



Gilbert Deléage

Professeur de bioinformatique

Université Claude Bernard  Lyon 1



IBCP-CNRS

7, passage du Vercors

69367 Lyon cedex 07

Tél: +33 (0)4 -72-72-26-55

fax: +33 (0)4 -72-72-26 -01

mel: g.deleage@ibcp.fr

<http://www.gdeleage.fr/prof/BIS.pdf>

© Ce cours ne peut être reproduit ni diffusé sans le consentement de l'auteur

X-ray crystallography	NMR spectroscopy	2D & single particle electron microscopy	electron tomography	immuno-electron microscopy	chemical cross-linking	affinity purification mass spectroscopy
subunit structure	subunit structure				subunit structure	
subunit shape	subunit shape	subunit shape	subunit shape		subunit-subunit contact	subunit-subunit contact
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact		subunit proximity	subunit proximity
subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity		
subunit stoichiometry	subunit stoichiometry			assembly symmetry		
assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry			
assembly shape	assembly shape	assembly shape	assembly shape			
assembly structure	assembly structure					

FRET	site-directed mutagenesis	yeast two-hybrid system	gene/protein arrays	protein structure prediction	computational docking	bioinformatics
				subunit structure		
				subunit shape		
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact		subunit-subunit contact	subunit-subunit contact
subunit proximity		subunit proximity	subunit proximity			

(d'après A. Sali)

- ✓ Introduction générale – Rappels – Connexion avec le cours de licence
- ✓ Relation séquence, des structures secondaires, structure 3D
- ✓ Les bases de données structurales
 - ❖ PDB (comment interroger?)
 - ❖ CATH
 - ❖ SCOP
 - ❖ HSSP
 - ❖ MOLMOV
- ✓ Analyse des structures de macromolécules biologiques
- ✓ Outils avancés de visualisation
 - ❖ Démonstrations Rasmol, PyMol, Discovery Studio, VMD, Deep view
 - ❖ Avantages/inconvénients des outils
- ✓ Comparaison /superposition de structures 3D
 - ❖ RMSD local,
 - ❖ RMSD global
 - ❖ Outils (Dali, CE, superposition multiple)
- ✓ Energétique moléculaire
 - ❖ Mécanique moléculaire
 - ❖ Éléments des champs de force
 - ❖ Minimisation d'énergie
 - ❖ Dynamique moléculaire
- ✓ Modélisation moléculaire
 - ❖ Méthodes (Enfilages, homologie, alphabets structuraux, *ab initio*)
 - ❖ Outils (Modeller, SwissModel, Geno3D)
- ✓ Mise en œuvre - Choix de stratégies
 - ❖ Recherche d'empreinte(s) - Apport des prédictions de structures secondaires
 - ❖ Alignement de la séquence d'intérêt avec le(s) empreinte(s)
 - ❖ Méthodes de construction (Progressive ou globale par géométrie des distances)
 - ❖ Validation – comparaison des modèles
- ✓ Cas pratiques
 - ✓ Modélisations de capsides virales, GP41, PGP
- ✓ Interactions protéines-ligands
- ✓ Conclusions

Modélisation Moléculaire = regroupement de différentes techniques pour



Construire



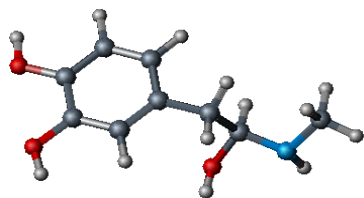
Visualiser



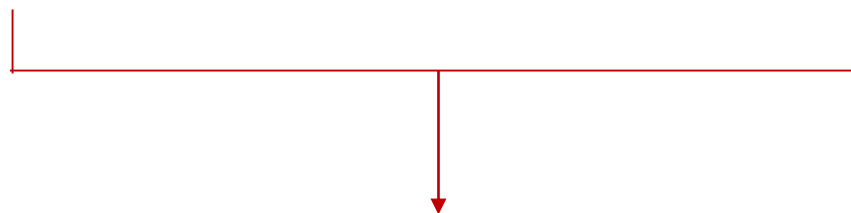
Manipuler/calculer



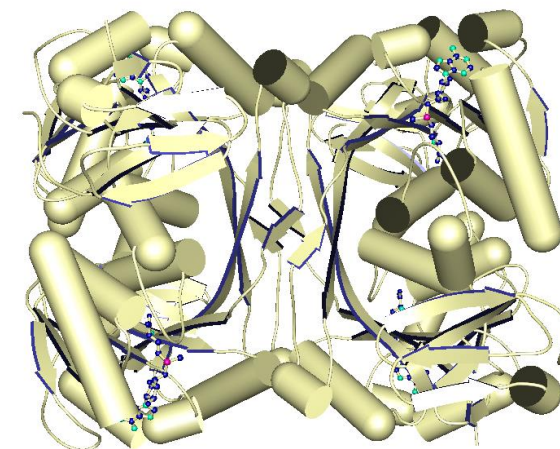
Analyser-comprendre

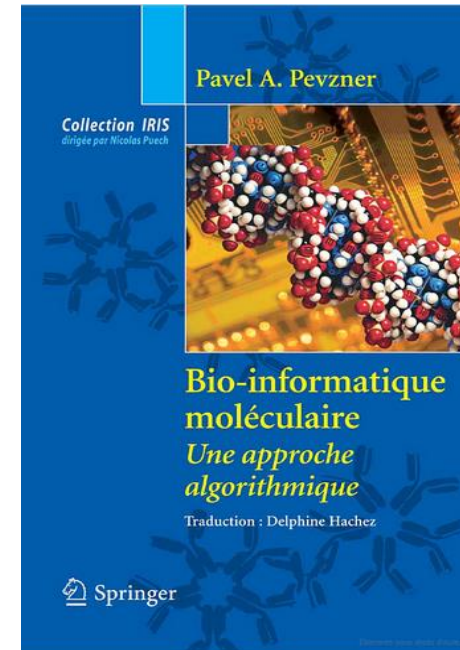
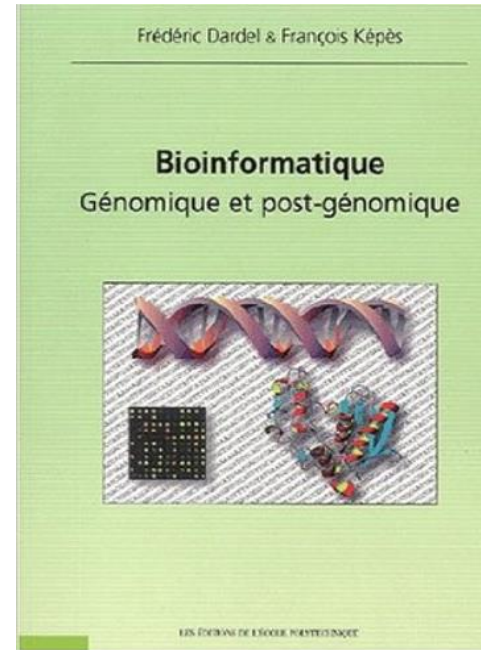
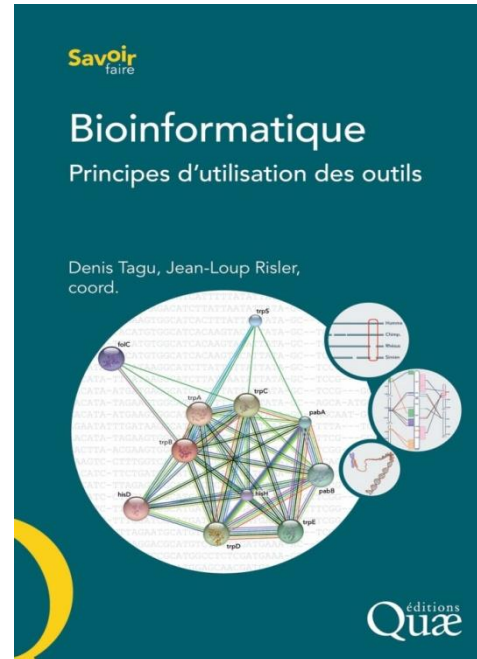
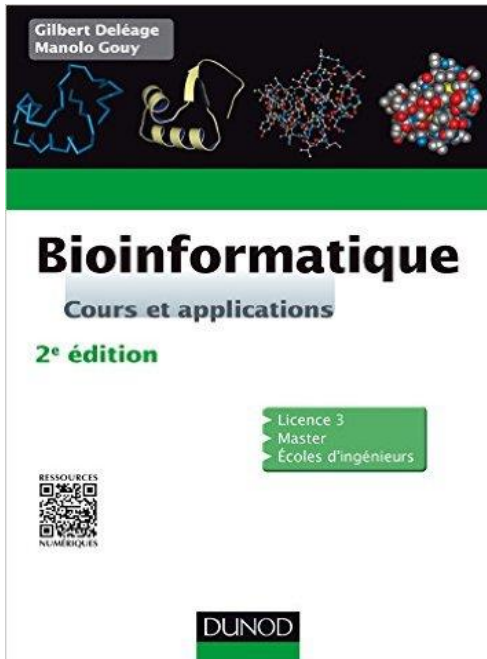


calcul



graphisme





Avant la **Bio-informatique** (=>1990)

Activité
Biologique
connue

Etude
Biochimique
Structure 3D

Séquence
Protéine

Gène
Mutagénèse

Relations
structure-
activité

BIO-INFORMATIQUE

Banques de données
Prédiction des gènes

Identification de protéines
Prédiction sites/signatures
Prédiction de structure
Modélisation moléculaire

Stockage
Classification
Intégration
Criblage

Séquences
génomiques

Séquences
Protéiques

Prédiction
Activités
biologiques

Etudes
Biochimiques
Structures 3D

Protéomique

Génomique
structurale

Aujourd'hui (depuis les programmes de
séquençages massifs et la **Bio-informatique**)

A	Alanine	Ala
C	Cysteine	Cys
D	Aspartic Acid	Asp
E	Glutamic Acid	Glu
F	Phenylalanine	Phe
G	Glycine	Gly
H	Histidine	His
I	Isoleucine	Ile
K	Lysine	Lys
L	Leucine	Leu
M	Methionine	Met
N	Asparagine	Asn
O	Pyrrolysine	Pyl
P	Proline	Pro
Q	Glutamine	Gln
R	Arginine	Arg
S	Serine	Ser
T	Threonine	Thr
U	Sélénocystéine	Sec
V	Valine	Val
W	Tryptophane	Trp
Y	Tyrosine	Tyr
B		Asn/Asp
Z		Gln/Glu
X	Inconnu	

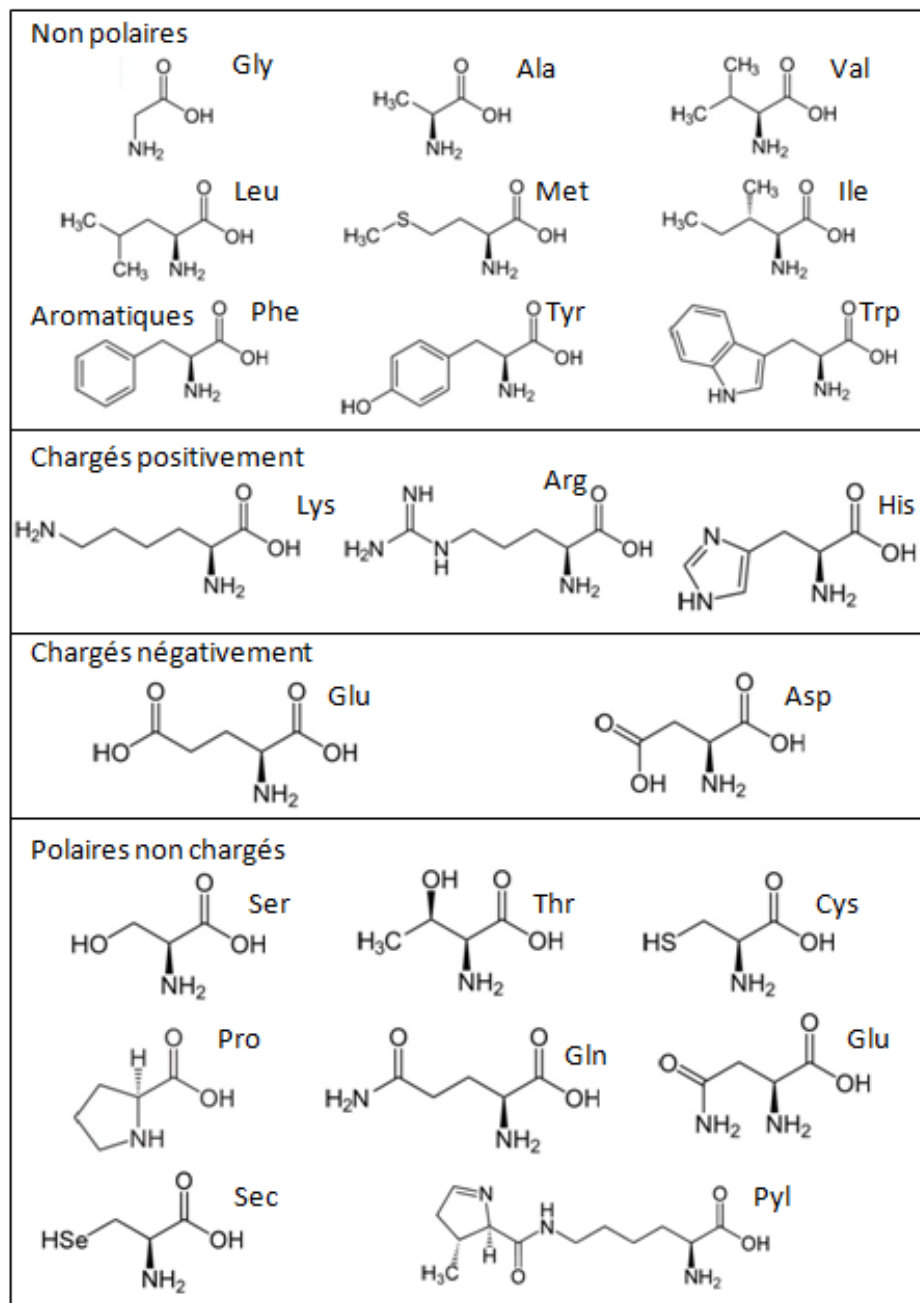
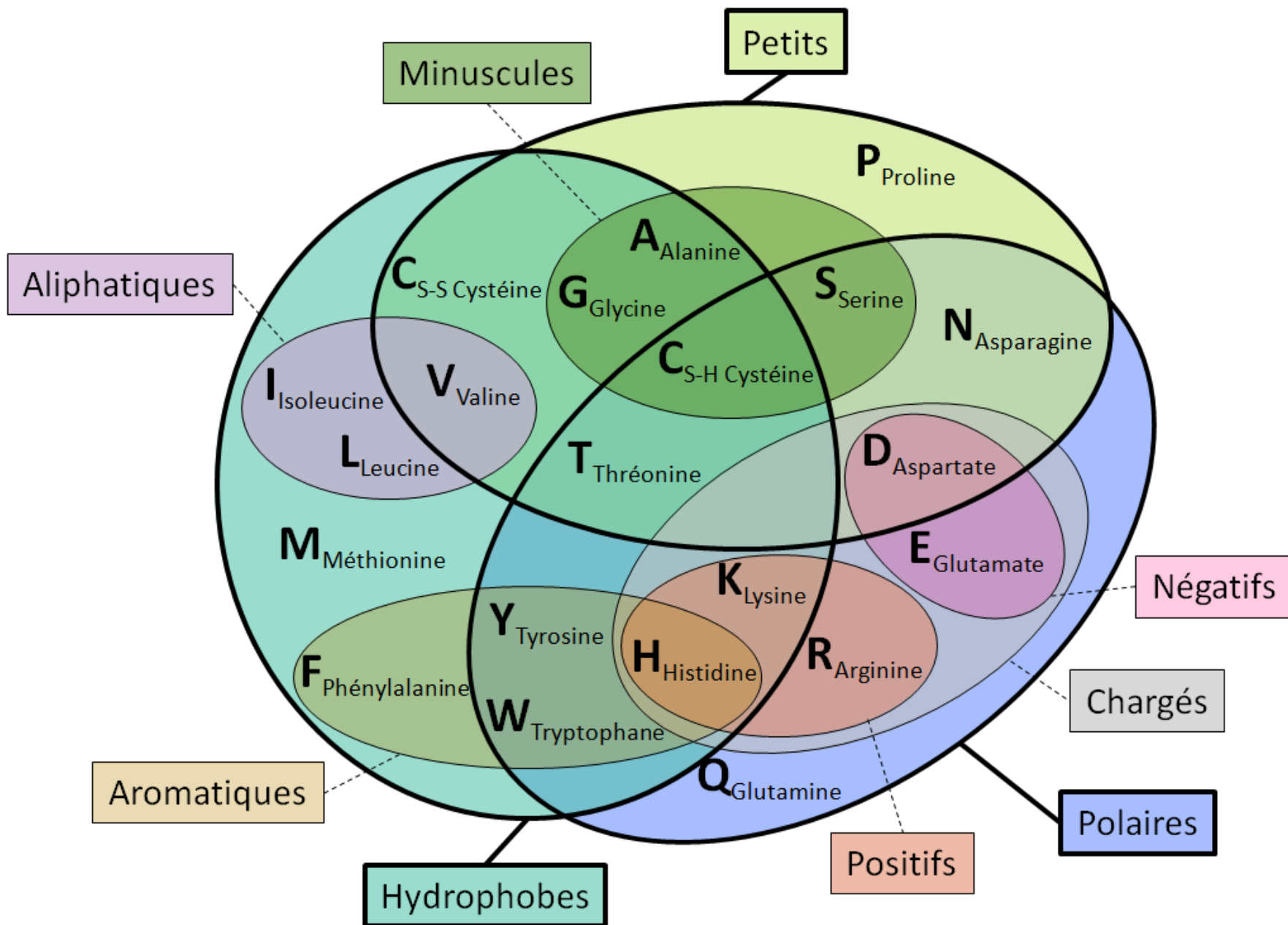
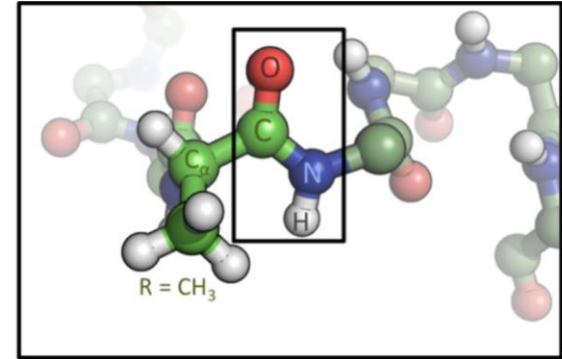


Diagramme de Venn des acides aminés



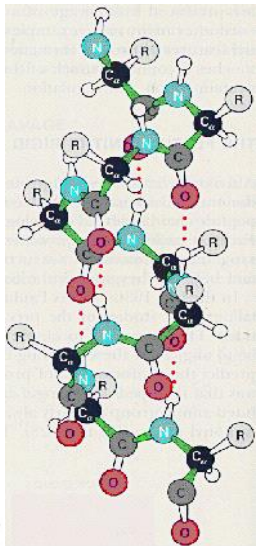
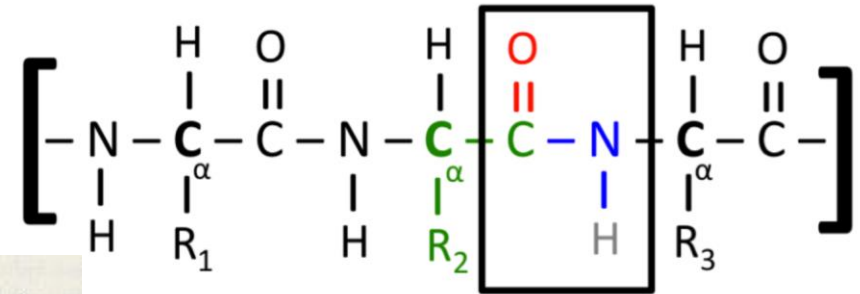
- Structure primaire ou séquence : la séquence des acides aminés dans la protéine.

```
MNGTEGPNFYVPFSNKTGVVRS PFEAPQYYLAEPWQFSMLAA YMFL I VL  
GFPINFLTLY VTVQHKKLR TPLNY ILLNLAVADLFMVFGGFTTTL Y TSLH  
GYFVFGPTGCNLEGFFA TLGGEIALW S LVV LA IERY VVCKPMSNFR FGE  
NHA IMGVAFTWVMALACA APPLVGWSRY I PQGMQCS CGALY FTLKPEINN
```

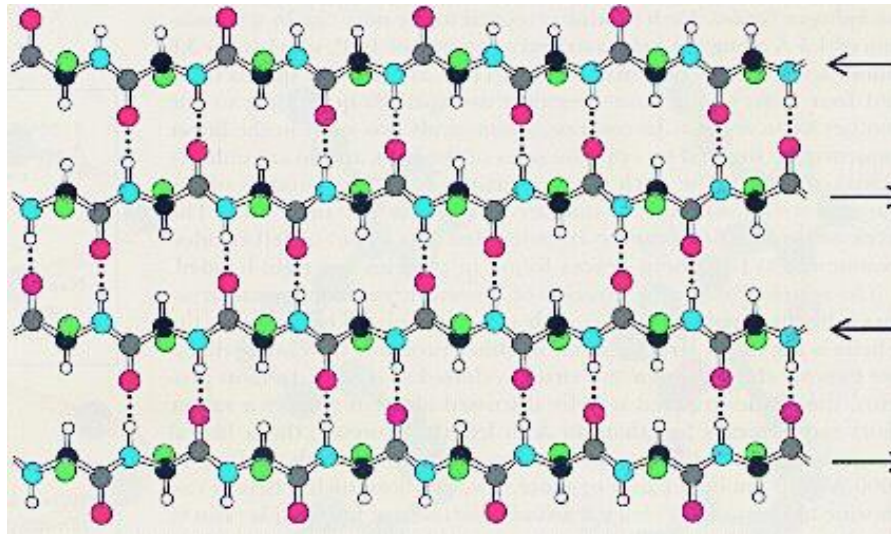


- Structure secondaire : ensemble des régions présentant un arrangement régulier (périodique) stabilisé par des liaisons hydrogènes impliquant les atomes de la chaîne principale (backbone)

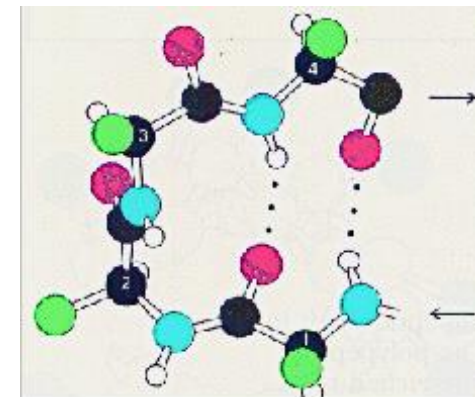
- *i.e.* hélice α , brins β , feuilletts β et coudes β .



30%



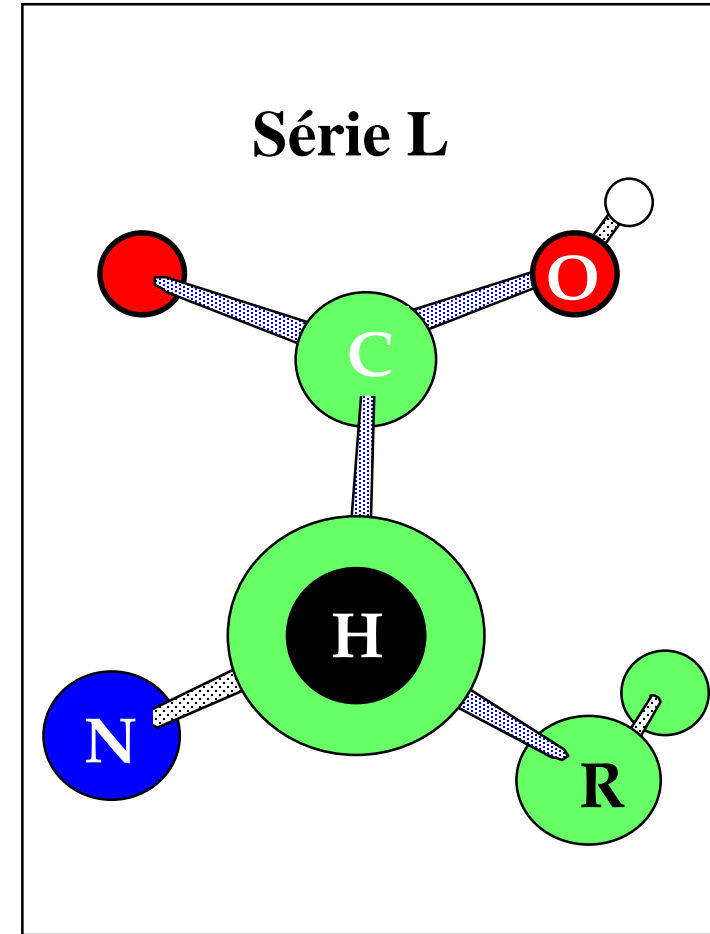
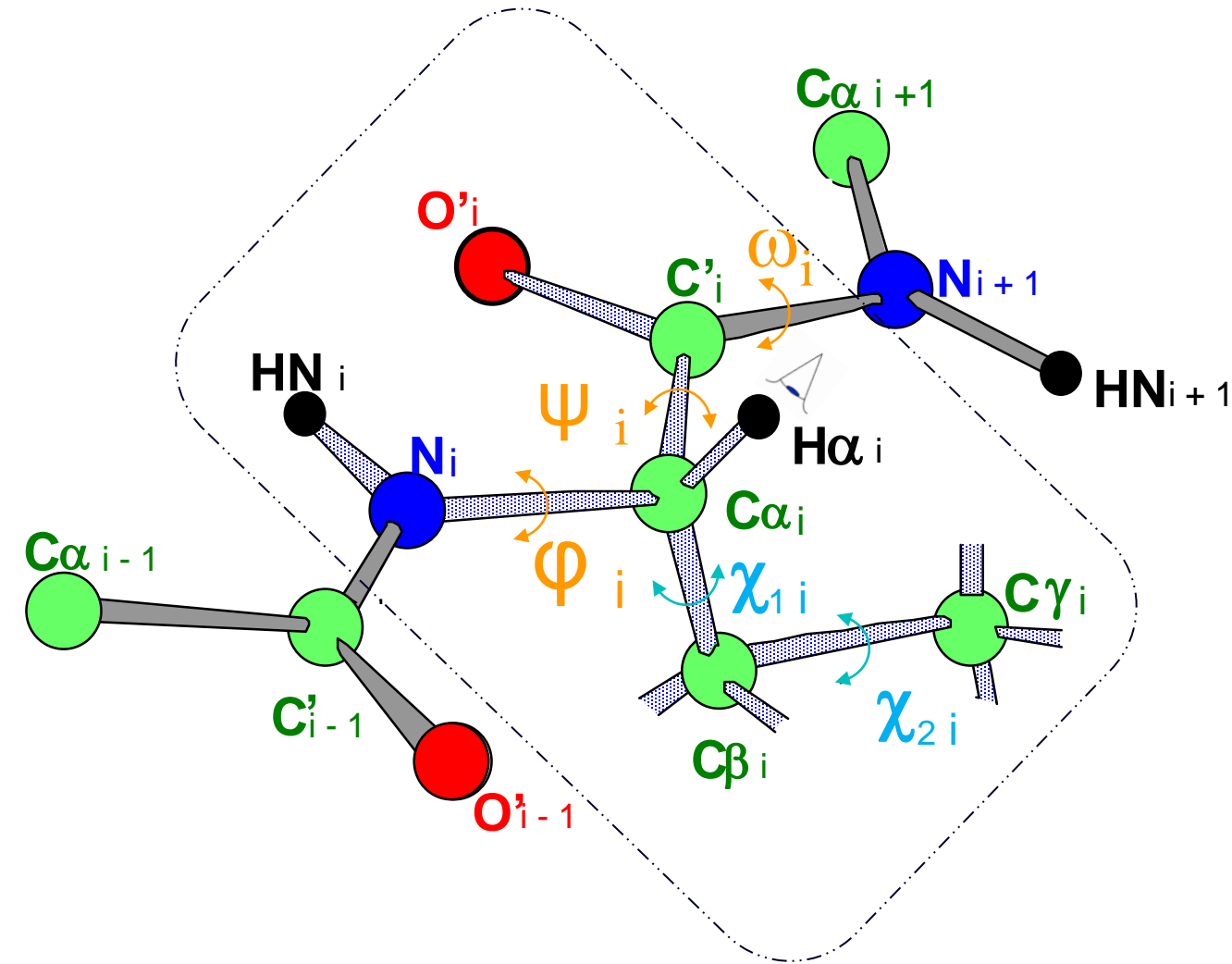
20%



25%

Géométrie du squelette protéique

Degrés de liberté dans la chaîne protéique



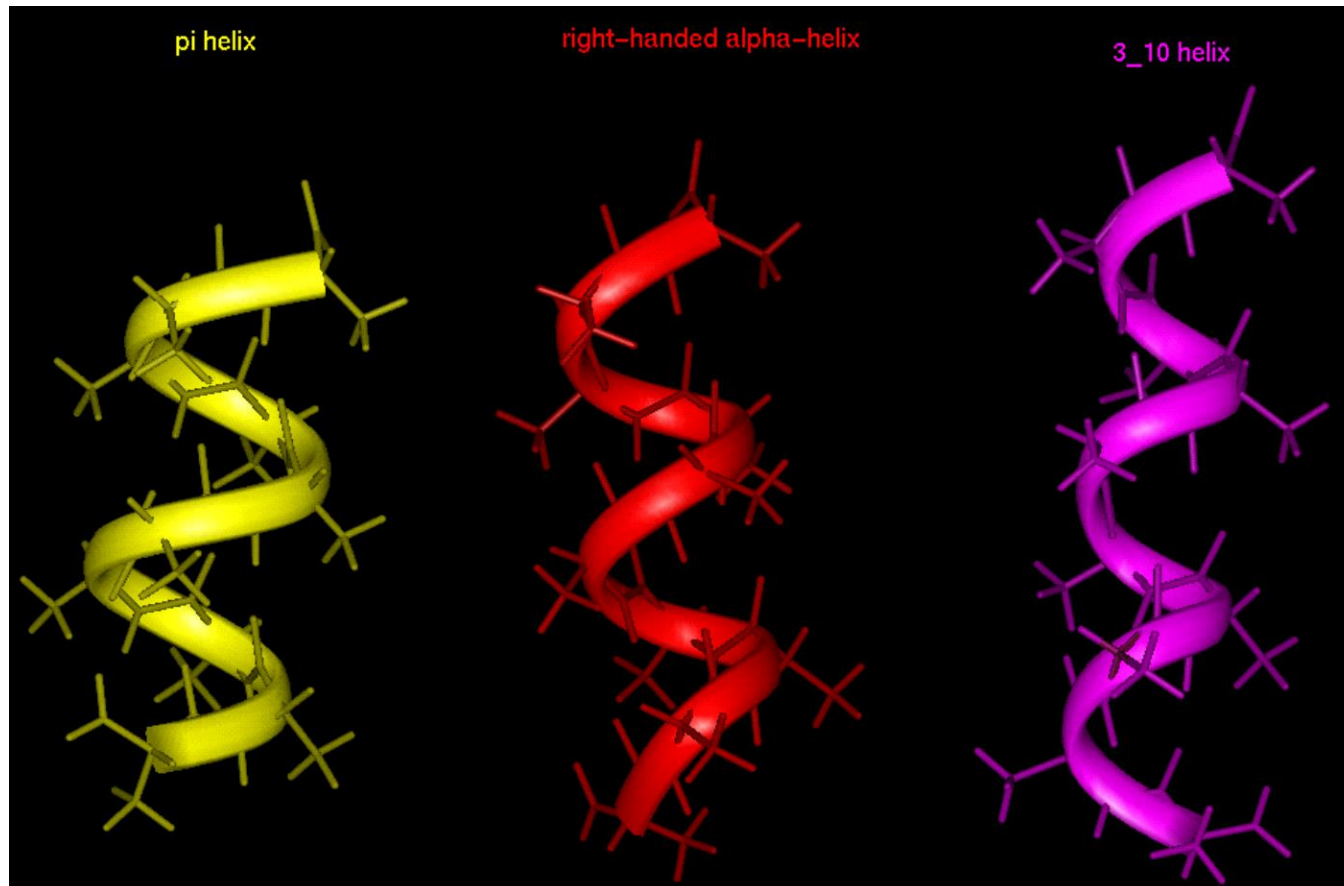
Lire « **C O R N** »

φ = l'angle dièdre définit par: $CO(i-1)-NH(i)-Ca(i)-CO(i)$

ψ = l'angle dièdre définit par: $NH(i)-Ca(i)-CO(i)-NH(i+1)$

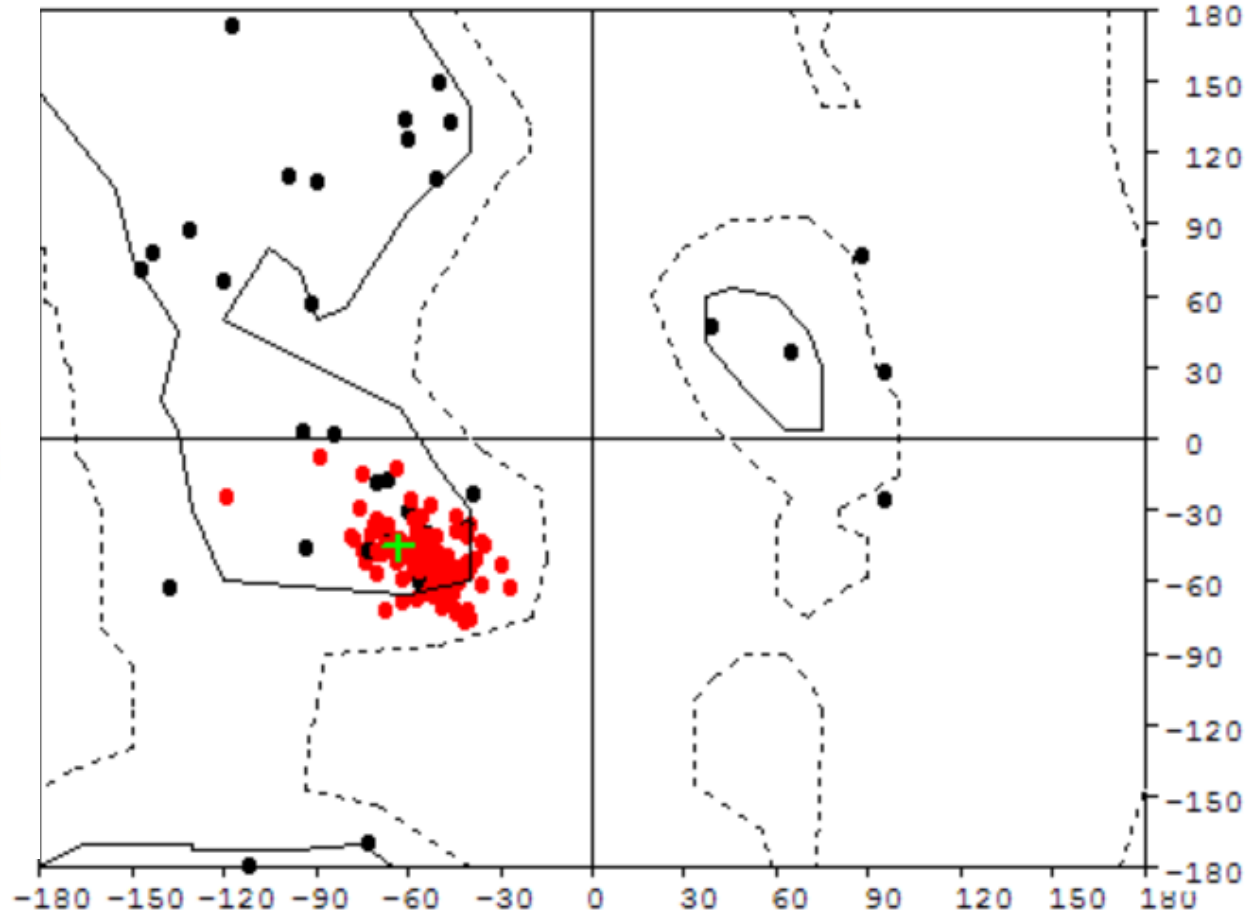
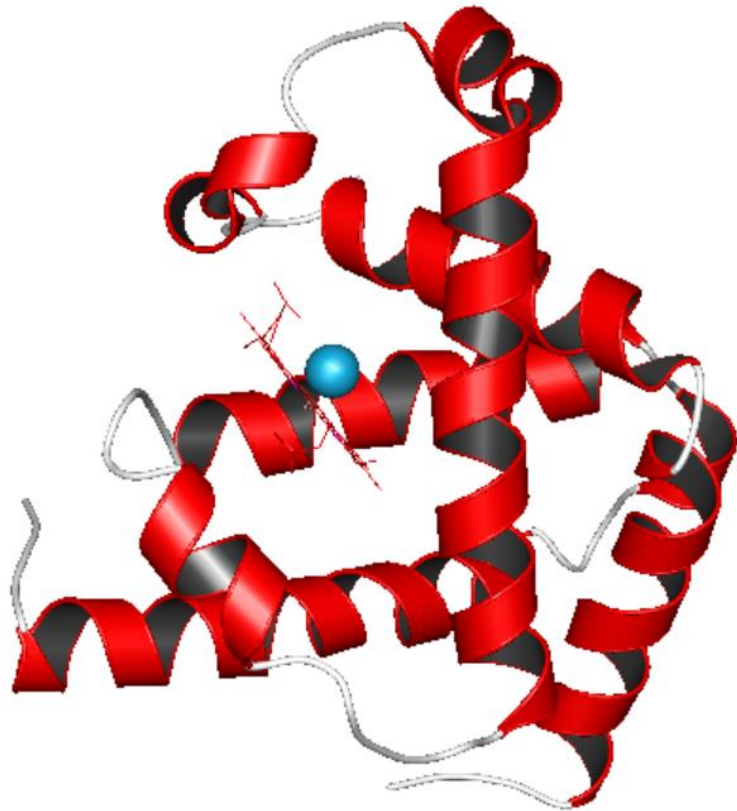
Les couples d'angles ϕ ψ dans les hélices

Structure	ϕ	ψ	Liaison H
Hélice α de pas droit	-57 °	-47.0 °	i+4
Hélice pi	-57.1 °	-69.7 °	i+5
Hélice 3-10 i+3	-49 °	-26 °	i+3



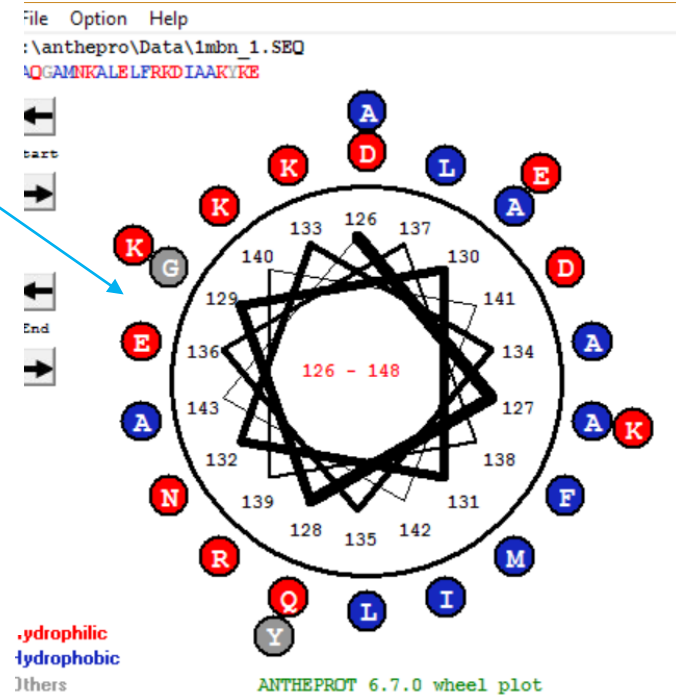
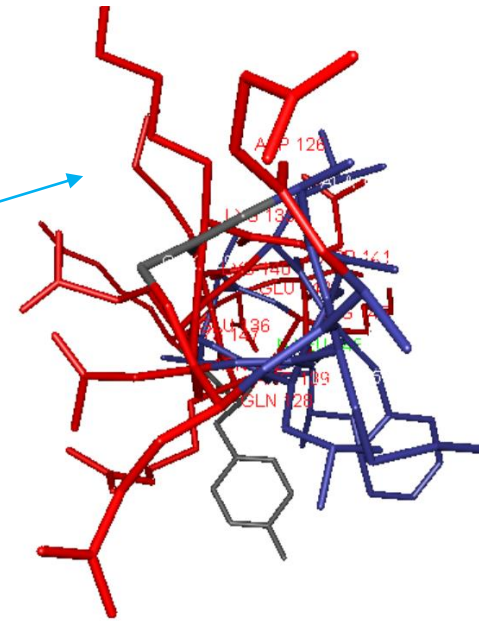
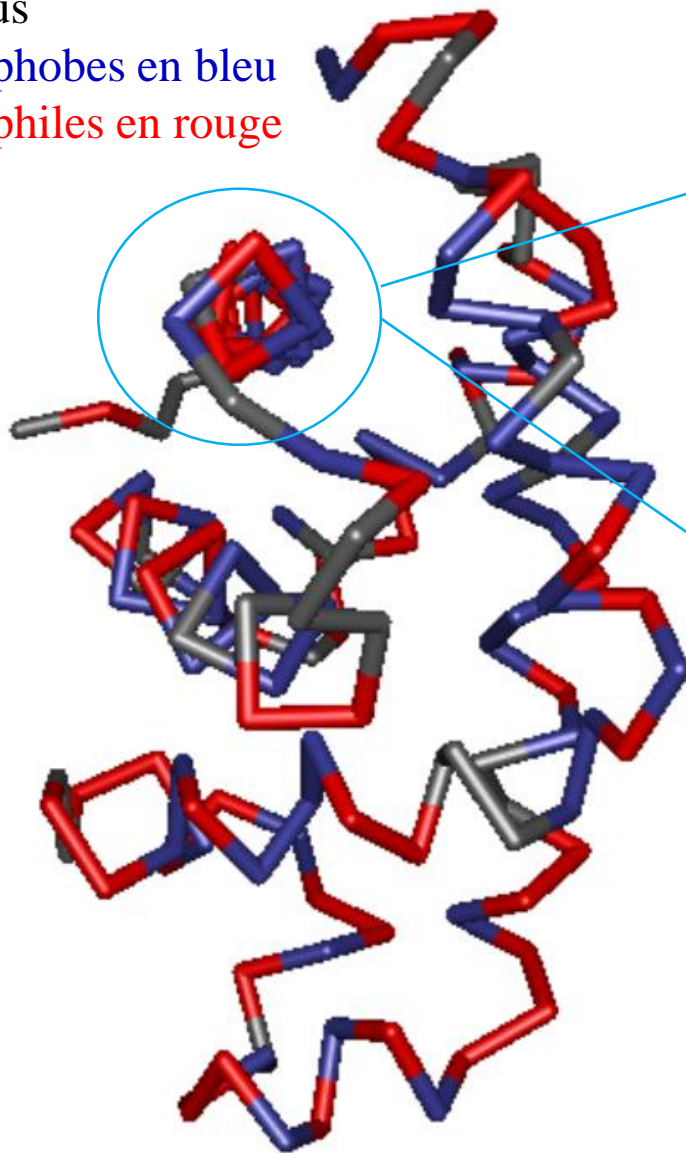
7 TRP Phi: -63,16 Psi: -45,77

LSEGEWQIVIHVWAKVEADVAGHG
CCFHHFHHEHHHHHHHCCCHHHH

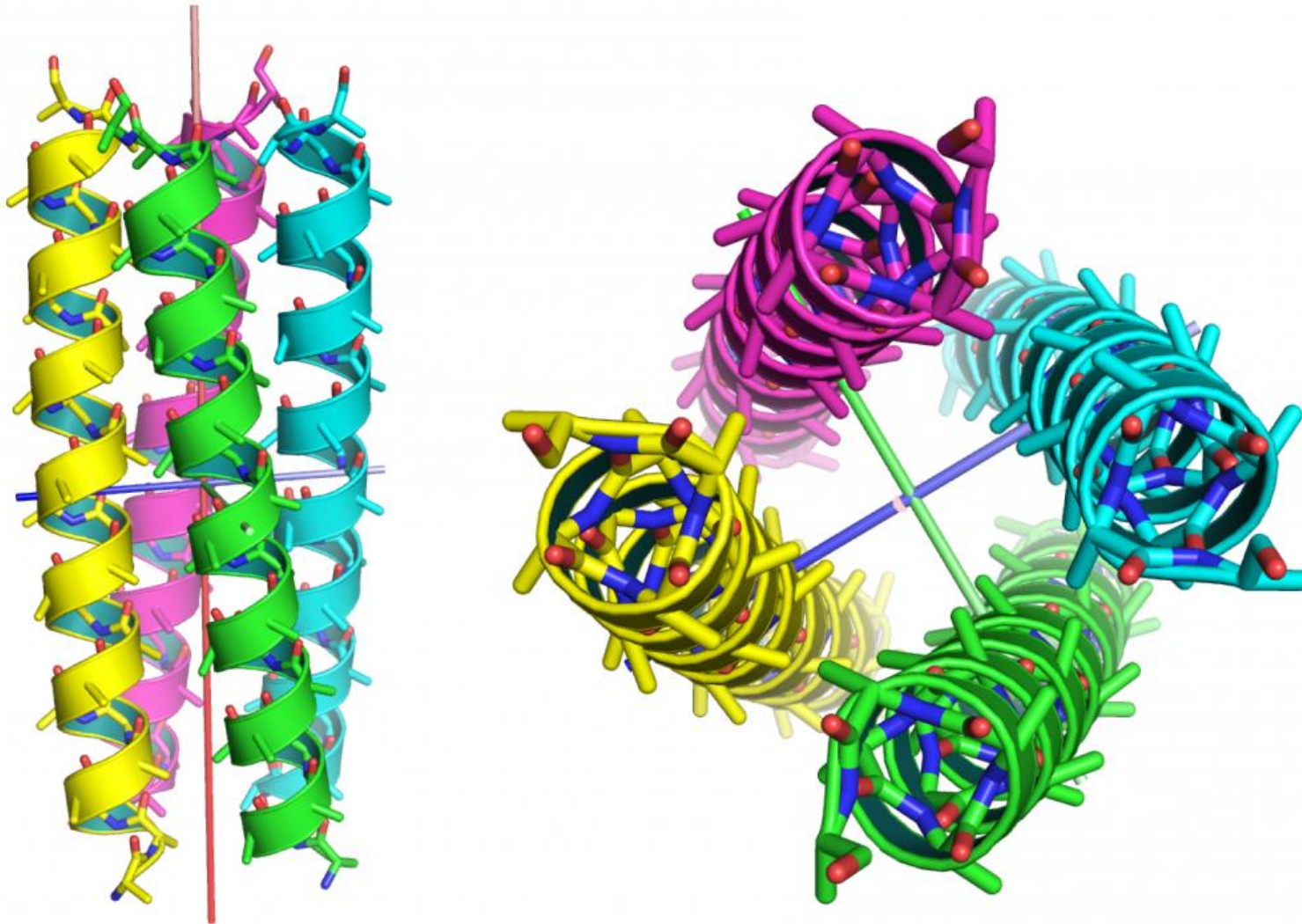


ANTHEPROT Ramachandran for Lovell et al. 2003 regions

Résidus
hydrophobes en bleu
hydrophiles en rouge

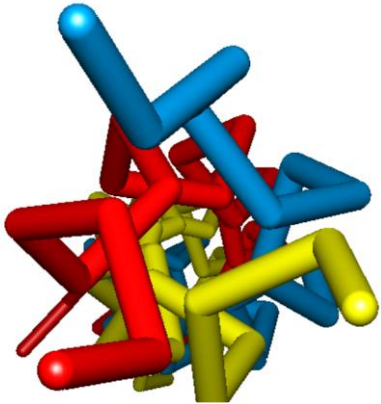


Polymère d'hélices α de pas droit surenroulé dans un pas gauche

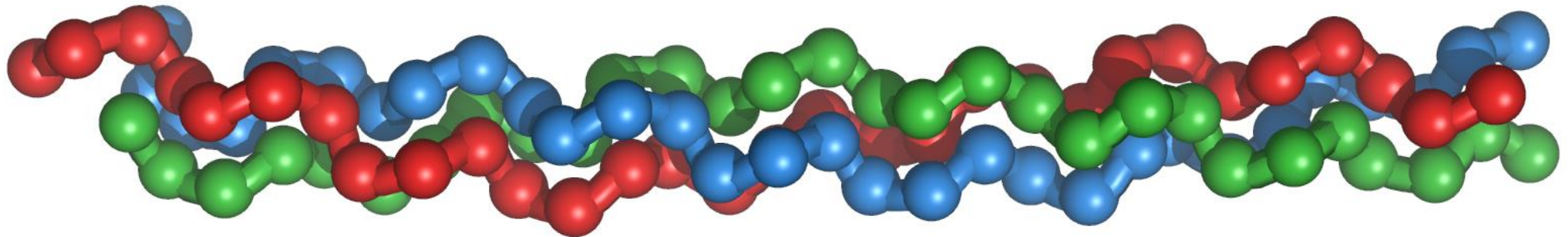


Existe de dimère à heptamère

Structure	ϕ	ψ
PolyPro type I	-83°	158°
PolyPro Type II	-78°	$+149^\circ$
Poly Gly II	-80°	$+150^\circ$



