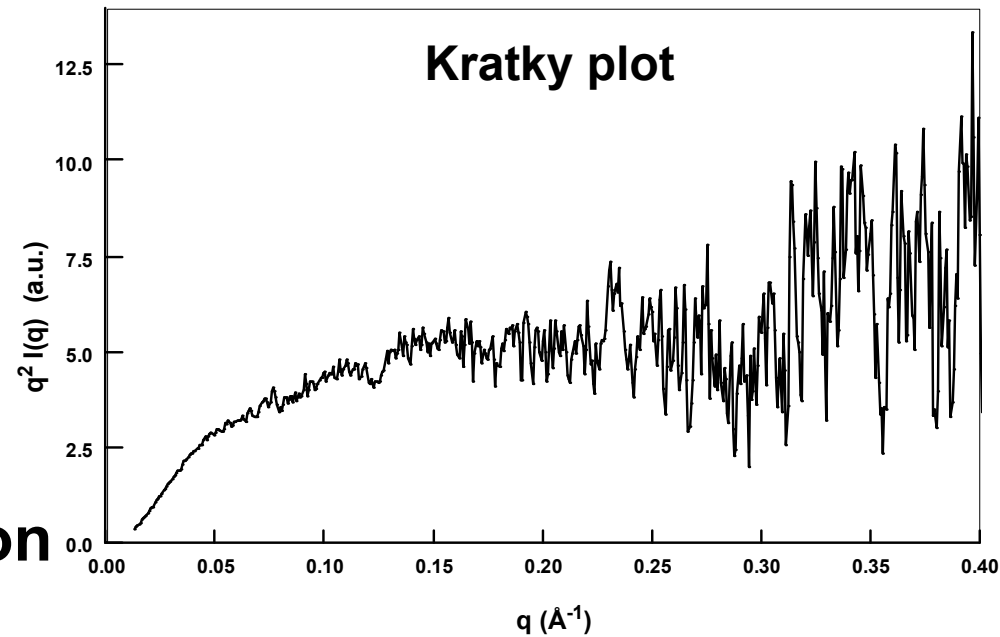
Guinier :