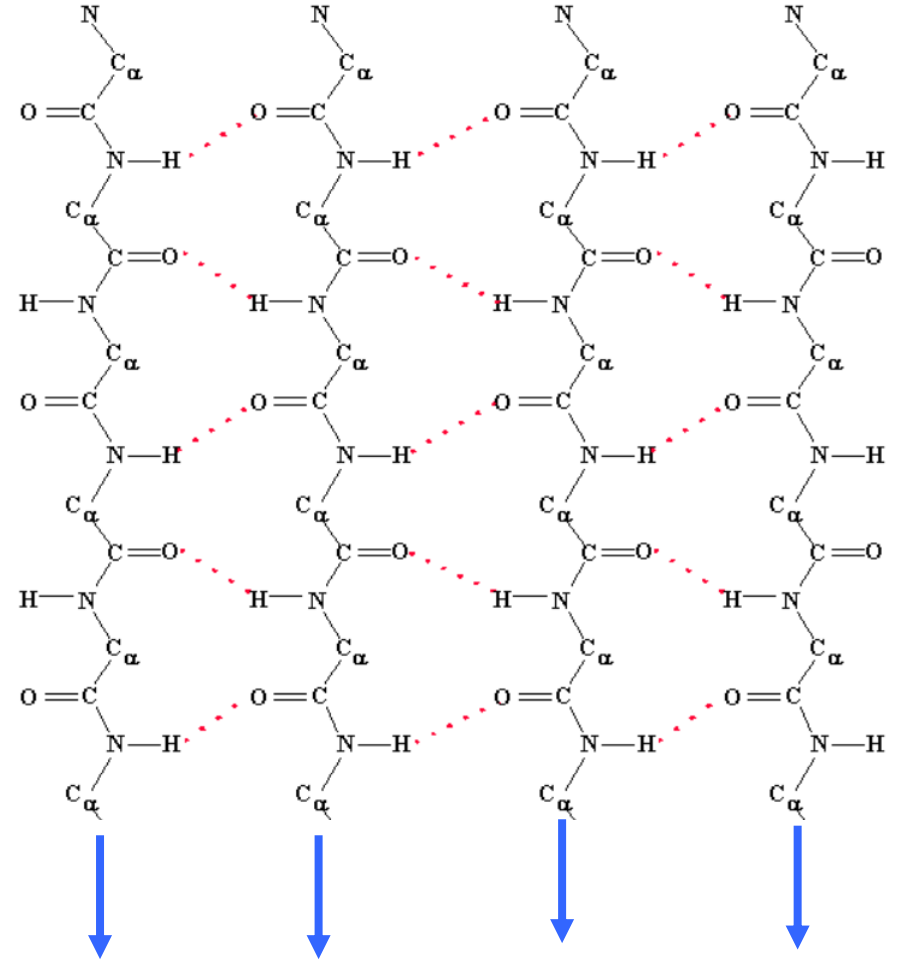
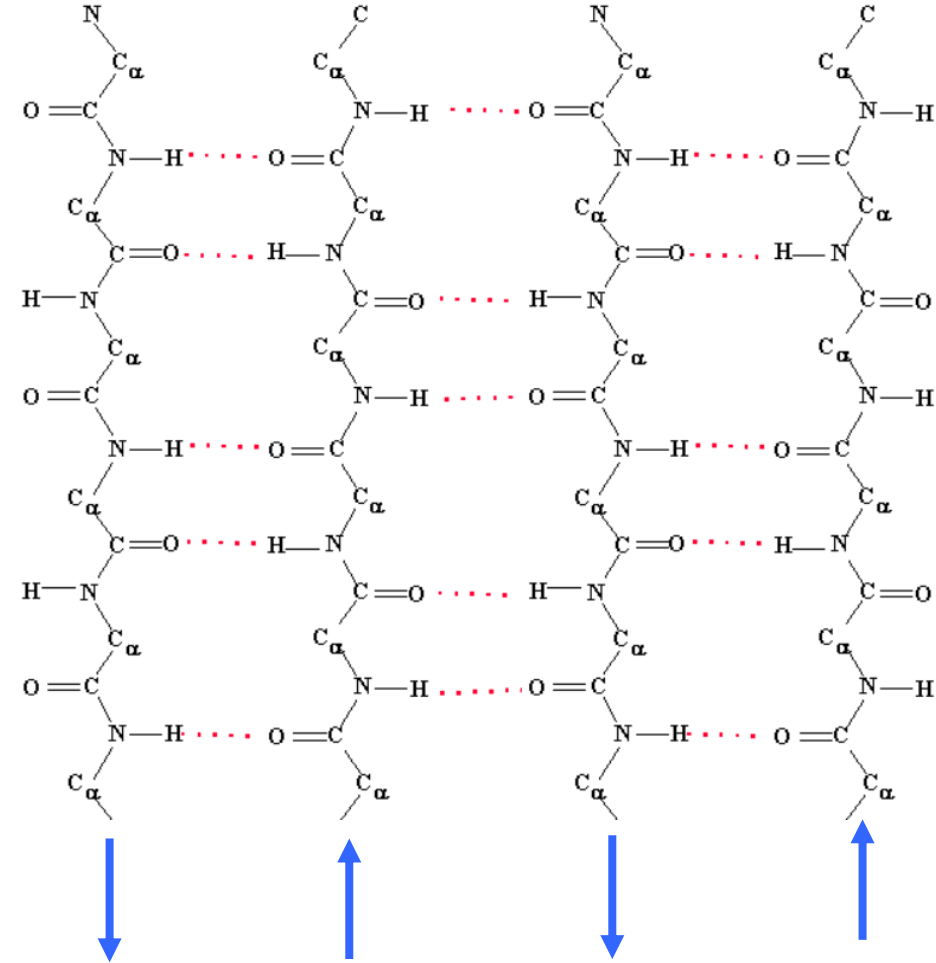
Trimère d'hélices de pas gauche surenroulé avec un pas droit



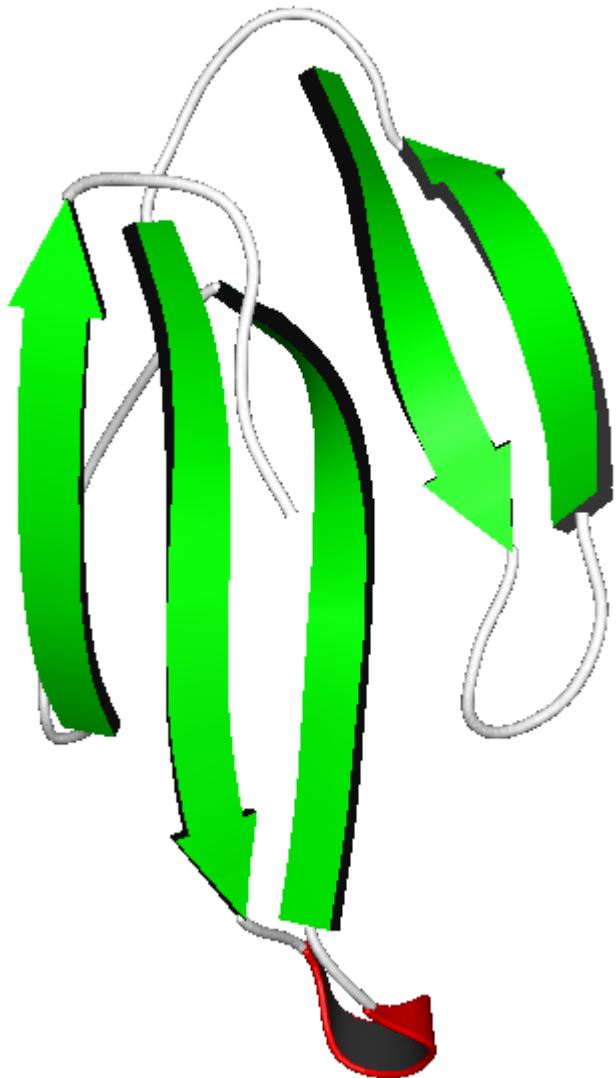
Hélice de tropocollagène (1BKV)

Feuillet de 4 brins anti parallèles

Feuillet de 4 brins parallèles



Structure	φ	ψ
Brins β anti parallèles	-139°	$+135^\circ$
Brins β parallèles	-119°	$+113^\circ$

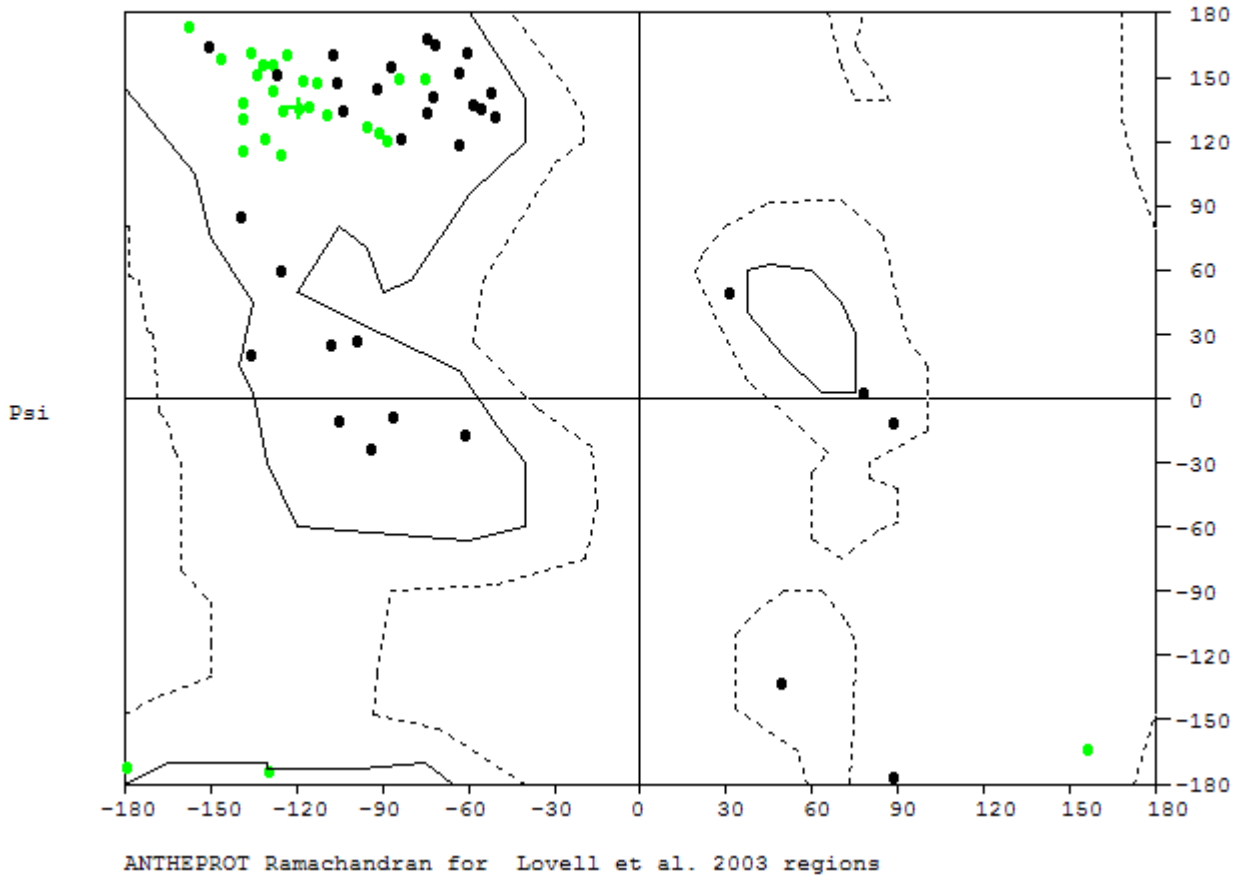


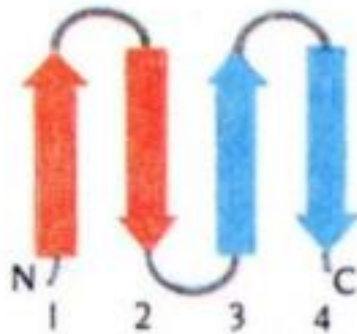
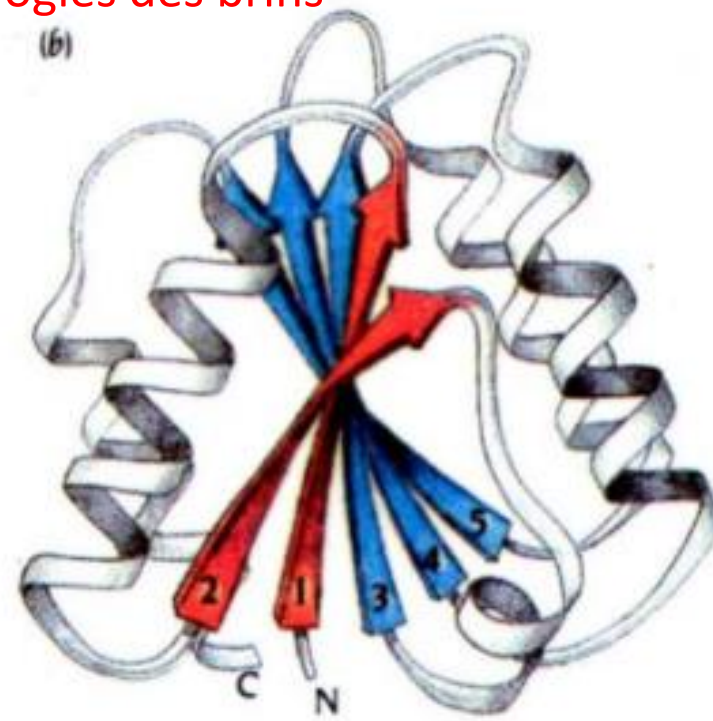
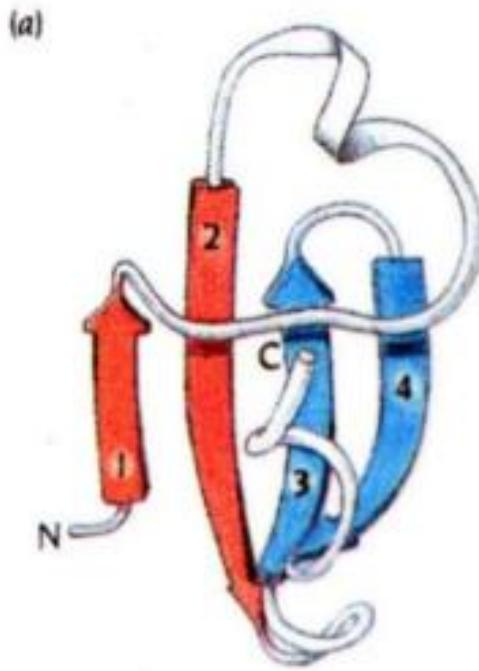
File Options

2 ILE Phi: -118,84 Psi: 135,54

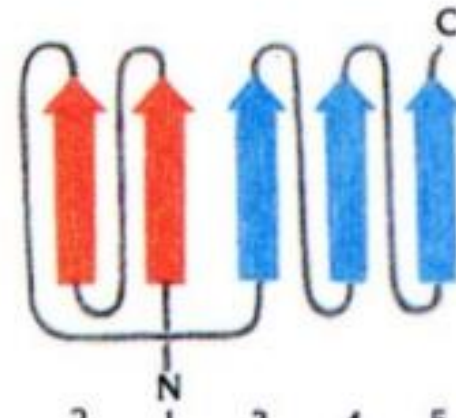
ICFNHQGSQPQITKTCSPG

EEEEEEEEEEEEEE



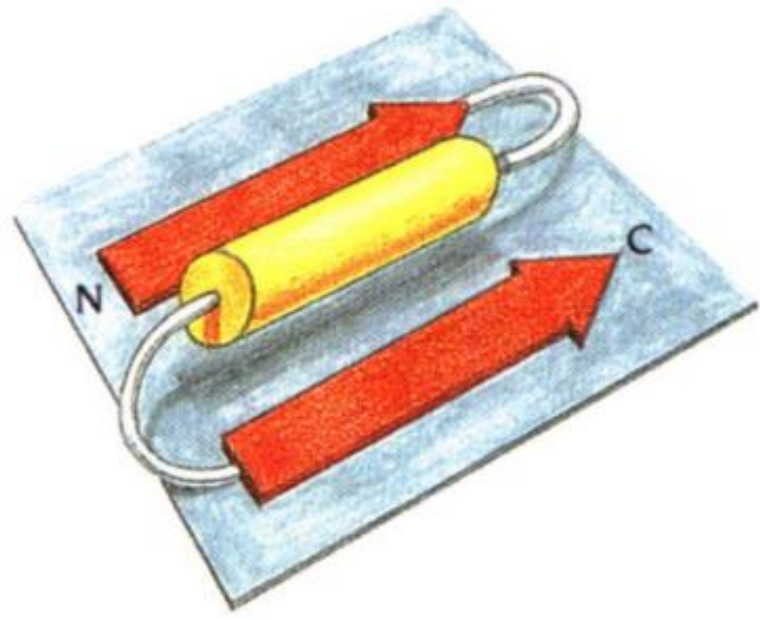


Brins antiparallèles

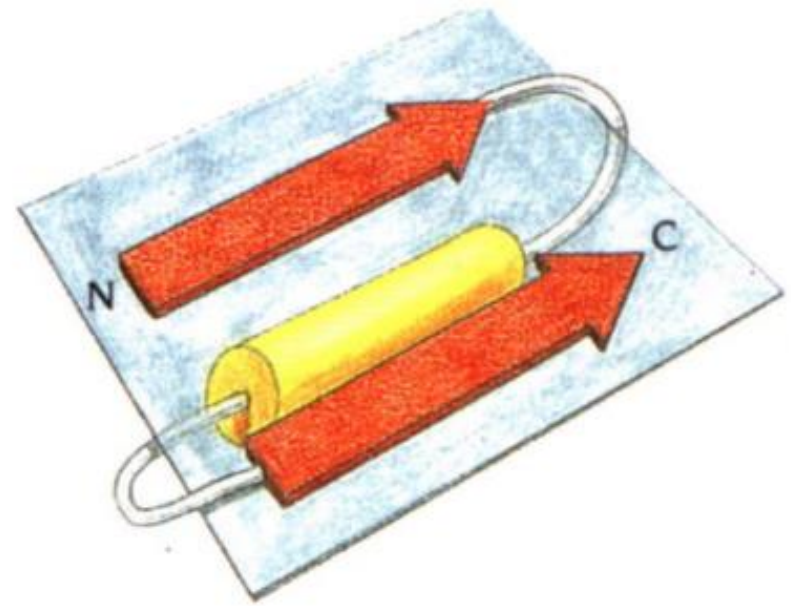


Brins parallèles

Motif β - α - β

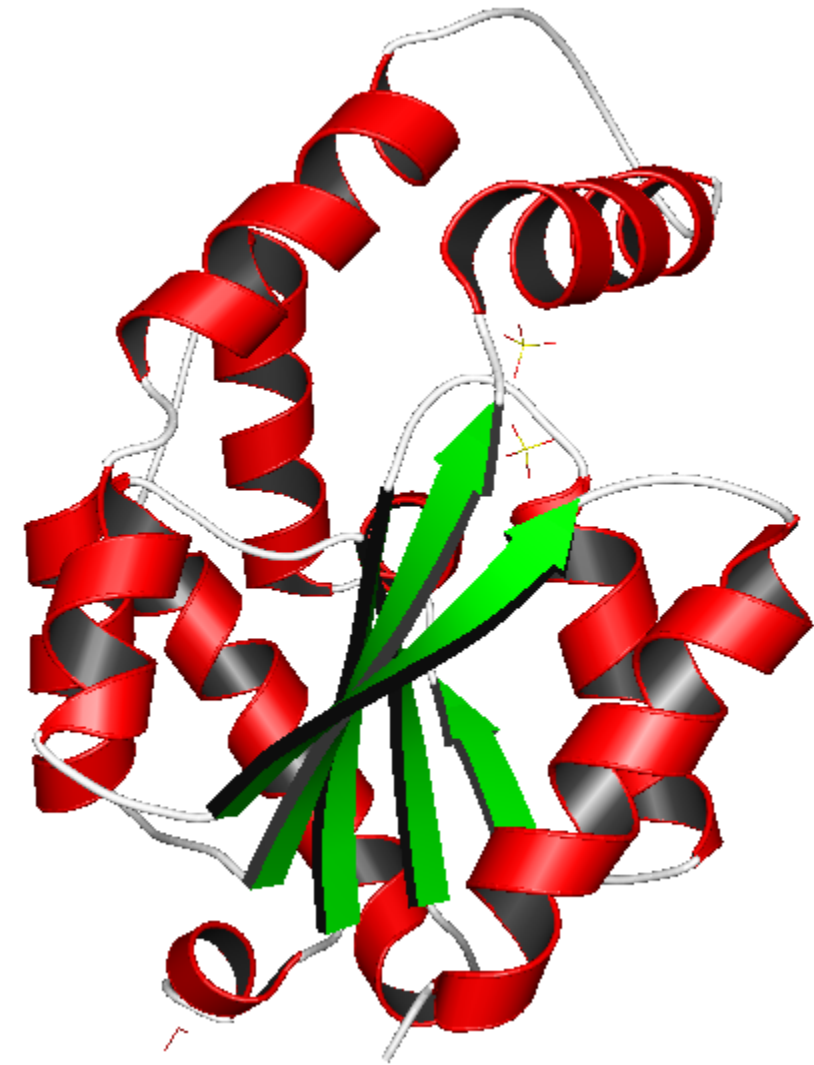


Droite



Gauche

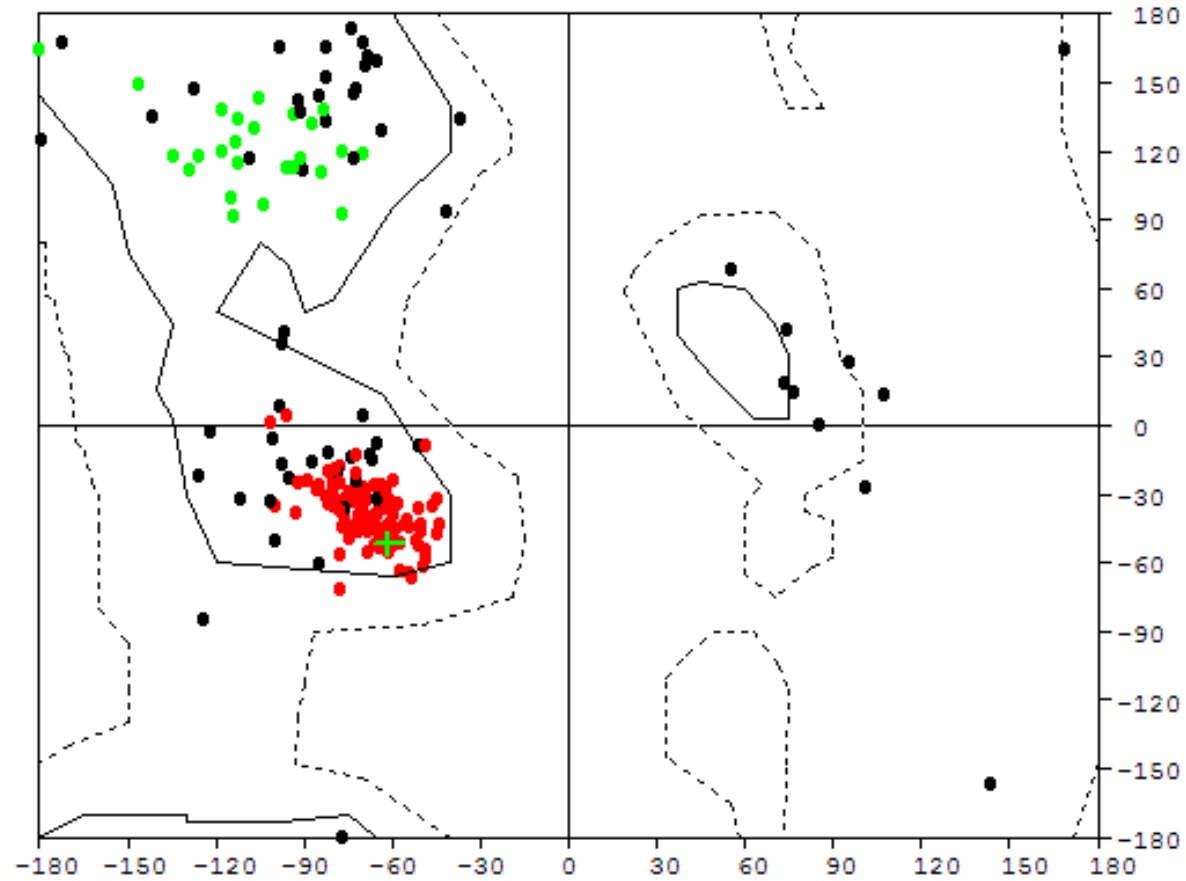
Alternance hélice-brin



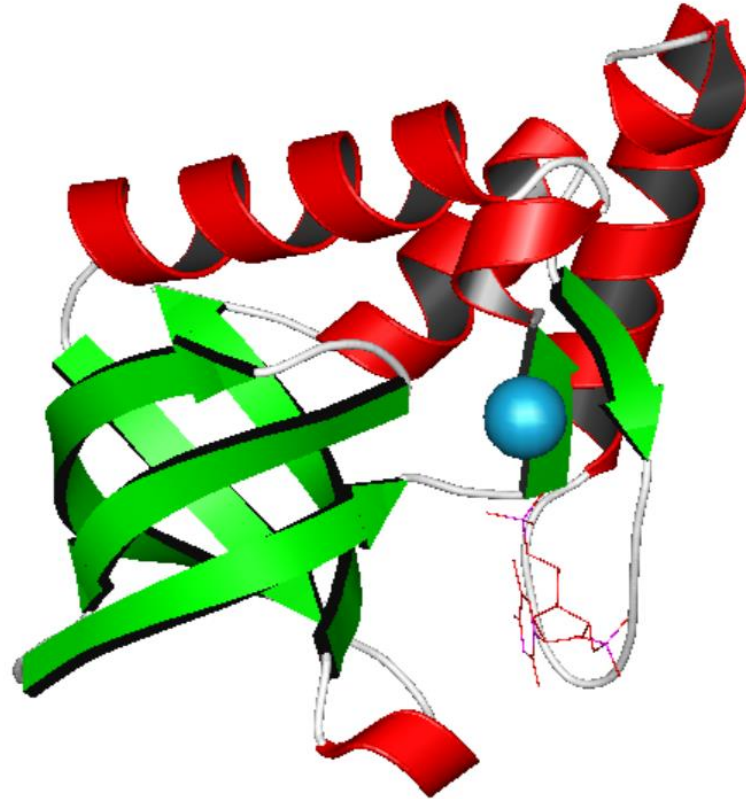
3adk

Σ LU Phi: -61,15 Psi: -51,66

EEKLLKSKIIIFVVGPGSG
HHHHHCCCCHEEEECCECCC

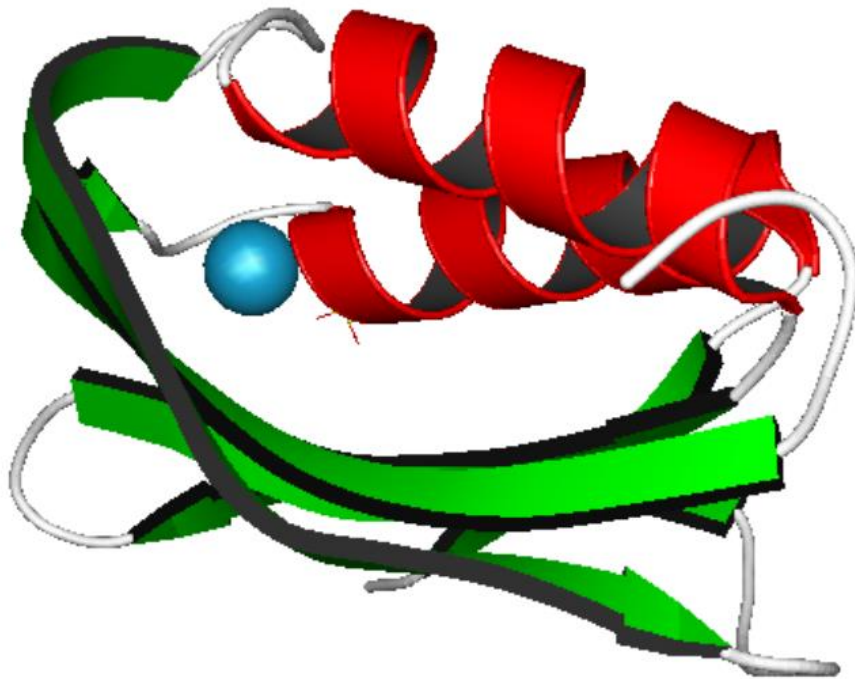


ANTHEPROT Ramachandran for Lovell et al. 2003 regions

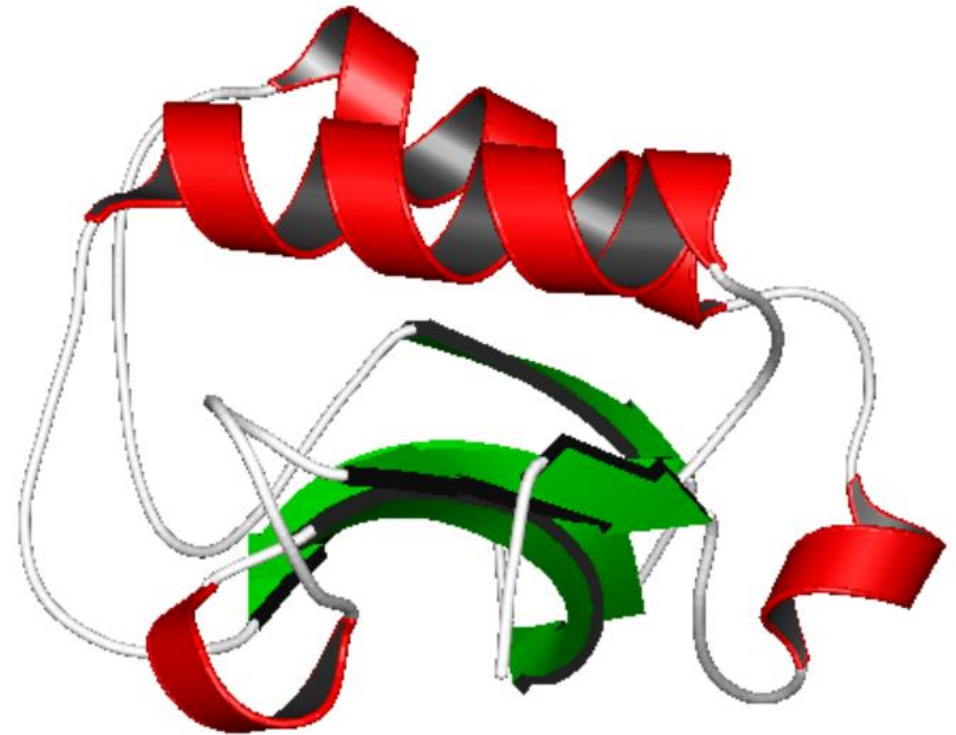


5kyi

Architectures identiques 2 hélices et 5 brins mais topologies différentes



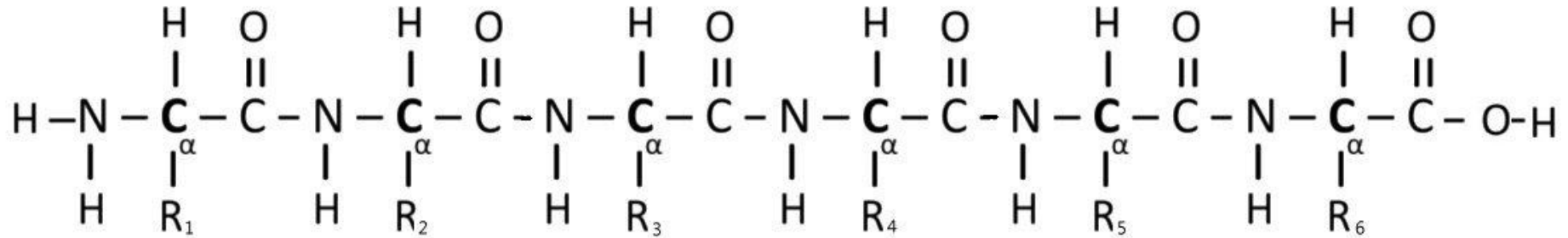
pdb : 2ACY



pdb : 1KPT



Soit une chaîne polypeptidique:



Une chaîne polypeptidique de n acides aminés en hélice α met en jeu une énergie:

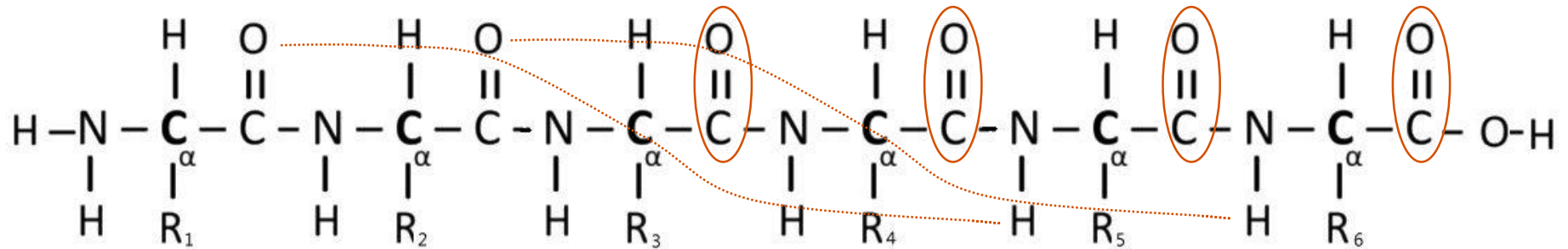
$$\Delta G = (n-4)\Delta H_r - (n-2)T \Delta S_r$$

Avec :

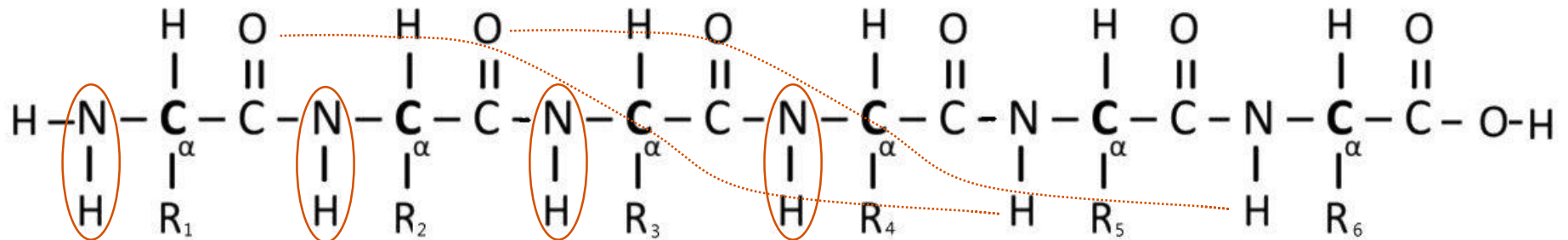
ΔH_r Variation d'enthalpie par résidu impliqué dans l'hélice

ΔS_r Variation d'entropie par résidu impliqué dans l'hélice

Si n acides aminés en hélice n-4 CO impliqués dans une liaison hydrogène

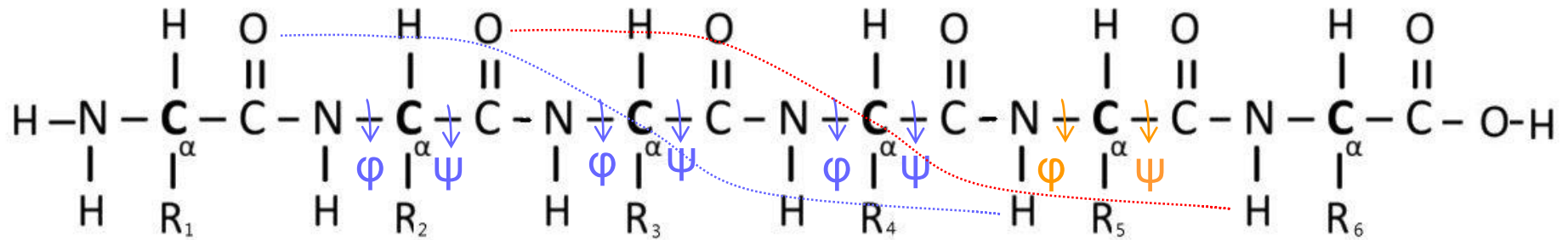


Si n acides aminés en hélice n-4 NH impliqués dans une liaison hydrogène

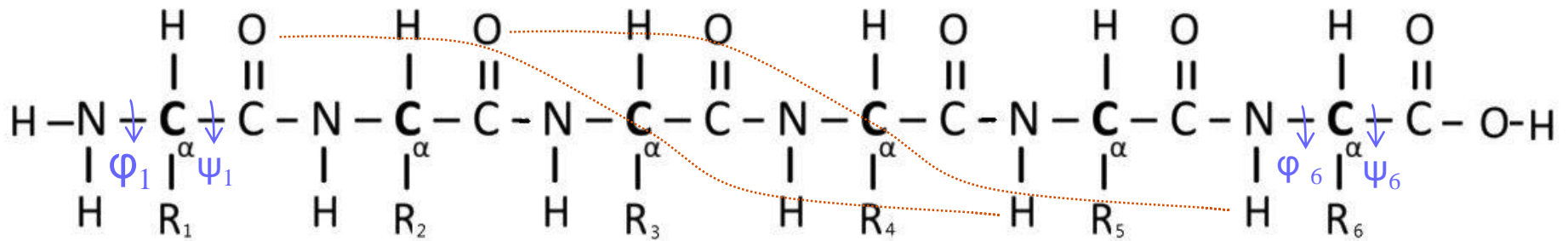


Si n acides aminés en hélice 4 CO+4NH ~4 AA non impliqués dans une liaison hydrogène

Si n acides aminés en hélices $n-2$ angles phi et $n-2$ angles psi imposés



Si n acides aminés en hélice 2 angles phi et 2 angles psi non imposés soient 2 AA



- **Hétérogénéité et nombre >2000 BD biologiques**

- ✓ Séquences nucléiques 200 000 000
- ✓ Séquences protéiques 500,000 annotées
- ✓ Structures 3D 150,000
- ✓ Structures 3D différentes (<25% Id) 20,000

- **Qualité variable**

- ✓ Erreurs
- ✓ Propagées par l'annotation automatique

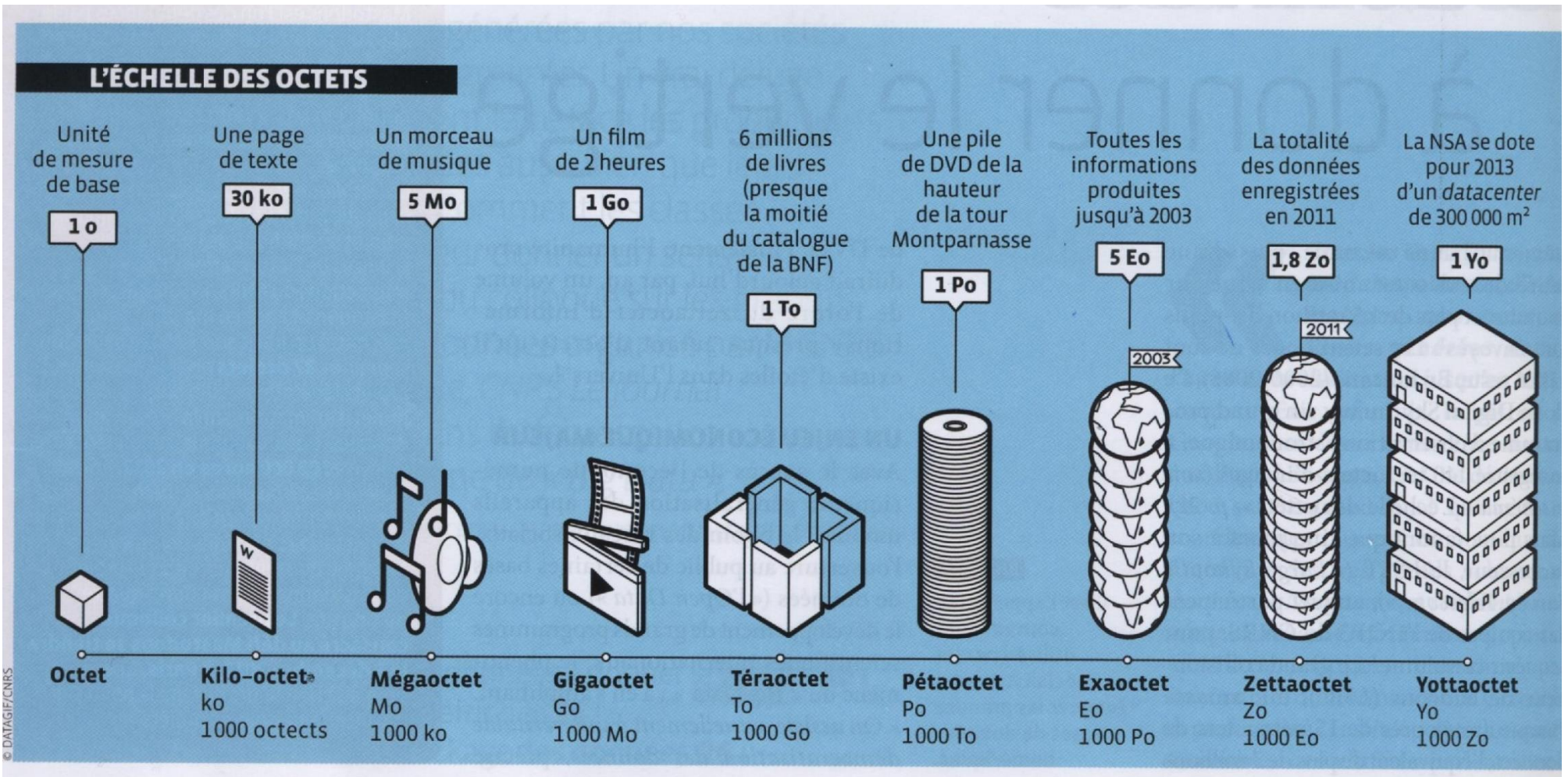
- **Biais**

- ✓ 20 espèces = 21% des entrées de SWISS-PROT
- ✓ Redondance
- ✓ Génome humain <http://www.ensembl.org>

- **Volume « faible » croissance exponentielle**

- ✓ Formats, traitements
- ✓ Le génome du jour...

<http://www.genomesonline.org/>



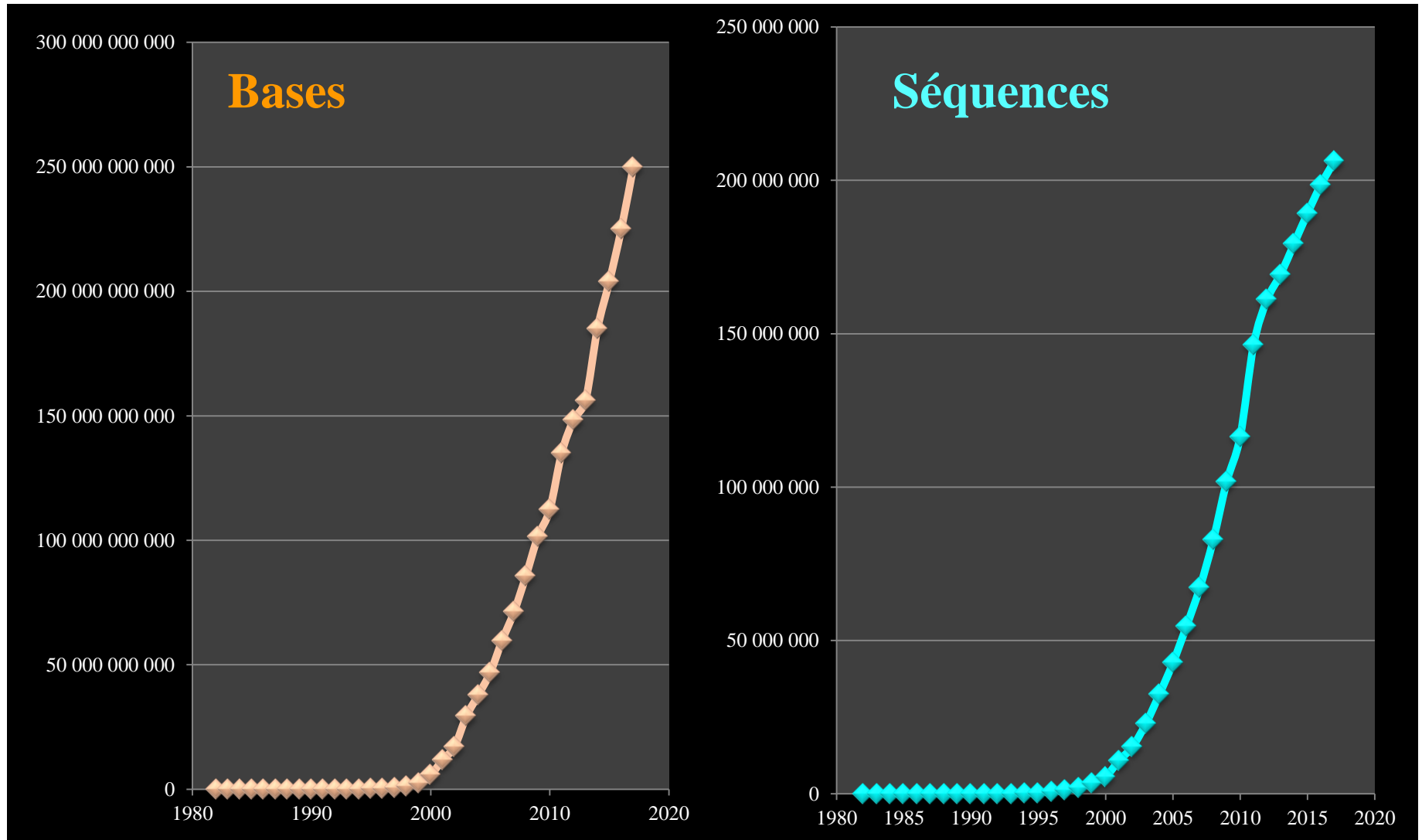
Le journal du CNRS 2012

1 octet = 8 bits soit 256 valeurs



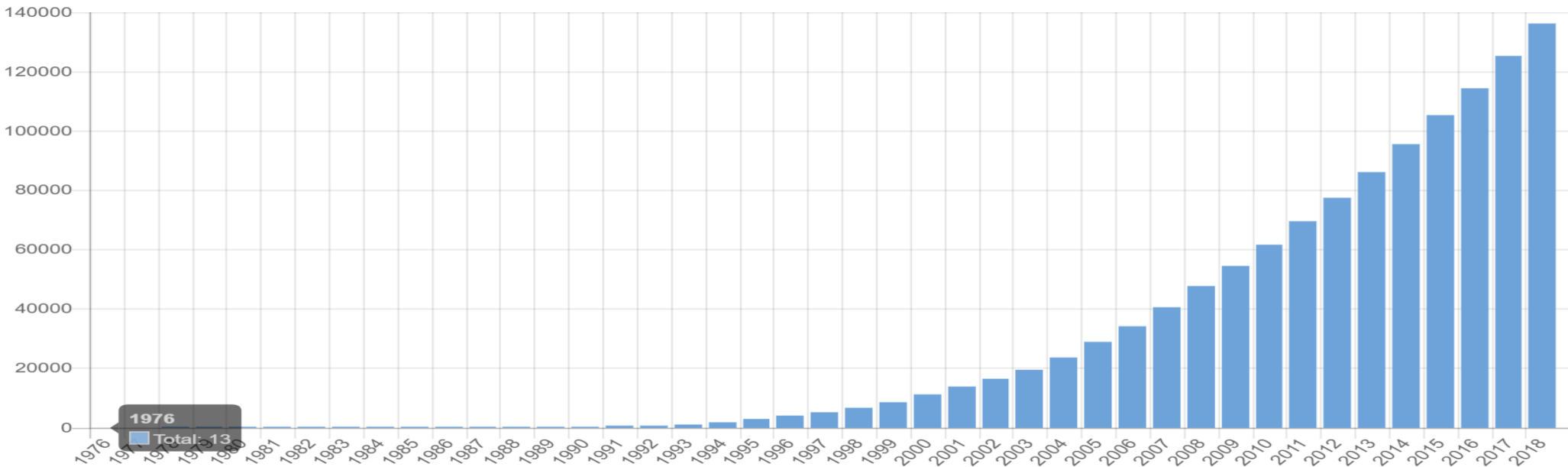
Croissance de genbank (1982-2017)

Banques de séquences d'acides nucléiques

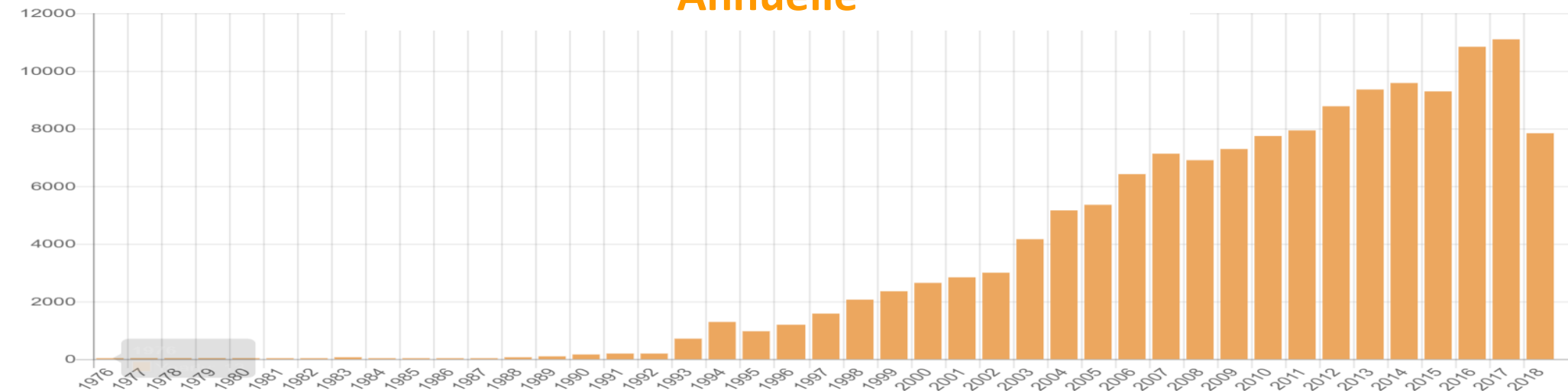


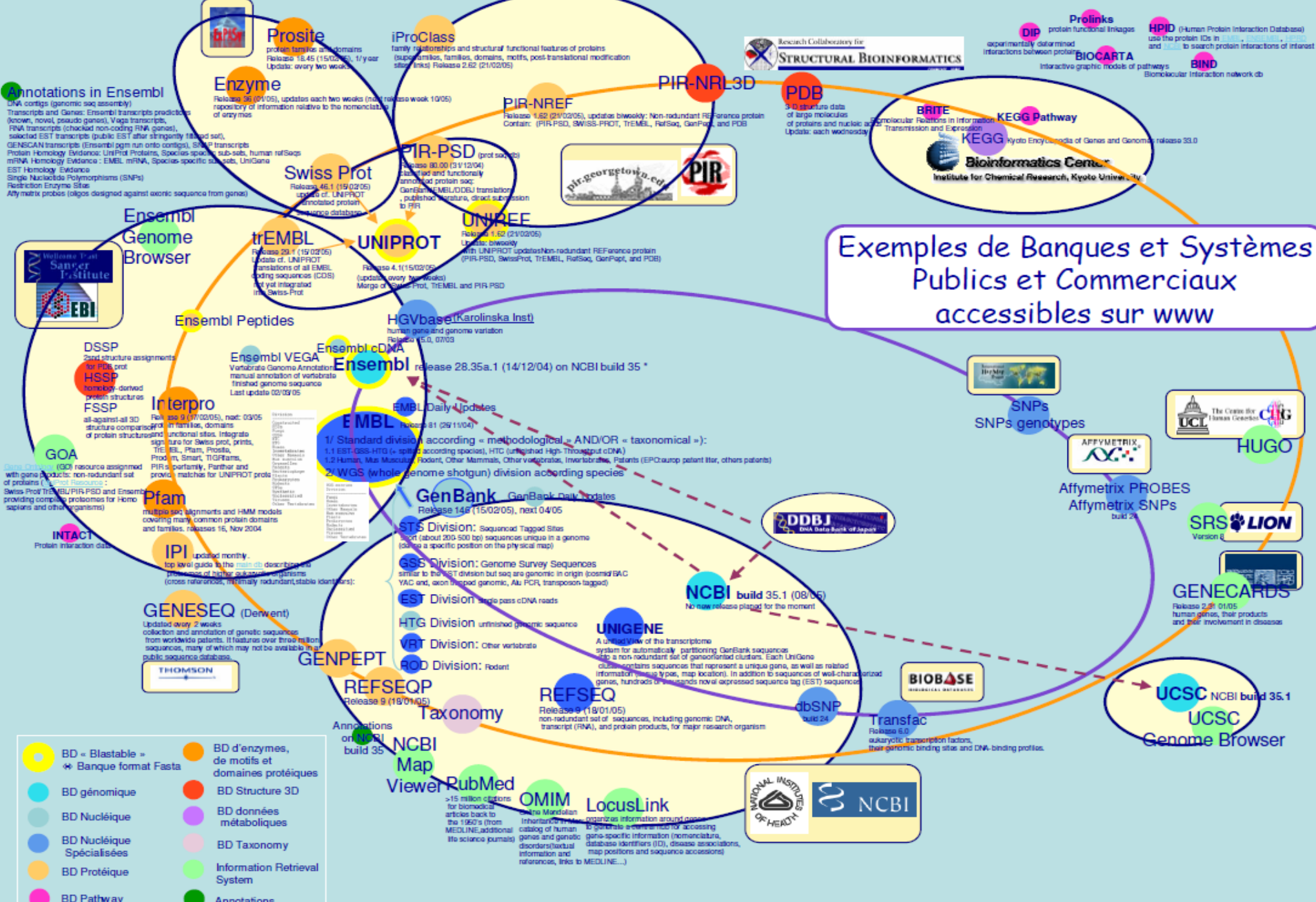
http://www.pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html

Cumulée



Annuelle





Exemples de Banques et Systèmes Publics et Commerciaux accessibles sur www

- BD « Blastable »
- BD d'enzymes, de motifs et domaines protéiques
- BD génomique
- BD Nucléique
- BD Nucléique Spécialisées
- BD Protéique
- BD Pathway
- BD Structure 3D
- BD données métaboliques
- BD Taxonomie
- Information Retrieval System
- Annotations

* : 28 is for API and annotation versions, 35 is for the ncbi build version, a is for the ncbi build version evolution, 1 is for API web server version

Bases de données nucléiques

Genbank 239 (08/2020)

<http://www.ncbi.nlm.nih.gov/genbank>

9 236 443 421 310 bases

1 901 329 611 séquences

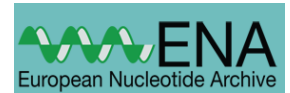


European Nucleotide Archive 140 (08/2020)

<http://www.ebi.ac.uk/ena/>

8 327 10⁹ bases

2 621 700 000 séquences



DDBJ 120 (06/2020)

<http://www.ddbj.nig.ac.jp/>

9,253,936,453,958 bases

2,414,499,799 séquences



Bases de données protéiques

<http://www.uniprot.org>

UniProt/TrEMBL(08/2020)

186 961 949 séquences

UniProt/Swiss-Prot (08/2020)

563 082 sequences annotées

202 799 066 acides aminés



Structures 3D

<http://www.rcsb.org/pdb/>

PDB RCSB (07/09/2020) 168358 Structures dont 50% avec des séquences <>

Protéines avec <30% identité

~35000 Groupes

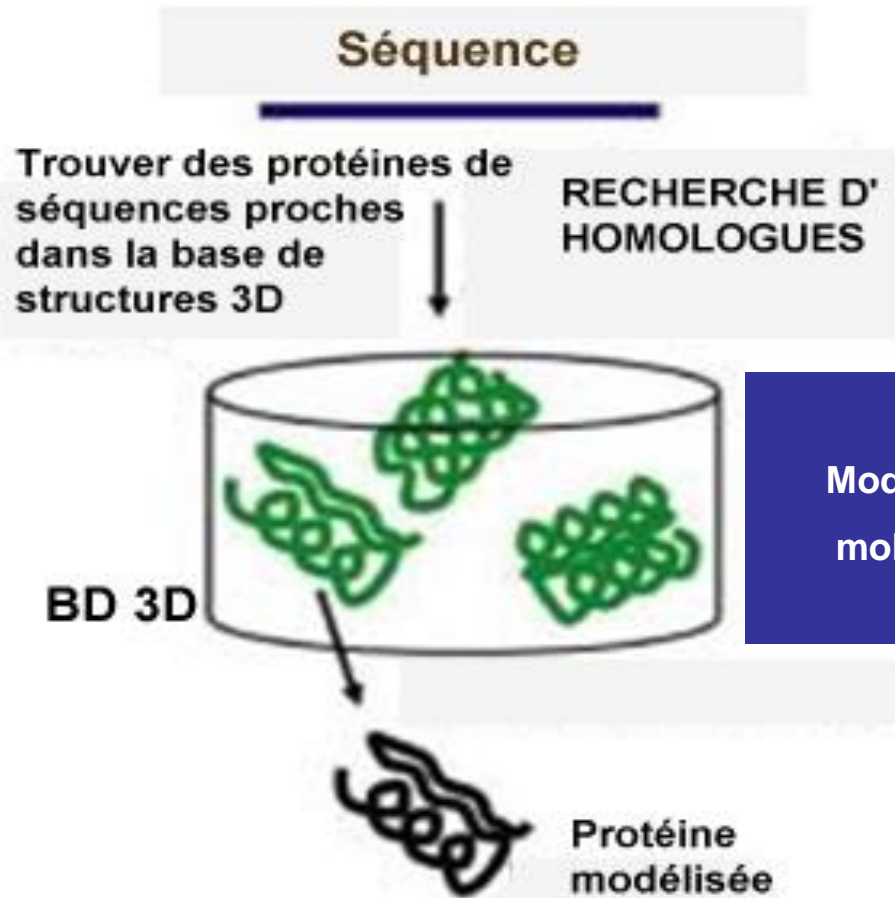




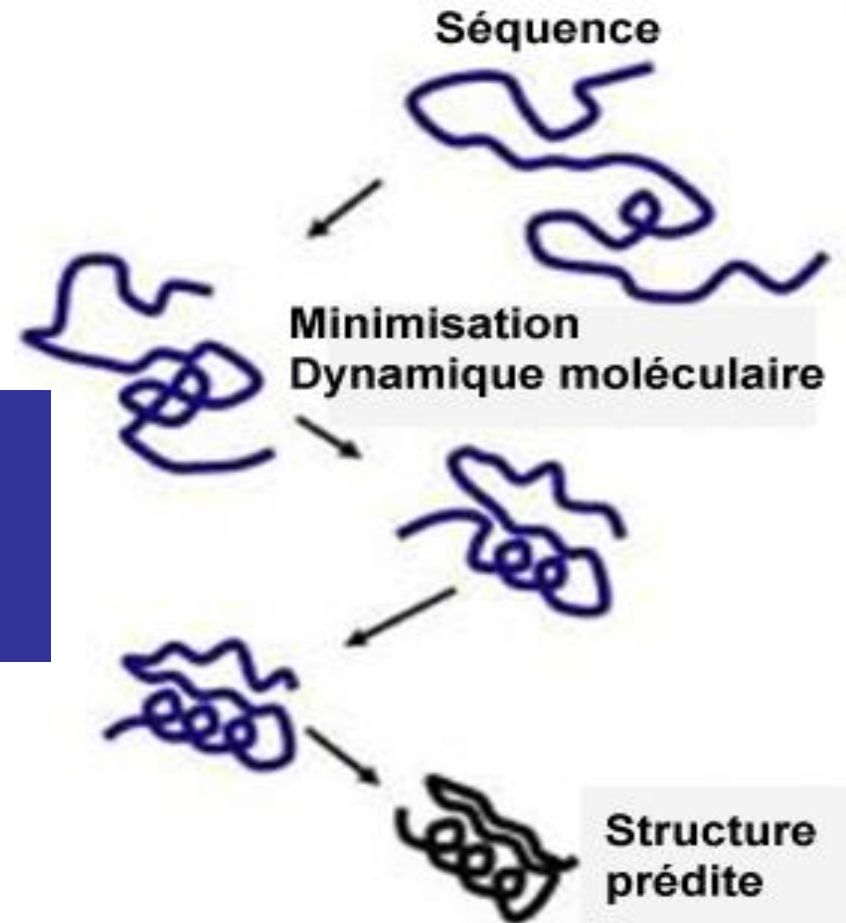
Molecular Type ↓↑	X-ray ↓↑	NMR ↓↑	EM ↓↑	Multiple methods ↓↑	Neutron ↓↑	Other ↓↑	Total ↓↑
Protein (only)	132181	11451	3880	160	67	32	147771
Other	7964	92	447	6	0	4	8513
Protein/NA	7016	265	1354	3	0	0	8638
Nucleic acid (only)	2081	1300	47	5	2	1	3436
Total	149242	13108	5728	174	69	37	168358

Method	Description	# of Clusters
blast	100% identity	68591
blast	95% identity	55426
blast	90% identity	52481
blast	70% identity	45807
blast	50% identity	38824
blast	40% identity	33936
blast	30% identity	28372

Comparative (par homologie)



Prédiction *ab initio*

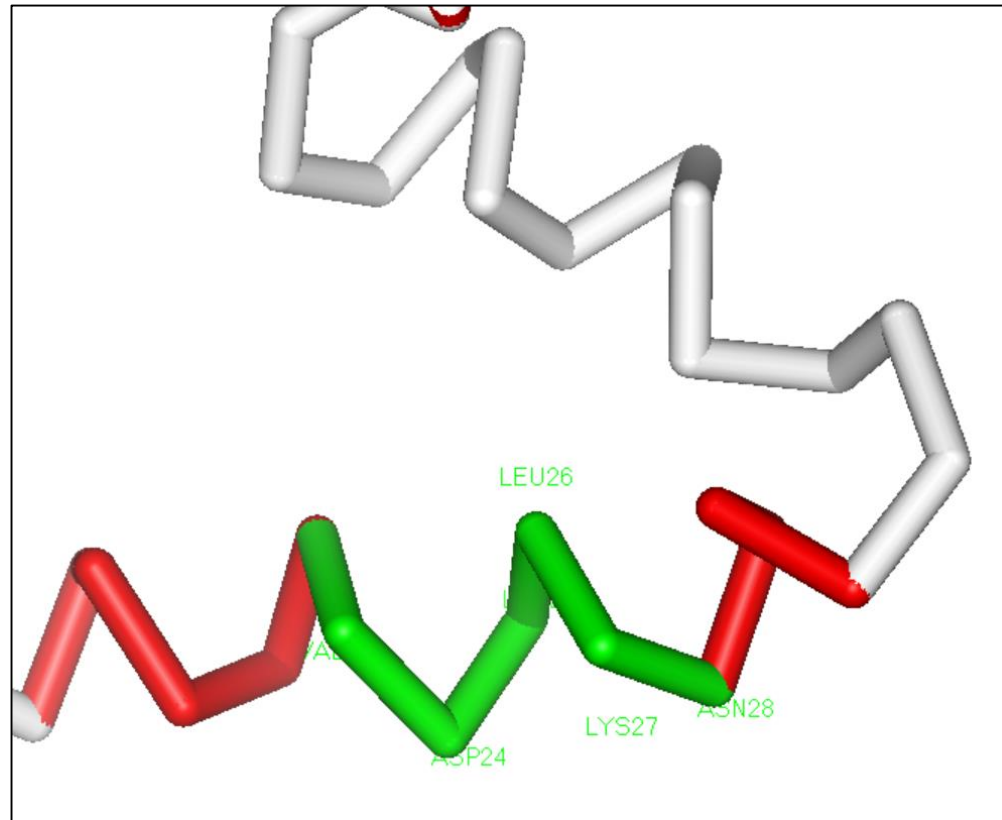


2 protéines homologues (même ancêtre commun),
partagent le même repliement (structures 3D proches)

VDLLKN

23

28

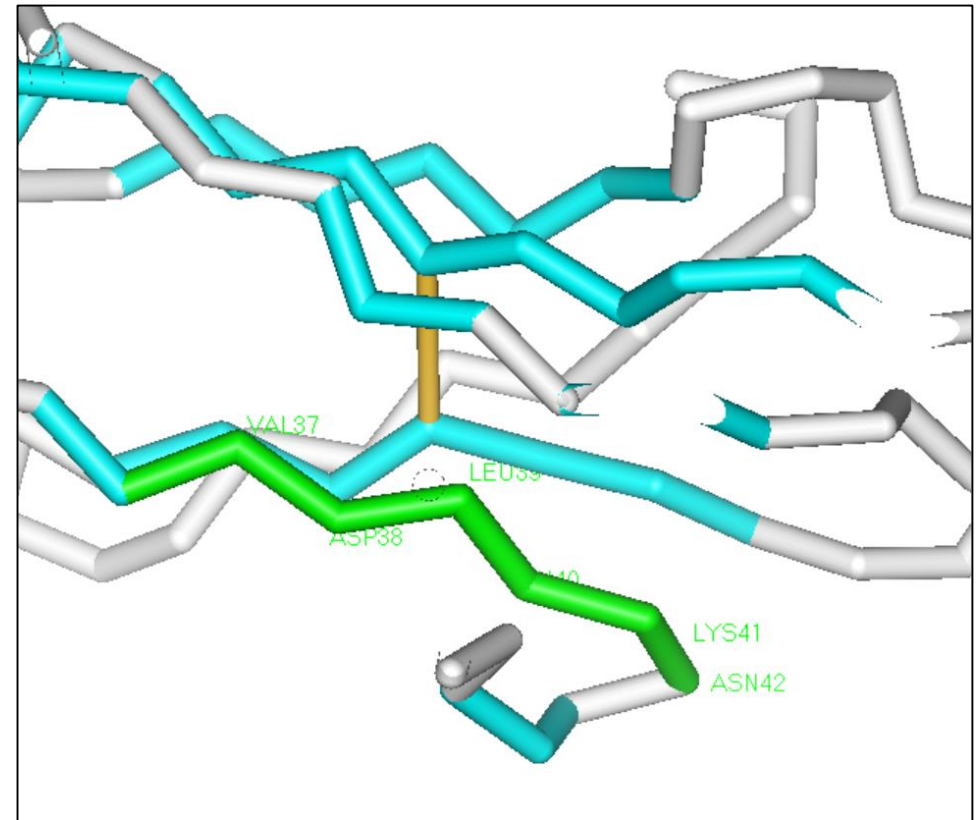


pdb : 1CO0 A

VDLLKN

37

42



pdb : 1B0G A

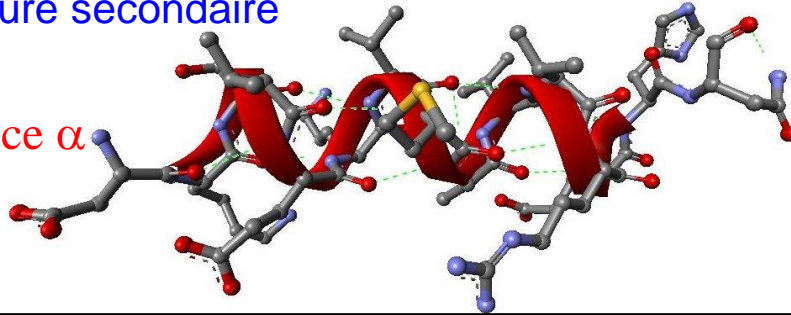
Structure primaire = séquence = mot écrit avec un alphabet de 20 lettres

MKLD**E**IARLAGVSRRTTASYVINGKAKQYRVSDKTVEKVMMAVVREHNYHPNAVAAGLRAGR

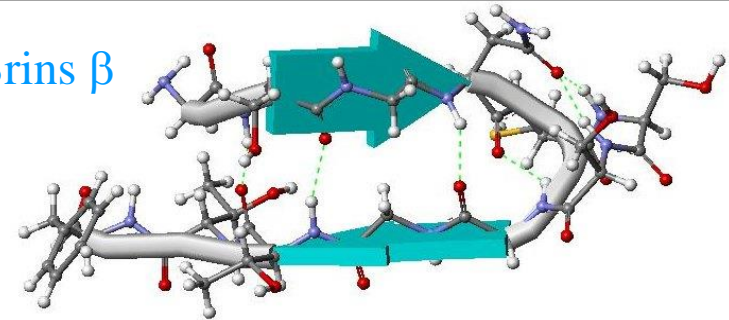


Structure secondaire

Hélice α



Brins β

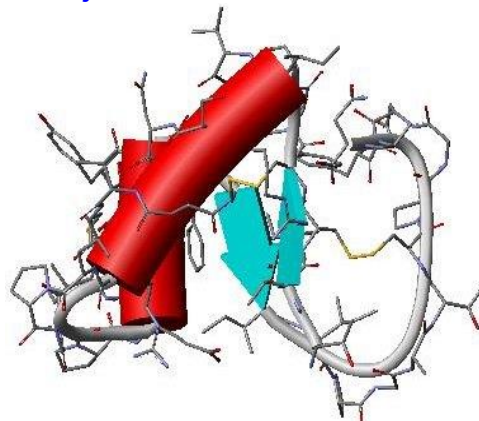


Structure secondaire = mot alphabet de 3 à 10 lettres

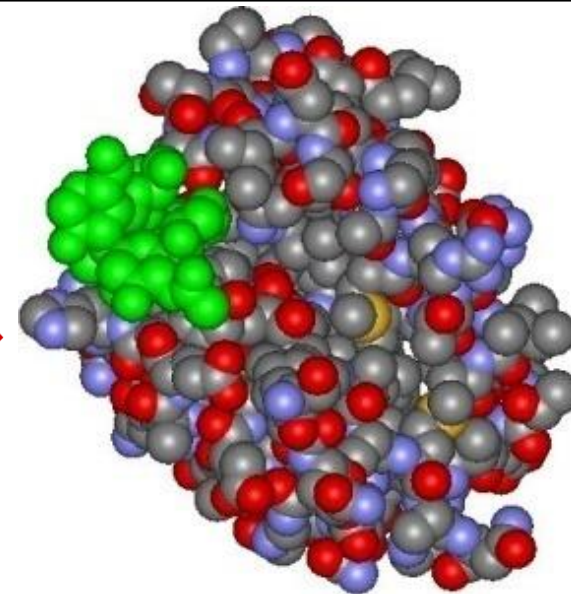
CCHHHHHHHHHHCCCEEEETTTEEEEECCCCHHHHHHHHHHHCCHHHHHHHHHHGCCCC



Structure tertiaire = objet 3D



Fonction



1. Dill K.A., Chan H.S., Nat. Stru. Biol. 1997, 4, 10-19 From Levinthal to pathways to funnels
2. Daggett V., Fersht A.R., TIBS 2003, 28, 18-25 Is there a unifying mechanism for protein folding ?
3. Radford S.E., Dobson C.M., Evans P.A., Nature 1992 358,302-358., The folding of hen lysozyme involves partially structured intermediates and multiple pathways.
4. Gilquin B., Guilbert C. Perahia D., Proteins 2000, 41, 58-74 Unfolding of hen egg lysozyme by molecular dynamics simulation at 300K: Insight into the role of the interdomain interface.
5. Hünenberger P.H., Mark A.E., van Gunsteren W.F. Proteins 1995, 21, 196-213 Computational approaches to study protein unfolding: hen egg white lysozyme as a case study.
6. Kazmirski S.L., Daggett V. J Mol Biol. 1998, 284,793-80. Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme.
7. Paci E., Karplus M., PNAS 2000, 97,6521-6526, Unfolding proteins by external forces and temperature : the importance of topology and energetics.
8. Boczko E.M., Brooks III C.L. Science 1995, 269, 393-396 First-principles calculation of the folding free energy of a three-helix bundle protein.
9. Garcia A.E., Onuchic J.N., PNAS 2003, 100, 13898-13903 Folding a protein in a computer : an atomic description of the folding/unfolding of protein A.



Christian Anfinsen (1916-1995)

**Ribonucléase native, active,
thermodynamiquement stable**

**Ribonucléase, renaturée active,
thermodynamiquement stable**

**+ UREE et
 β Mercaptoethanol**



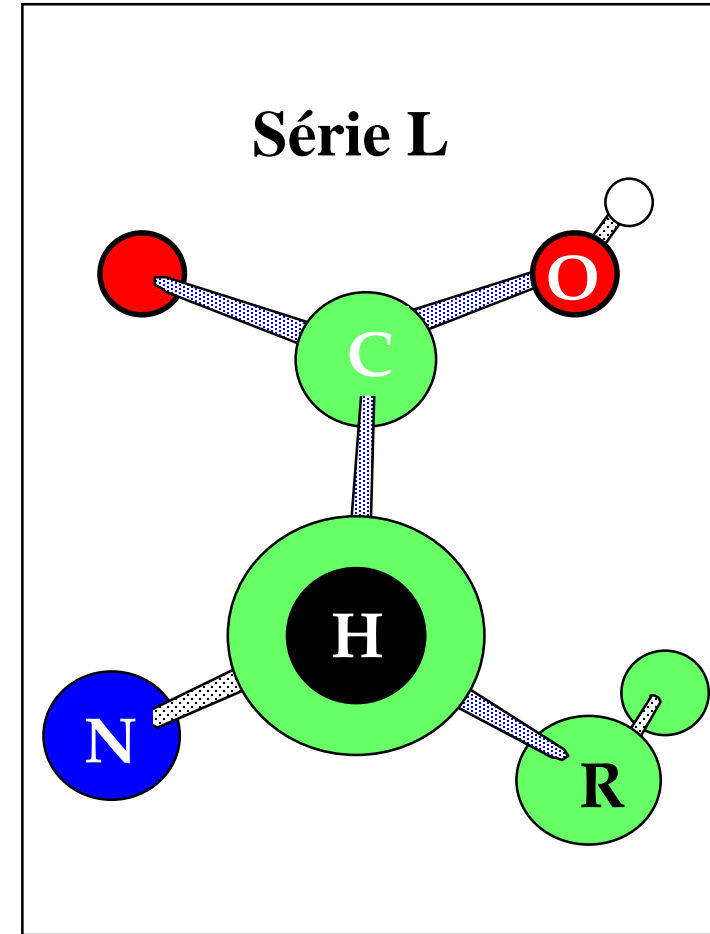
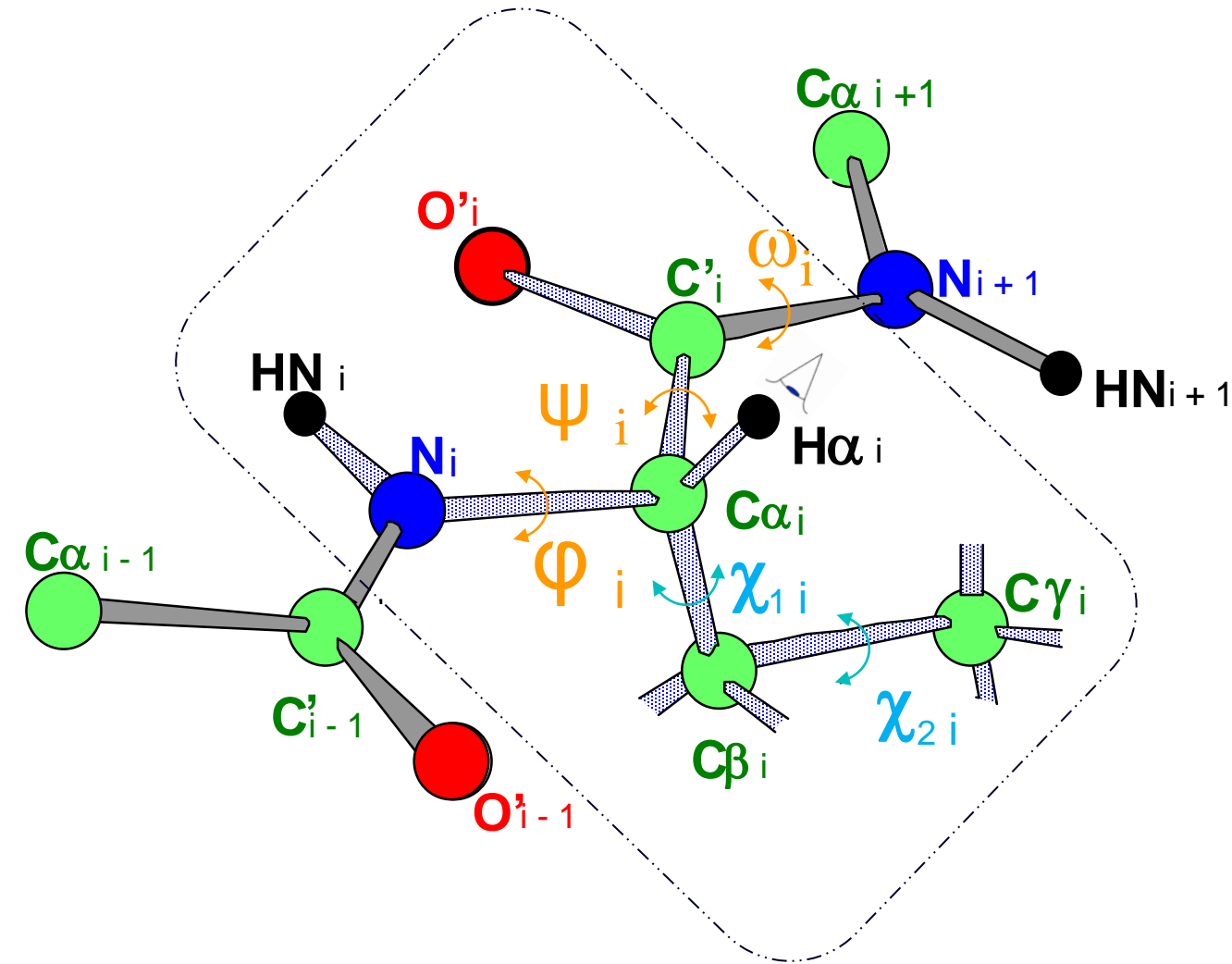
**Quelques ms à 48h
- UREE et
 β Mercaptoethanol**



Structure instable
Plus d'activité
Information de séquence
Attention exceptions

Géométrie du squelette protéique

Degrés de liberté dans la chaîne protéique

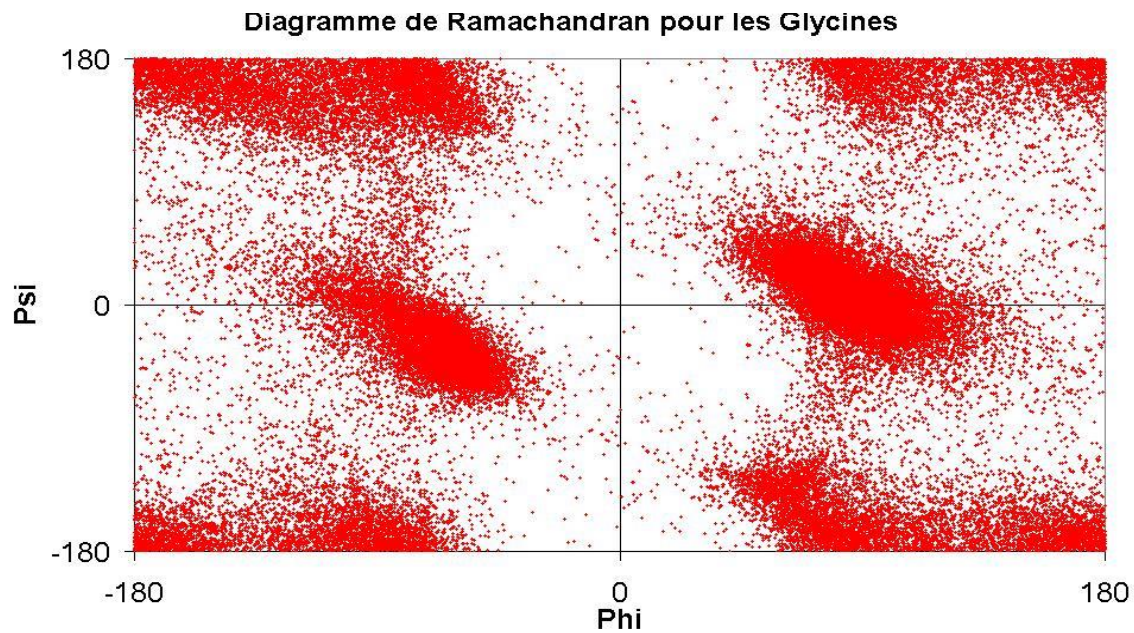
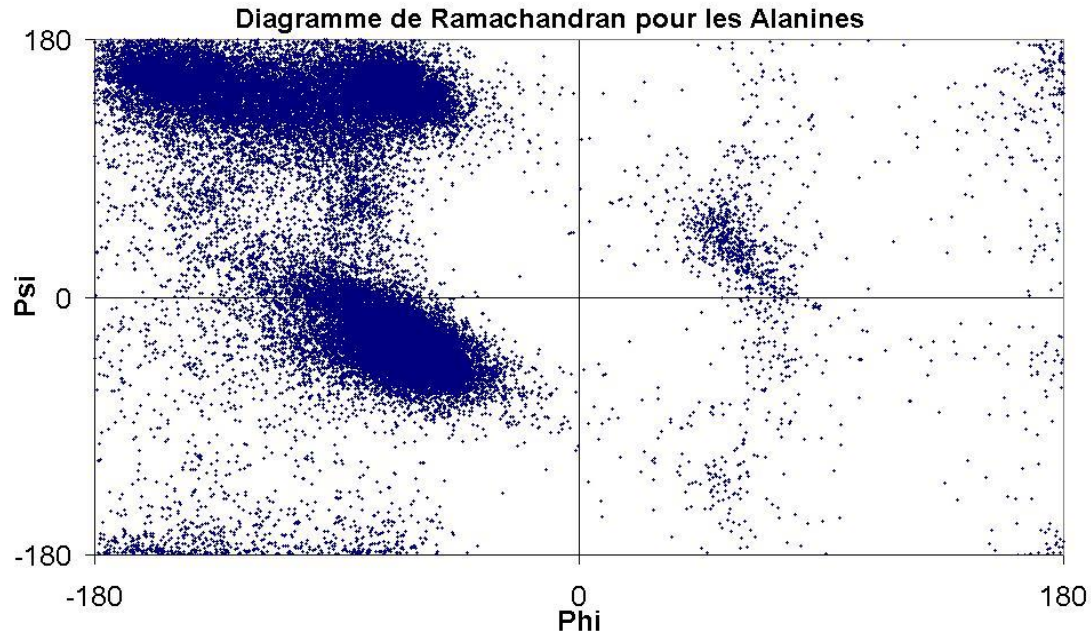


Lire « **C O R N** »

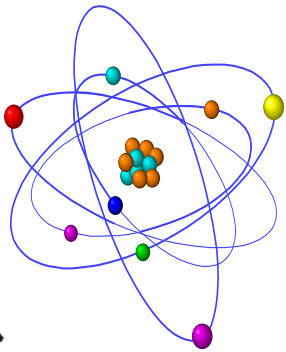
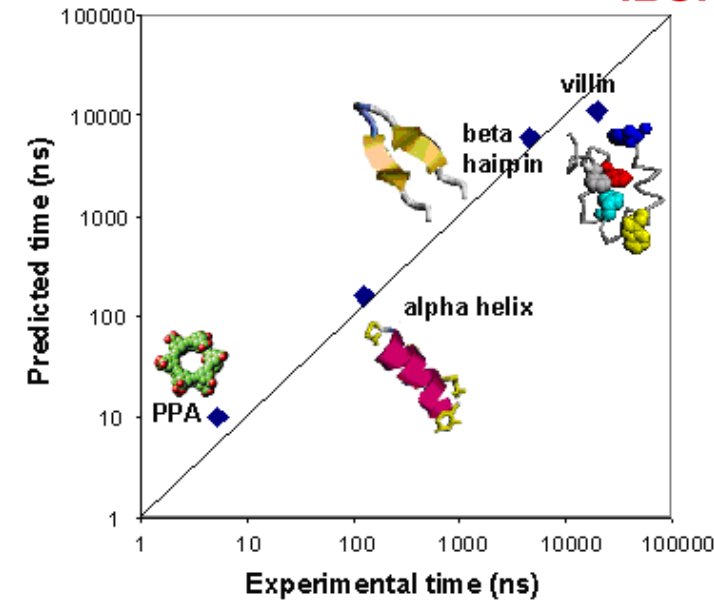
φ = l'angle dièdre définit par: $CO(i-1)-NH(i)-Ca(i)-CO(i)$

ψ = l'angle dièdre définit par: $NH(i)-Ca(i)-CO(i)-NH(i+1)$

Diagramme de Ramachandran



- Soit une petite protéine (100 aa)
- 20^{100} séquences théoriquement possibles
- Supposons 10 conformations par aa (10^{100} conformations)
- même si 2 conformations par aa
 - hélice ou non hélice
 - $2^{100} 10^{30}$ conformations
- Durée de vie d'une conformation 0,1 ps
 - | ceci donne 10^{17} s pour que la protéine se replie soit 3 milliards d'années
 - | et en admettant le calcul de l'énergie de 10^{10} molécules/seconde il faudrait 10^{20} s pour simuler par minimisation d'énergie soit 30 milliards de siècles!



Nombre d'atomes dans l'univers $\approx 10^{80}$...



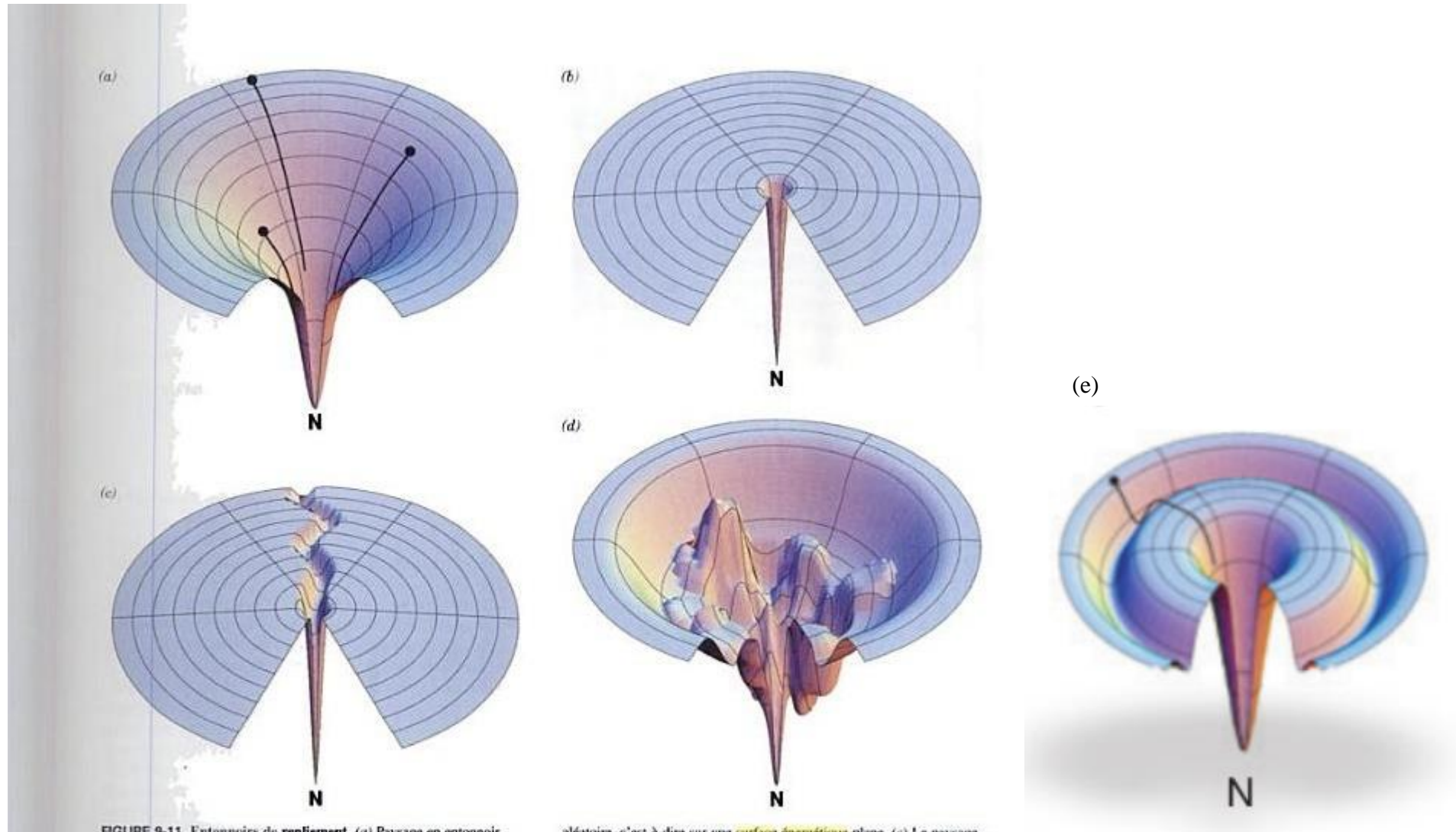
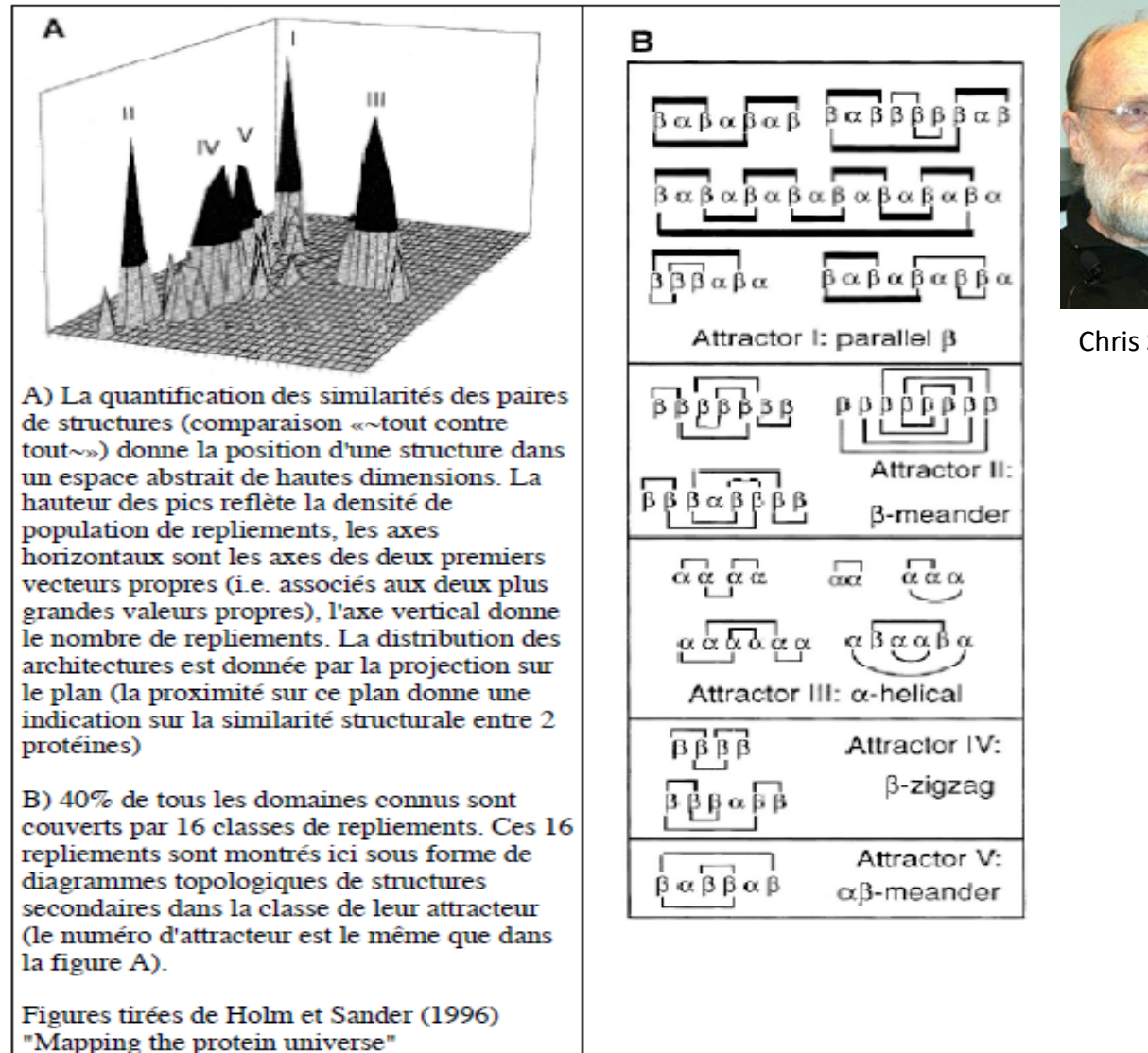


FIGURE 9-11 Entonnoirs de repliement. (a) Paysage en entonnoir idéalisé. À mesure que la chaîne polypeptidique établit un nombre croissant de contacts intracaténaux, son énergie libre interne (sa hauteur au dessus de l'état natif, N) décroît ainsi que sa liberté conformationnelle (la largeur de l'entonnoir). Des polypeptides de conformations différentes (points noirs) suivent différentes voies (lignes noires) pour atteindre l'état natif. (b) Le paysage de Levinthal en « terrain de golf » dans lequel la chaîne polypeptidique doit trouver son état natif (le trou) par recherche

aléatoire, c'est-à-dire sur une surface énergétique plane. (c) Le paysage de repliement classique où la chaîne se déplace au hasard sur une surface énergétique plane jusqu'à ce qu'elle rencontre une gorge qui conduit à l'état natif. (d) Surface énergétique accidentée présentant des minima locaux dans lesquels un polypeptide en cours de repliement peut être piégé momentanément. On pense que les entonnoirs de repliement des protéines naturelles ont cette topographie là. [Avec la permission de Ken Dill, University of California at San Francisco.]

On peut estimer selon les différentes classifications qu'il y a environ 1500 familles de structures. En considérant les protéines comme des points dans un espace à haute dimension - espace des «formes» où les distances structurales entre protéines sont respectées, Holm et Sander (Holm and Sander, 1996) ont mis en évidence cinq «attracteurs» pour toutes les structures.