$$R_g(0) = 33.0 \pm 1.5 \text{ \AA}$$

$$M_w = 11,1 \text{ kDa (monomer)}$$

Distance Distribution Function

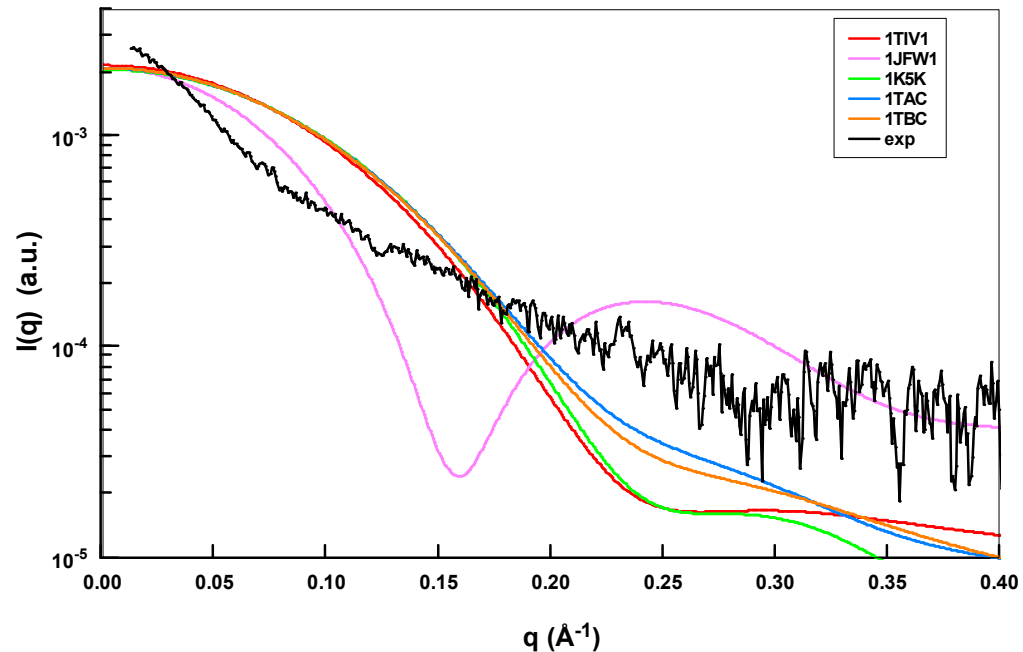
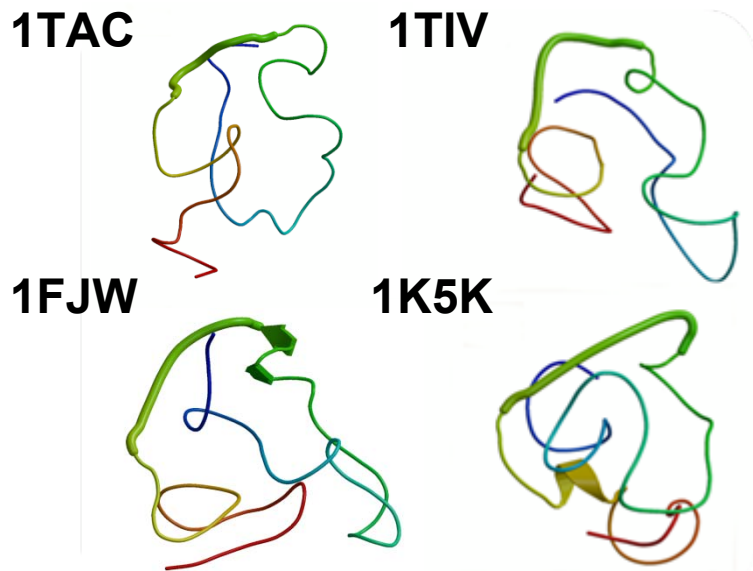
$$D_{\max} = 110-120 \text{ \AA}$$



Random coil (101 aa) : $R_g \sim 32-34 \text{ \AA}$
(Millett *et al.*, 2002)

HIV-1 Tat is strongly disordered and extended

SAXS vs NMR structures



Program CRY SOL

NMR Structures : $R_g \sim 12 \text{ \AA}$ $\chi^2 \sim 12$

SAXS data : $R_g = 33.0 \pm 1.5 \text{ \AA}$

Conclusions

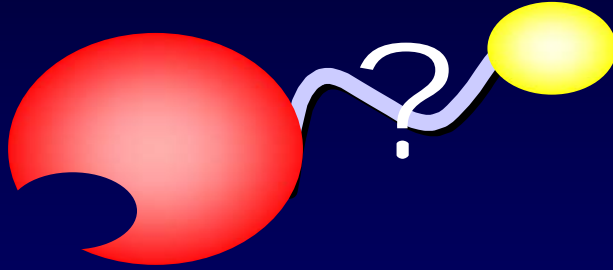


HIV-1 Tat is a Random Coil

- ✓ Induced folding (β -turn or α -helix) upon binding to TAR ?
- ✓ Induced folding upon binding to its other partners (Cyclin T1, etc.) ?
- ✓ Among Intrinsically Disordered Proteins, there probably exists a continuum of possible “disorder”
- ✓ Residual structures are not a prerequisite for induced folding

**The interactions between IUPs and their partners
can be very complex**

Cellulases



Pseudoaltermonas haloplanktis

Psychrophilic bacteria from Antarctic



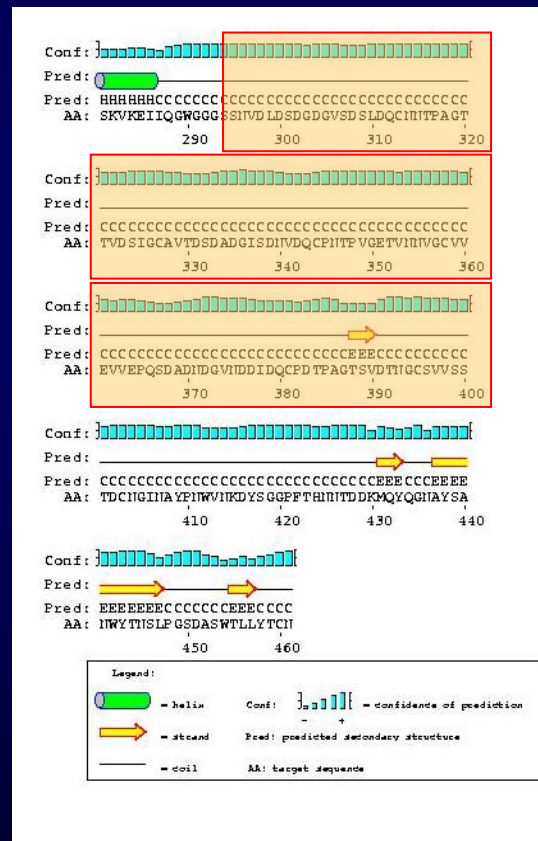
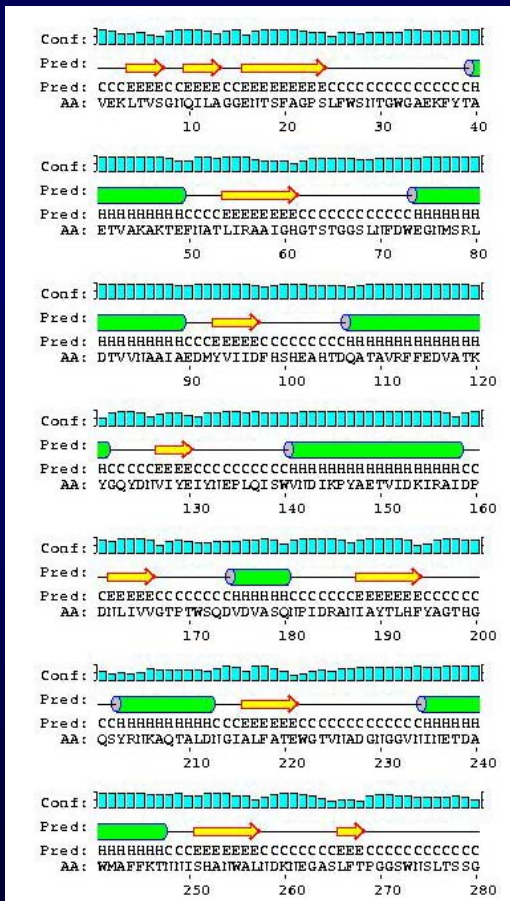
- ◆ Family 5 : Endoglucanase Cel5G
 - Non-glycosylated linker
 - Very long linker : 109 residues

➔ No cold adaptation specificity revealed by the crystal structure of the catalytic domain (66% identity with mesophilic *E. chrysanthemii*)