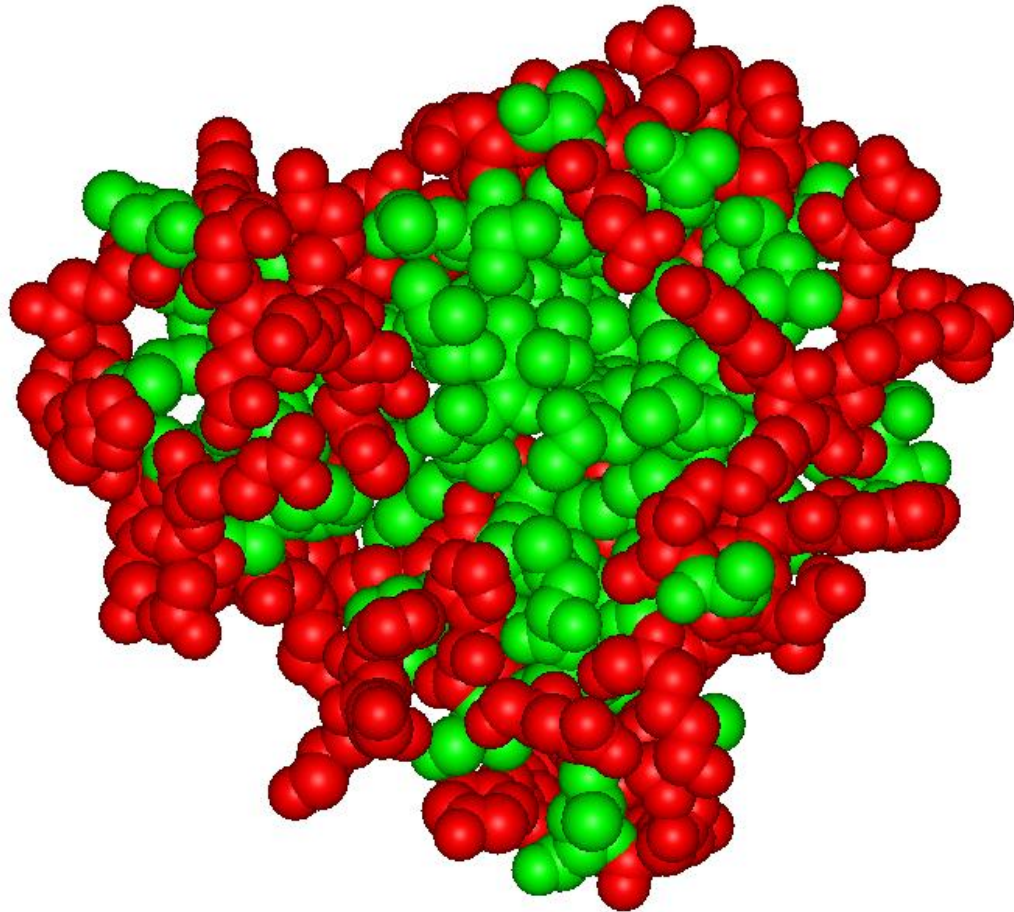
Le nombre de repliements est fini estimé de 1000 à 8000 selon les études (Chothia, 1992; Orengo et al., 1994; Wang, 1996; Wang, 1998; Govindarajan et al., 1999).



Chris Sander

Protéine	N aa	Conformations
Met-enkephaline	5	$5,9 \times 10^4$
Glucagon	29	$4,7 \times 10^{27}$
Insuline	51	$4,6 \times 10^{48}$
Cytochrome C	104	$1,7 \times 10^{99}$
Myogloïne	153	$9,9 \times 10^{145}$
Chymotrypsine	241	$9,4 \times 10^{229}$

Le confinement des résidus **hydrophobes** au cœur de la protéine est une des forces motrices du repliement.



hydrophobes

hydrophiles

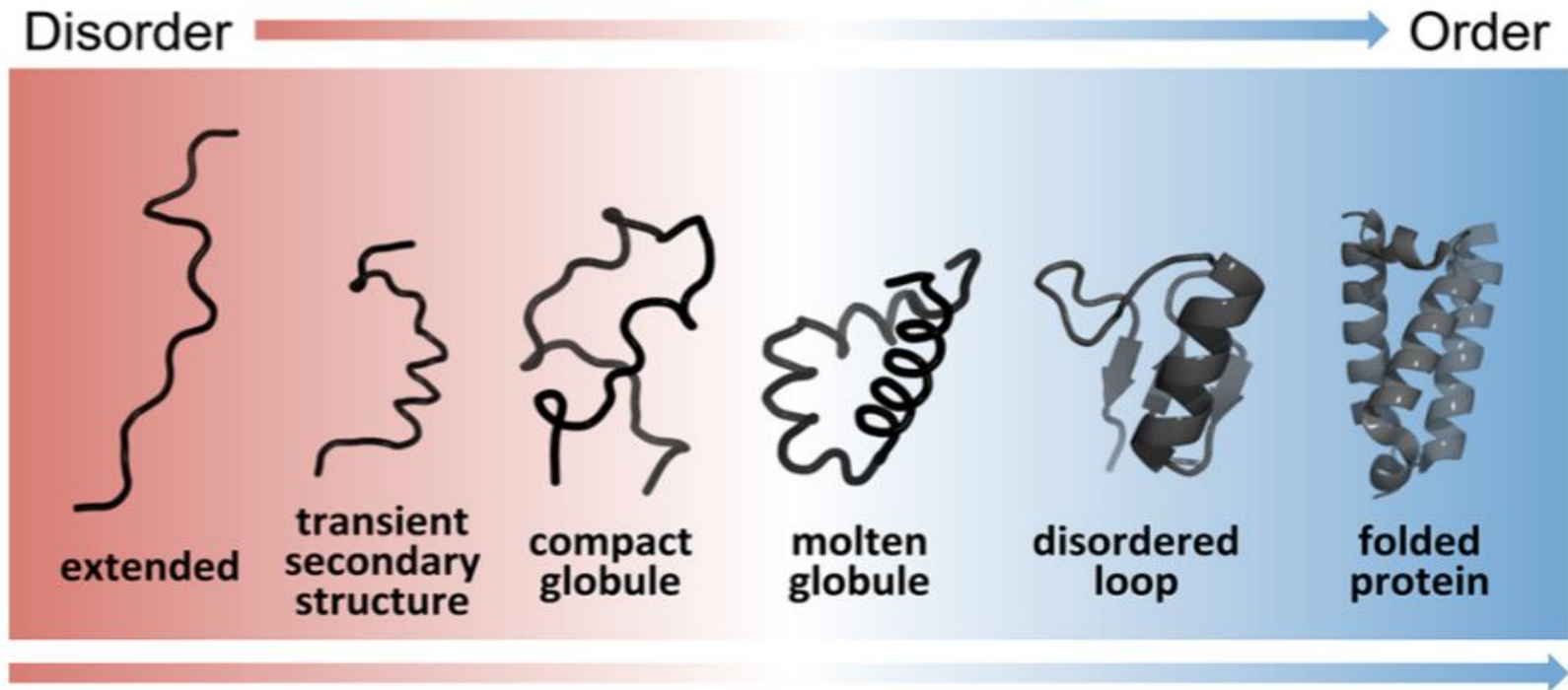
Le rapport entre résidus hydrophiles et Hydrophobes est à peu près constant ($h/H=0.36$ écart type : 0.06).

Les protéines enfouissent une fraction constante de leur surface totale

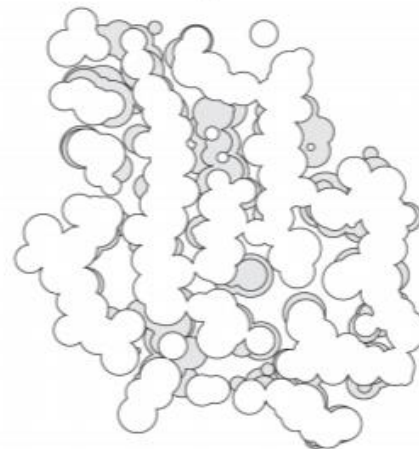
- ✓ C & S sont enfouis à 86%
- ✓ O & N neutres (liaisons H) sont enfouis à 40%
- ✓ O & N chargés sont enfouis à 32%

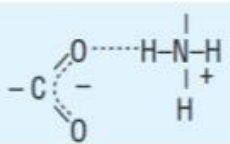
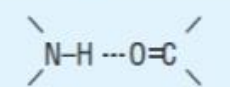
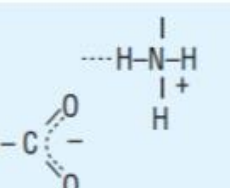
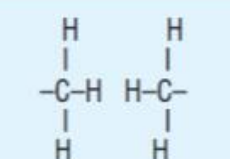
Les liaisons H internes à la protéine contribuent à diminuer la présence de groupes polaires (O, H) à l'intérieur des protéines.

L'intérieur des protéines est aussi dense que des cristaux de petites molécules organiques



Compaction



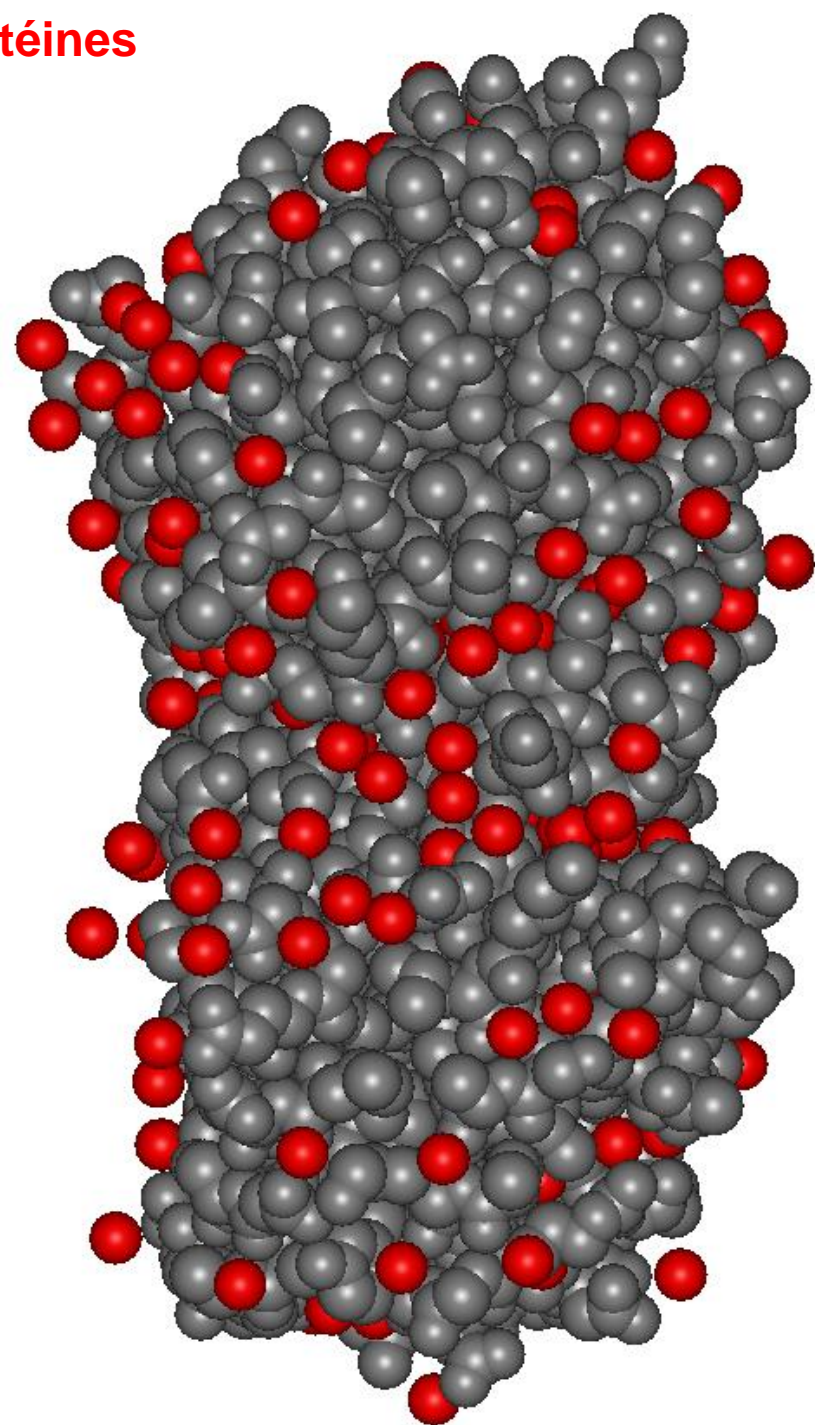
Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	$-C_{\alpha}-C-$	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	$-Cys-S-S-Cys-$	-	2.2 Å	167 kJ/mole
Salt bridge		Donor (here N), and acceptor (here O) atoms <3.5 Å	2.8 Å	12.5–17 kJ/mole; may be as high as 30 kJ/mole for fully or partially buried salt bridges (see text), less if the salt bridge is external
Hydrogen bond		Donor (here N), and acceptor (here O) atoms <3.5 Å	3.0 Å	2–6 kJ/mole in water; 12.5–21 kJ/mole if either donor or acceptor is charged
Long-range electrostatic interaction		Depends on dielectric constant of medium. Screened by water. $1/r$ dependence	Variable	Depends on distance and environment. Can be very strong in nonpolar region but very weak in water
Van der Waals interaction		Short range. Falls off rapidly beyond 4 Å separation. $1/r^6$ dependence	3.5 Å	4 kJ/mole (4–17 in protein interior) depending on the size of the group (for comparison, the average thermal energy of molecules at room temperature is 2.5 kJ/mole)

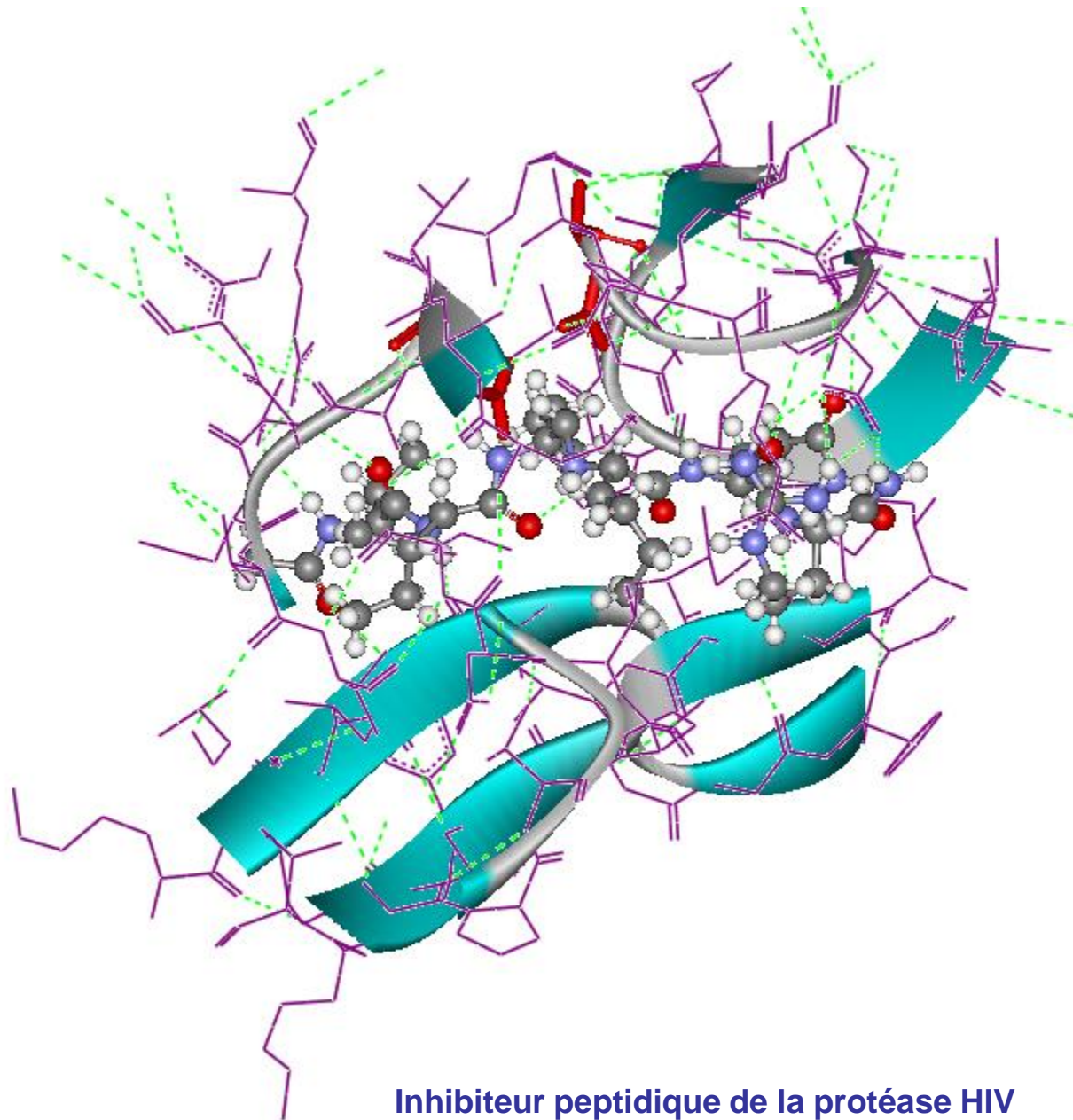
Les forces qui stabilisent la forme repliée:

- ✓ Liaisons hydrogènes intramoléculaires
- ✓ Entropie de déshydratation
- ✓ Interactions hydrophobes
- ✓ Ponts salins
- ✓ Interactions dipolaires

Les forces qui déstabilisent la forme repliée

- ✓ Liaisons hydrogènes avec l'eau
- ✓ Perte d'entropie configurationnelle

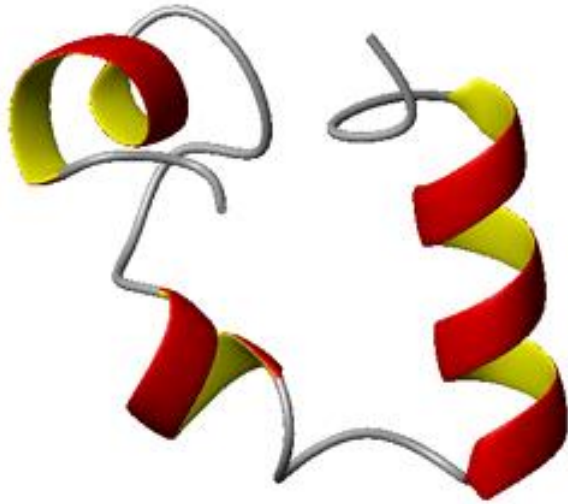




Inhibiteur peptidique de la protéase HIV



Vijay Pande



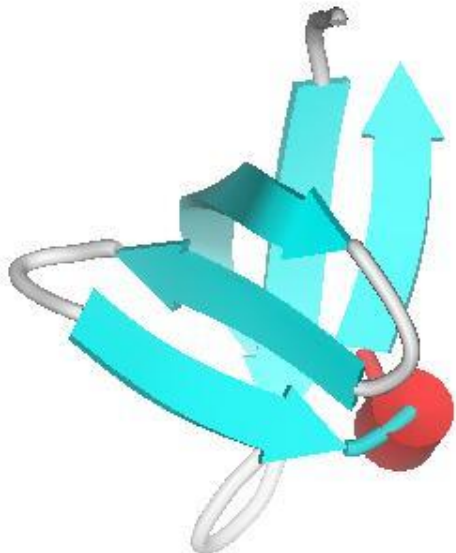
La villine

Simulation du repliement de la villine

Formation d'une hélice

HIV intégrase

Simulation de la dénaturation de HIV intégrase



Cours sur les méthodes

<http://www.gdeleage.fr/prof/PSSP.pdf>

Méthodes d'analyse des séquences

<http://www.gdeleage.fr/prof/IBIS.pdf>

✓ **Chou et Fasman (1974 =>1980)**

- ✓ Peu de paramètres (≈ 120 coefficients)
- ✓ La première méthode utilisable par les biologistes
- ✓ Méthode elle
- ✓ Non reproductible
- ✓ Difficile à implémenter
- ✓ Qualité 52%



Gerald Fasman

✓ **Garnier *et al.* (GOR I, II, III, IV) (1978 =>1989)**

- ✓ Utilise la théorie de l'information
- ✓ Prise en compte de l'environnement séquentiel ($17 \times 4 \times 20 \approx 1360$ coefficients)
- ✓ Méthode automatique non ambiguë
- ✓ Rapide (instantanée)
- ✓ Méthode insensible à l'homologie
- ✓ Qualité 56% à 65%

✓ **Double Prédiction DPM (Deléage et Roux, 1987)**

- ✓ Confrontation de la prédiction de la classe structurale et des % de structures
- ✓ Qualité 60%

✓ **Discrimination linéaire DSC (King et Sternberg, 1996)**

- ✓ Hydrophobie, effets de terminaison, propensions aa, filtrage
- ✓ Utilisation des alignements multiples
- ✓ Qualité 68,5%



Michael Sternberg

- ✓ **Plus proches voisins (Levin *et al.*, 1986; 1988)**
 - ✓ Comparaison de peptides courts ($\approx 10^{10}$ pour 500 acides aminés)
 - ✓ Paramètres (Matrice, seuil, longueur des peptides)
 - ✓ Méthode automatique
 - ✓ Facile à implémenter
 - ✓ Temps de calcul
 - ✓ Qualité 62%
- ✓ **Méthodes auto-optimisées Geourjon et Deléage (SOPM, 1994 ; SOPMA, 1995)**
 - ✓ Méthode automatique non ambiguë
 - ✓ Temps de calcul assez long (5' / séquence)
 - ✓ Méthode sensible à l'homologie
 - ✓ Prise en compte des alignements de protéines homologues
 - ✓ Qualité 65-72%
- ✓ **Réseaux de neurones (PHD, Rost et Sander 1993->1999 ; HNN 1997, Nnpredict, Predator)**
 - ✓ Méthodes optimales
 - ✓ Paramétrage délicat (couches, neurones, jeu test, ré-apprentissage)
 - ✓ Qualité 72-75%
- ✓ **Prise en compte des familles de protéines homologues (PHD, SIMPA96, SOPMA)**
 - ✓ Méthodes modulaires
 - ✓ FASTA, BLAST, CLUSTALW, Prédiction
 - ✓ SWISS-PROT, PDB

Forte évolutivité des méthodes de similarité

- ✓ PHD <https://www.predictprotein.org/>
- ✓ JPRED 4 <http://www.compbio.dundee.ac.uk/~www-jpred/>
- ✓ PSIPred 4 <http://bioinf.cs.ucl.ac.uk/psipred/>
- ✓ NPS@ <https://npsa-prabi.ibcp.fr/>
 - ✓ [SOPM](#) (Geourjon and Deléage, 1994)
 - ✓ [SOPMA](#) (Geourjon and Deléage, 1995)
 - ✓ [HNN](#) (Guermeur, 1997)
 - ✓ [MLRC](#) (Guermeur *et al.*, 1999)
 - ✓ [DPM](#) (Deléage and Roux, 1987)
 - ✓ [DSC](#) (King and Sternberg, 1996)
 - ✓ [GOR I](#) (Garnier *et al.*, 1978)
 - ✓ [GOR III](#) (Gibrat *et al.*, 1987)
 - ✓ [GOR IV](#) (Garnier *et al.*, 1996)
 - ✓ [PHD](#) (Rost and Sander, 1993)
 - ✓ [PREDATOR](#) (Frishman and Argos, 1996)
 - ✓ [SIMPA96](#) (Levin, 1997)
- ✓ Logiciel client/serveur
 - ✓ ANTHEPROT (<http://antheprot-pbil.ibcp.fr>)

| Homologie

- | 2 protéines sont homologues ssi elles ont un ancêtre commun => Structures 3D proches
- | Il est possible d'observer la ressemblance résiduelle entre les séquences originelles après l'évolution, ce qui permet d'inférer l'homologie.
- | En général, pour des séquences de longueur standard, on peut inférer l'homologie entre 2 protéines si leurs séquences présentent 30% ou plus d'identités résiduelles mais...
- | Il existe des séquences homologues avec moins de 30% d'identité dans ce cas là:
 - | Homologie est transitive
 - | si A homologue à B
 - | et B homologue à C
 - | alors A homologue à C même si A et C ont peu de similarités
 - | Utilisation des structures secondaires (recherche d'homologues distants)

HOMOLOGIE

Ancêtre 3D commun
Myoglobine de cachalot
Leghemoglobine de lupin
15% identité de séquences
Fonctions différentes



ANALOGIE

Ancêtres différents
Structures 3D souvent différentes
Fonctions identiques
Protéases



Les structures 3D évoluent moins vite que les séquences
Les structures 3D sont plus préservées par l'évolution que les séquences
La pression de l'évolution est plus forte sur les structures que sur les séquences
Les structures 3D s'accommodent des séquences (plasticité)

Evolution divergence
Structures 3D proches

Evolution convergente
Structures 3D différentes



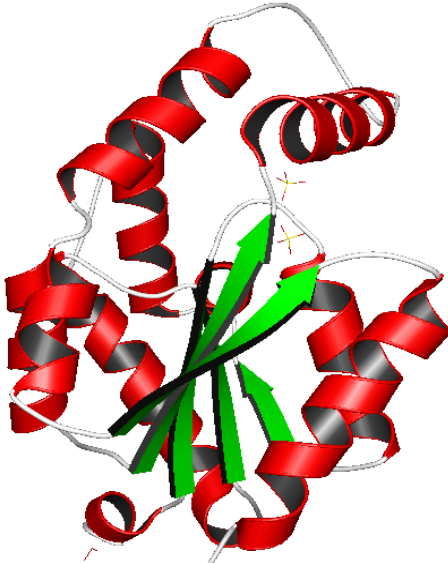
Michael Levitt

- Trouver des protéines représentatives de l'ensemble des protéines
 - Protéines membranaires?
 - Que faire des complexes et des ligands?
 - Structures différentes de la même protéine.
- Assignation automatique des structures 2D à partir structures 3D
 - Critères géométriques (angles phi et psi , Ramachandran) Levitt & Greer
 - Critères énergétiques (liaisons hydrogènes) **DSSP** Kabsch & Sander
 - Combinaison **Psea** (Colloch et al.), **KAKSI** (Martin et al.)
- Fixer un seuil de similarité entre les protéines (50% ou 25% d'identité)
 - Chou & Fasman, 1978, 29 protéines (totalité)
 - Kabsch et Sander (1983) 60 protéines (Id<50%)
 - Geourjon & Deléage (1994) 234 protéines (Id<50%)
 - Rost & Sander (1993) 126 protéines (Id< 25%)
 - Hobohm et Sander (1998) 700 protéines (Id<25%)
 - PDB 2015 24000 protéines (id<30%)
- Comment intégrer les familles de séquences de protéines?
 - Alignements multiples (qualité, exhaustivité, mise à jour)
- Combien d'états?

Prédiction (prévision) de structures secondaires: Utilisation des séquences ~75%

Séquence
MREDKL => Structure prédite
CCHHHC

Assignation (déduction) de structures secondaires : Déduction à partir des structures 3D ~85%

Structure expérimentale	Structure secondaire « observée »	Séquence
	<p>=> CCHHHH</p>	<p>~ MREDKL</p>

```
>1ACX-1ANTIBACTERIALPROTEIN17-DE
APAFSVSPASGASDGQSVSVSVAAGETYIYAQCAPVGGQDACNPATATSFTTASGAASFSTVRKSYAGQTPSGTPVGS
VDCATDACNLGAGNSGLNLGHVALTF*
CEEEEECCCCCCCCCEEEEEEECCCEEEEEEECEETTEECCCTTTCCEEECCCCCCCCCEEECCCEEEEEECTTCCEEE
EETTTCCCEEEEECCCCCCCCCCCC*

>1AK3-ATRANSFERASE (PHOSPHOTRANSFERASE) 17-JA
RLLRAIMGAPGSGKGTVSSRITKHFELKHLSSGDLRLDNMLRGTEIGVLAKTFIDQGKLIIPDDVMTRLVLHELKNLTQYN
WLLDGFPRTLPQAEALDRAYQIDTVINLNVPFVVIKQRLTARWIHPGSGRVYNIENPPKTMGIDDLTGEPLVQREDDRP
ETVVKRLKAYEAQTEPVLEYRKKGVLETFSGTETNKIWPVHYAFLQTKLPQRS*
CEEEEECCCTTCCHHHHHHHHHHHCCCEEEHHHHHHHHHTTTCHHHHHHHHHHTTCCCCHHHHHHHHHHHHTTCCCC
EEEECCCCCHHHHHHHHTTCCCCEEEEEECCHHHHHHHHTCEEETTTEEEETTCCCCCTTCTTTCCCCCCTTCC
HHHHHHHHHHHHHHHHHHHHHTTCEEEEECCCHHHHHHHHHHHHTTCCCC*

>1AZU-1ELECTRONTRANSPORT (COPPERBINDING) 04-AU
SVDIQGNDQMVFNTNAITVDKSKQFTVNL SHPGNL PKNVMGHNWVLSTAADMQGVVTDGMASGLDKDYLPDDSRVIAH
TKLIGSGEKDSVTFDVS KLKEGEQYMFCTFPGHSALMKGTLTL*
CEEECCCCCCCCCECCCCCEEEEEEECCCCCTTCCCCCEEEETTTHHHHHHHHHHCHHHHCCCCCTTCCCC
CCCCCTTCEEEEEEECCCCCCCCCEEECCCTTTTTTCEEEEC*

>1BBP-ABILINBINDING19-SE
NVYHDGACPEVKPVDNFDWSNYHGKWEVAKYPNSVEKYGKCGWAEYTPGKSVKVSNYHVIHGKEYFIEGTAYPVGDSX
XXKIGKIYHKLTYGGVTKENVFNVLSTDNKNYIIGYYCKYDEDKKGHQDFVWVLSRSKVL TGEAKTAVENYLIGSPVXD
SQKLVYSDFSEAACKVN*
CEEECCCCCCCCCCCCCHHHCCCEEEEEEECCCTTTTCEEEEEEECCCEEEEEEEETTEEEEEEEEEEECCCTC
CCTCEEEEEEEETTEEEEEEEEEEECCCEEEEEEEETTTEEEEEEECCCCCHHHHHHHHHHHHCCCC
HHHCECCCHHHHCC*

>1BDS-1ANTI-HYPERTENSIVEANTI-VIRALPROTEIN14-NO
AAPCFCSGKPGRDLWILRGTCPPGGYGYTSNCYKWPNICCYPH*
CCCCCCCCCCCCCEEECCCCCTTCCCCEEETTEEECC*

>1BMV-1VIRUS09-OC
SISQQTVWNQMATVRTPLNFDSSKQSFQFSVDLLGGGISVDKTGDWITLVQNSPISNLLRVAAWKKGCLMVKVMSGNA
```

Quelle Méthode?

- ✍ DSSP (Kabsch & Sander, 1983) basé sur le réseau de liaison hydrogène
- ✍ Define (Richards & Kundrot, 1988) basé sur les distances $C\alpha$ pour des structures secondaires idéales
- ✍ P-Curve (Sklenar & al) basé sur les propriétés de l'axe hélicoïdal
- ✍ Stride : extension de DSSP avec prise en compte des angles de torsion.
- ✍ Psea (Labesse et al) : basé sur les angles α et τ décrivant respectivement 3 et 4 $C\alpha$ consécutifs.

Combien d'états?

- **Kabsch & Sander: 8 états**
 - H: α -helix
 - G: 3_{10} helix
 - I: π -helix
 - E: β -strand
 - B: bridge
 - T: β -turn
 - S: bend
 - C: coil
- **Classique 3 états**
 - H: α -helix
 - E: β -strand
 - C: coil

L'assignation des structures secondaires:

Calcul de l'accord?

Analyse automatique des structures 3D par le C3

AA	WDKYAQEVYEMNFG EKPEGDITQV
DSSP	CCC HHHHHHHHHHH CCCCCCCCC
STRIDE	CC HHHHHHHHHHH CCCCCCCCC
PSEA	CC HHHHHHHH CCCCCCCCCCCCC
PCURVE	CC HHHHHHHHHH CCCCCCCCCCCCC

$$C_3 = \frac{N_C}{N_T}$$



DSSP	CCC HHHHHHHHHHH CCCCCCCCC
STRIDE	CC HHHHHHHHHHH CCCCCCCCC
	#***##*****

$$C_3^{DSSP/STRIDE} = \frac{21}{24} = 87,5\%$$

	DSSP	STRIDE	DEFINE	PCURVE	SECSTR	XTLSSTR	KAKSI	SEGNO	PSEA
STRIDE	95,05								
DEFINE	56,26	55,88							
PCURVE	86,33	87,34	68,19						
SECSTR	94,23	92,88	55,41	85,59					
XTLSSTR	86,14	86,81	54,04	82,21	84,95				
KAKSI	87,07	89,16	87,07	85,52	87,08	87,20			
SEGNO	88,11	89,00	56,19	86,13	86,78	85,92	85,78		
PSEA	86,71	87,09	59,38	85,97	85,33	82,77	86,04	86,69	

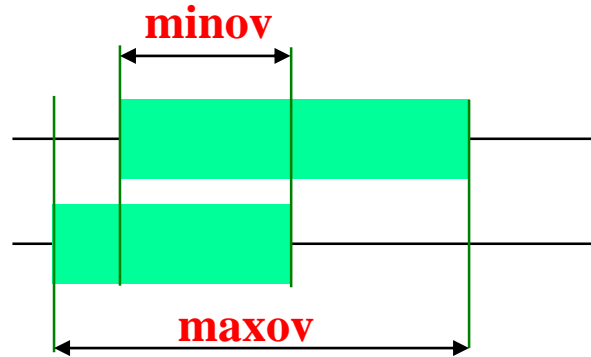
ZZ : Zhou & Zhou (*Protein Sci.*, 2003)

- Sov coefficient (Structural Overlap) (Rost *et al.*, 1994 ; Zemla *et al.*, 1999)

$a^2 + b^2 = c^2$

$$\text{Sov} = 100 \times \left[\frac{1}{N} \sum_{i \in [H,E,C]} \sum_{S(i)} \frac{\text{minov}(s_q, s_t) + \delta(s_q, s_t)}{\text{maxov}(s_q, s_t)} \times \text{len}(s_q) \right]$$

- minov* : longueur de la structure secondaire chevauchante entre la source S_q et la cible S_t
- maxov* : longueur maximale des structures secondaires chevauchantes entre la source S_q et la cible S_t



Burkhard Rost

- δ est défini par :

$$\delta(s_q, s_t) = \min \left\{ \begin{array}{l} (\text{maxov}(s_q, s_t) - \text{minov}(s_q, s_t)); \text{minov}(s_q, s_t); \\ \text{int}(\text{len}(s_q/2)); \text{int}(\text{len}(s_t/2)) \end{array} \right\}$$

DSSP	CCC HHHHHH C HHHH CCCCCCCCC
STRIDE	C HHHHHH CC HHHHHHH CCCCCCC
	***** *****
	---- ----

Exemple: $sov(H) = 86,7\%$

Méthode*	SOV
SECSTR	92,59
STRIDE	91,69
PSEA	90,51
XTLSSTR	82,29
SEGNO	83,57
KAKSI	73,92
PCURVE	70,51
DEFINE	64,06



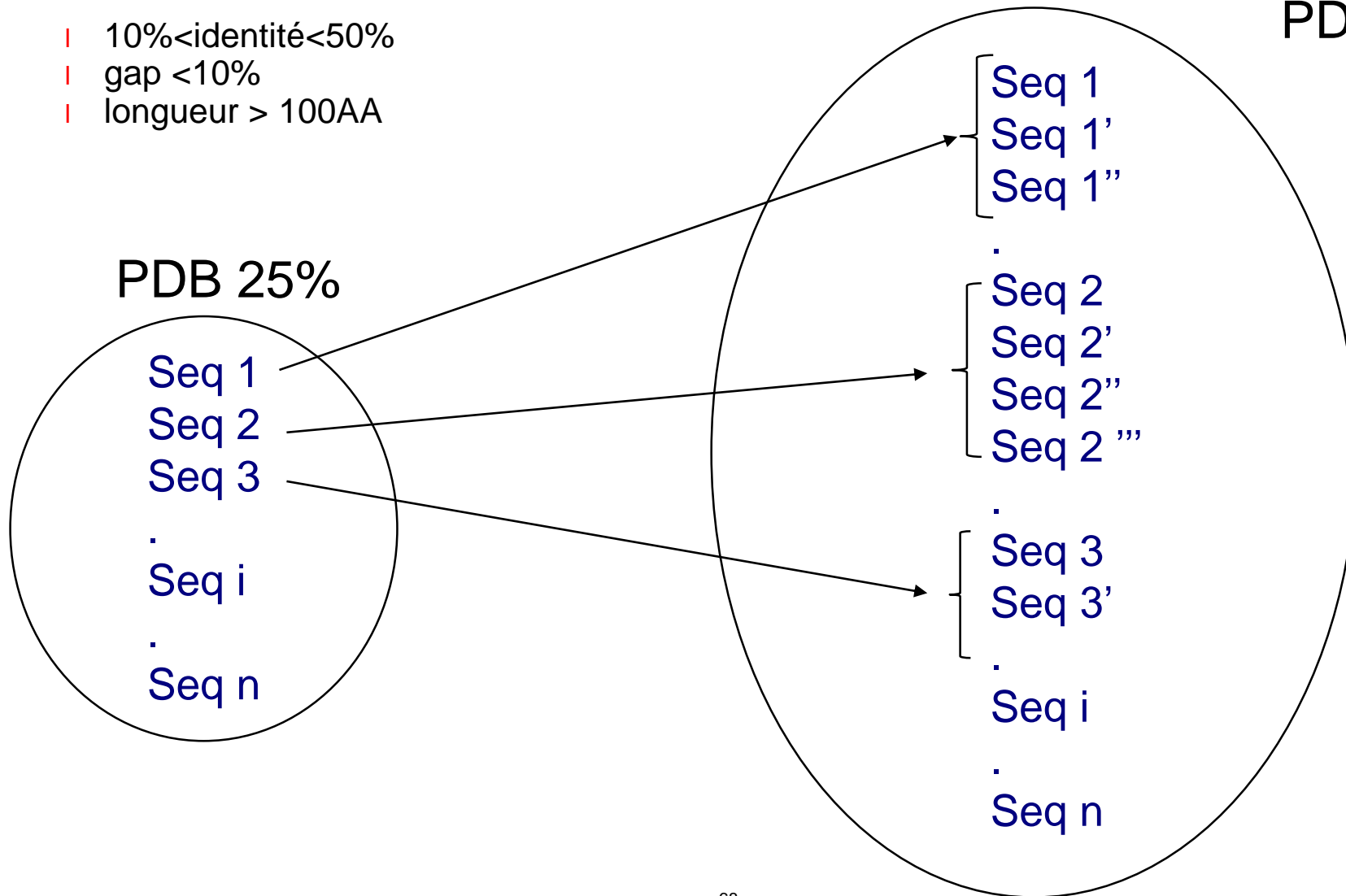
Relation séquence, structures secondaires, structure 3D

Ssearch (E=10)

- Recherche de similarité entre paires (pdb_select_25 vspdb_select_95)
- Sélection des fragments

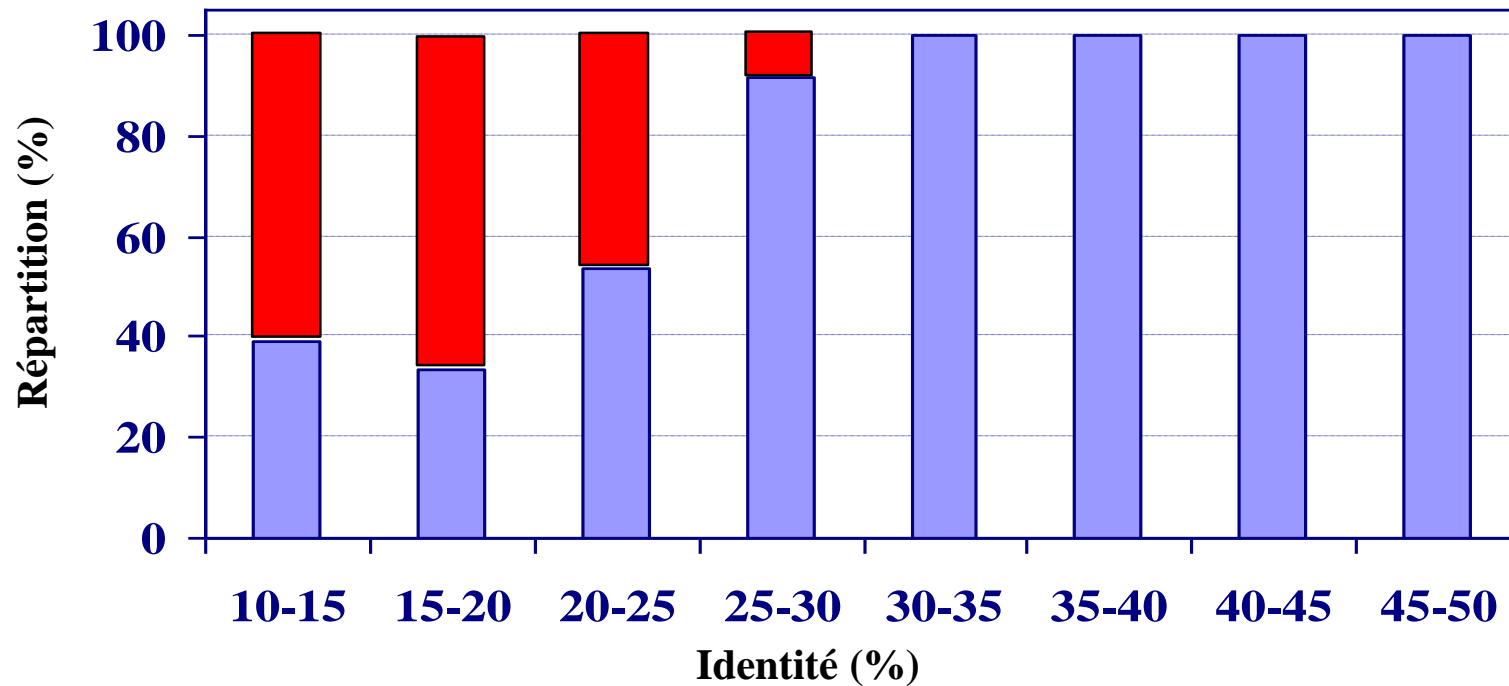
- 10% < identité < 50%
- gap < 10%
- longueur > 100AA

PDB 95%



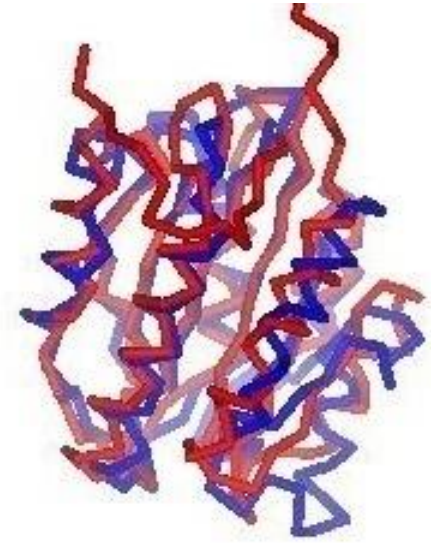
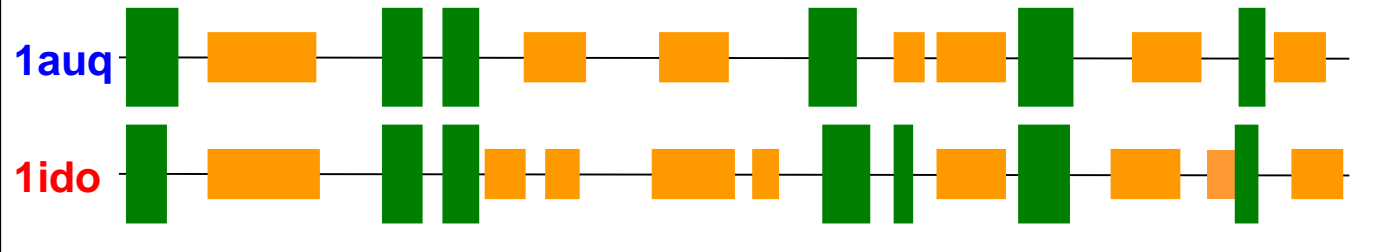
- Classification selon la structure 3D en paire de structures « proches » ou « différentes » au niveau structural

Ssearch (E=10)

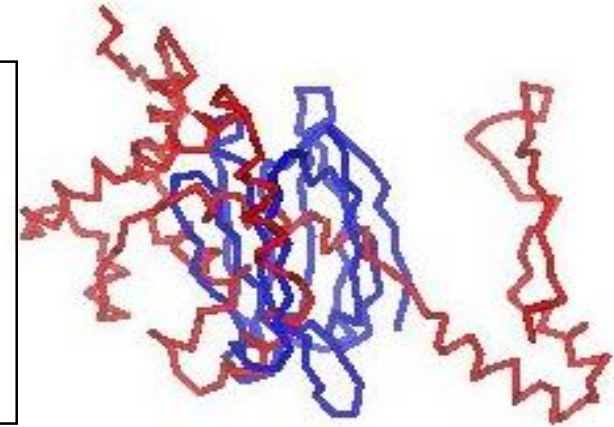
Geourjon, C., *et al.* Protein Science (2001) 10, 788-797

>1ACX-1ANTIBACTERIALPROTEIN17-DE
APAFSVSPASGASDGQSVSVSVAAGETYIYAQCAPVGGQDACNPATATSFTTASGAASFSTVRKSYAGQTPSGTPVGS
VDCATDACNLGAGNSGLNLGHVALTF*
CEEEEECCCCCCCCCEEEEEEECCCEEEEEEECEETTECCCTTTCCEEECCCCCCCCCEEEEECCCEEEEEECTTCCEEE
EETTTCCCEEEEECCCCCCCCCCCC*
>1AK3-ATRANSFERASE (PHOSPHOTRANSFERASE) 17-JA
RLLRAIMGAPGSGKGTVSSRITKHFELKHLSSGDLRLDNMLRGTEIGVLAKTFIDQGKLIIPDDVMTRLVLHELKNLTQYN
WLLDGFPRTLPQAEALDRAYQIDTVINLNVPFVVIKQRLTARWIHPGSGRVYNIENPPKTMGIDDLTGEPLVQREDDRP
ETVVKRLKAYEAQTEPVLEYRKKGVLETFSGTETNKIWPVHYAFLQTKLPQRS*
CEEEEECCCTTCCHHHHHHHHHHHCCCEEEHHHHHHHHHTTTCHHHHHHHHHHTTCCCCHHHHHHHHHHHHTTCCCC
EEEECCCCCHHHHHHHHTTCCCCEEEEEECCHHHHHHHHTCEEETTTEEEETTCCCCCTTCTTTCCCCCCTTCC
HHHHHHHHHHHHHHHHHHHHHTTCEEEEECCCHHHHHHHHHHHHTTCCCC*
>1AZU-1ELECTRONTRANSPORT (COPPERBINDING) 04-AU
SVDIQGNDQMVFNTNAITVDKSKQFTVNL SHPGNL PKNVMGHNWVLSTAADMQGVVTDGMASGLDKDYLPDDSRVIAH
TKLIGSGEKDSVTFDVS KLKEGEQYMFCTFPGHSALMKGTLTL*
CEEECCCCCCCCCECCCCCEEEEEEECCCCCTTCCCCCEEEETTTHHHHHHHHHHCHHHHCCCCCTTCCCC
CCCCCTTCEEEEEEECCCCCCCCCEEECCCTTTTTTCEEEEC*
>1BBP-ABILINBINDING19-SE
NVYHDGACPEVKPVDNFDWSNYHGKWWEVAKYPNSVEKYGKCGWAEYTPGKSVKVSNYHVIHGKEYFIEGTAYPVGDSX
XXKIGKIYHKLTYGGVTKENVFNVLSTDNKNYIIGYYCKYDEDKKGHQDFVWVLSRSKVL TGEAKTAVENYLIGSPVXD
SQKLVYSDFSEAACKVN*
CEEECCCCCCCCCCCCCHHHCCCEEEEEEECCCTTTTCEEEEEEECCCEEEEEEEETTEEEEEEEECCTC
CCTCEEEEEEEETTEEEEEEEEEECCCEEEEEEEETTTEEEEEEECCCCCHHHHHHHHHHHHCCCC
HHHCEEECCCHHHCC*
>1BDS-1ANTI-HYPERTENSIVEANTI-VIRALPROTEIN14-NO
AAPCFCSGKPGRDLWILRGTCPPGGYGYTSNCYKWPNICCYPH*
CCCCCCCCCCCCCEEECCCCCTTCCCCCEEEETTEEEEC*
>1BMV-1VIRUS09-OC
SISQQT VWNQMATVRTPLNFDSSKQSFQFSVDLLGGGISVDKTGDWITLVQNSPISNLLRVAAWKKGCLMVKVMSGNA

15,9% identity ; Close structures (RMSD 2Å)



16% identity ; Different structures (RMSD = 20 Å)

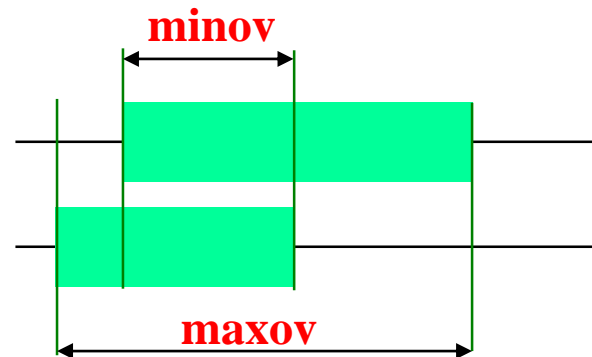


- Sov coefficient (Structural Overlap) (Rost *et al.*, 1994 ; Zemla *et al.*, 1999)

$a^2 + b^2 = c^2$

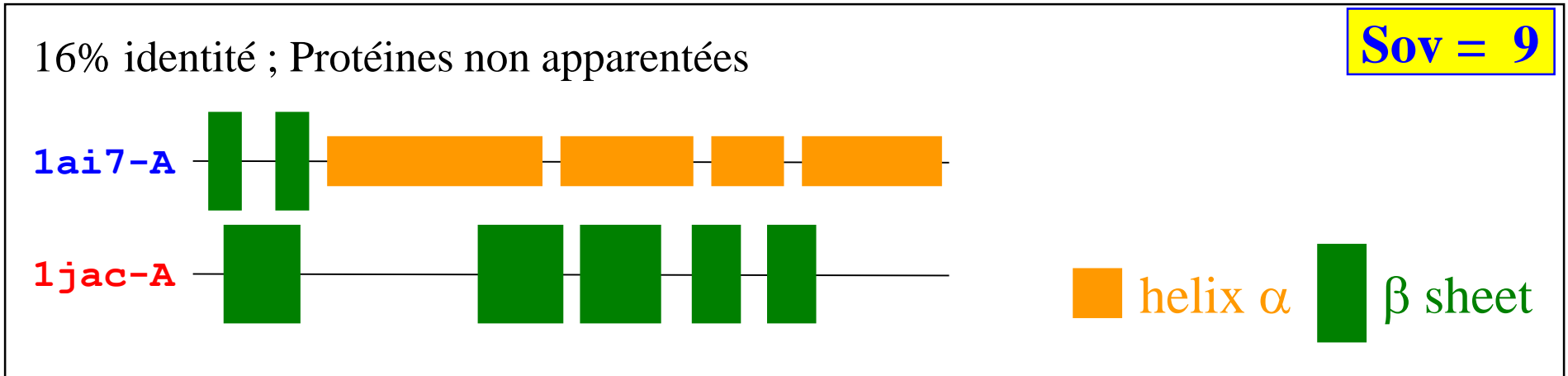
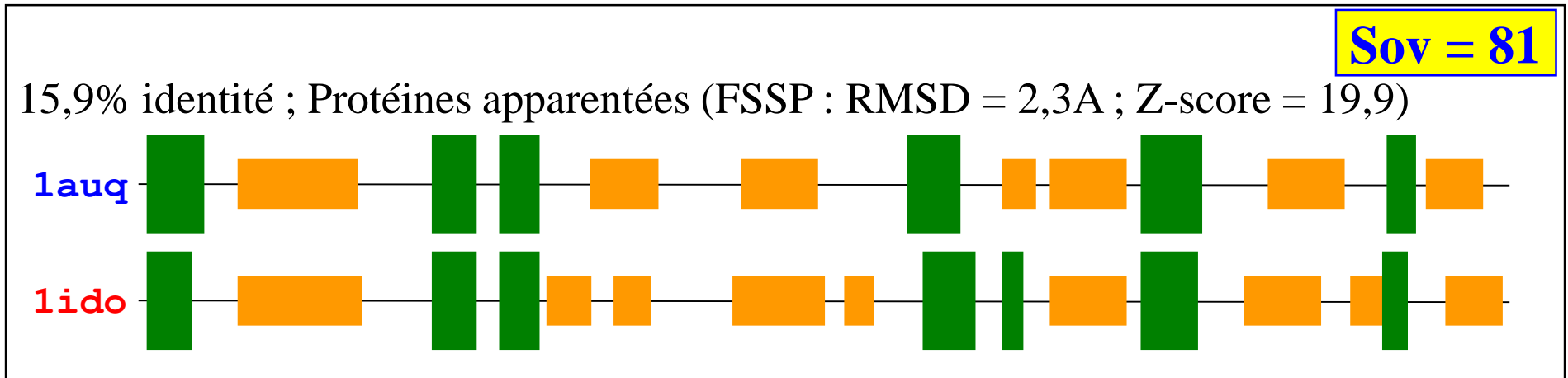
$$\text{Sov} = 100 \times \left[\frac{1}{N} \sum_{i \in [H, E, C]} \sum_{S(i)} \frac{\text{minov}(S_q, S_t) + \delta(S_q, S_t)}{\text{maxov}(S_q, S_t)} \times \text{len}(S_q) \right]$$

- *minov* : longueur de la structure secondaire chevauchante entre la source S_q et la cible S_t
- *maxov* : longueur maximale des structures secondaires chevauchantes entre la source S_q et la cible S_t



- δ est défini par :

$$\delta(S_q, S_t) = \min \left\{ \begin{array}{l} (\text{maxov}(S_q, S_t) - \text{minov}(S_q, S_t)); \text{minov}(S_q, S_t); \\ \text{int}(\text{len}(S_q/2)); \text{int}(\text{len}(S_t/2)) \end{array} \right\}$$



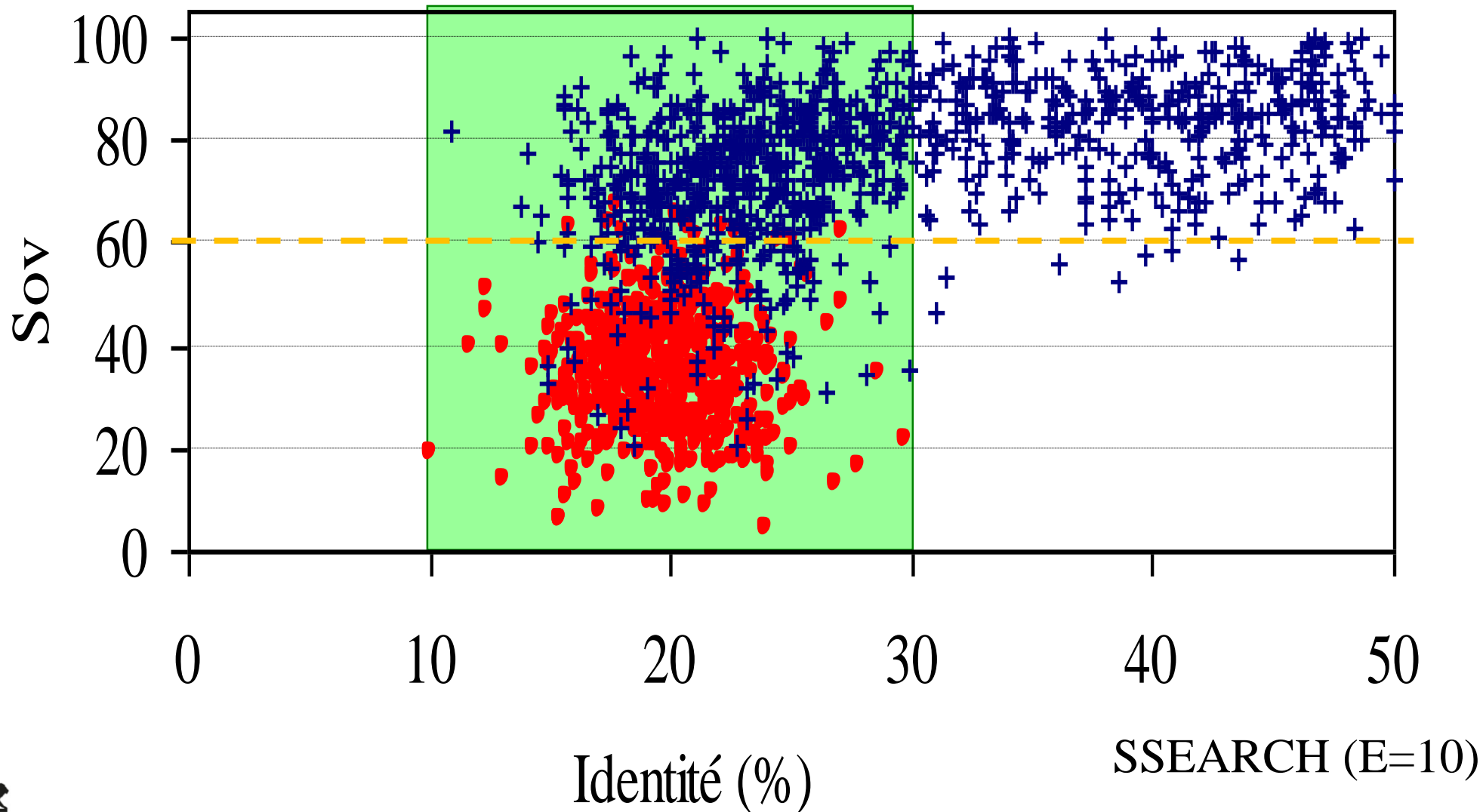
Structures secondaires observées plus conservées que les séquences dans la « zone floue » de 10-30%



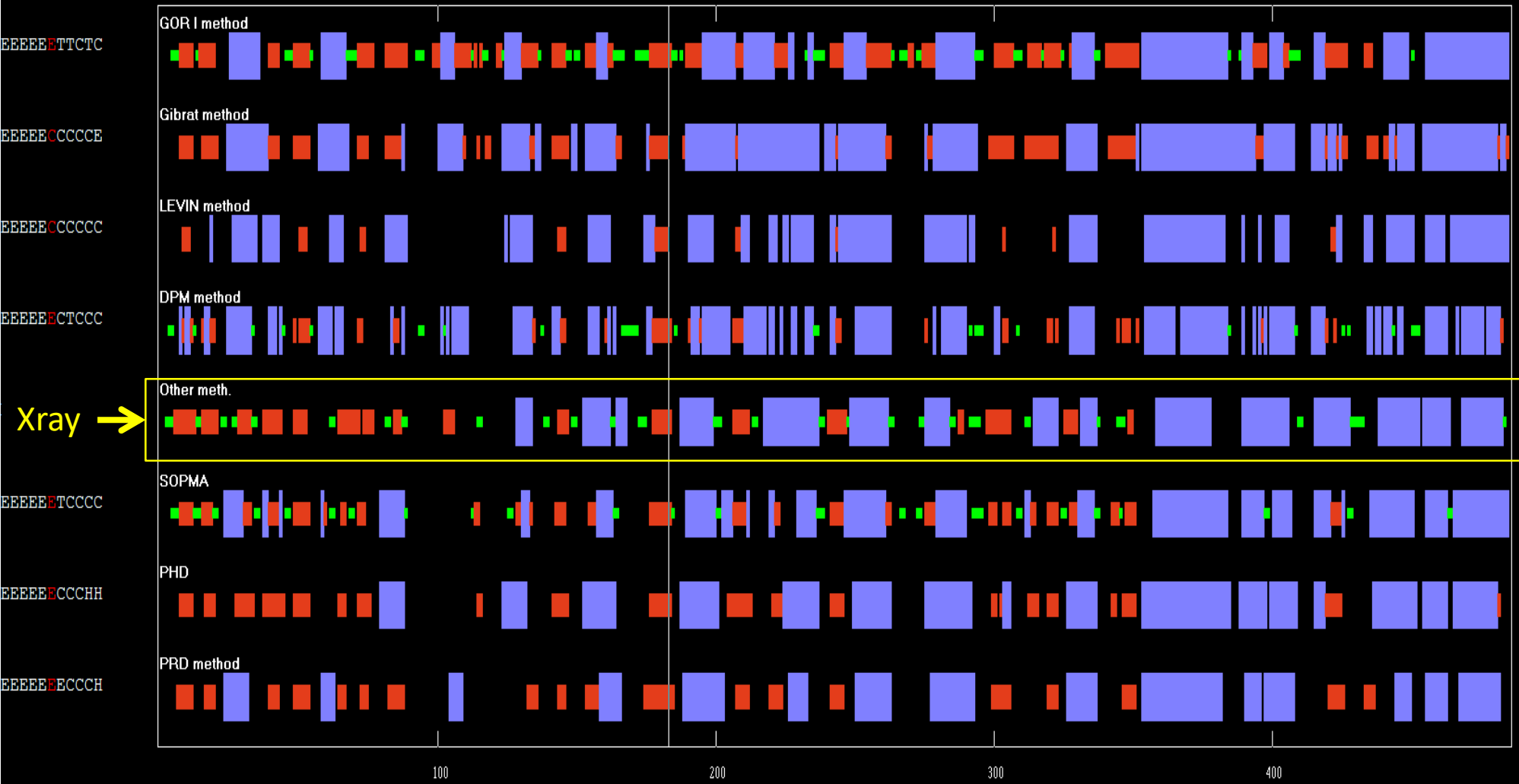
● Différentes

+ Proches

Discrimination selon la structure secondaire



N° 183 AA: ESKKLYCYVAVGQKRSTVAQLV



	Coil	Helix	Sheet	Q3	Sov
SOPMA	75,5	75,3	62,1	72,5	66,7
DSC	78,0	64,5	56,2	68,5	61,5
PHD	74,9	74,3	64,8	72,5	67,8
Consensus	80,1	72,9	59,4	72,8	67,9

SOPMA : Geourjon & Deléage, 1995

DSC : King & Sternberg, 1996

PHD : Rost & Sander, 1993

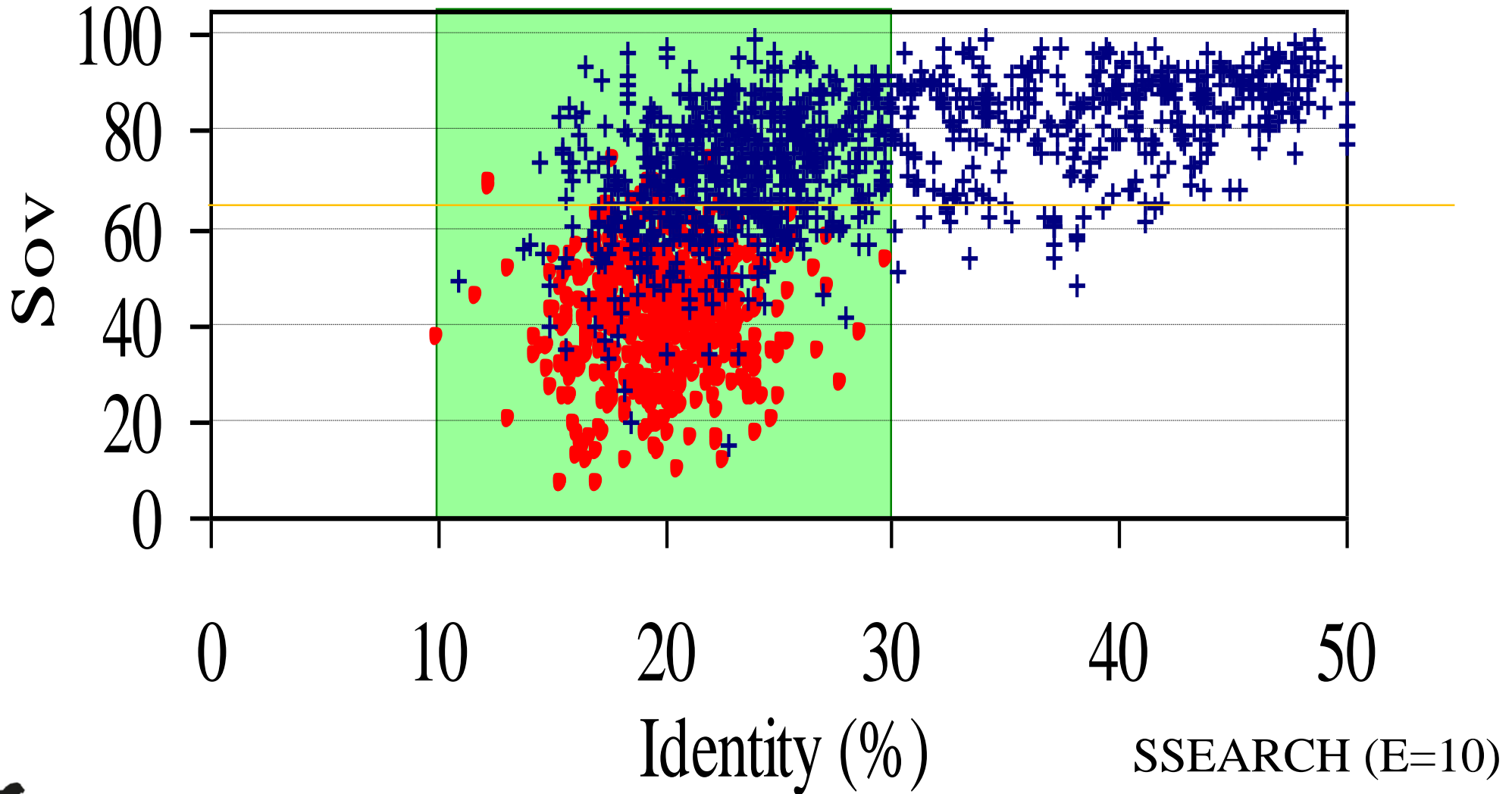
Consensus : Consens prediction calculée sur le serveur web NPS@

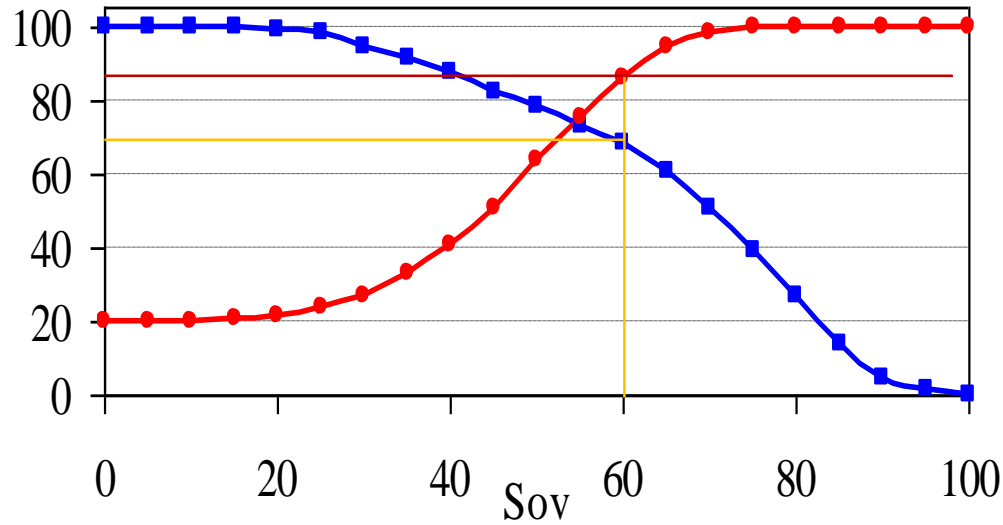
72,8% de qualité de prediction sur les 1106 protéines pdb_select_25

NPS@ Web server ==> <http://npsa-pbil.ibcp.fr>

- Différentes
- + Proches

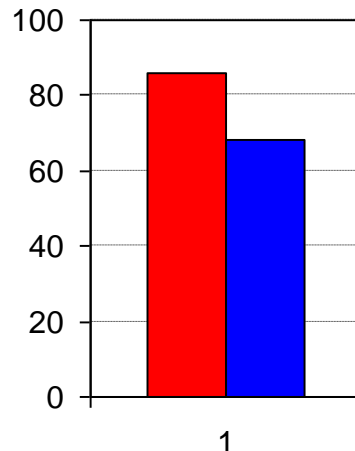
Discrimination selon la prédiction de structure secondaire

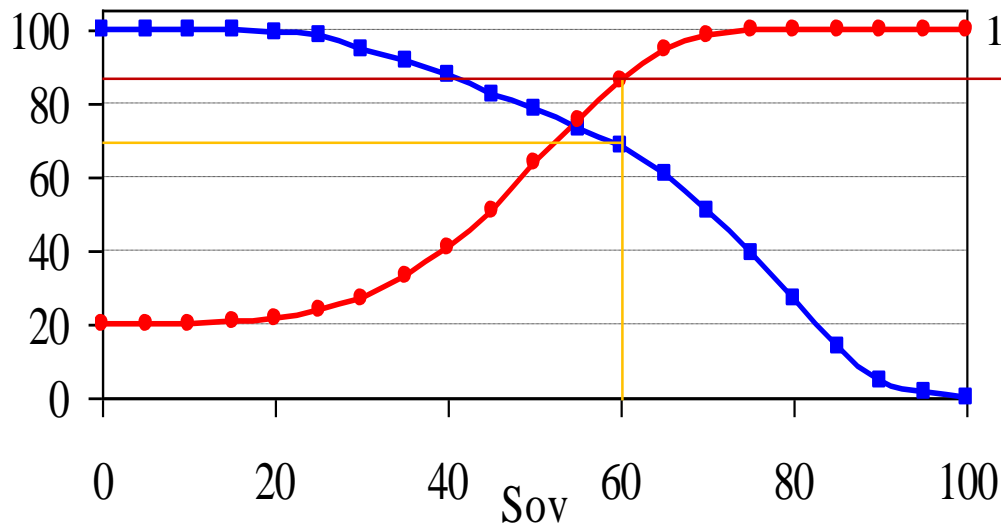




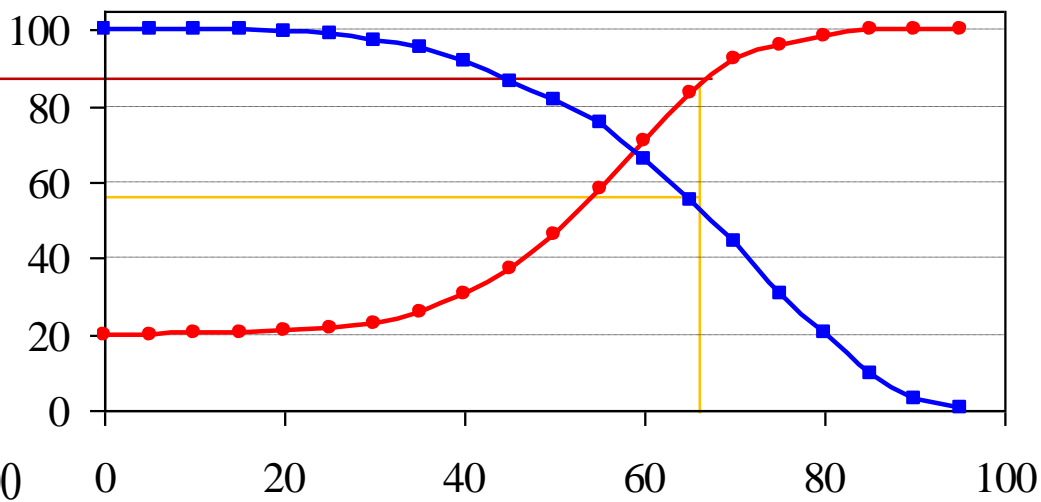
Structure observée des 2 protéines de la paire

Sov > 60
Obs-obs





Structure observée des 2 protéines de la paire

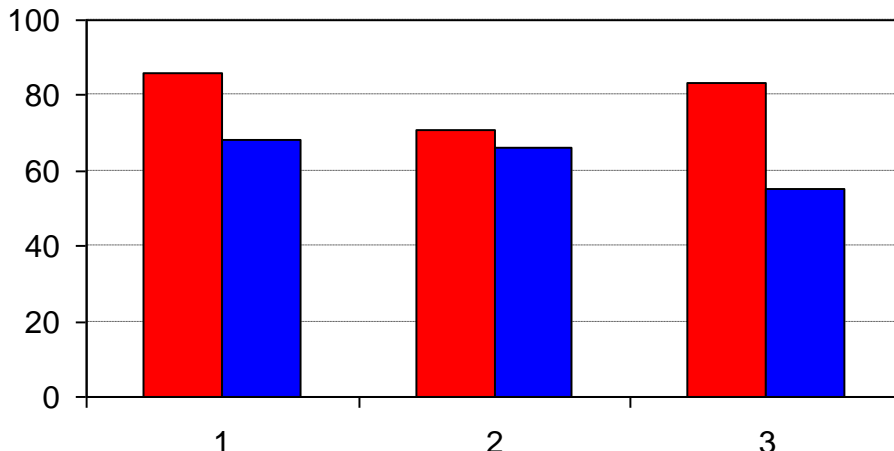


Prédiction des 2 protéines de la paire

**Sov>60
Obs-obs**

**Sov>60
Pred-pred**

**Sov>65
Pred-pred**



```
FASTA (3.43 Nov 2001) function [optimized, BL50 matrix (15:-5)] ktup: 2
  join: 37, opt: 25, gap-pen: -12/-2, width: 16
  Scan time: 2.520
```

The best scores are:

					opt	bits	E(103258)
sw	ATPA_TOBAC	(P00823)	ATP synthase alpha chain	(507)	3142	652	1.1e-186
sw	ATPA_SPIOL	(P06450)	ATP synthase alpha chain	(507)	2995	622	1.2e-177
sw	ATPA_ARATH	(P56757)	ATP synthase alpha chain	(507)	2980	619	1e-176
sw	ATPA_PEA	(P08215)	ATP synthase alpha chain (E	(501)	2909	605	2.4e-172
sw	ATPA_MARPO	(P06283)	ATP synthase alpha chain	(507)	2814	585	1.7e-166
sw	ATPA_MAIZE	(P05022)	ATP synthase alpha chain	(507)	2776	577	3.8e-164
sw	ATPA_ORYSA	(P12084)	ATP synthase alpha chain	(507)	2770	576	8.9e-164
sw	ATPA_WHEAT	(P12112)	ATP synthase alpha chain	(504)	2728	568	3.4e-161
sw	ATPA_PINTH	(P41602)	ATP synthase alpha chain	(494)	2697	561	2.7e-159
sw	ATPA_CHLVU	(P56294)	ATP synthase alpha chain	(506)	2570	535	1.8e-151
sw	ATPA_MESVI	(Q9MUT2)	ATP synthase alpha chain	(505)	2552	532	2.3e-150
sw	ATPA_NEPOL	(Q9TL16)	ATP synthase alpha chain	(501)	2540	529	1.3e-149
sw	ATPA_EUGGR	(P30392)	ATP synthase alpha chain	(506)	2512	524	6.8e-148
sw	ATPA_PORPU	(P51242)	ATP synthase alpha chain	(504)	2457	512	1.7e-144
sw	ATPA_CYAPA	(P48080)	ATP synthase alpha chain	(505)	2456	512	1.9e-144
sw	ATPA_GUITH	(O78475)	ATP synthase alpha chain	(502)	2450	511	4.4e-144
sw	ATPA_SYNP1	(Q05372)	ATP synthase alpha chain	(503)	2447	510	6.8e-144
sw	ATPA_ANTSP	(Q02848)	ATP synthase alpha chain	(505)	2442	509	1.4e-143
sw	ATPA_CHLRE	(P26526)	ATP synthase alpha chain	(507)	2426	506	1.3e-142
sw	ATPA_SYNP6	(P08449)	ATP synthase alpha chain	(505)	2407	502	2e-141
sw	ATPA_GALSU	(P35009)	ATP synthase alpha chain	(505)	2357	492	2.4e-138
sw	ATPA_ODOSI	(Q00820)	ATP synthase alpha chain	(503)	2352	491	4.8e-138

E() value faible



-
-
-

sw	VATA_PHAU	(P13548)	Vacuolar ATP synthase cat	(622)	230	57	2.5e-07
sw	ATP2_HEVBR	(P29685)	ATP synthase beta chain,	(562)	228	56	3.1e-07
sw	VATA_GOSHI	(P31405)	Vacuolar ATP synthase cat	(623)	228	56	3.4e-07
sw	VATA_BETVU	(Q39442)	Vacuolar ATP synthase cat	(623)	227	56	3.9e-07
sw	VATA_BORBU	(O51121)	V-type ATP synthase alpha	(575)	225	56	4.9e-07
sw	VAA2_HUMAN	(P38607)	Vacuolar ATP synthase cat	(615)	225	56	5.1e-07
sw	VAA1_DROME	(P48602)	Vacuolar ATP synthase cat	(614)	224	56	5.9e-07
sw	VATA_PLAFA	(Q03498)	Vacuolar ATP synthase cat	(611)	223	55	6.8e-07
sw	VATA_MANSE	(P31400)	Vacuolar ATP synthase cat	(617)	222	55	7.8e-07
sw	VATA_CYACA	(P48414)	Vacuolar ATP synthase cat	(587)	220	55	1e-06
sw	VATA_PYRAB	(Q9UXU7)	V-type ATP synthase alpha	(1017)	223	56	1e-06
sw	VAA2_DROME	(Q27331)	Vacuolar ATP synthase cat	(614)	219	55	1.2e-06
sw	VATA_AEDAE	(O16109)	Vacuolar ATP synthase cat	(615)	219	55	1.2e-06
sw	VAA1_PIG	(Q29048)	Vacuolar ATP synthase catal	(617)	219	55	1.2e-06
sw	VATA_NEUCR	(P11592)	Vacuolar ATP synthase cat	(607)	218	54	1.4e-06
sw	ATPB_DROVI	(Q24751)	ATP synthase beta chain,	(228)	211	53	1.7e-06
sw	ATPB_CYTLY	(P13357)	ATP synthase beta chain ((502)	215	54	1.8e-06
sw	VATA_CANTR	(P38078)	Vacuolar ATP synthase cat	(1088)	209	53	7.7e-06
sw	VATA_YEAST	(P17255)	Vacuolar ATP synthase cat	(1071)	204	52	1.5e-05
sw	ATPB_BACFR	(P13356)	ATP synthase beta chain ((505)	189	48	7.2e-05
sw	VATB_DROME	(P31409)	Vacuolar ATP synthase sub	(490)	178	46	0.00034
sw	VATB_CAEEL	(Q19626)	Probable vacuolar ATP syn	(491)	178	46	0.00034
sw	VATB_YEAST	(P16140)	Vacuolar ATP synthase sub	(517)	178	46	0.00035
sw	VAB1_ACEAT	(Q38681)	Vacuolar ATP synthase sub	(492)	176	46	0.00045
sw	VATB_MANSE	(P31401)	Vacuolar ATP synthase sub	(494)	172	45	0.00079
sw	VATB_HELVI	(P31410)	Vacuolar ATP synthase sub	(494)	172	45	0.00079
sw	VAB2_ACEAT	(Q38680)	Vacuolar ATP synthase sub	(492)	171	45	0.00091
sw	ATPB_ASPND	(O03063)	ATP synthase beta chain ((284)	167	44	0.001
sw	RHO_DEIRA	(P52153)	Transcription termination	(426)	168	44	0.0012
sw	VAB2_BOVIN	(P31408)	Vacuolar ATP synthase sub	(511)	168	44	0.0014
sw	VAB2_HUMAN	(P21281)	Vacuolar ATP synthase sub	(511)	168	44	0.0014

E() value intermédiaire

Significatif ?



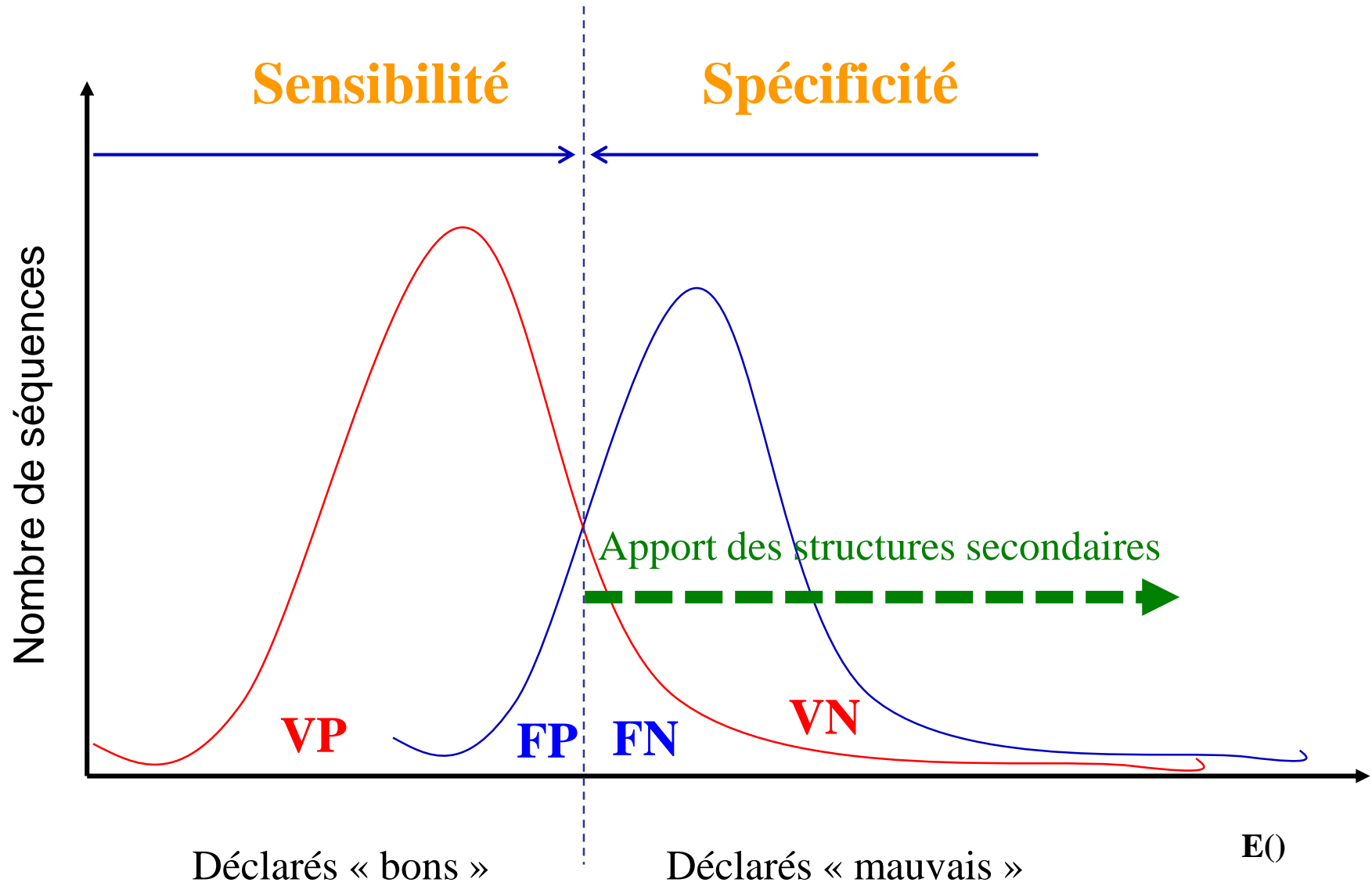


sw	ATPB_LONHI	(O03073)	ATP synthase beta chain	(208)	148	40	0.012
sw	ATPB_MICPL	(O03074)	ATP synthase beta chain	(208)	147	40	0.014
sw	VATB_THETH	(Q56404)	V-type ATP synthase beta	(478)	151	41	0.015
sw	RHO_CHRVI	(P52152)	Transcription termination	(433)	150	40	0.016
sw	RHO_AQUAE	(O67031)	Transcription termination	(436)	144	39	0.038
sw	RHO_MICLU	(P52154)	Transcription termination	(690)	146	40	0.041
sw	ATPB_OSMCI	(O03077)	ATP synthase beta chain	(220)	137	37	0.06
sw	RHO_THEMA	(P38527)	Transcription termination	(427)	140	38	0.066
sw	RHO_PSEFL	(P52155)	Transcription termination	(419)	137	38	0.099
sw	RHO_RHOSH	(P52156)	Transcription termination	(422)	137	38	0.1
sw	ATPB_HYPHO	(O03070)	ATP synthase beta chain	(208)	130	36	0.16
sw	RHO_BORBU	(P33561)	Transcription termination	(419)	132	37	0.2
sw	RHO_BUCAI	(P57652)	Transcription termination	(419)	131	36	0.23
sw	RHO_NEIGO	(Q06447)	Transcription termination	(419)	128	36	0.36
sw	RHO_STRLI	(P52157)	Transcription termination	(707)	130	36	0.4
sw	RHO_HELPJ	(Q9ZLS9)	Transcription termination	(438)	127	36	0.42
sw	RHO_HELPY	(P56466)	Transcription termination	(438)	127	36	0.42
sw	RHO_RICPR	(Q9ZD24)	Transcription termination	(457)	126	35	0.51
sw	AG43_ECOLI	(P39180)	Antigen 43 precursor (AG4	(1039)	130	37	0.54
sw	RHO_ECOLI	(P03002)	Transcription termination	(419)	125	35	0.54
sw	RHO_SALTY	(P26980)	Transcription termination	(419)	125	35	0.54
sw	RHO_HAEIN	(P44619)	Transcription termination	(420)	124	35	0.63
sw	RHO_TREPA	(O83281)	Transcription termination	(519)	125	35	0.64
sw	ATPB_DENPU	(O03068)	ATP synthase beta chain	(215)	119	34	0.76
sw	YLC7_YEREN	(P21212)	Hypothetical protein in L	(58)	109	31	1.1
sw	DCDA_ZYMMO	(Q9Z661)	Diaminopimelatedecarboxy	(421)	119	34	1.3
sw	RHO_BACSU	(Q03222)	Transcription termination	(427)	119	34	1.3
sw	ATPA_BRYMA	(P26965)	ATP synthase alpha chain	(29)	102	30	1.8
sw	SR54_METJA	(Q57565)	Signal recognition 54 kDa	(451)	117	34	1.8
sw	TPO_MOUSE	(P40226)	Thrombopoietin precursor	(356)	115	33	2
sw	FMT_PASMU	(P57949)	Methionyl-tRNA formyltrans	(317)	112	32	2.8
sw	RECA_CHLAU	(O52394)	RecA protein (Recombinase	(351)	111	32	3.4
sw	Z198_HUMAN	(Q9UBW7)	Zinc finger protein 198	(1377)	117	34	4.3
sw	TRBB_RHISN	(P55395)	Probable conjugal transfe	(325)	105	31	7.6
sw	YI73_AQUAE	(O67720)	Hypothetical protein AQ_1	(407)	105	31	9.1
sw	SYA_HELPY	(P56452)	Alanyl-tRNA synthetase (EC	(847)	109	32	9.1
sw	FMT_HAEIN	(P44787)	Methionyl-tRNA formyltrans	(318)	103	31	9.9

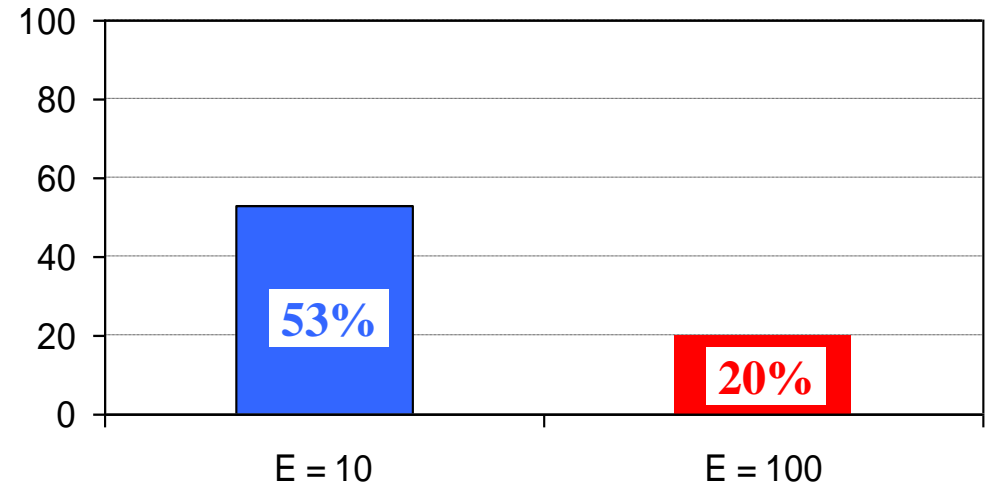
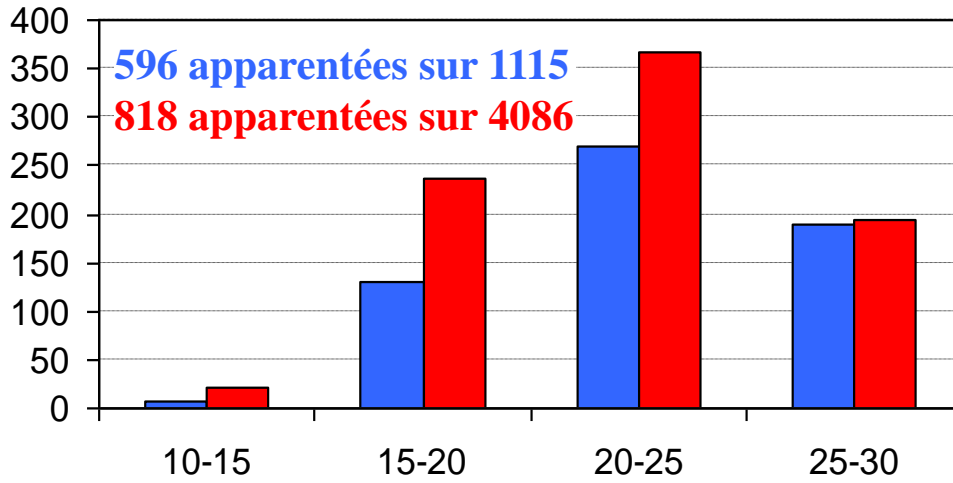
E() value « forte »

Pas significatif

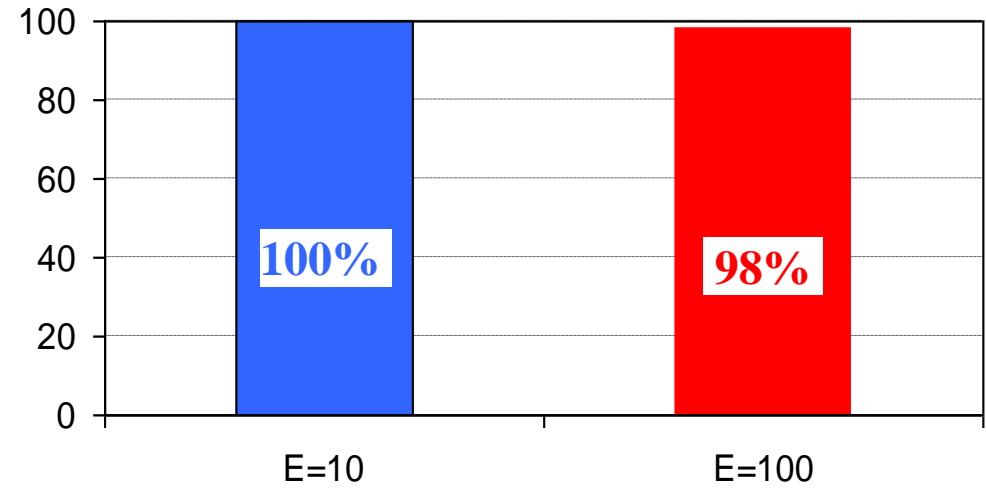
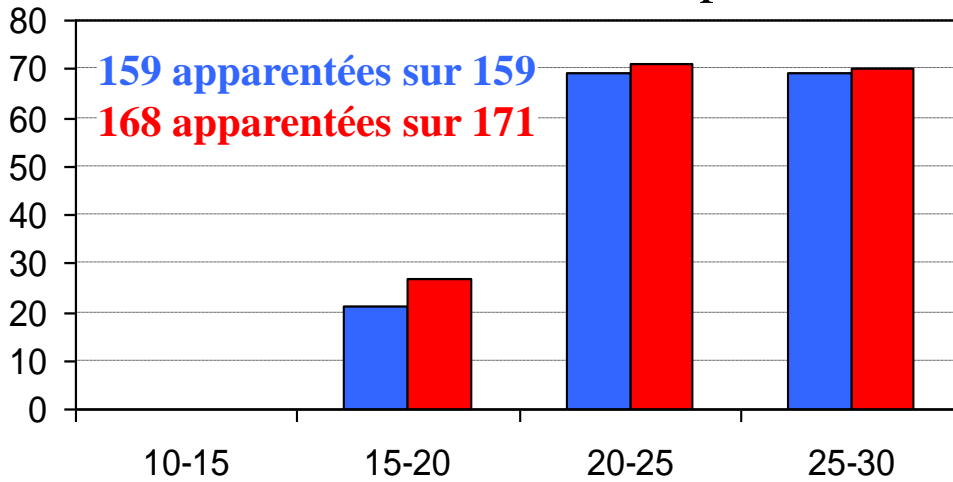




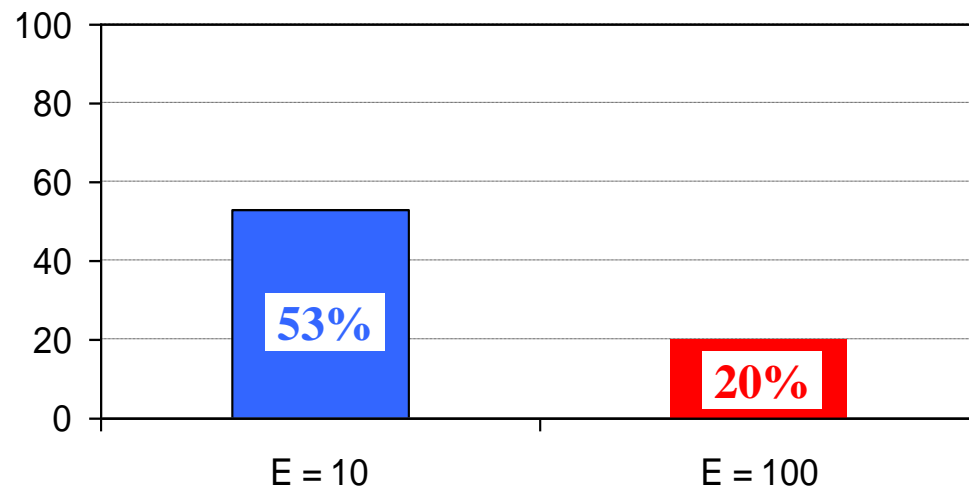
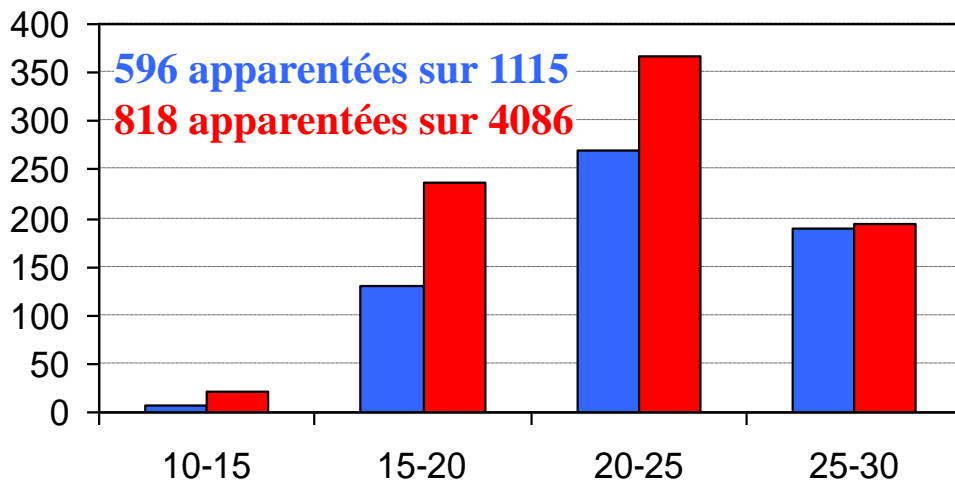
- Sans structure secondaire



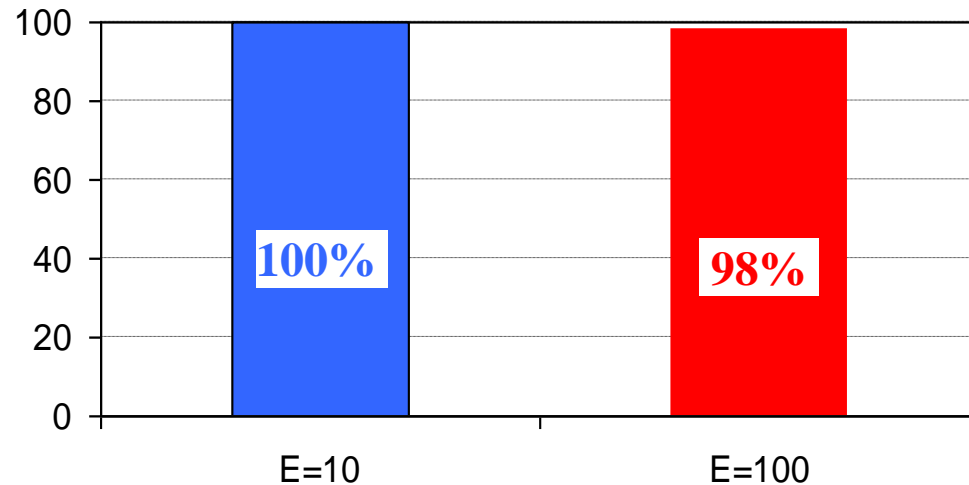
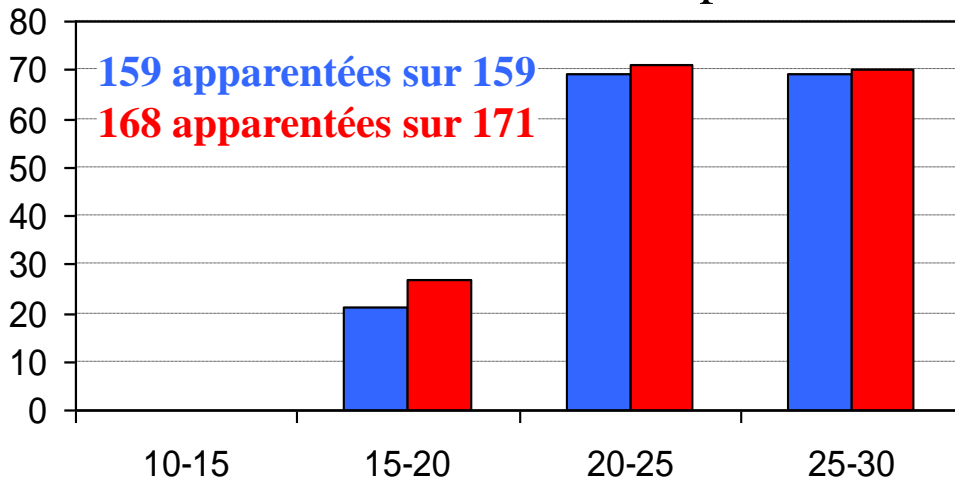
- Avec structure secondaire prédite et $SOV > 65$



• Sans structure secondaire



• Avec structure secondaire prédite



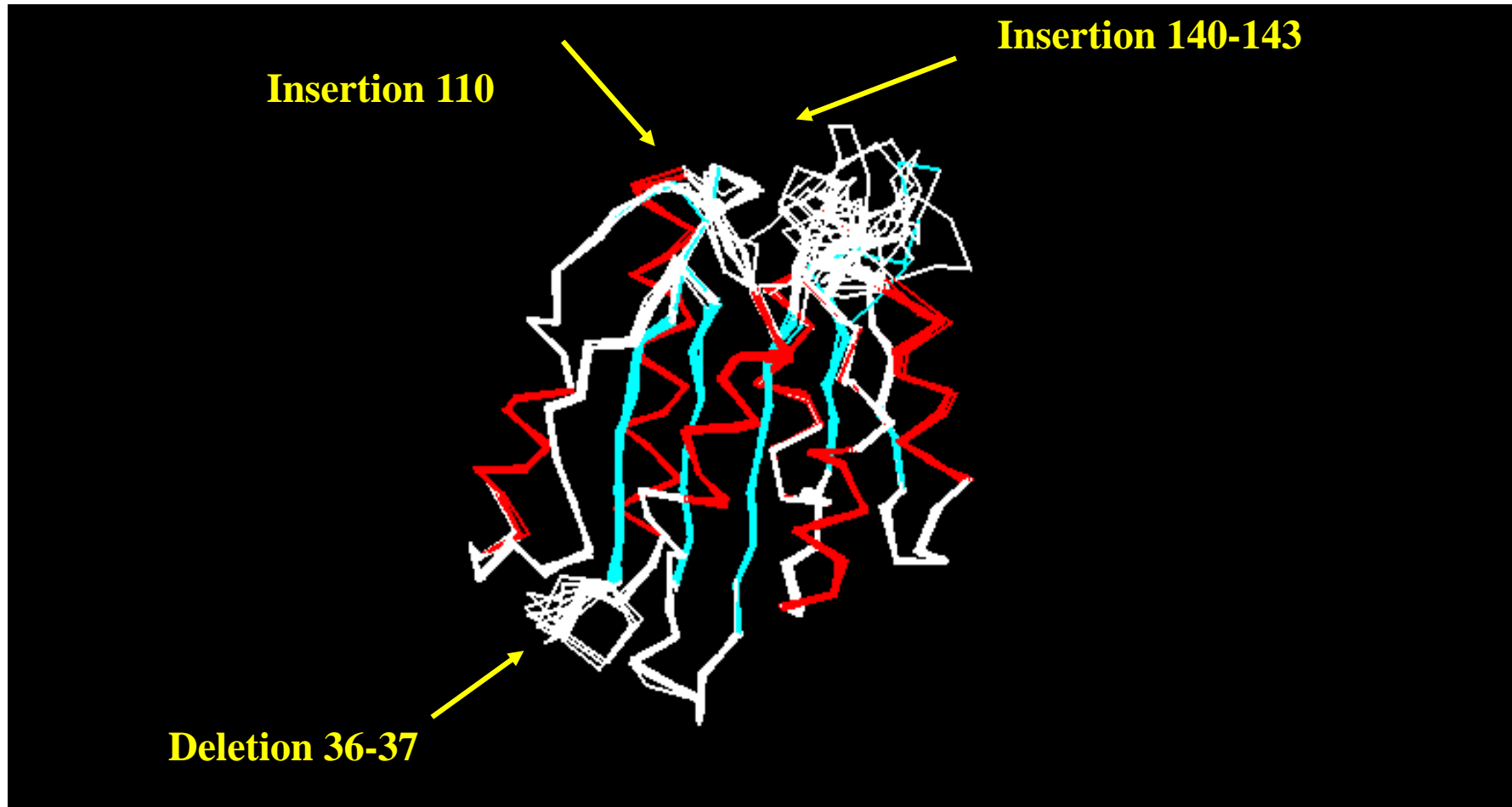
- Sov permet de sélectionner les paires apparentées (100% et 98% de confiance pour E=10 et E=100)
- Sov augmente le nombre de protéines apparentées détectées (159 et 168 protéines détectées avec E=10 et E=100)
- Gain de 6%.