Analyse de la séquence de Cel5G

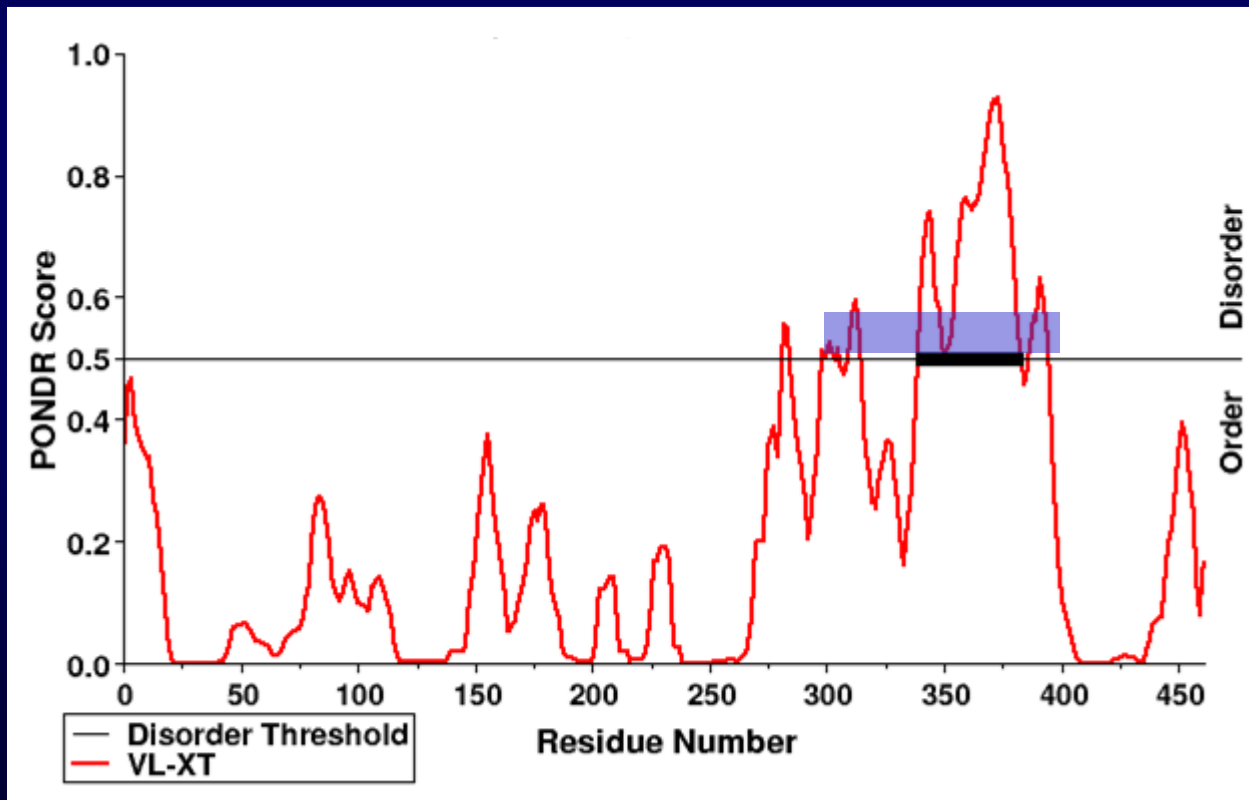
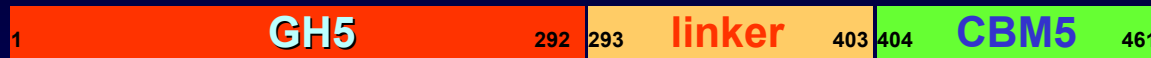


Prédiction de structure secondaire (PSI-PRED)

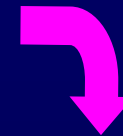


Le linker est prédit random coil

PONDR (Prediction of Natural Disordered Regions)



Long disordered region :
residues 339-382

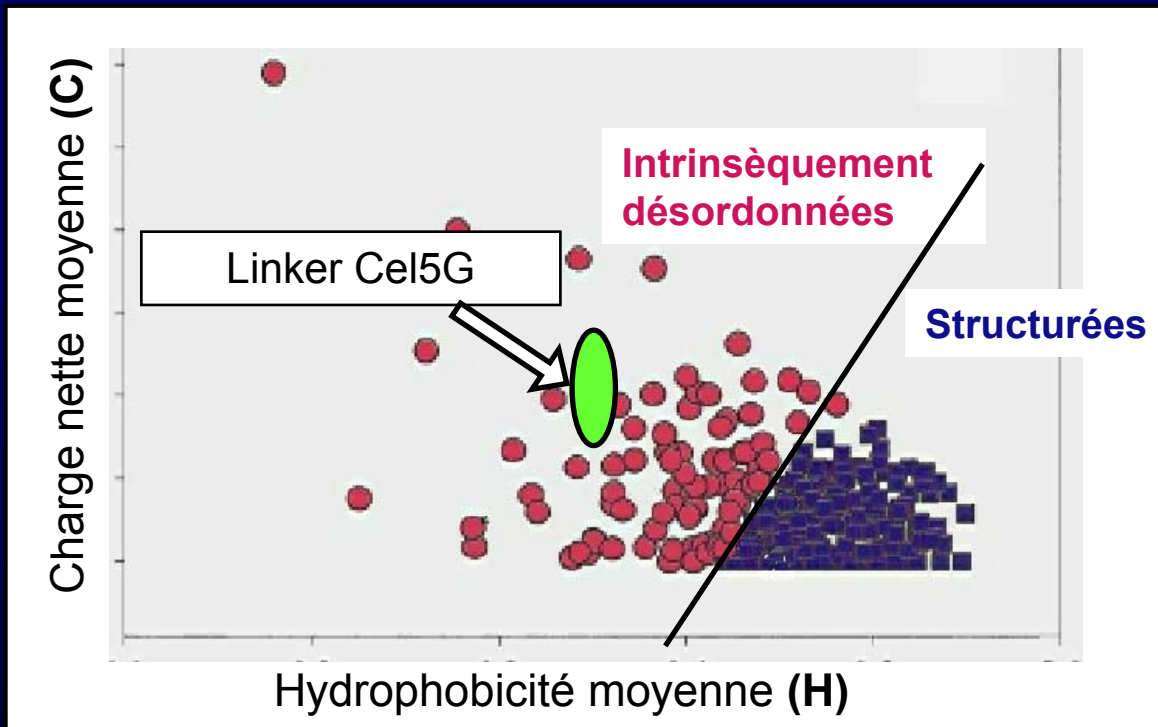


Le linker est prédit
comme
intrinsèquement
désordonné

Composition en acides aminés du linker

Propriétés de séquence des protéines intrinsèquement désordonnées

Pour se replier, une protéine a besoin d'un cœur hydrophobe, et de répulsions électrostatiques pas trop élevées



Le linker est prédit comme intrinsèquement désordonné

Pseudoaltermonas haloplanktis Cel5G

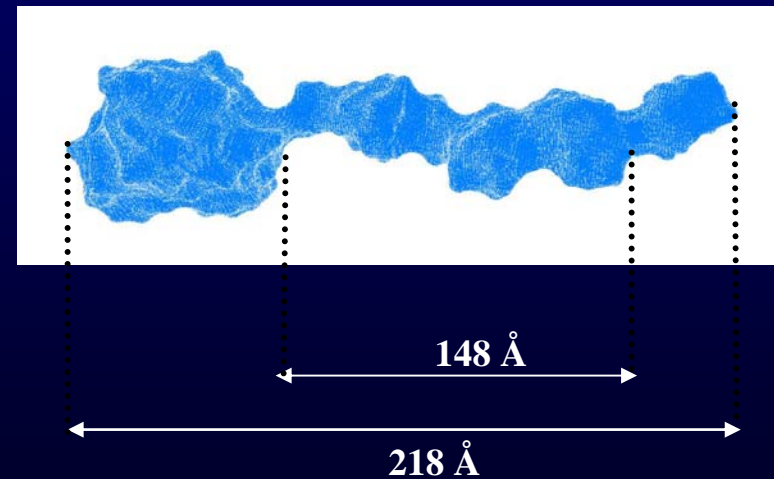
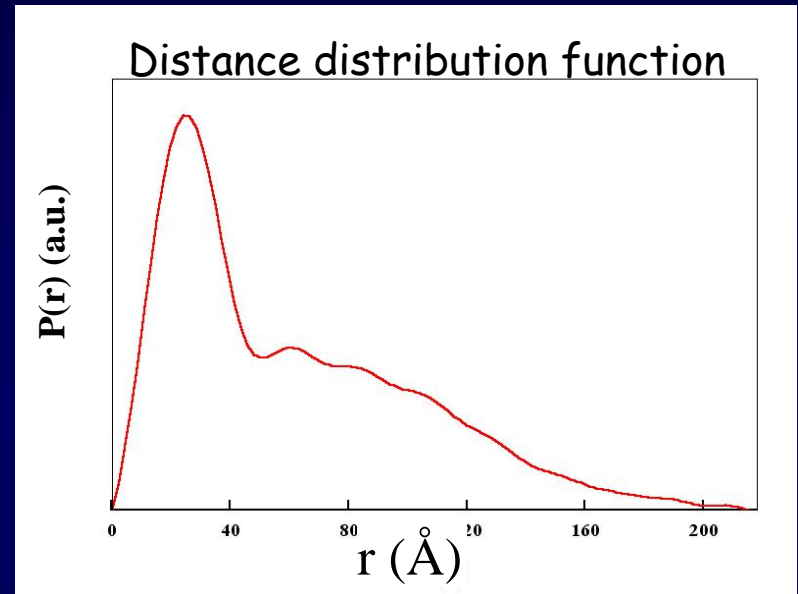
$M_w = 53$ kDa $N = 292+109+61$ aa