1auq : A1 domain of Von Willebrand factor

1ido : Integrin CR3

	10	20	30	40	50	60
1.1auq					
.predicted	DLVFLLDGSSRLSEAEFEVLKAFVVDMMERLRISQKWVRVAVVEYHDGSHAYIGLKDRKR					
2.1ido	DIAFLIDGSGSIIPHDFRRMKEFVSTVMEQLKKS--TLFSLMQYSEEFRIHFTFKEFQN					
.predicted	ceeeeeeeccccchhhhhhhhhhhhhhhhhhhccccccccceeeeeccceeeeeccccccc					
Homology	*:.**:*:. : :*. :* ** :**:*: *: . .:::* : : : :*:. :					
	70	80	90	100	110	120
1.1auq					
.predicted	PSELRRIASQVKYAGSQVASTSEVLKYTLFQIFSKIDRPEASRIALLMASQ-EPQRMRSR					
2.1ido	NPNRSLVKPITQLLGRTHATGIRK-VVRELFNITNGARKNAFKILVITDGEKFGDPL					
.predicted	ccchhhhhccccccccchhh					
Homology	.: * :.. : . :.. ::: : * . : ::* . : . . : :*:: : : * .					
	130	140	150	160	170	180
1.1auq					
.predicted	NFVRYVQGLKKKKVIVIPVGIG----PHANLKQIRLIEKQAPENKAFVLSVDELEQQRD					
2.1ido	GYEDVIPEADREGVIRYVIGVGDAFRSEKSRQELNTIASKPPRDHVFQVNNFEALKTIQN					
.predicted	chhhhhhhhhhhccccccccce????ccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh					
Homology	.: : .:: ** :** .. . :::: * . :.*:::* :... :* : :					

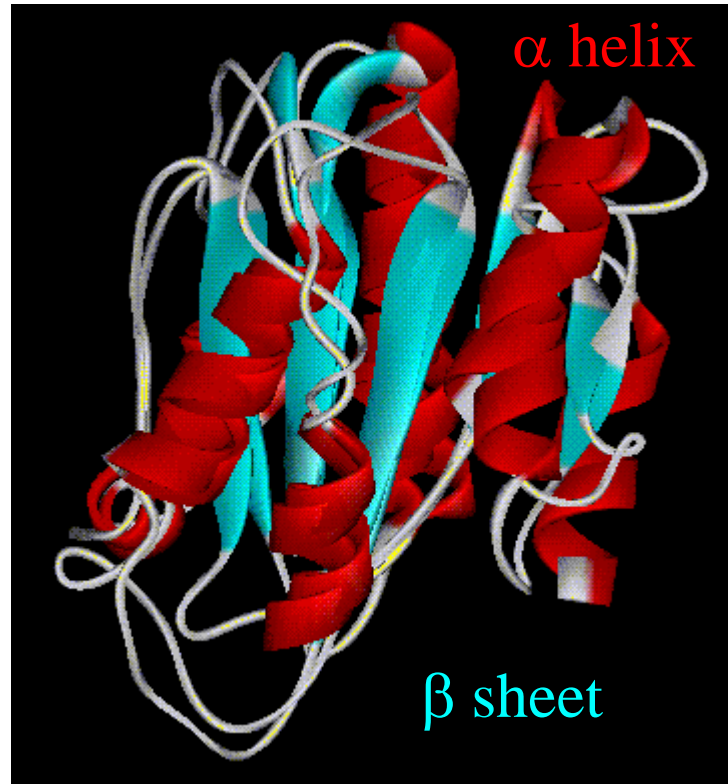
15,9 % d'identité ; Blast E-value = 0,05 ; Sov = 81,7



Structures géométriquement correctes, sans violation.
Energie = $-547 \text{ kcal mol}^{-1}$; RMSD = 0.36 \AA

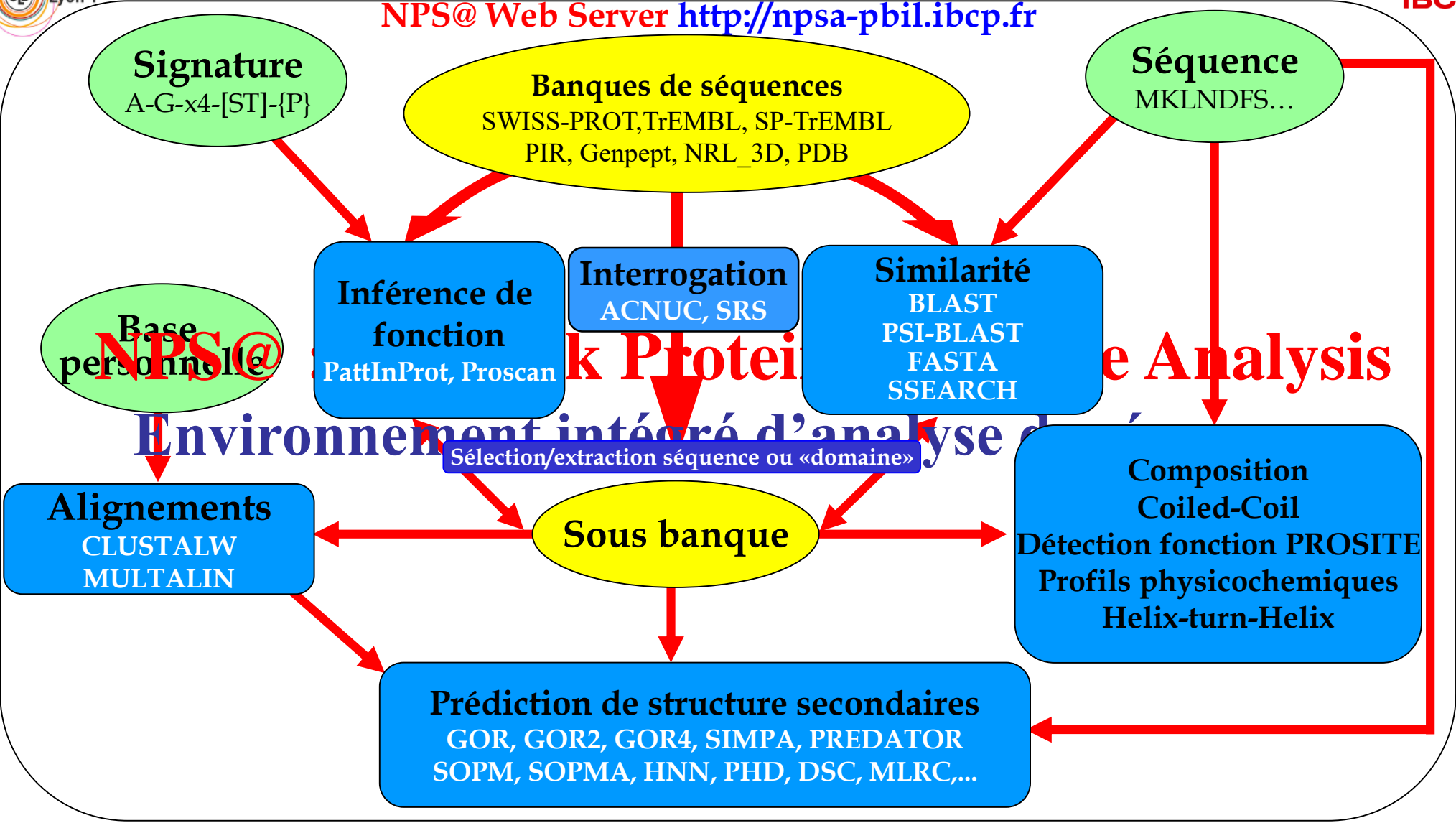
lido structure

lido modèle



RMSD entre modèle et structure $C\alpha$ 3,3 Å
(Fssp RMSD entre 1auq et lido = 2,3 Å)

NPS@ Web Server <http://npsa-pbil.ibcp.fr>



INTERNET

Logiciels clients/serveurs: ANTHEPROT, MPSA

B Network Protein Sequence @analysis
NPS@ is the IBCP contribution to PBL in Lyon, France

[HOME] [HELP] [REFERENCES] [NEWS] [CONTACT US] [MPSA] [ANTHEPROT] [PBL] **NEW**

Work with 1UXD proteic sequence :

View sequence in: [MPSA (Mac UNIX) , About...] [AnTheProt (PC) , Download...] [HELP]

HOMOLOGY SEARCH with sequence with BLAST

SCAN PROSITE

SEARCH PATTERN

CONSENSUS secondary structure prediction

PRIMARY ANALYSIS on sequence with PHYS.-CHEM. PROFILES

Three-dimensional links for 1UXD

PDB FILE SUMMARY at IGH (FR)

SECONDARY STRUCTURE from 3D by DSSP on full PDB file 1uxd

GET COORDINATES of Full PDB file 1uxd

VIEW full PDB file 1uxd with Rasmol

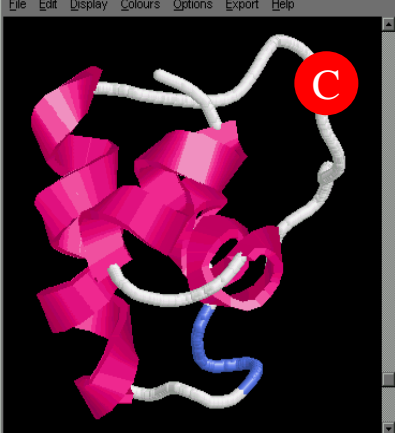
3DEE Database of Protein Domain Definitions

CATH Class Architecture Topology and Homol

FSSP Fold classification based on Structure-S

SCOP Structural Classification Of Proteins

MOLMOV molecular movement



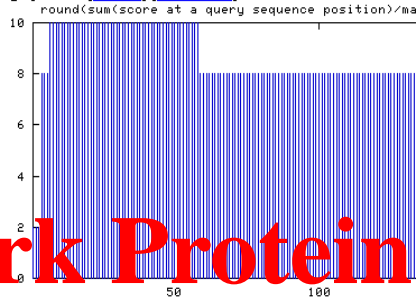
A Network Protein Sequence @analysis
NPS@ is the IBCP contribution to PBL in Lyon, France

[HOME] [HELP] [REFERENCES] [NEWS]

BLASTP results for : Unknown sequence

View BLASTP in: [MPSA (Mac UNIX) , About...] [AnTheProt (PC) , Download...] [HELP]

View graphic in: [MPSA] [AnTheProt]



first step : SELECT (selection of proteins to be aligned)

- with E threshold lesser or equal to $5e-5$
- and matching any residue from 1 to 270 of our database

You can also select/unselect sequence

sequences producing significant results

- NPSA_3DSE01pdb1lbgA
- NPSA_3DSE01pdb2puaA
- NPSA_3DSE01pdb1gpzA
- NPSA_3DSE01pdb2pugA
- NPSA_3DSE01pdb1luxd

second step : EXTRACT (for selected sequences)

- Full sequences from database
- Partial sequences from blast

For partial extraction : extract sequence with at least 270 residues in match

- Add query sequence to created database.
- Remove identical sequences in created database.

CLUSTALW multiple alignment

Abstract Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. (1994) Nucleic Acids Research, 22, 4673-4680 [HELP]

SHOW alignment with width 364

all residues.

Conservation level of 0 % (works with "using conservation level of" option in list above).

Only secondary consensus

F Alignment : C:\WINDOWS\TEMP\1J728SA.ALN using ClustaW

DSC (King and Stern) DPM (Deleage and Doolittle) GOR I (Garnier et al.) GOR II (Garnier et al.) GOR III (Gibrat et al.) GOR IV (Garnier et al.) HNN (Guermeur, 1994) SIMPA96 (Levin et al.) MLRC (Guermeur et al.)

When MLRC is checked, the following options are available:

- PRED TOP (Garnier et al., 1994)
- PRED BOT (Garnier et al., 1994)
- SOPMA (Geourjon and Delaunoy, 1998)

A prediction method with a weight of 1000 can be added.

The other methods can be kept if they have been computed or can be added after. All methods can be displayed at the same time in the alignment.

```
1.pdb1gpzA  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
2.pdb2puaA  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
3.pdb2pugA  ---TIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
4.pdb1lbgA  ---TIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
5.pdb1luxd  ---MKLDEIARLAGVSTTASYVINGKAKQYRVSQKTRFVAVETRNNAVWAAIKELHYSPPSAVARSLLKVNHTK
Prim.cons.  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
```

G MPSA (BCP-Lyon) 0.86b; Sequence Display

```
1.pdb1gpzA  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
2.pdb2puaA  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
3.pdb2pugA  ---TIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
4.pdb1lbgA  ---TIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
5.pdb1luxd  ---MKLDEIARLAGVSTTASYVINGKAKQYRVSQKTRFVAVETRNNAVWAAIKELHYSPPSAVARSLLKVNHTK
Prim.cons.  ---ATIKDVAKRANVSTTVSHVIN---KTRFVAEETRNAVWAAIKELHYSPPSAVARSLLKVNHTK
```

NPS@ : Network Protein Sequence Analysis

Démonstration

Structures 3D



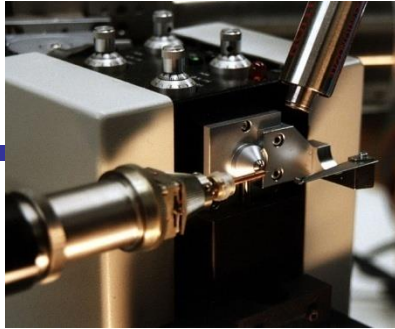
Molecular Type ↓↑	X-ray ↓↑	NMR ↓↑	EM ↓↑	Multiple methods ↓↑	Neutron ↓↑	Other ↓↑	Total ↓↑
Protein (only)	132181	11451	3880	160	67	32	147771
Other	7964	92	447	6	0	4	8513
Protein/NA	7016	265	1354	3	0	0	8638
Nucleic acid (only)	2081	1300	47	5	2	1	3436
Total	149242	13108	5728	174	69	37	168358

Interrogations sur le RCSB
<http://www.rcsb.org>

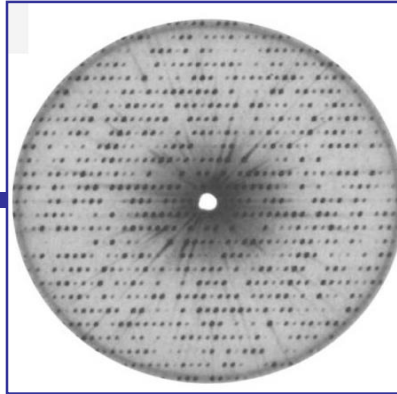


Obtention de cristaux

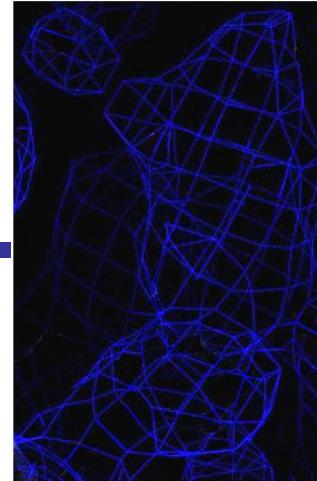
Cristallisation de
Protéine purifiée



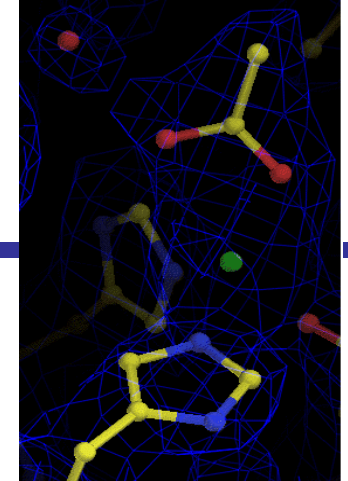
Diffractomètre
Rayons X



Cliché de diffraction



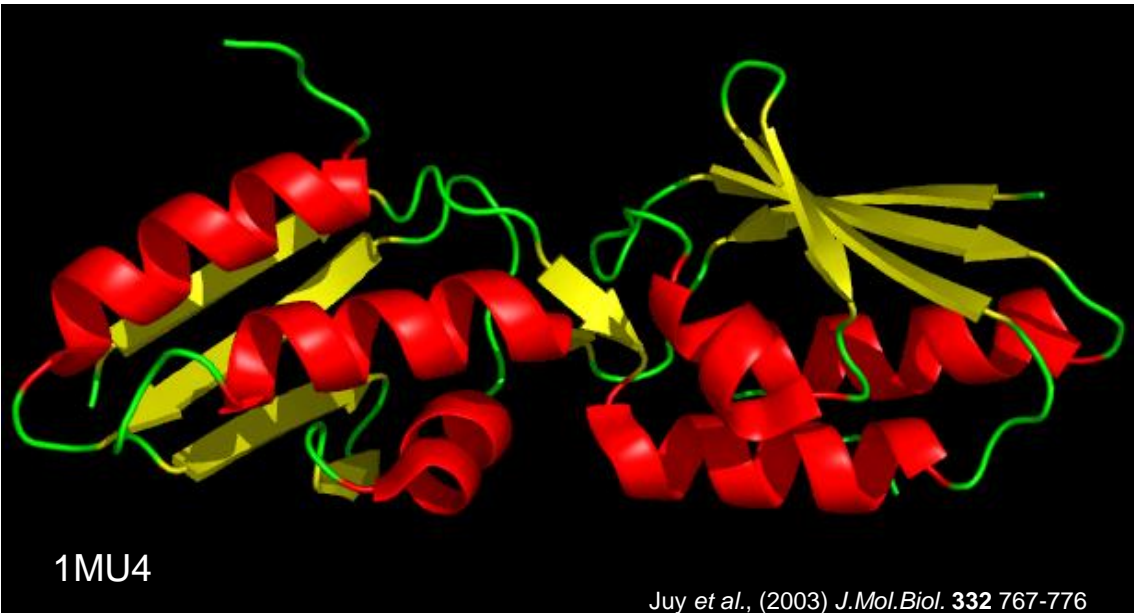
Densité électronique



Optimisation

Modélisation moléculaire

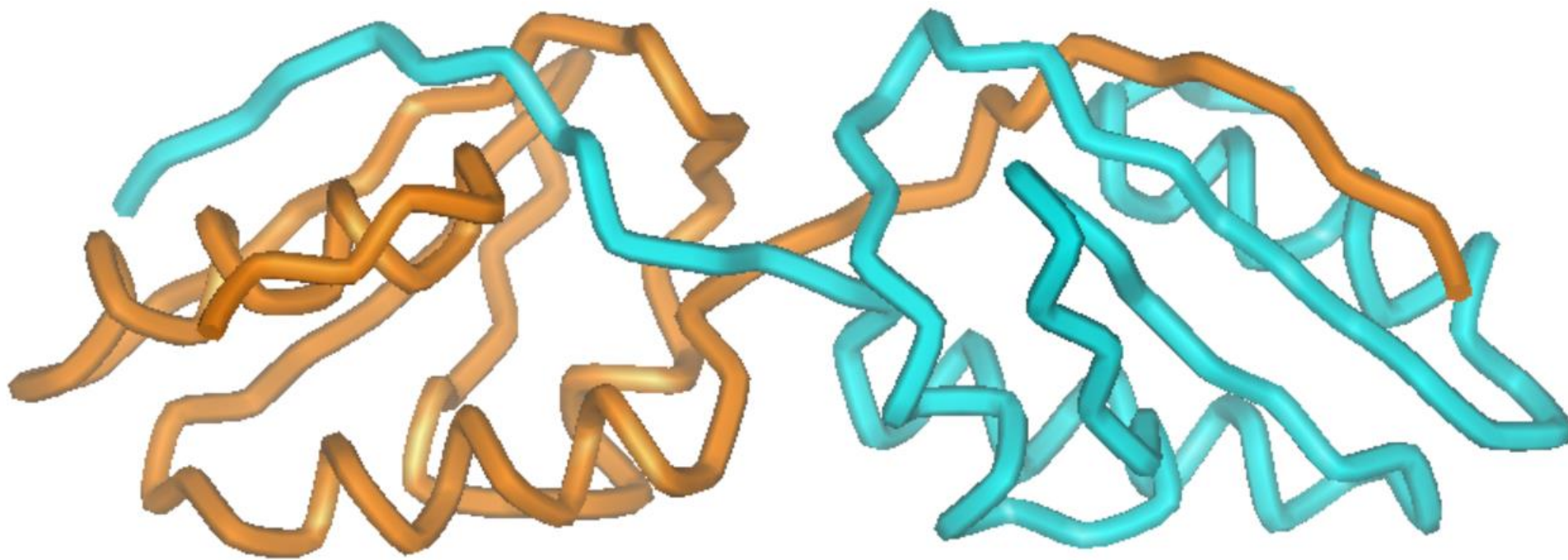
Structure 3D



Juy et al., (2003) *J.Mol.Biol.* **332** 767-776

CRH *Bacillus subtilis* Catabolite Repression HPr

Echanges de domaine « domain swapping »



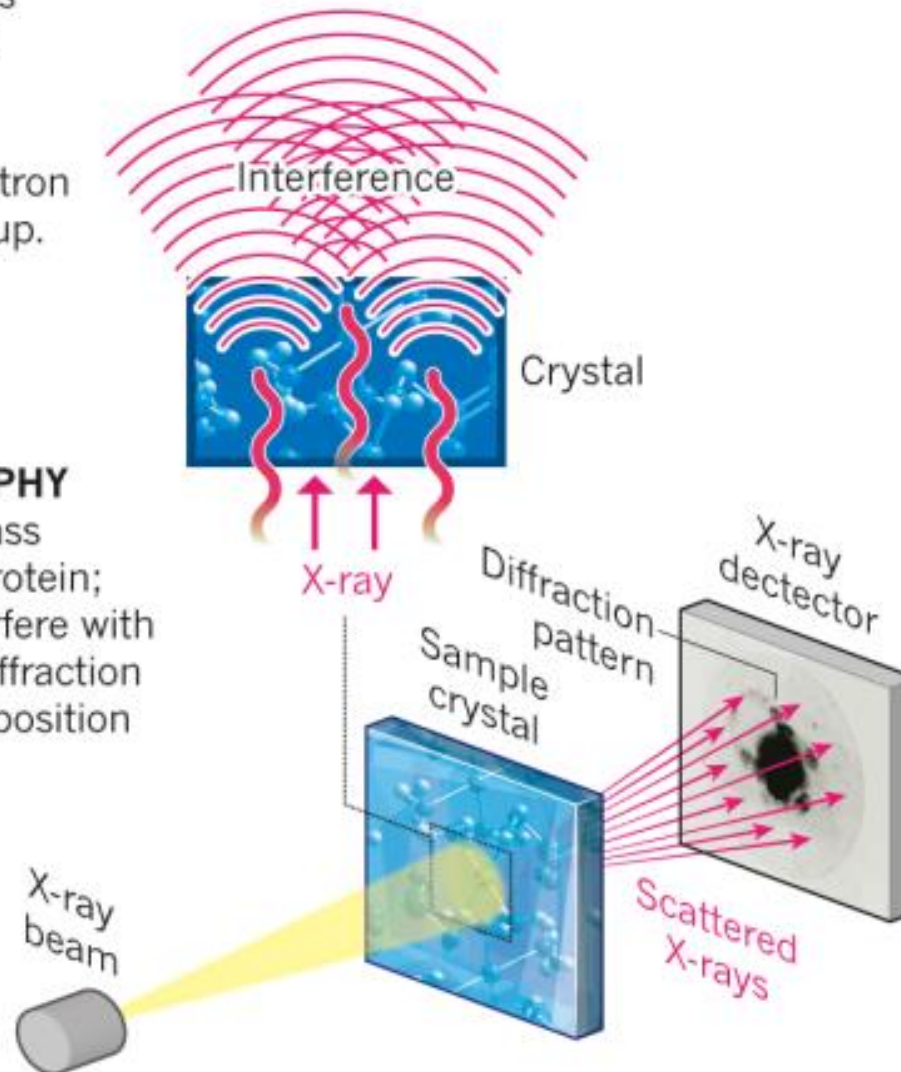
1MU4

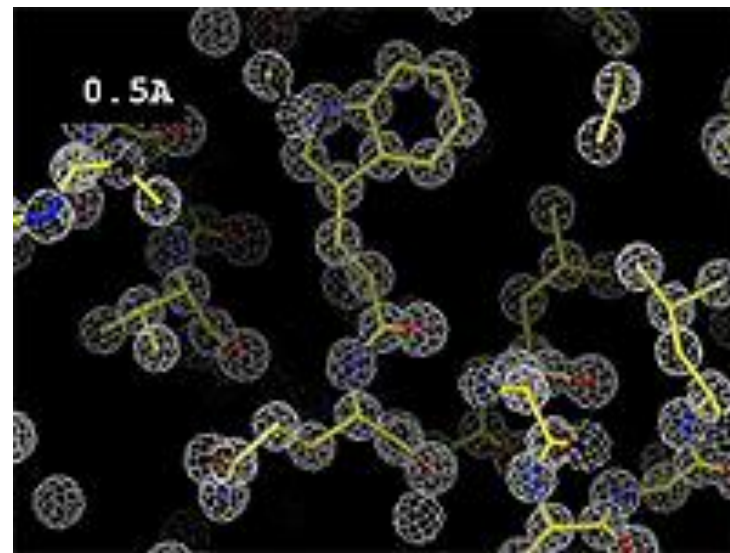
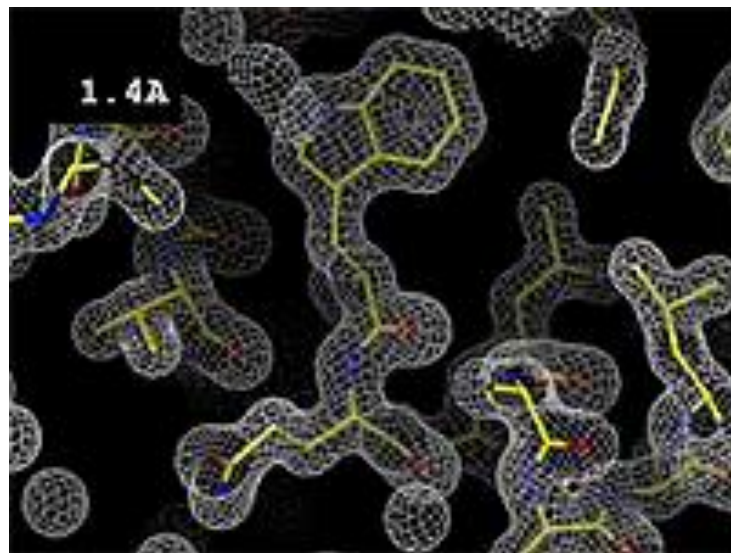
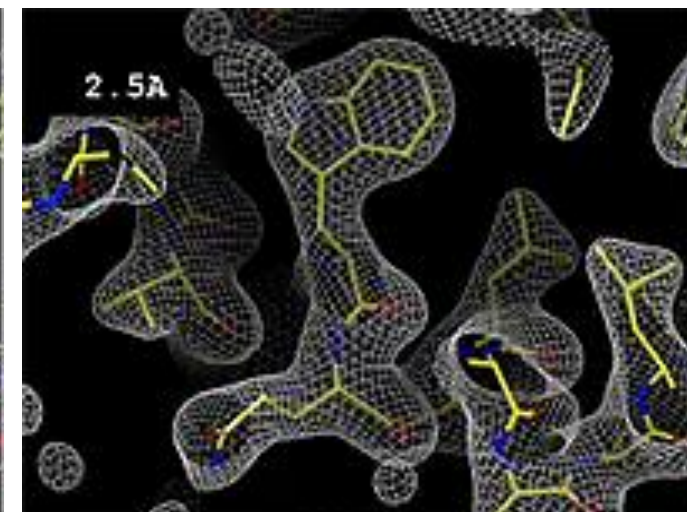
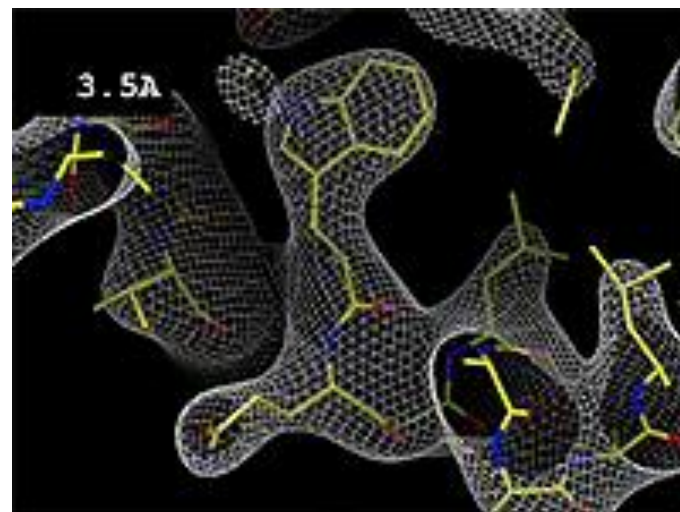
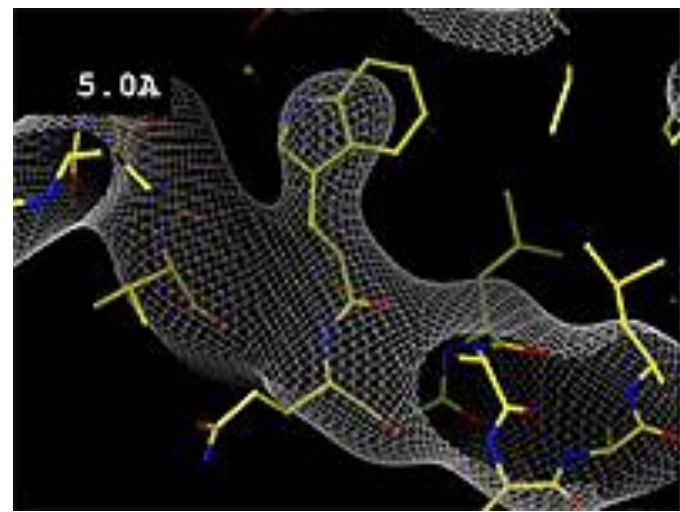
CRH *Bacillus subtilis* Catabolite Repression HPr

X-ray crystallography has long been the dominant method for deducing high-resolution protein structures, but cryo-electron microscopy is catching up.

X-RAY CRYSTALLOGRAPHY

X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.

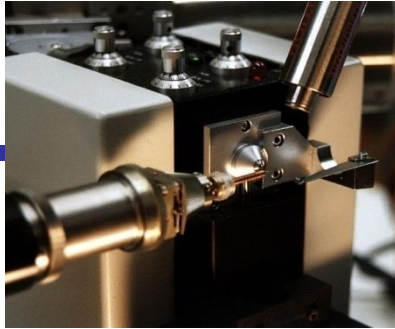




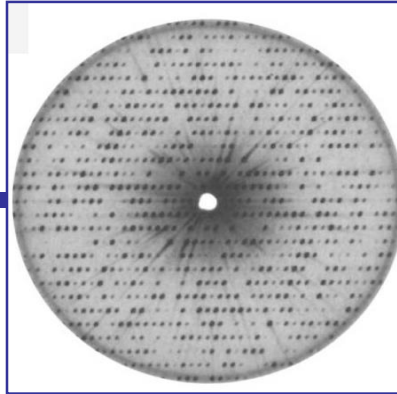


Obtention de cristaux

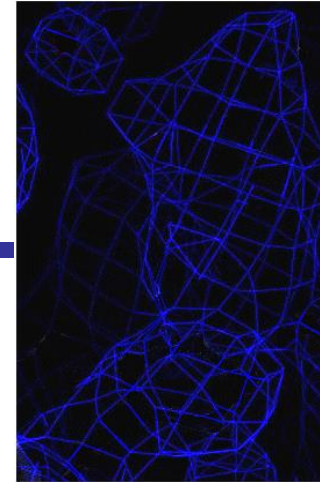
Cristallisation de
Protéine purifiée



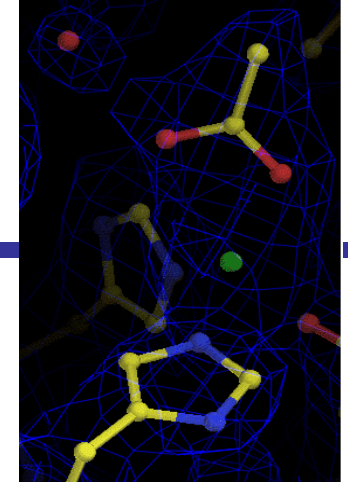
Diffractomètre
Rayons X



Cliché de diffraction



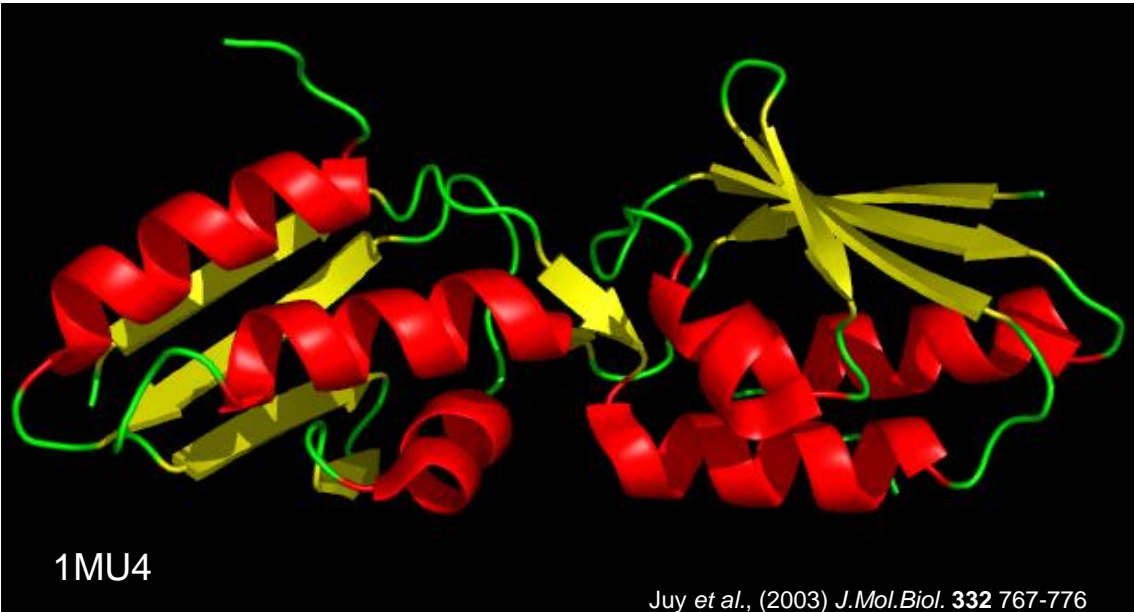
Densité électronique



Optimisation

Modélisation moléculaire

Structure 3D



Juy et al., (2003) *J.Mol.Biol.* **332** 767-776

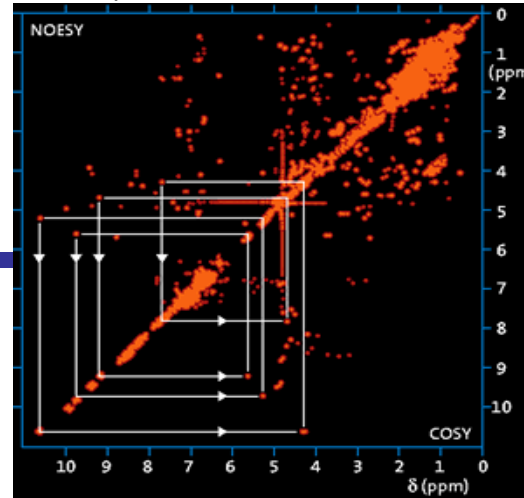
CRH *Bacillus subtilis* Catabolite Repression HPr

Spectromètre RMN à haut champ (ici 500 MHz)

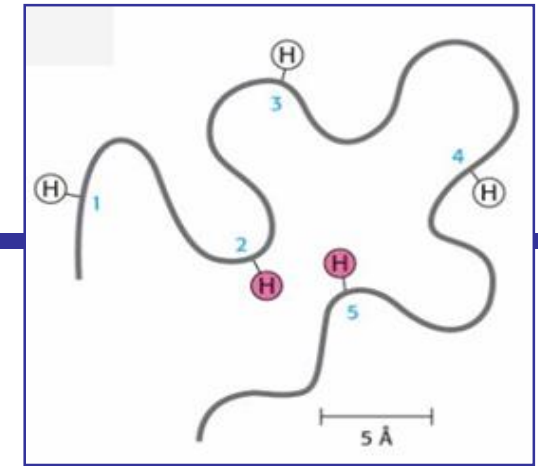


Protéine purifiée
en solution
très concentrée

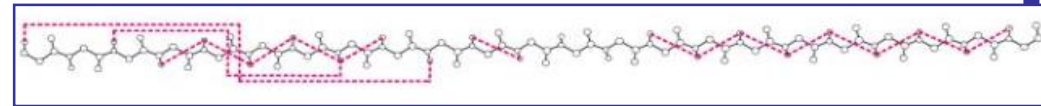
Spectres 2D NOESY TOCSY



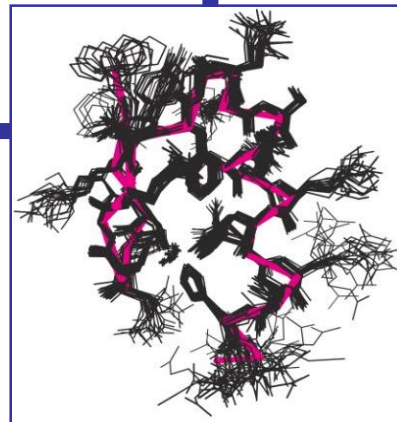
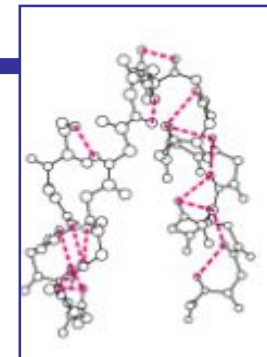
Proximités spatiales



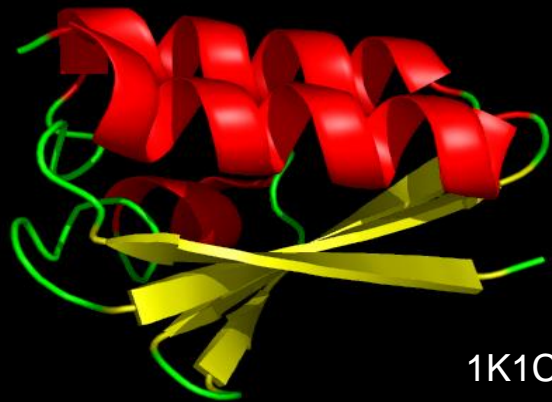
Attribution



Modélisation moléculaire



«Fagot» de structures



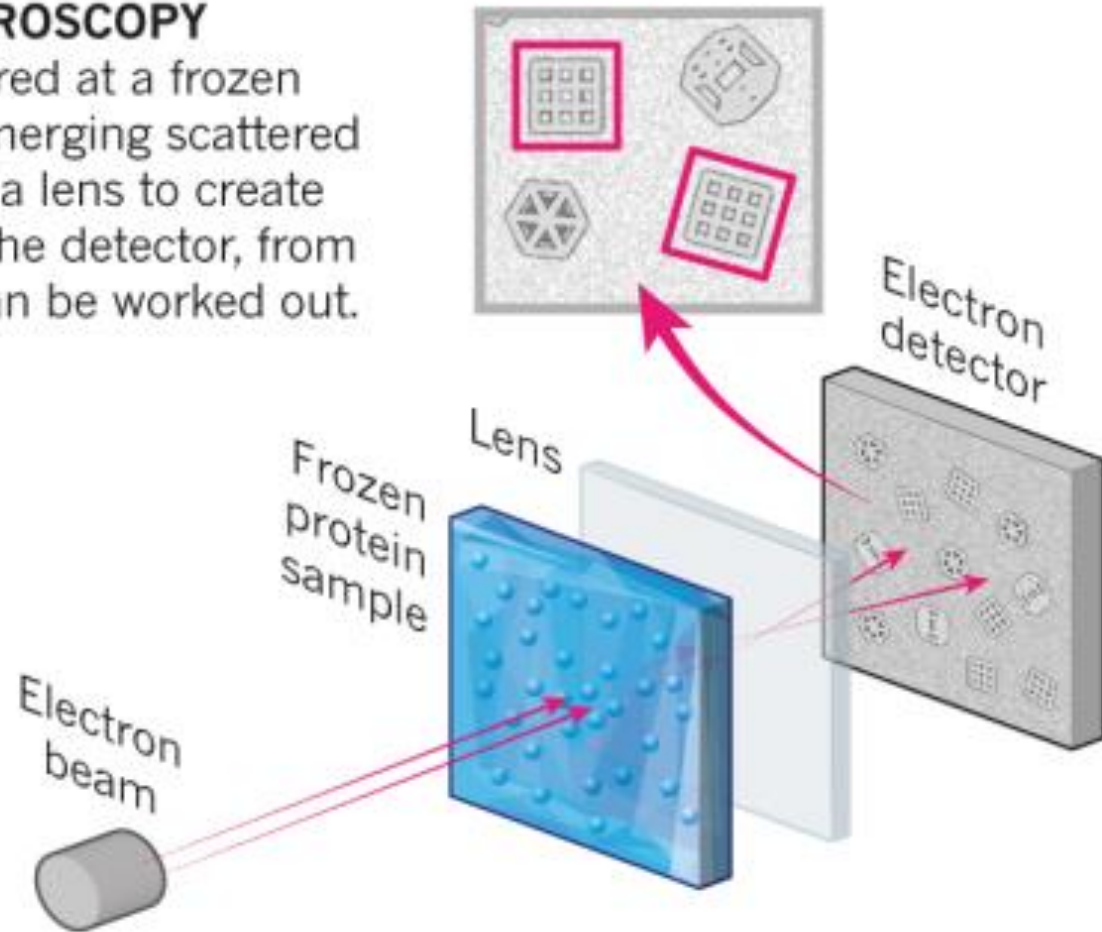
1K1C

Favier et al. (2002) *J.Mol.Biol.* **317**, 131-144

CRH *Bacillus subtilis* Catabolite Repression HPr

CRYO-ELECTRON MICROSCOPY

A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.