$$R_g = 53.2 \pm 1.3 \text{ \AA}$$

$$D_{\max} = 218 \pm 2 \text{ \AA}$$

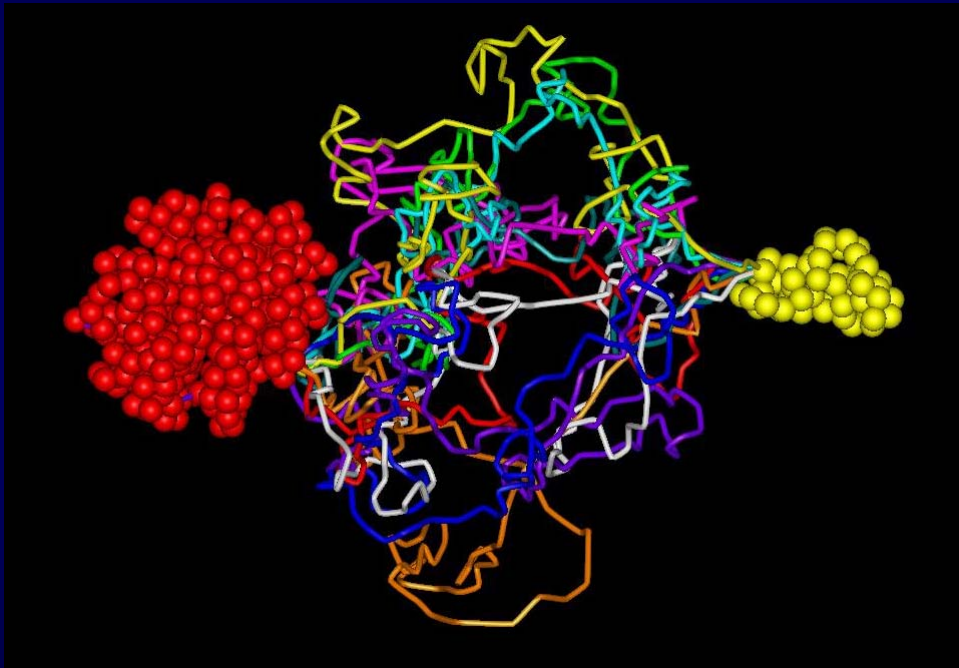
GGGSSNVDLSDGDGVSDSLDQCNNTPAGTT
VDSIGCAVTDSADGISDNVDQCPNTPVGETV
NNVGCVEVVEPQSDADNDGVNDDIDQCPDT
PAGTSVDTNGCSVVSST

 Linker : ≈ 0.75 residue/ \AA ?