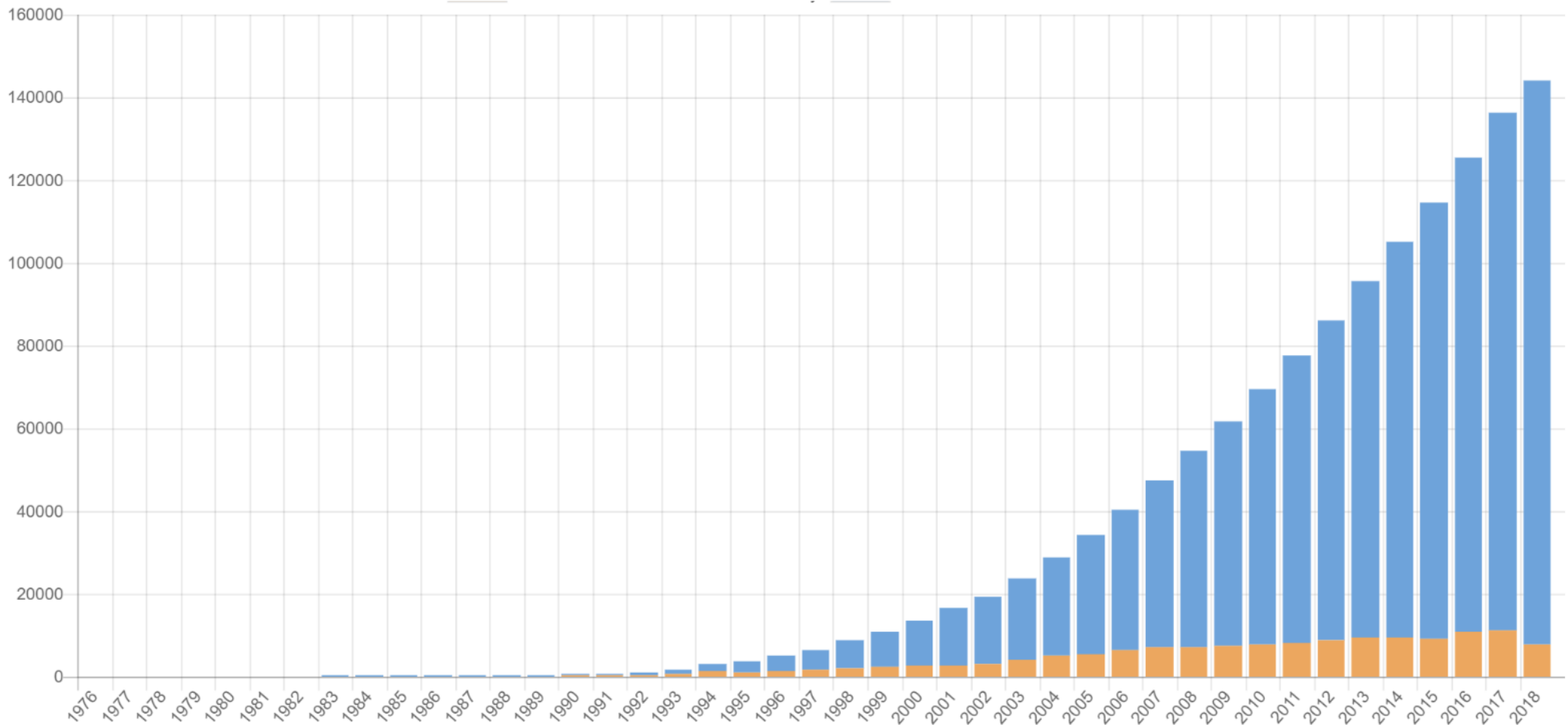


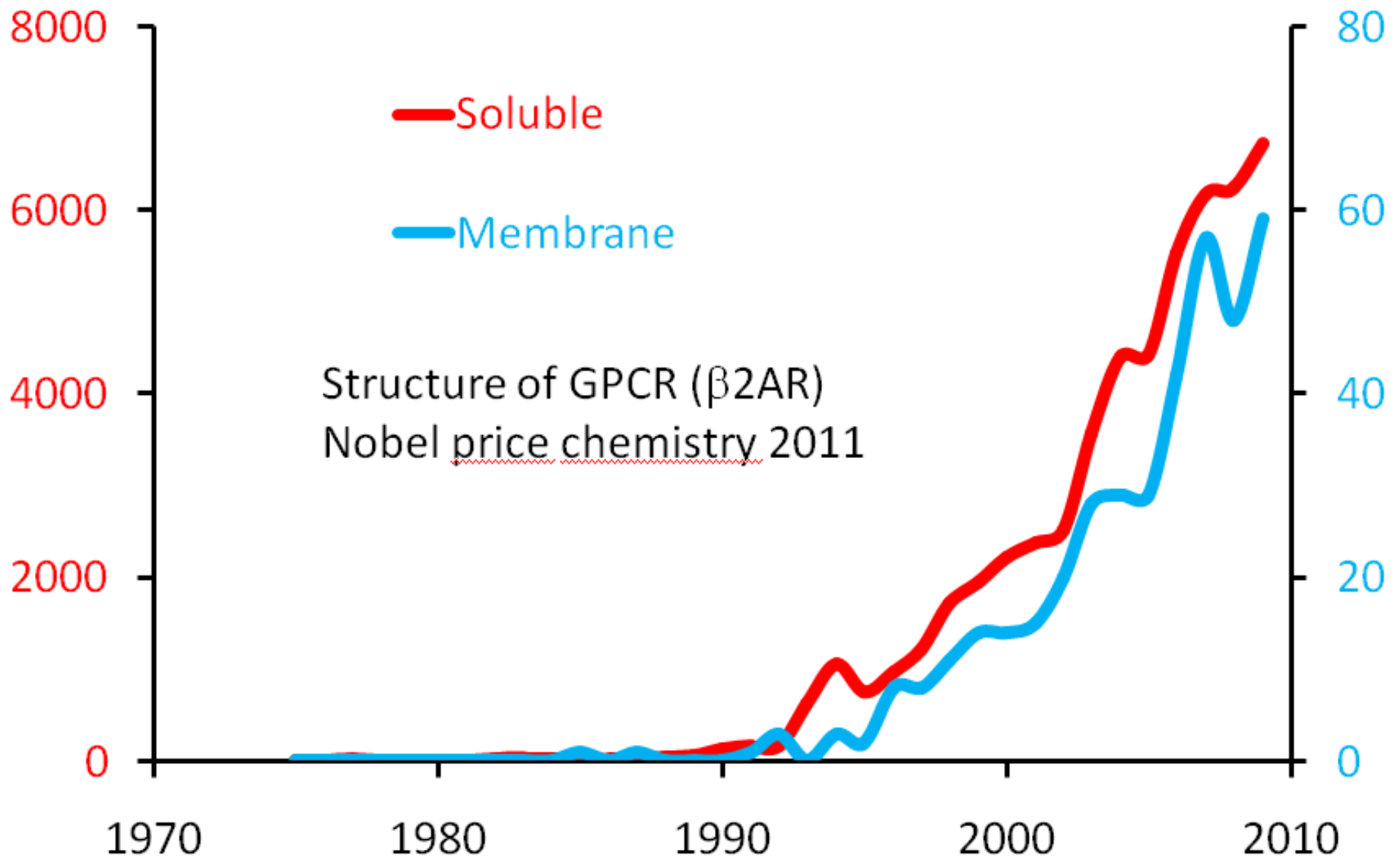
© nature

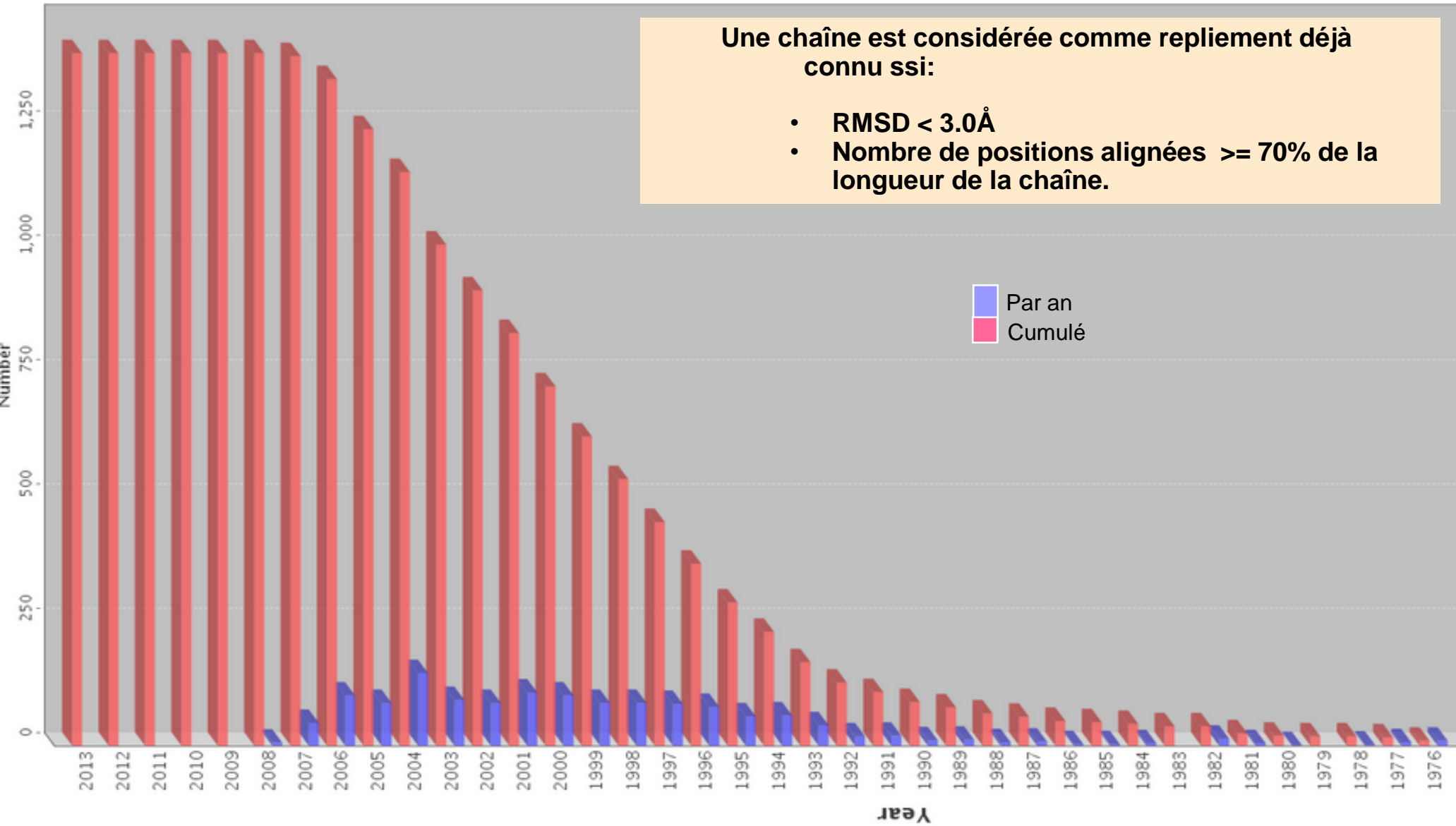
	Coût moyen	Change de succès
Protéines solubles bactériennes	\$140000	35%
Protéines humaines solubles (kinases, protéases, ...)	\$450000	35%
Protéines membranaires bactériennes	\$1,5 million	10%
Protéines membranaires humaines	\$2,5 million	10%

R.C. Stevens, Drug Discovery, 2003

Coût n'incluant pas les développements technologiques ni l'amortissement des équipements lourds

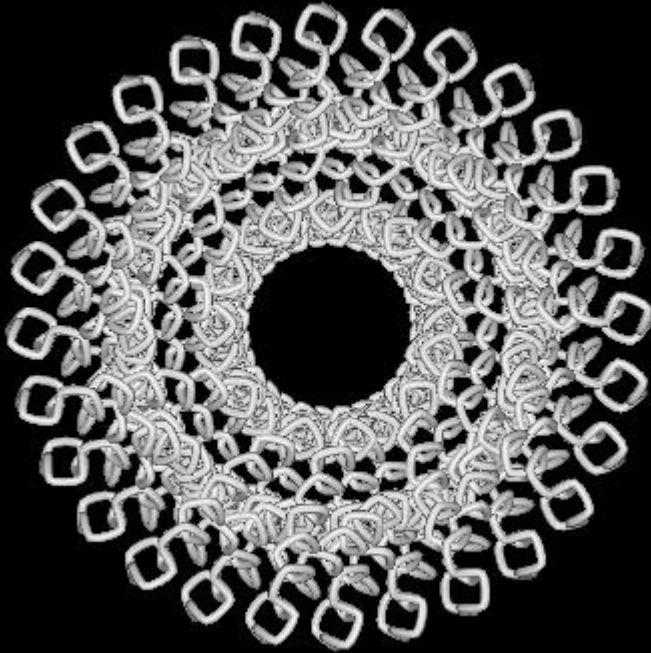




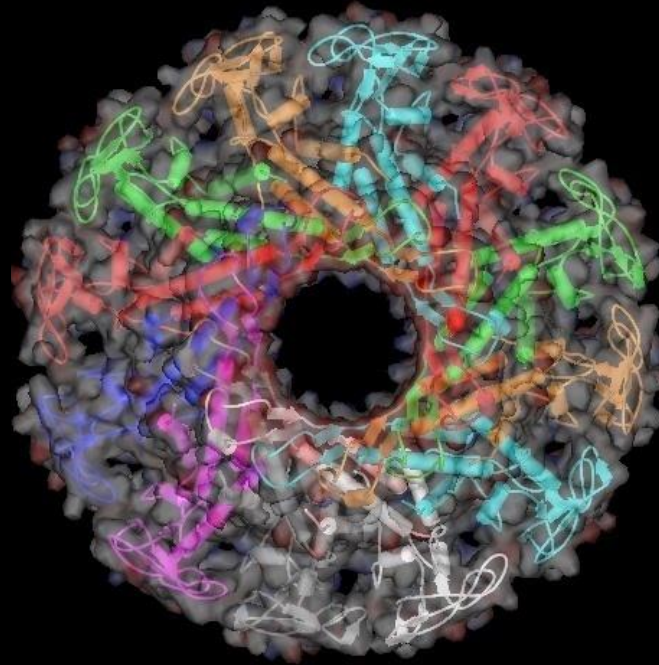


Formes régulières dans les protéines

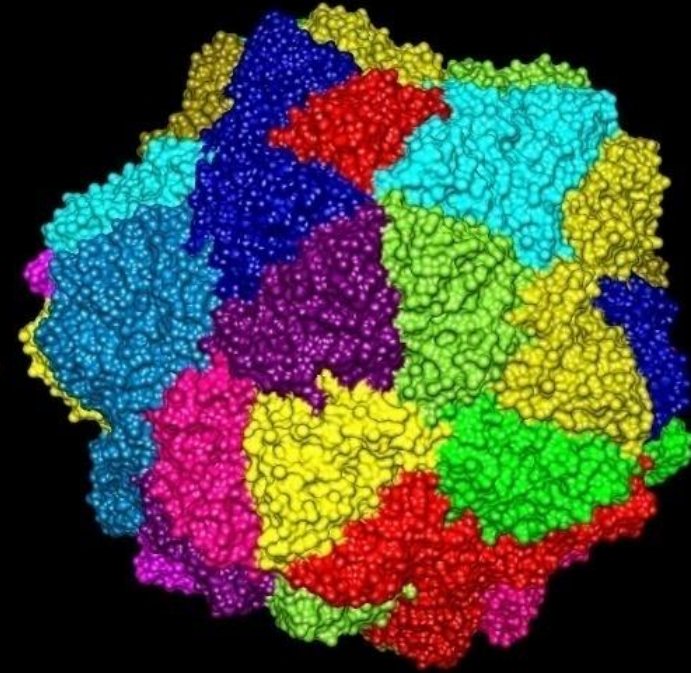
Formes cylindriques



2v6l : seringue d'injection bactérienne

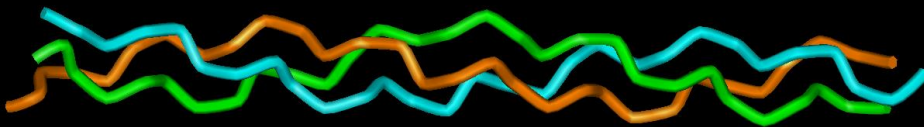


1fou : connecteur du motor du bacteriophage T4



2buk : capsid virale

Formes allongées

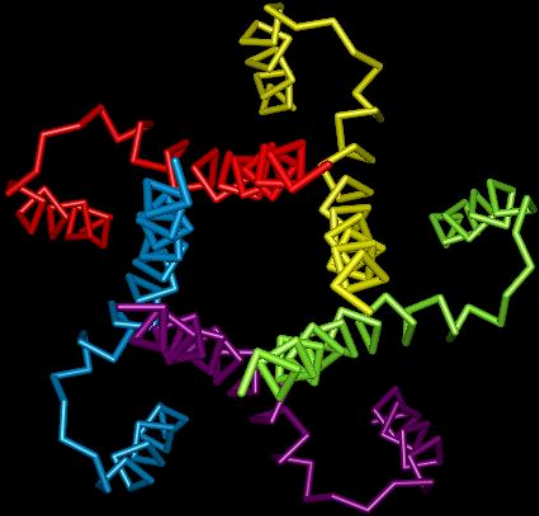


1cag: triple hélice de collagène

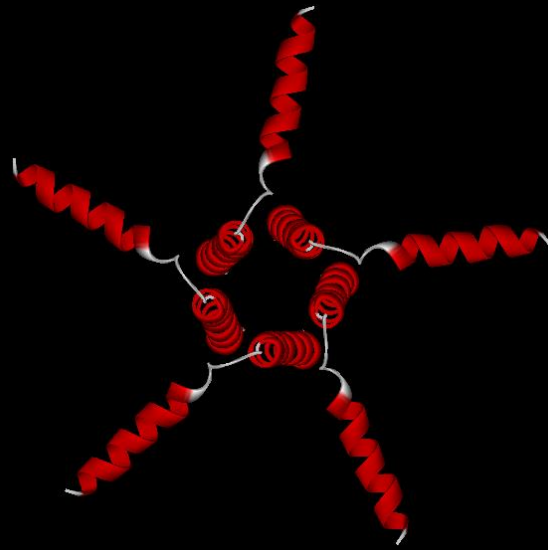


1X03: endophiline
senseur de courbure de membrane

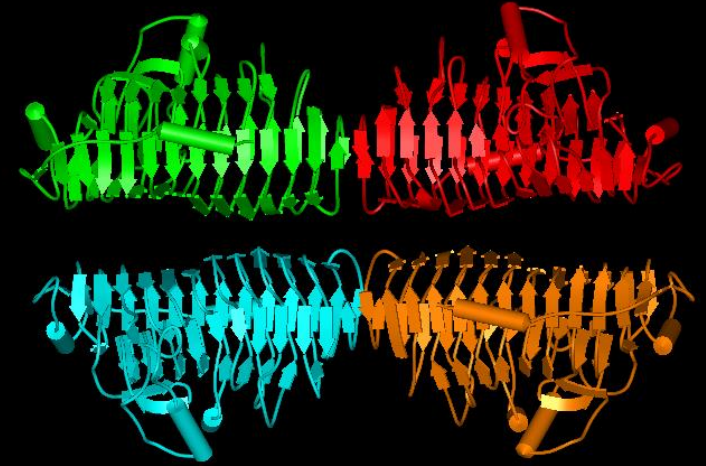
Formes atypiques dans les protéines



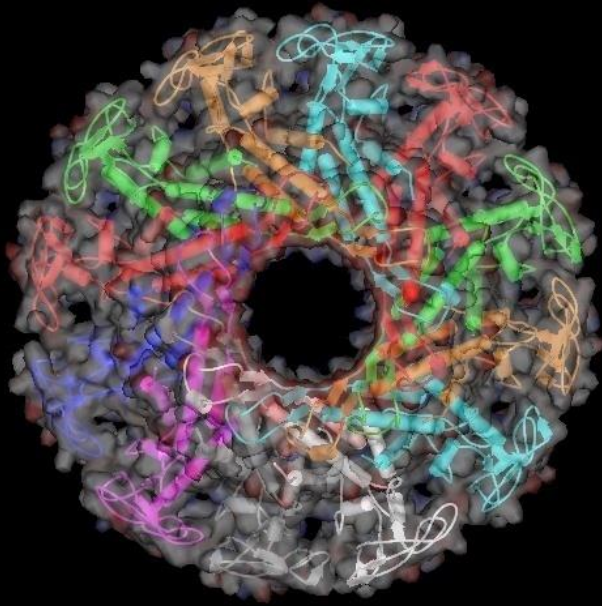
1zll : régulateur calcique



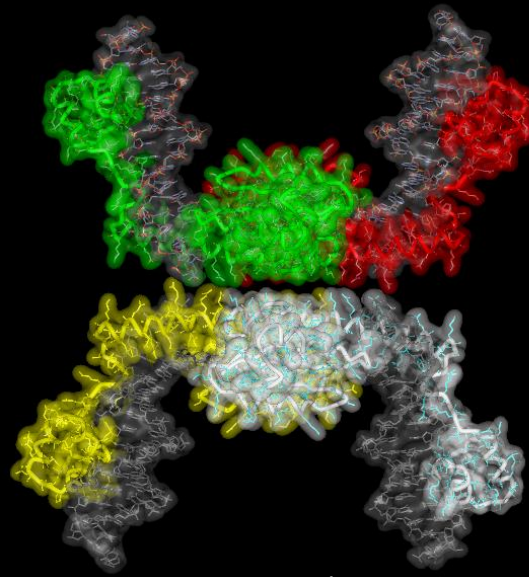
2kyv : inhibiteur de ATPase
calcique sarcolplasmique



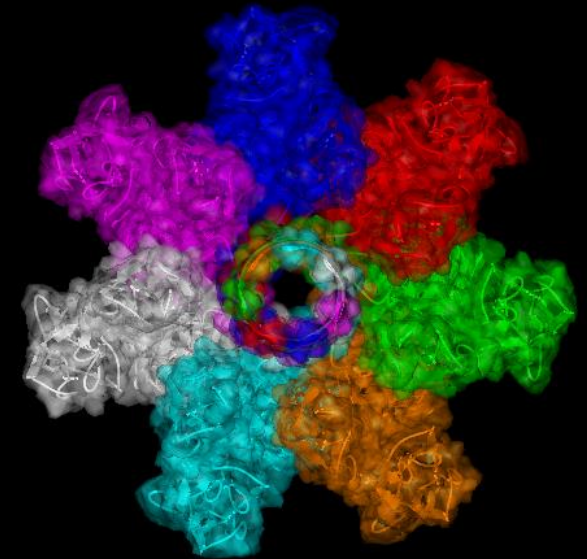
3jur : exopolygalacturonase
Thermotoga maritima



1fou : connecteur du motor du bacteriophage T4



2gm4 : resolvase



3O44 : cytotoxine *V. cholera*

- Archive mondiale des données structurales des macromolécules biologiques
- Créée en 1971 aux Brookhaven National Laboratories : 7 structures
- Depuis 1999, la PDB est sous la responsabilité du RCSB (Research Collaboratory for Structural Bioinformatics)
 - Rutgers university
 - San Diego Supercomputer Center (SDSC)
 - National Institute of Standards and Technology (NIST)
- Conséquences :
 - Utilisation de technologies plus modernes
 - Capture efficace de données
 - Curation des données
 - Introduction d'un nouveau format : mmCIF (système relationnel de représentation des données)





COMMODITY
GRAPHICS CLUSTER
in dual 12.5x12 racks
100 of dual 12.5x12 racks, no million polygons!

14

13

12

11

10



IBM 800
BLUE HORIZON
TERACTIVE NODES



5700SP

Overview

The HEADER record uniquely identifies a PDB entry through the idCode field. This record also provides a classification for the entry. Finally, it contains the date the coordinates were deposited at the PDB.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"HEADER"	
11 - 50	String(40)	classification	Classifies the molecule(s)
51 - 59	Date	depDate	Deposition date. This is the date the coordinates were received by the PDB
63 - 66	IDcode	idCode	This identifier is unique within PDB

Example

```

1           2           3           4           5           6           7           8
1234567890123456789012345678901234567890123456789012345678901234567890
HEADER      COMPLEX (PROTEASE/INHIBITOR)                02-OCT-97    1AWF
  
```


Overview

The ATOM records present the atomic coordinates for standard residues. They also present the occupancy and temperature factor for each atom. Heterogen coordinates use the HETATM record type. The element symbol is always present on each ATOM record; segment identifier and charge are optional.

Record Format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM "	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real (8.3)	x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real (8.3)	y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real (8.3)	z	Orthogonal coordinates for Z in Angstroms.
55 - 60	Real (6.2)	occupancy	Occupancy.
61 - 66	Real (6.2)	tempFactor	Temperature factor.
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

1. Section titre

HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL, REMARK, REMARK 1, REMARK 2, REMARK 3, REMARK 4 - 999

2. Section structure primaire

DBREF, SEQADV, SEQRES, MODRES

3. Section hétérogène

HET, HETNAM, HETSYN, FORMUL

4. Section structure secondaire

HELIX, SHEET, TURN

5. Section annotation de connectivité

SSBOND, LINK, HYDBND, SLTBRG, CISPEP

6. Section autres caractéristiques

SITE

7. Section cristallographie et transformation de coordonnées

CRYST1, ORIGXn, SCALEn, MTRIXn, TVECT

8. Section des coordonnées

MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL

9. Section connectivité

CONECT

10. Section comptable

MASTER, END

PDB Chain Sequences:

- Chain 'A':

Sequence, based on **SEQRES** records: ([download](#))

```
>1umoA (A:)
mekvpgemeierrenseelseaerkavqamwarlyansedvgvailvrffvnfpsakqyf
sqfkhmedplemerspqlrkhasrvmgalntvvenlhdpdkvssvlalvgkahalkhkve
pvyfkilsgvilevvaeeffasdfppetqrawaklrgliyshvtaaykevgwvqvpnatt
ppatlpssgp
```

Sequence, based on observed residues (**ATOM** records): ([download](#))

```
>1umoA (A:)
elseaerkavqamwarlyansedvgvailvrffvnfpsakqyfsqfkhmedplemerspq
lrkhasrvmgalntvvenlhdpdkvssvlalvgkahalkhkvepvyfkilsgvilevvaee
ffasdfppetqrawaklrgliyshvtaaykevgw
```

Attention à la numérotation !!

REMARK FILENAME="V_MINI_5.PDB"

ATOM	1	CA	MET	1	-0.399	20.462	-11.874	1.00	0.00
ATOM	2	HA	MET	1	-0.827	19.859	-11.089	1.00	0.00
ATOM	3	CB	MET	1	0.010	19.571	-13.053	1.00	0.00
ATOM	4	HB1	MET	1	1.081	19.428	-13.041	1.00	0.00
ATOM	5	HB2	MET	1	-0.279	20.044	-13.980	1.00	0.00
ATOM	6	CG	MET	1	-0.683	18.212	-12.934	1.00	0.00
ATOM	7	HG1	MET	1	-1.020	17.895	-13.910	1.00	0.00
ATOM	8	HG2	MET	1	-1.532	18.296	-12.271	1.00	0.00
ATOM	9	SD	MET	1	0.483	16.996	-12.273	1.00	0.00
ATOM	10	CE	MET	1	-0.650	16.201	-11.107	1.00	0.00
ATOM	11	HE1	MET	1	-0.128	15.414	-10.579	1.00	0.00
ATOM	12	HE2	MET	1	-1.484	15.778	-11.644	1.00	0.00
ATOM	13	HE3	MET	1	-1.014	16.936	-10.402	1.00	0.00
ATOM	14	C	MET	1	0.810	21.246	-11.346	1.00	0.00
ATOM	15	O	MET	1	0.910	22.444	-11.535	1.00	0.00
ATOM	16	N	MET	1	-1.425	21.397	-12.428	1.00	0.00
ATOM	17	HT1	MET	1	-1.044	21.874	-13.270	1.00	0.00
ATOM	18	HT2	MET	1	-1.671	22.108	-11.709	1.00	0.00
ATOM	19	HT3	MET	1	-2.277	20.863	-12.692	1.00	0.00
ATOM	20	N	ASP	2	1.727	20.570	-10.686	1.00	0.00
ATOM	21	HN	ASP	2	1.615	19.607	-10.554	1.00	0.00
ATOM	22	CA	ASP	2	2.943	21.252	-10.133	1.00	0.00
ATOM	23	HA	ASP	2	3.567	20.537	-9.623	1.00	0.00
ATOM	24	CB	ASP	2	3.676	21.804	-11.352	1.00	0.00
ATOM	25	HB1	ASP	2	3.322	22.800	-11.564	1.00	0.00
ATOM	26	HB2	ASP	2	3.477	21.165	-12.195	1.00	0.00

Exemple de 3LC6 (PDB)



Pas d'indication de connectivité des atomes

Calcul de la connectivité

- Basé sur la distance (problème de la chimie incorrecte)
- Basé sur la reconnaissance des atomes, acides aminés (problème de la nomenclature et de la complétude)

Atomes manquants:

Parties manquantes (troncatures pour cristalliser)

Mauvaise résolution ou définition

Volonté délibérée (industrie pharmaceutique)

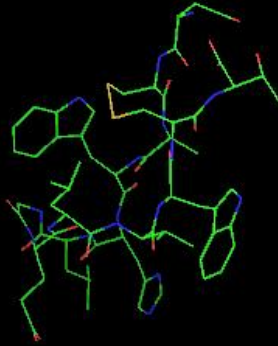
Atomes supplémentaires

Montage de surexpression (ajout d'une partie de séquence)

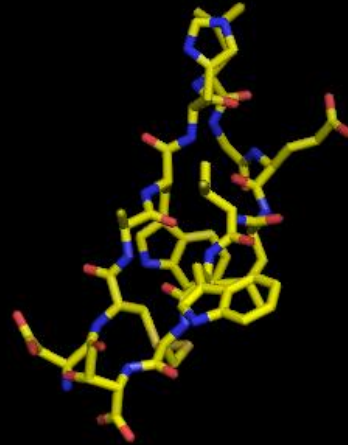
Modification chimique (aa non naturels)

Chimère pour les protéines membranaires

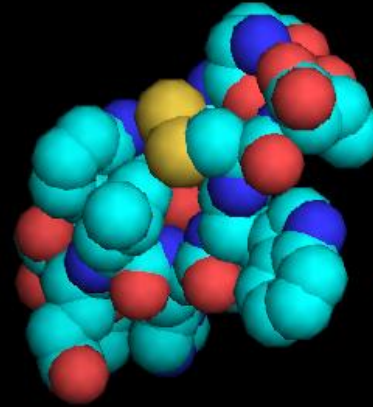
lines



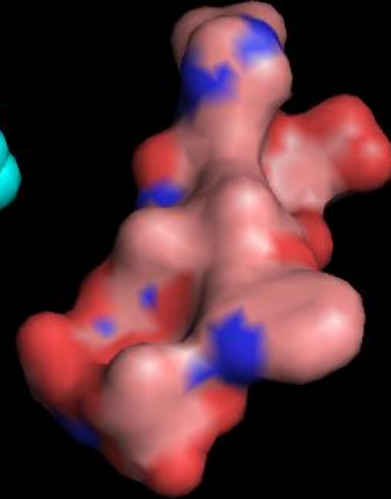
sticks



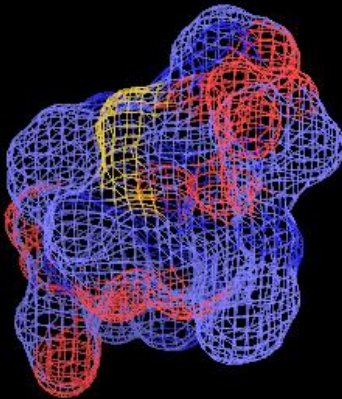
spheres



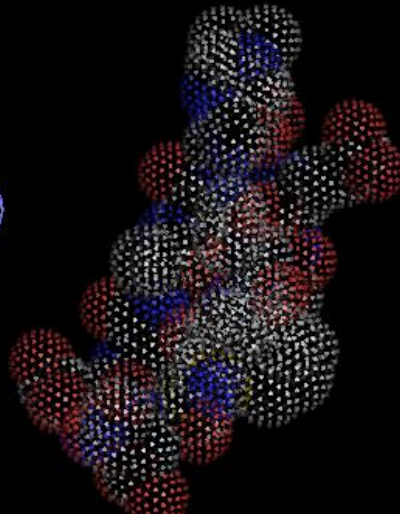
surface



mesh



dots



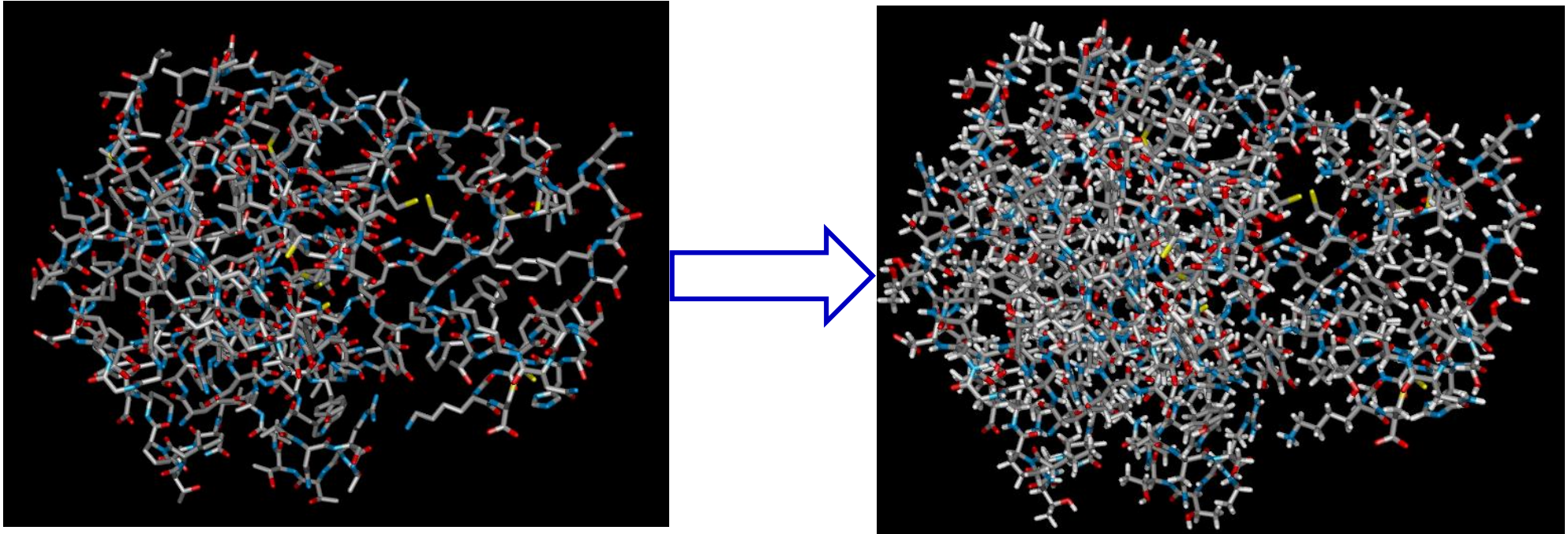
ribbon



cartoon



Les H représentent \sim la moitié des atomes

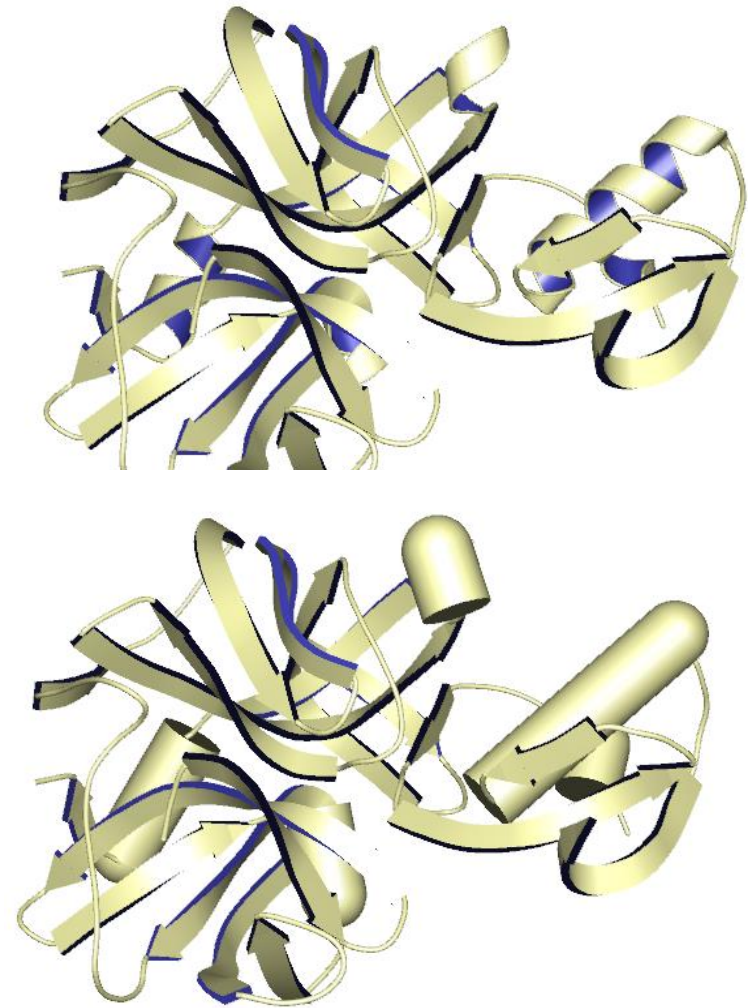
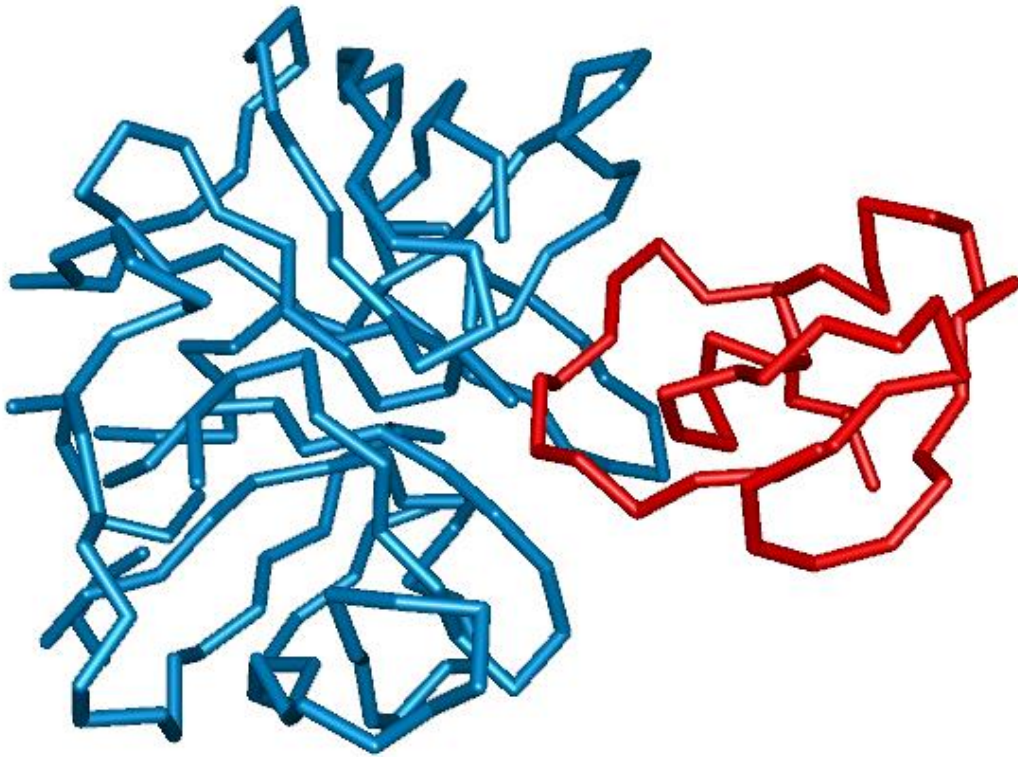


Ajouter les H avant tout calcul de surface, d'interaction ou d'énergie
Les enlever pour aider à mieux voir le repliement

Modes schématiques (pas de réalité chimique)

Modes trace ou carbonés α

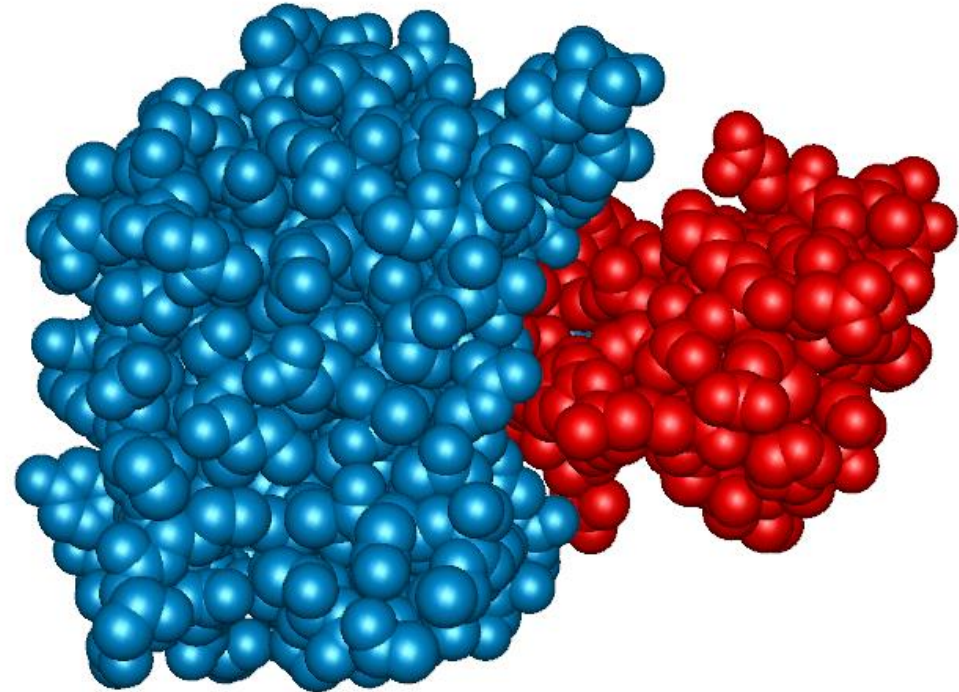
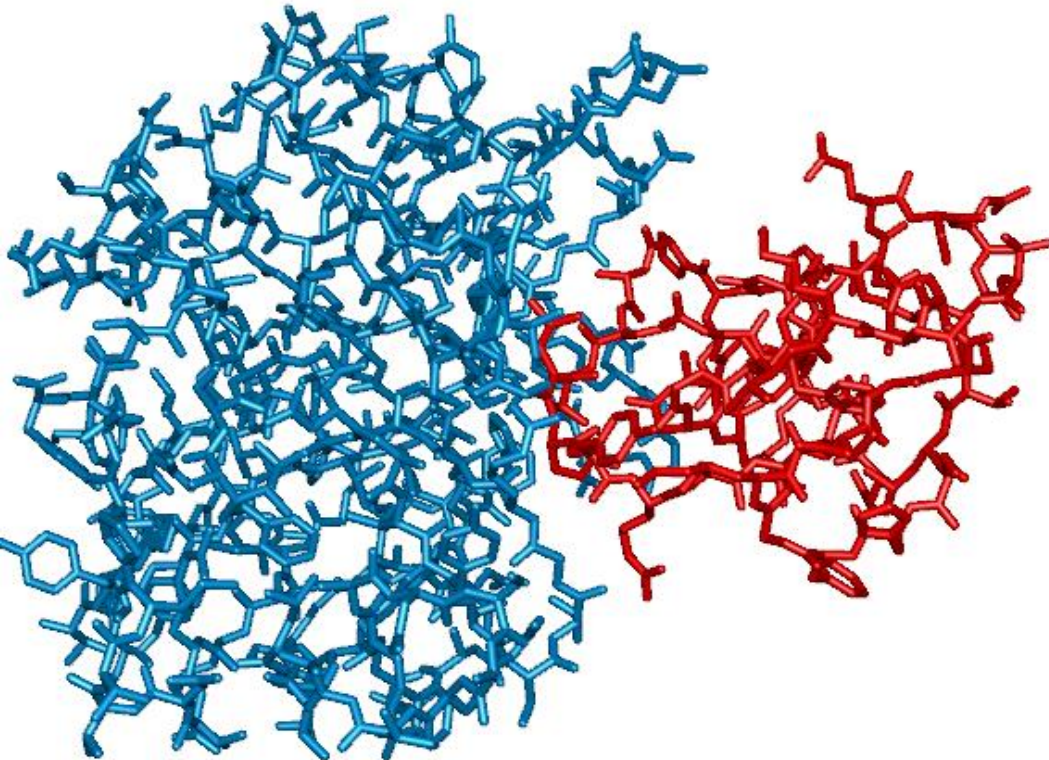
Rubans ou Cartoons



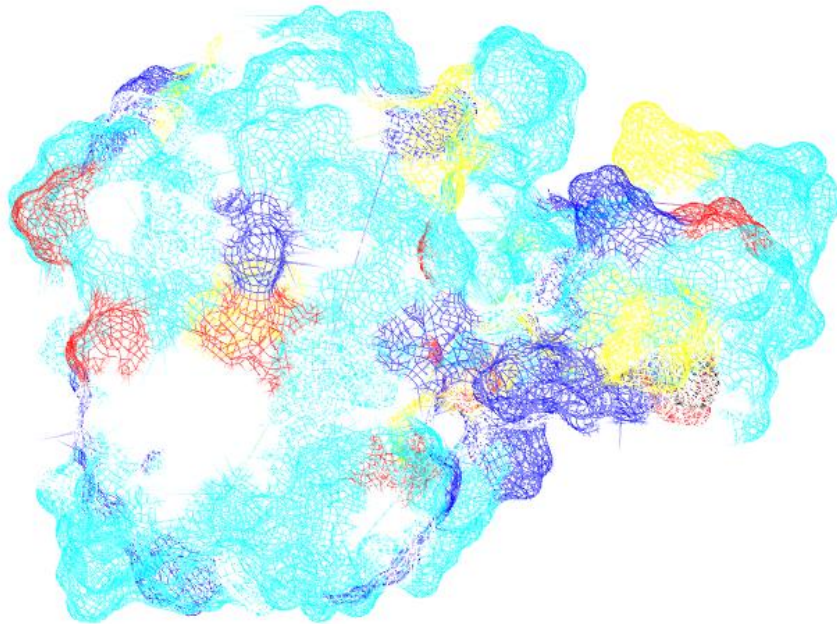
Fils de Fer, Dreiding, Sticks

Sphères pleines

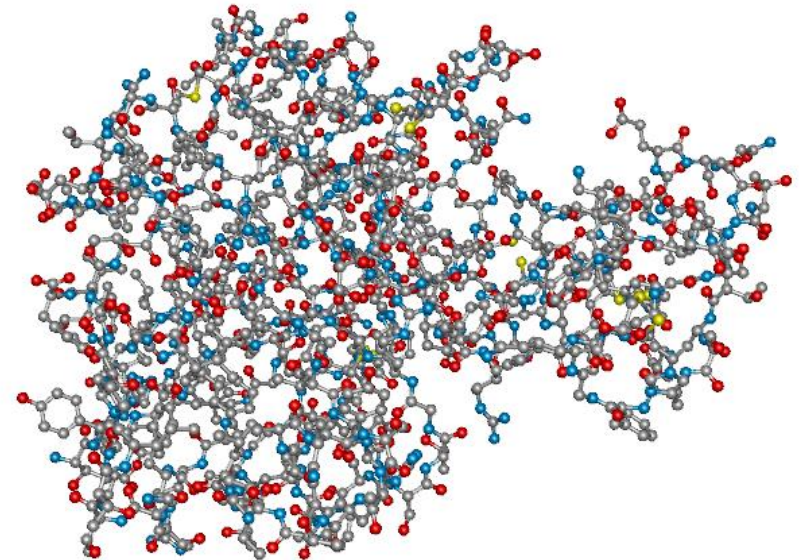
Codage en couleur chaîne

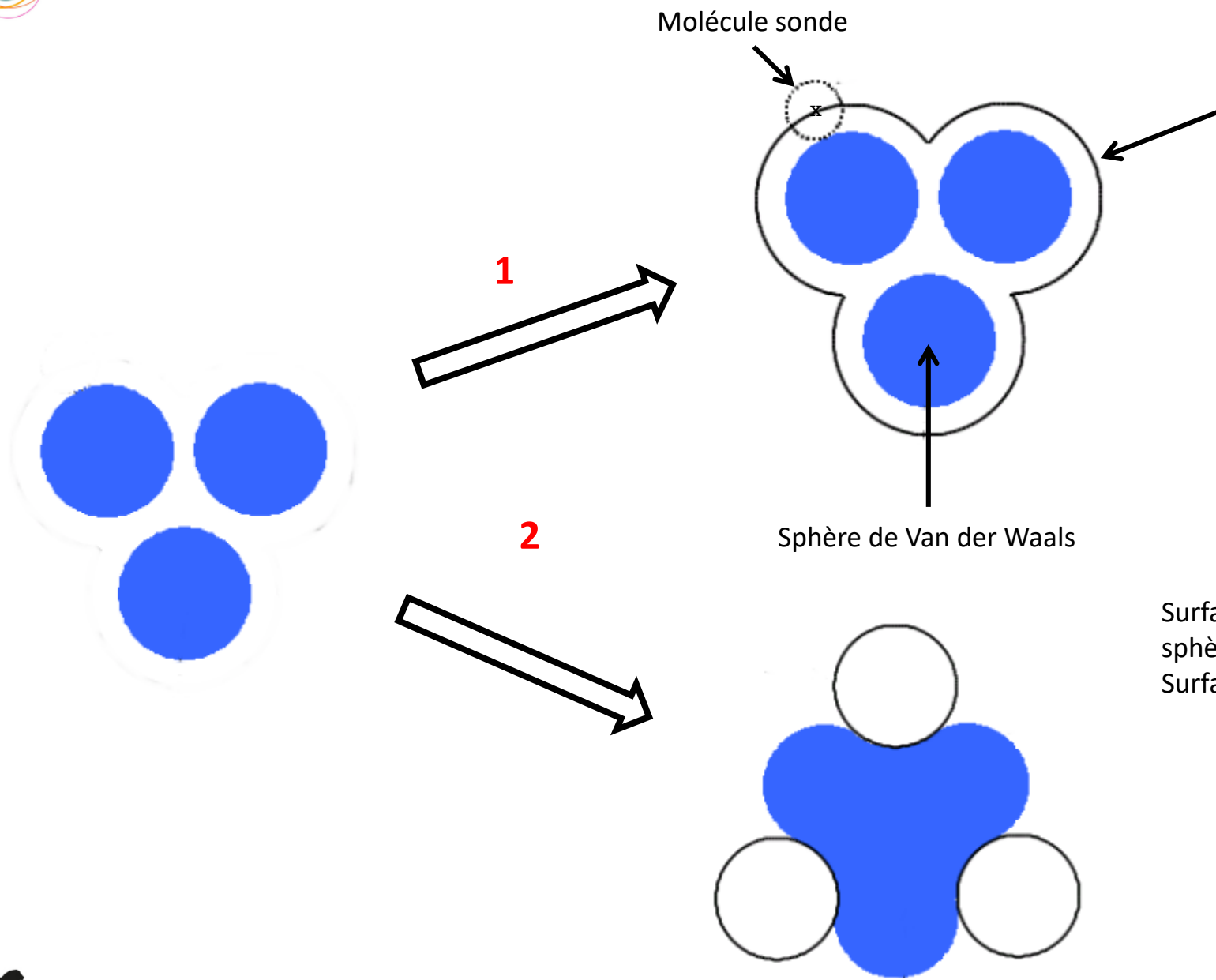


Surfaces moléculaires



Boules et bâtons ball and sticks





Surface accessible au solvant décrite par le déplacement du centre d'une sphère de rayon fixée à l'avance (SAS).