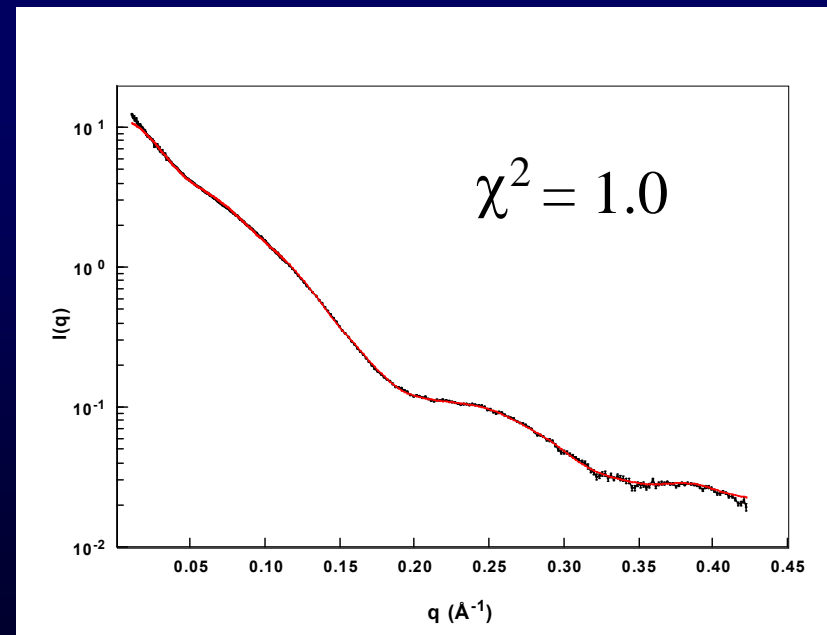
Conformation of the linker between the globular domains

è Programs CREDO + TURBO



Flexibility of the linker

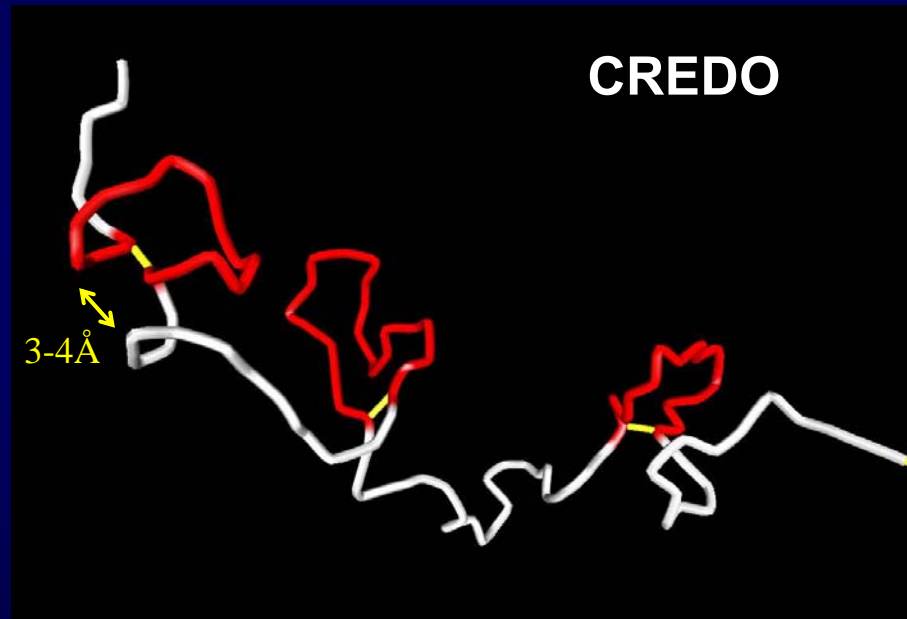
Several conformations



Loops in the linker

GGGSSNVDLSDGDGVSDSLDQ **CN**NTPAGTTVDSIG**C**AVTDSDADGISDNVDQ

CPNTPVGETVNNVG**C**VVEVVEPQSDADNDGVNDDIDQ **C**PDTPAGTSVDTNG**C**SVVSST



↪ Linker : ≈ 0.48 residue/ Å

Acknowledgement



Cellulases

- **Mirjam CZJZEK**
- **Georges FELLER**
- **Charles GERDAY**
- **Nushin AGHAJARI**
- **Richard HASER**

Station Biologique, Roscoff
University of Liège, Belgium
University of Liège, Belgium
IBCP, Lyon, France
IBCP, Lyon, France

HIV TAT

- **Patrice GOUET**

IBCP, Lyon, France

Synchrotrons

- **Pierre PANINE, Stéphanie FINET** ESRF, Grenoble, France
- **Dmitri SVERGUN, Manfred ROESSLE, Efstratios MYLONAS**
DESY, Hamburg, Germany