Taille de la sphère mimant une molécule d'eau $\sim 1.4 \text{ \AA}$

Surface de contact de la sphère avec les atomes
Surface de Connolly

T. Can et al. / Journal of Molecular Graphics and Modelling 25 (2006) 442–454

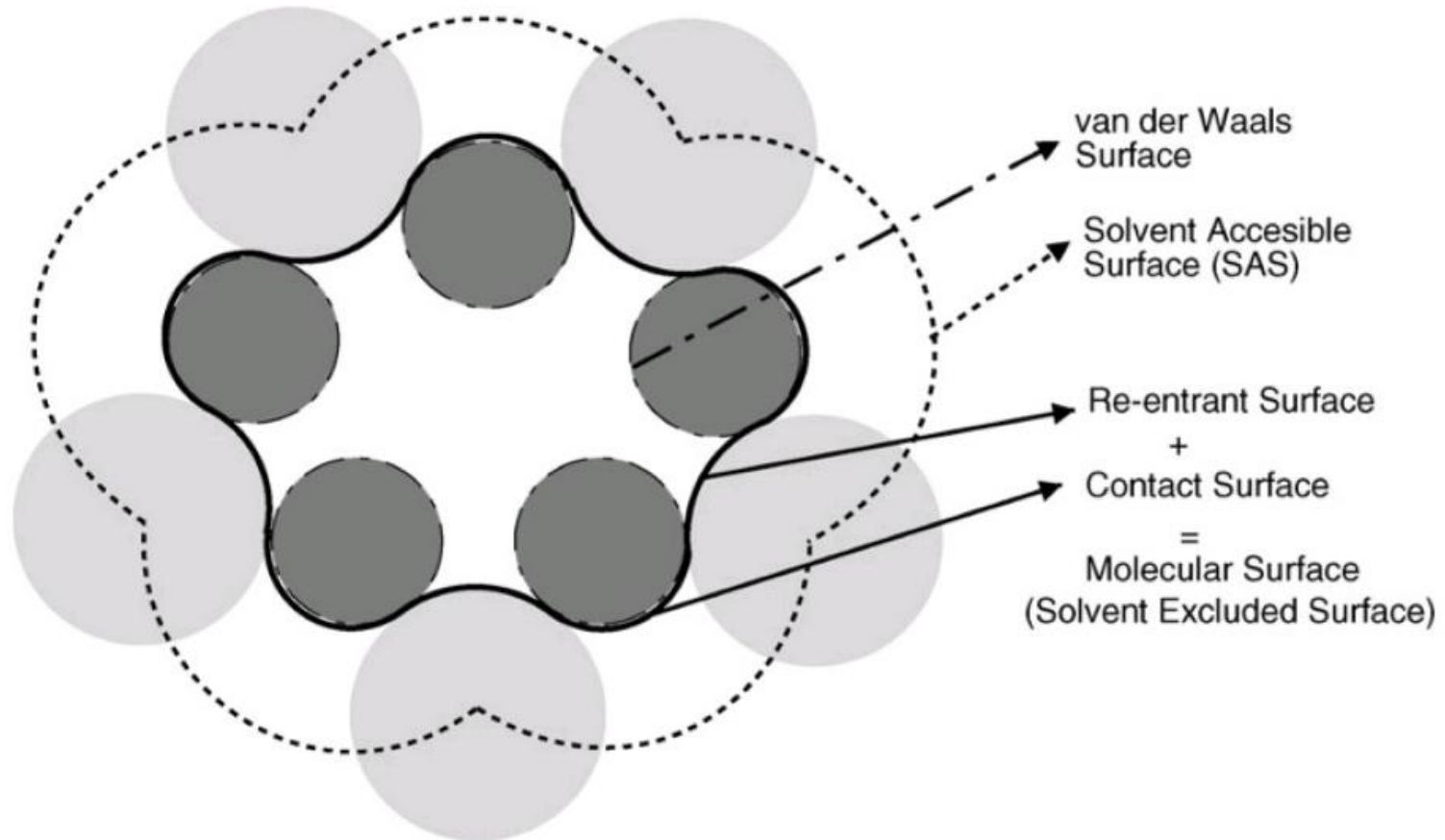
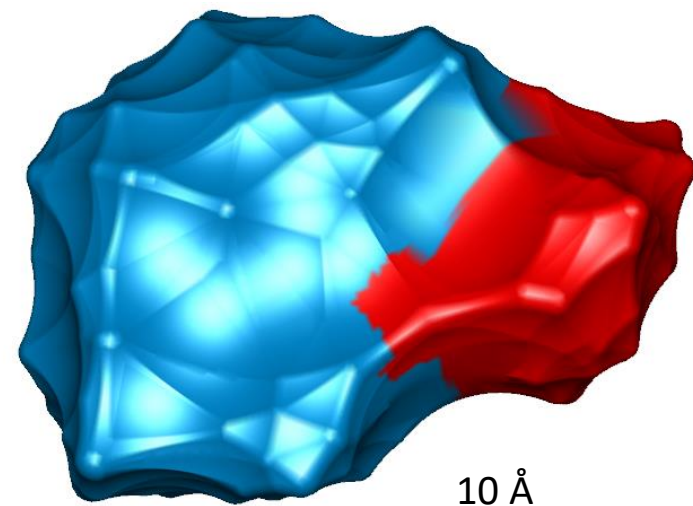
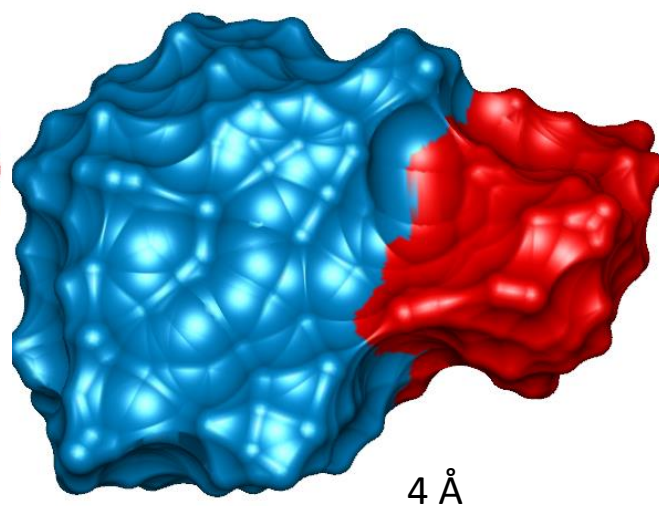
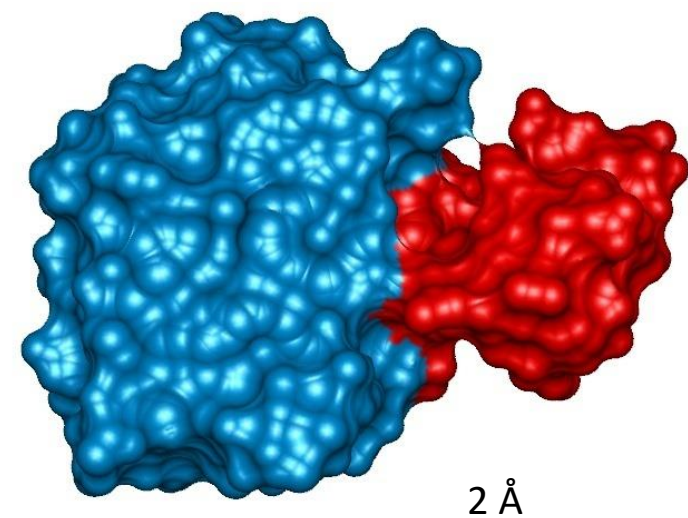
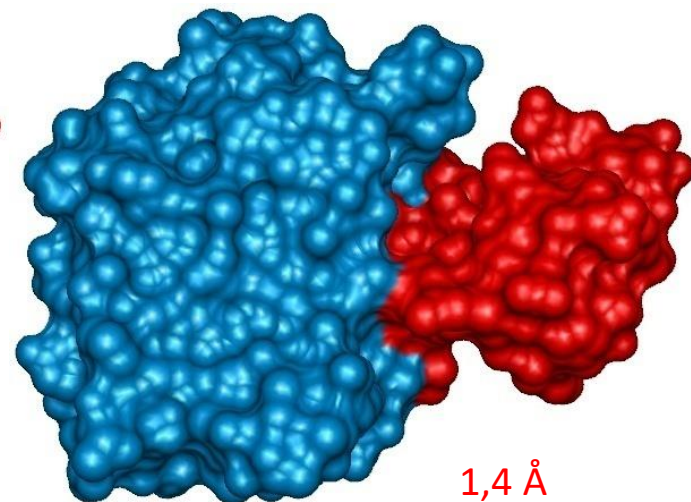
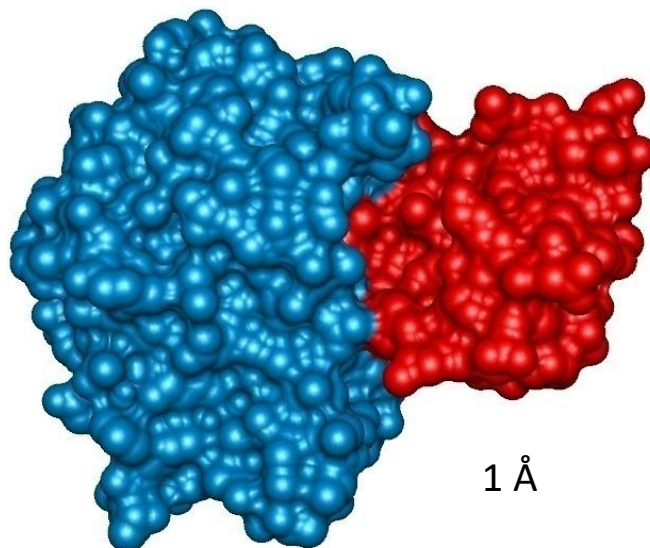
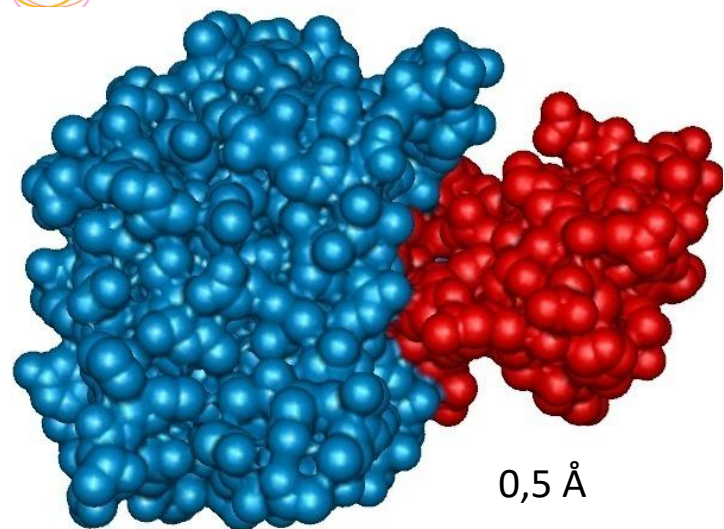
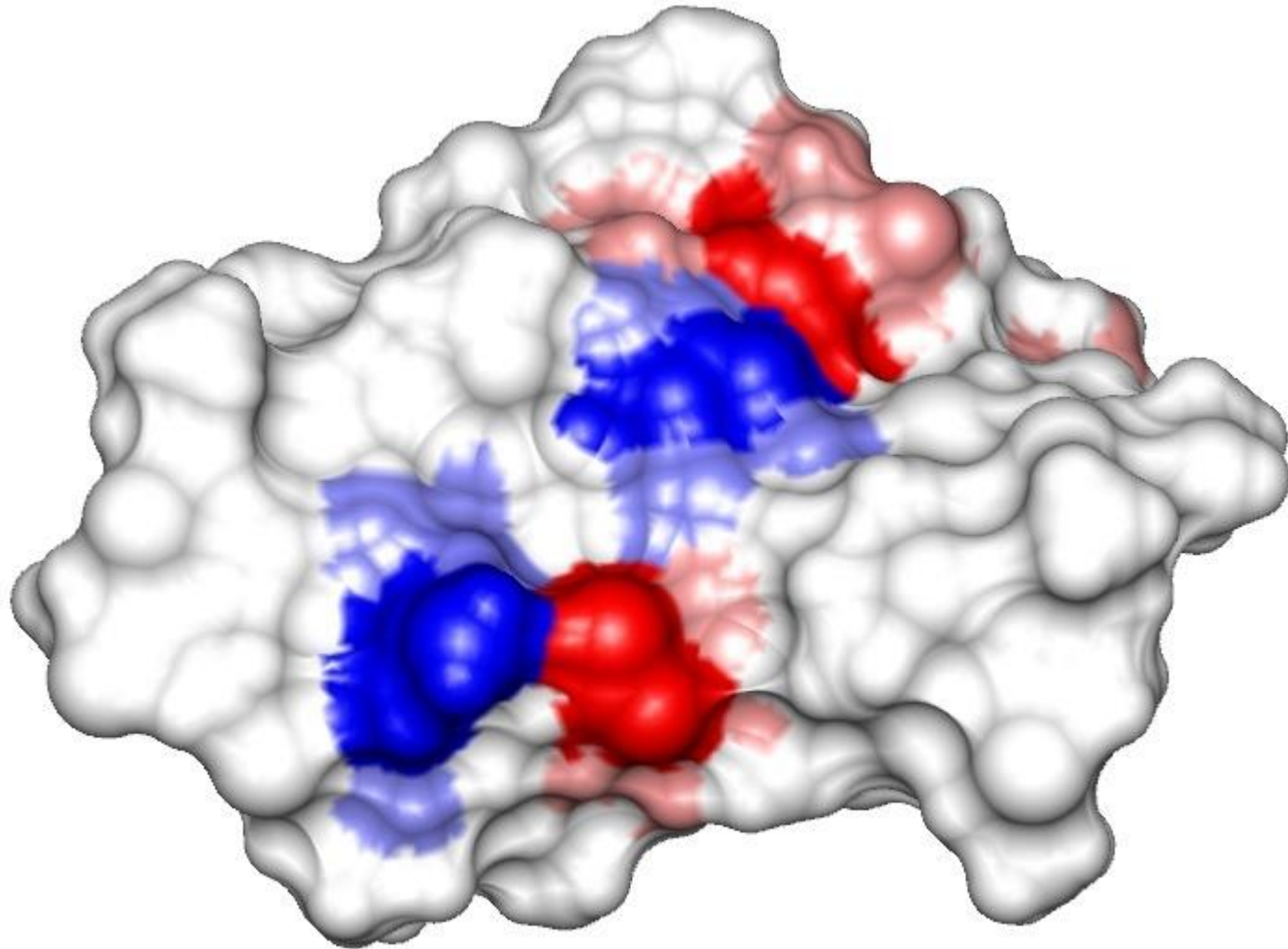


Fig. 1. A two-dimensional illustration of surface definitions.



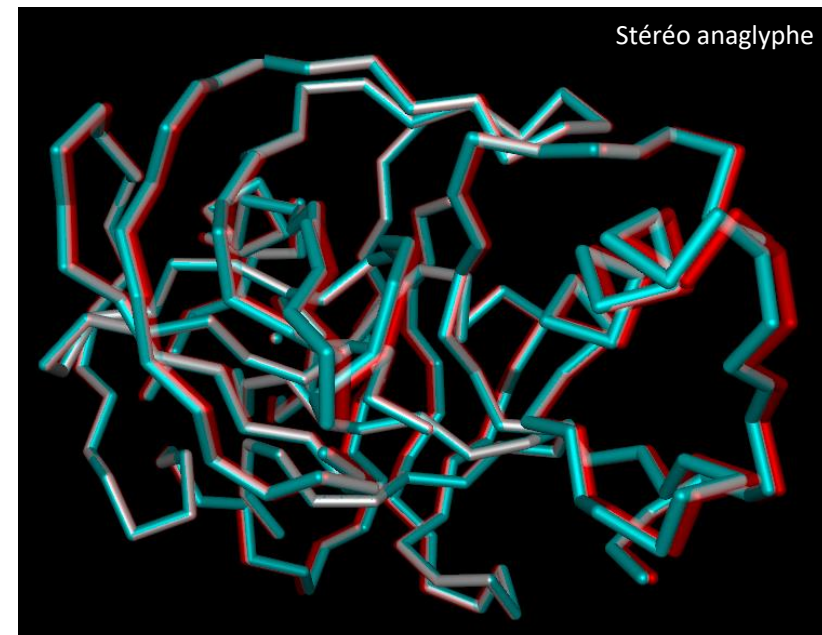
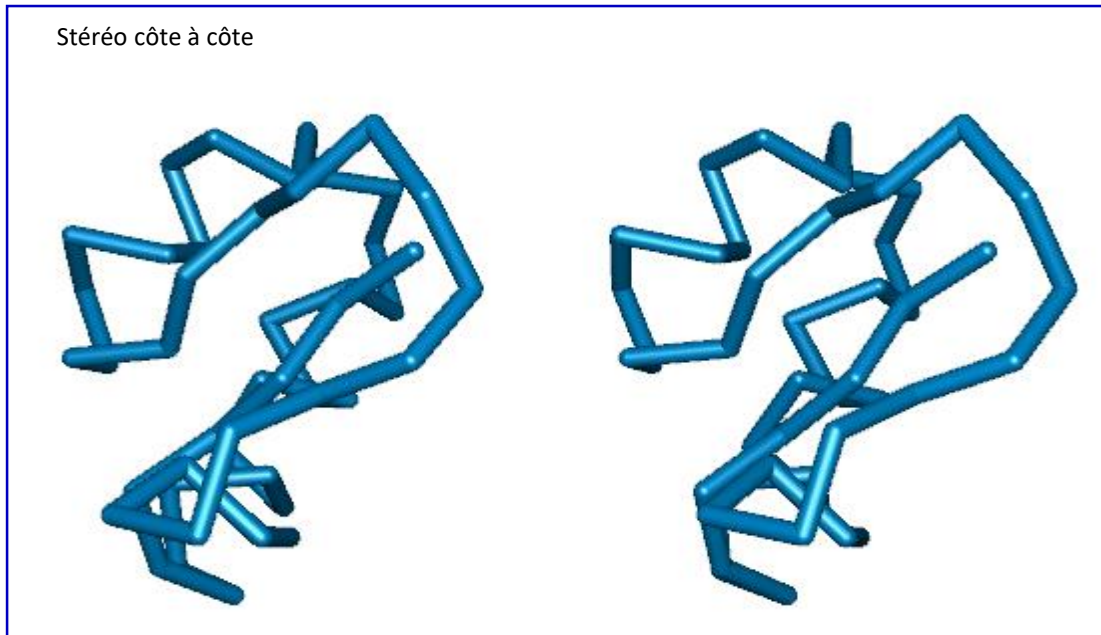


Pont salin

2 images décalées reproduisant la vision produite par l'écart entre les 2 yeux (~ 6 cm)

<http://www.stereoscopie.fr> <http://canarlake.org/stereo/>

- ✓ Anaglyphe (exemple une image rouge et une verte) (simple et peu couteux, bien pour le NB, pénible en couleur)
- ✓ Image côte à côte (5° de rotation) (simple, peu coûteux, pas d'effort de programmation, surtout bien pour les impressions, inconfortable, lunettes convergentes)
- ✓ Entrelacement horizontal (ou vertical) (une image sur les lignes impaires et l'autre sur les lignes paires (bien en couleur, nécessite des lunettes polarisées et un écran adapté, problème de distance par rapport à l'écran, nécessité langage graphique dédié OpenGL)
- ✓ Quad buffer (le meilleur système) (nécessite une carte et un écran 120Hz, OpenGL). Les 2 images sont stockées dans des mémoires tampons qui sont alternativement affichées (écran polarisé ou vidéoprojection)
- ✓ Lunettes actives (liaison IR entre un boîtier et les lunettes, problème de flickering et d'interférence avec les lumières alternatives)
- ✓ A venir les systèmes sans lunettes?



ViewerLite (ou Web Lab viewer) (Version allégée d'un produit Accelrys)

WLViewerLite32.exe

WLViewerLite40.exe

ViewerLite42.exe

ViewerLite50.exe

AnTheProt 3D

AntheProt_3D.exe

Rasmol 2.7.3

Raswin_2_7_1.exe

Raswin_2_7_3.exe

PyMol

pymol-0_99rc6-bin-win32.exe

VMD

software/vmd171.exe

Swiss-PDB Viewer

spdbv37.exe

Chimera

www.cgl.ucsf.edu/chimera/

Yasara

www.yasara.org/

Galleries : <http://antheprot-pbil.ibcp.fr/3D>
<http://www.ks.uiuc.edu/Research/vmd/gallery/>
<http://pymol.sourceforge.net/pmimag/screen.html>

ViewerLite (Version allégée d'un ex produit Accelrys)

ViewerLite ??

AnTheProt 3D

<http://antheprot-pbil.ibcp.fr>

Rasmol 2.7.3

<http://www.openrasmol.org/>

PyMol

<https://pymol.org/>

VMD

<https://www.ks.uiuc.edu/Research/vmd/>

Swiss-PDB Viewer

<https://spdbv.vital-it.ch/>

Chimera

<https://www.cgl.ucsf.edu/chimera/>

Yasara

www.yasara.org/

Moe (commercial)

<https://www.chemcomp.com/>

Galleries : <http://antheprot-pbil.ibcp.fr/3D>
<http://www.ks.uiuc.edu/Research/vmd/gallery/>
<http://pymol.sourceforge.net/pmimag/screen.html>

Visualisation, construction, optimisation de structures 3D

- ✓ Fil de fer, sphères ombrées, carbones α , chaîne principale, surfaces, «cartoons»
- ✓ Sélection par type d'atomes, acides aminés, chaîne, segment, SS-bonds, Hbonds
- ✓ Codage de couleur associé : atomes, chaîne, propriétés, etc..
- ✓ Etiquettes (atomes, acides aminés, ligands)

Différents logiciels de modélisation

- | | | |
|--------------------|-----------------------------|---|
| ✓ Rasmol | multi-plateforme | http://www.umass.edu/microbio/rasmol/ |
| ✓ AntheProt | Windows | http://antheprot-pbil.ibcp.fr |
| ✓ ViewerLite | MacOS et Windows | http://www.accelrys.com |
| ✓ Swiss-PDB viewer | Windows, Linux, MacOS, IRIX | http://www.expasy.ch/spdbv |
| ✓ PyMol | multi-plateforme | http://pymol.sourceforge.net/ |
| ✓ VMD | Windows, Unix, MacOS | http://www.ks.uiuc.edu/Research/vmd/ |
| ✓ Modeller | Web | http://salilab.org/modeller/modeller.html |
| ✓ Geno3D | Web | http://geno3d-pbil.ibcp.fr |
| ✓ SuMo | Web | http://sumo-pbil.ibcp.fr |

Géométrie, Ramachandran, chiralités

- ✓ Distances, angles, Φ, Ψ
- ✓ Diagrammes de Ramachandran
- ✓ Empêchements stériques
- ✓ Affichage des voisins
- ✓ Affichage des liaisons hydrogènes

Comparaison/superposition et évaluations des structures

- ✓ Mesure du RMSD (Ecart quadratique moyen des distances entre atomes)
- ✓ Superposition LOCALE ou GLOBALE
 - ✓ CE = Combinatorial Extension <http://cl.sdsc.edu/ce.html>
 - ✓ CL = Compound Likeness <http://cl.sdsc.edu/cl>
 - ✓ VAST (Vector Alignment Search Tool) <http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>
- ✓ Evaluation Procheck <http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>
- ✓ Whatif <http://www.cmbi.kun.nl/whatif/>

Classification des structures

- ✓ CATH http://www.biochem.ucl.ac.uk/bsm/cath_new
- ✓ SCOP <http://scop.mrc-lmb.cam.ac.uk/scop/>
- ✓ FSSP/HSSP <http://www.ebi.ac.uk/dali/fssp/>

✓ Pourquoi comparer?

- ✓ Regrouper les protéines par similarité structurale (classification)
- ✓ Estimer l'importance de certains résidus dans la structure (stabilité, mutabilité)
- ✓ Identifier des homologues distants vis à vis de familles de protéines (modélisation)
- ✓ Prédire la fonction des protéines ayant de faible taux de similarité avec d'autres protéines

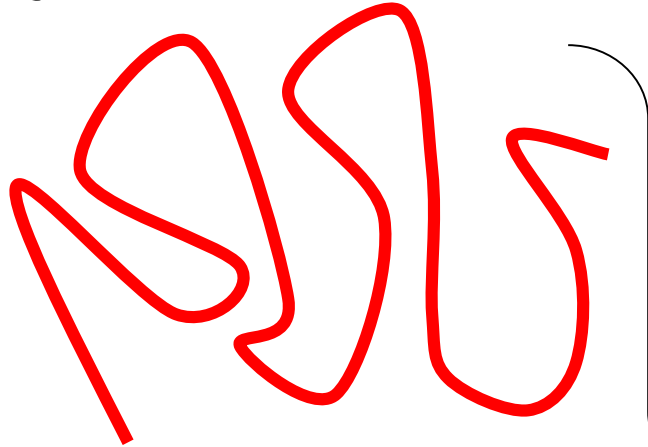
✓ Comment?

Comparaison de 2 objets = calcul du RMSD

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2 \right]}$$

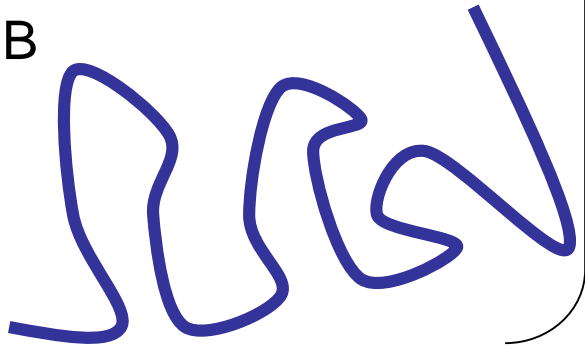
Superposition = positionnement des 2 structures de façon à minimiser le RMSD

Structure A

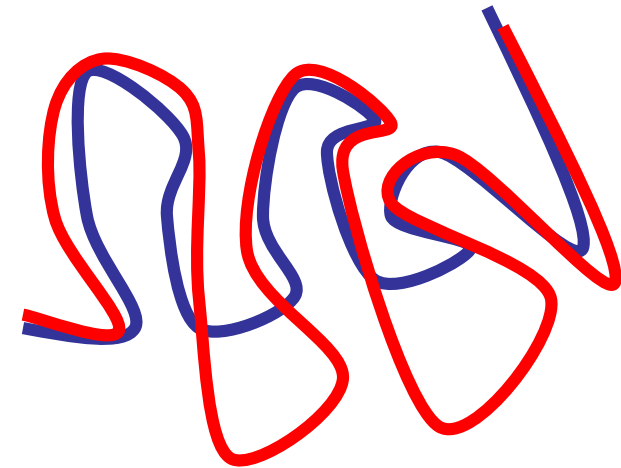
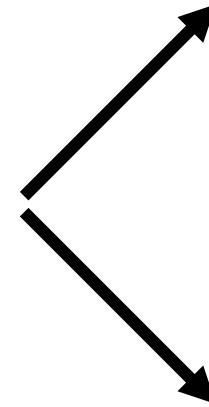
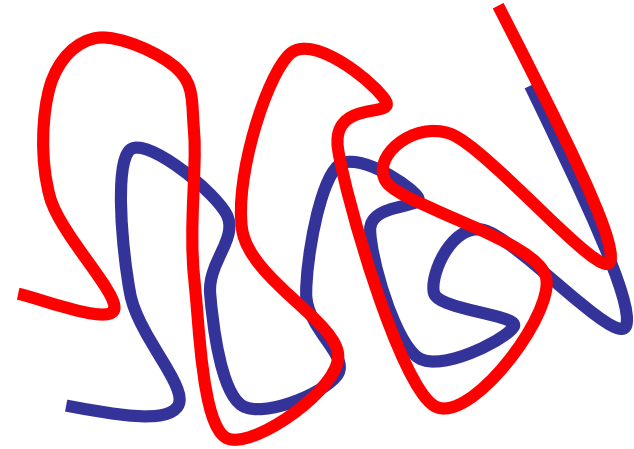


+

Structure B



Superposition globale trouver le maximum d'atomes alignés avec le plus faible écart. Meilleur compromis



Superposition locale trouver les régions les plus ressemblantes au niveau 3D

✓ Comment?

- ✓ Ecart quadratique moyen global des distances entre les n atomes i en correspondance dans les 2 structures

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2]}$$

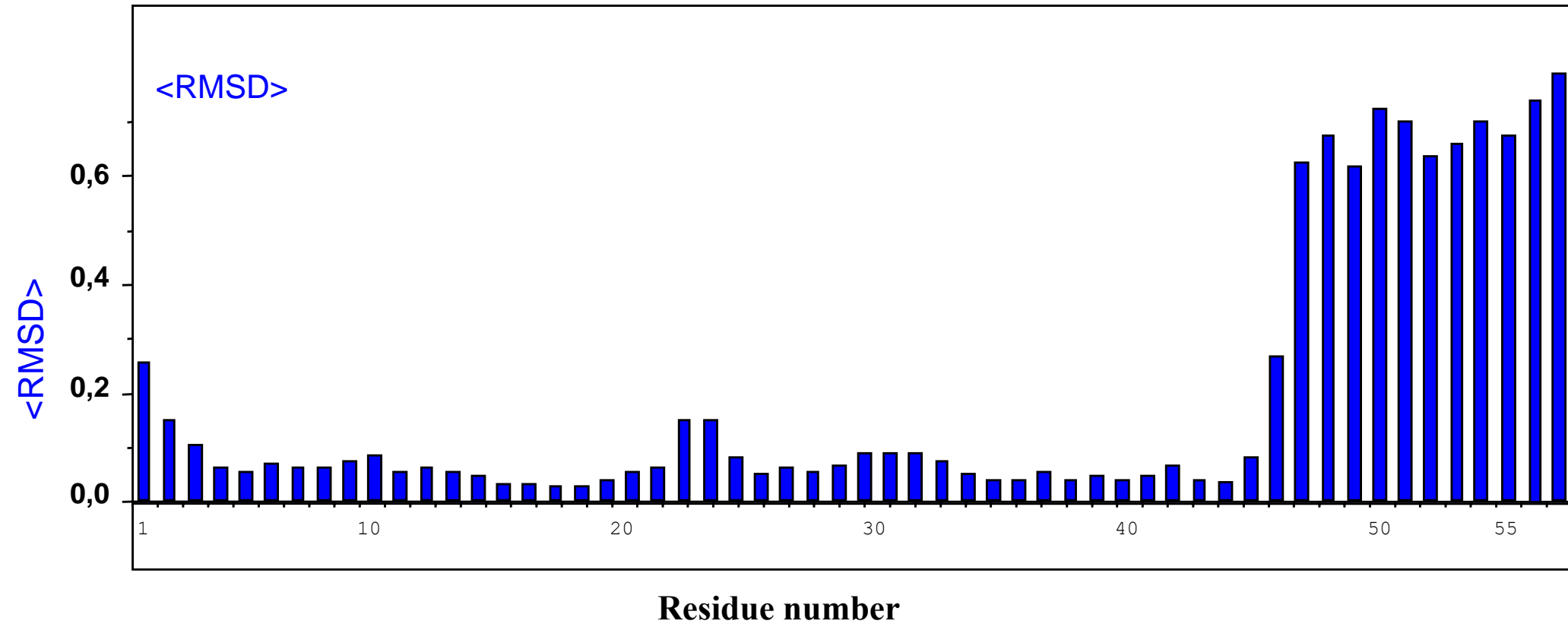
- ✓ Alignement optimal des structures revient à minimiser le RMSD

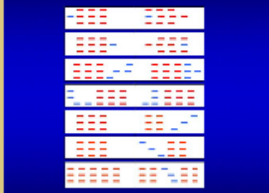
$$RMSD(j) = \sum_{J=1}^M \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2]}$$

- ✓ Idem mais local sur une fenêtre glissante
- ✓ Choix des atomes entre les 2 structures? Tous? Que ceux identiques après alignement? Le backbone?, Les C α (tous?, seuls ceux des structures secondaires)

Structure A THESESENTENCESALIGN--NICELY
 ||| || |||| ||||| |||||
 Structure B THE--SEQ-ENCE-ALIGNEDNICELY

$$RMSD(j) = \sum_{J=1}^M \sqrt{\frac{1}{N} \sum_{i=1}^N [(x_i - x'_i)^2 + (y_i - y'_i)^2 + (z_i - z'_i)^2]}$$



<http://pathway.rit.albany.edu/~cemc/><http://schubert.bio.uniroma1.it/CEMC/>

THE GUDA LAB

DMAPS HOME

ALGORITHM

REFERENCES

DOWNLOAD

CONTACT

CREDITS

FAQ'S

CEMC - Multiple Protein Structure Alignment Server

If you already have a Job ID, submit here to retrieve results.

Enter **6-character** PDB IDs (space-delimited, maximum 20)

Note: Enter ALL PDB IDs here, even if you are uploading local file(s)

Z-score Cutoff Distance Cutoff

Type e-mail address, to receive your results (Required)

Upload local coordinates files, if any (Optional)

Note: File name and data must be in standard PDB format. For example: "pdb1bbs.ent"

<input type="text"/>	<input type="button" value="Parcourir..."/>	<input type="text"/>	<input type="button" value="Parcourir..."/>
<input type="text"/>	<input type="button" value="Parcourir..."/>	<input type="text"/>	<input type="button" value="Parcourir..."/>
<input type="text"/>	<input type="button" value="Parcourir..."/>	<input type="text"/>	<input type="button" value="Parcourir..."/>
<input type="text"/>	<input type="button" value="Parcourir..."/>	<input type="text"/>	<input type="button" value="Parcourir..."/>

Your input parameters are Valid.

PDB IDs: = 1N1B:A 1N20:A 1N23:A 1N1Z:A

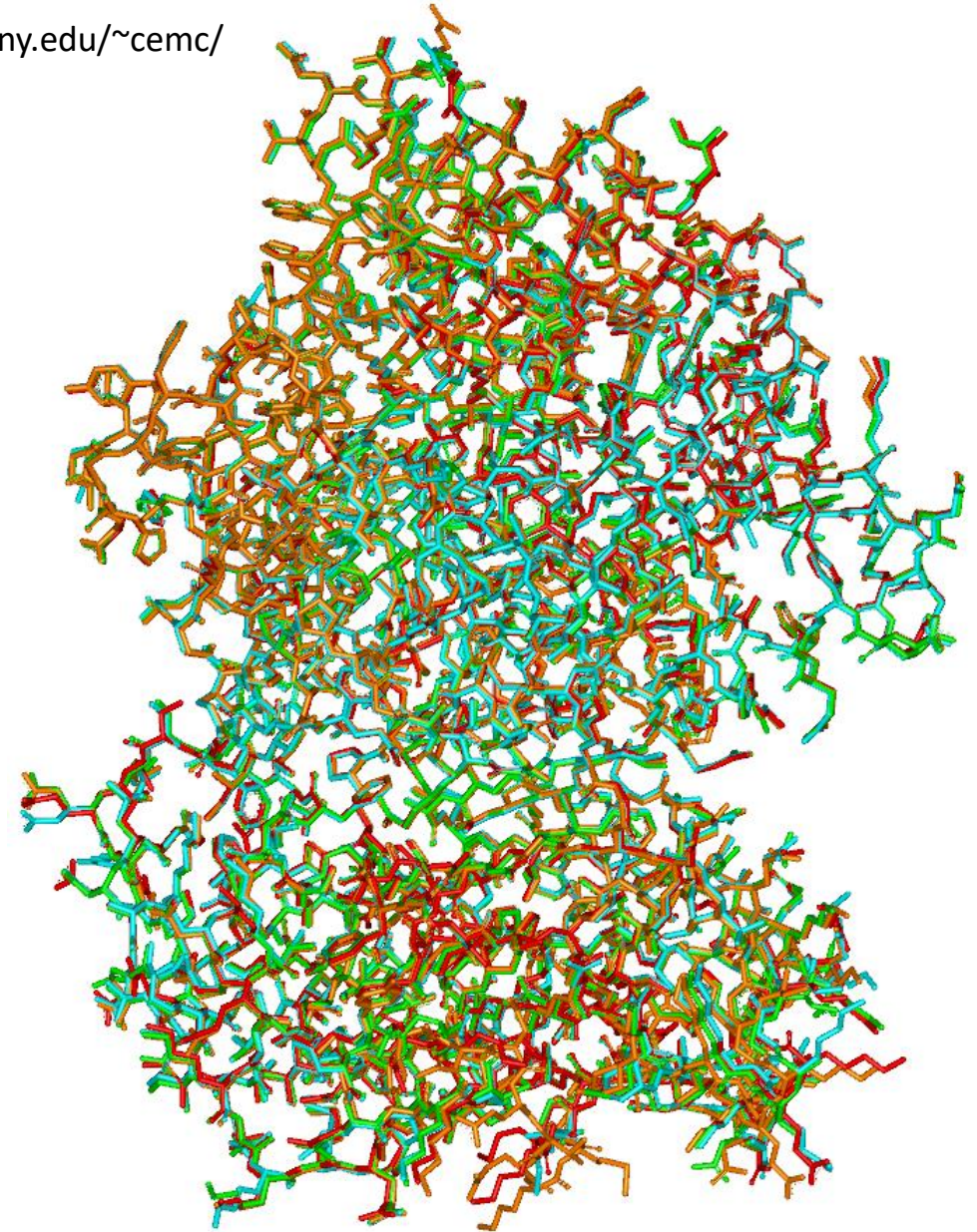
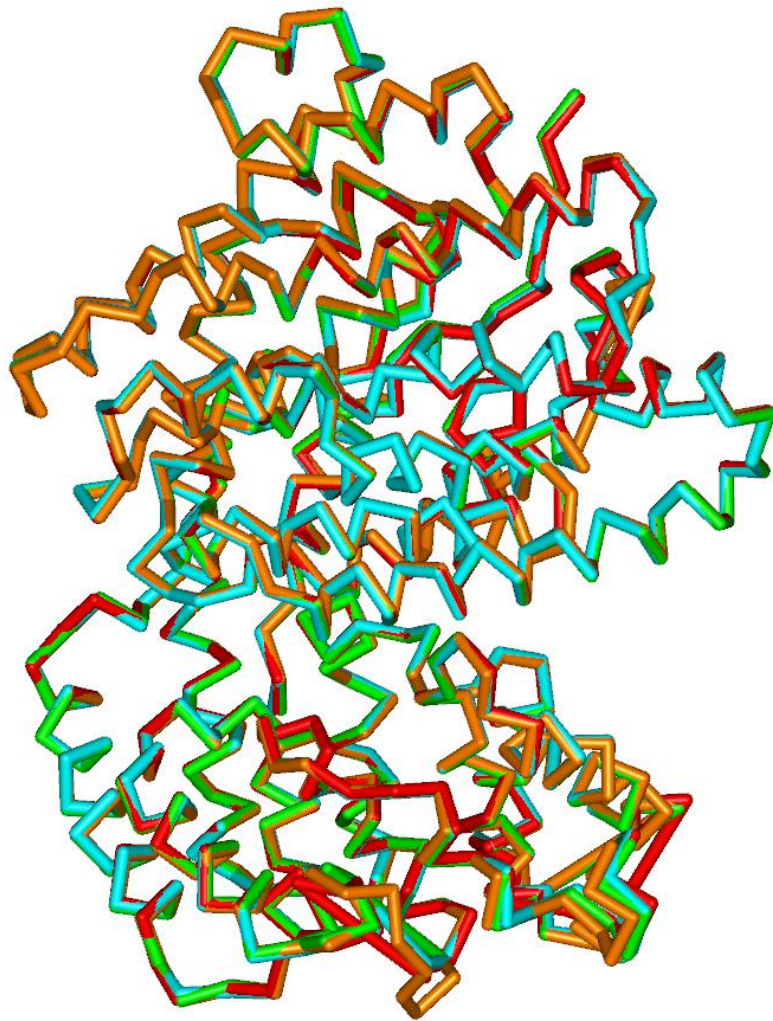
PDB files are loaded successfully! ...

Your JOB ID is

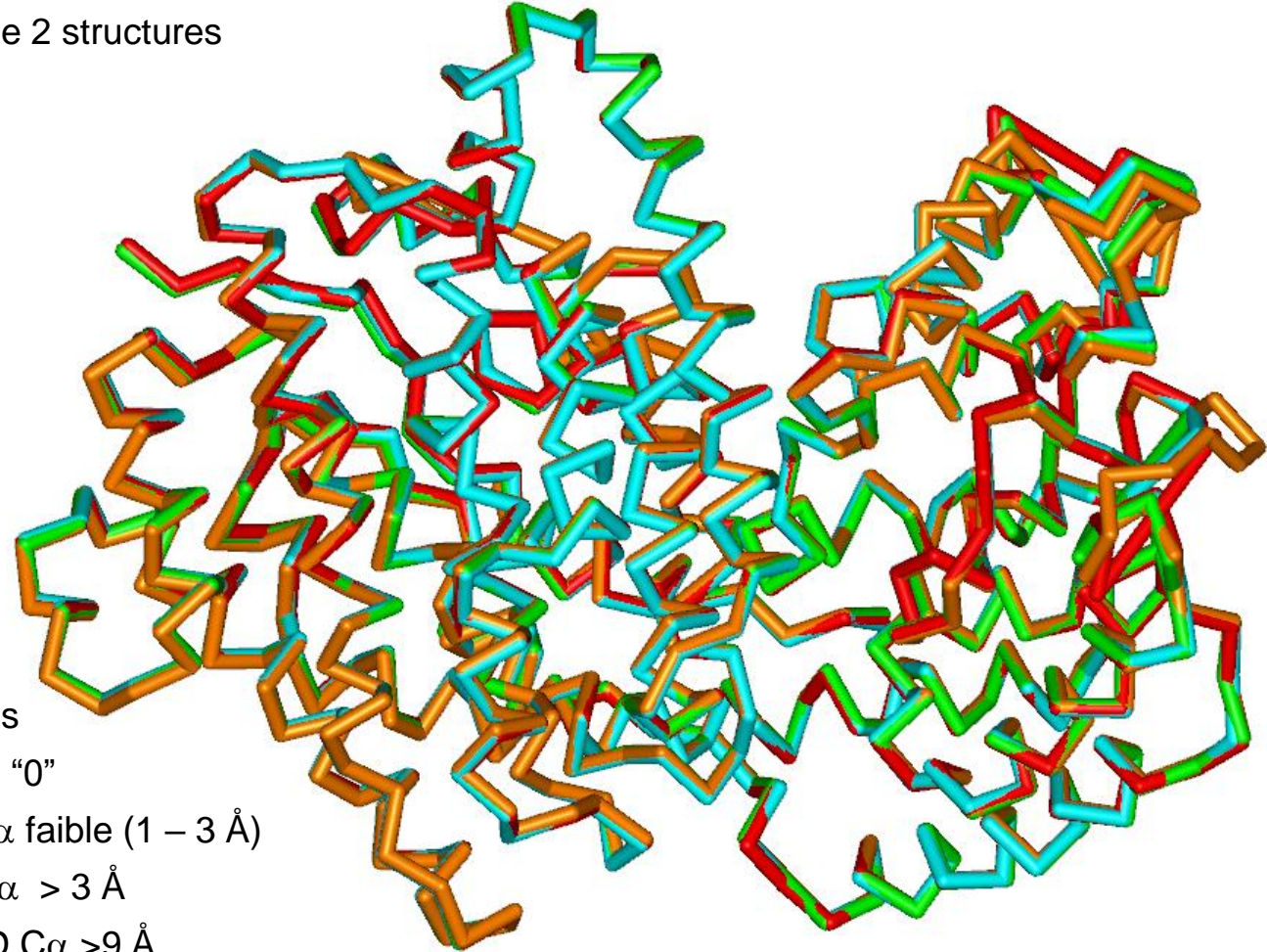
You will receive a schedule confirmation email shortly.

[University at Albany Home Page](#)[Contact UAlbany](#) | [Directories](#) | [Calendars](#) | [Visitors](#) | [Site I](#)
[Admissions](#) | [Academics](#) | [Research](#) | [IT Services](#) | [Librarie](#)

<http://pathway.rit.albany.edu/~cemc/>



- ✓ Traitement identique de tous les atomes (or en surface ils sont plus flexibles qu'au cœur)
- ✓ L'alignement structural le plus pertinent n'a pas forcément le RMSD le plus faible
- ✓ La signification du RMSD dépend de la taille des protéines
- ✓ La valeur du RMSD dépend du choix des atomes (Tous, $C\alpha$, structures secondaires)
- ✓ Superposition simultanée que de 2 structures
- ✓ Pas de flexibilité dans le calcul



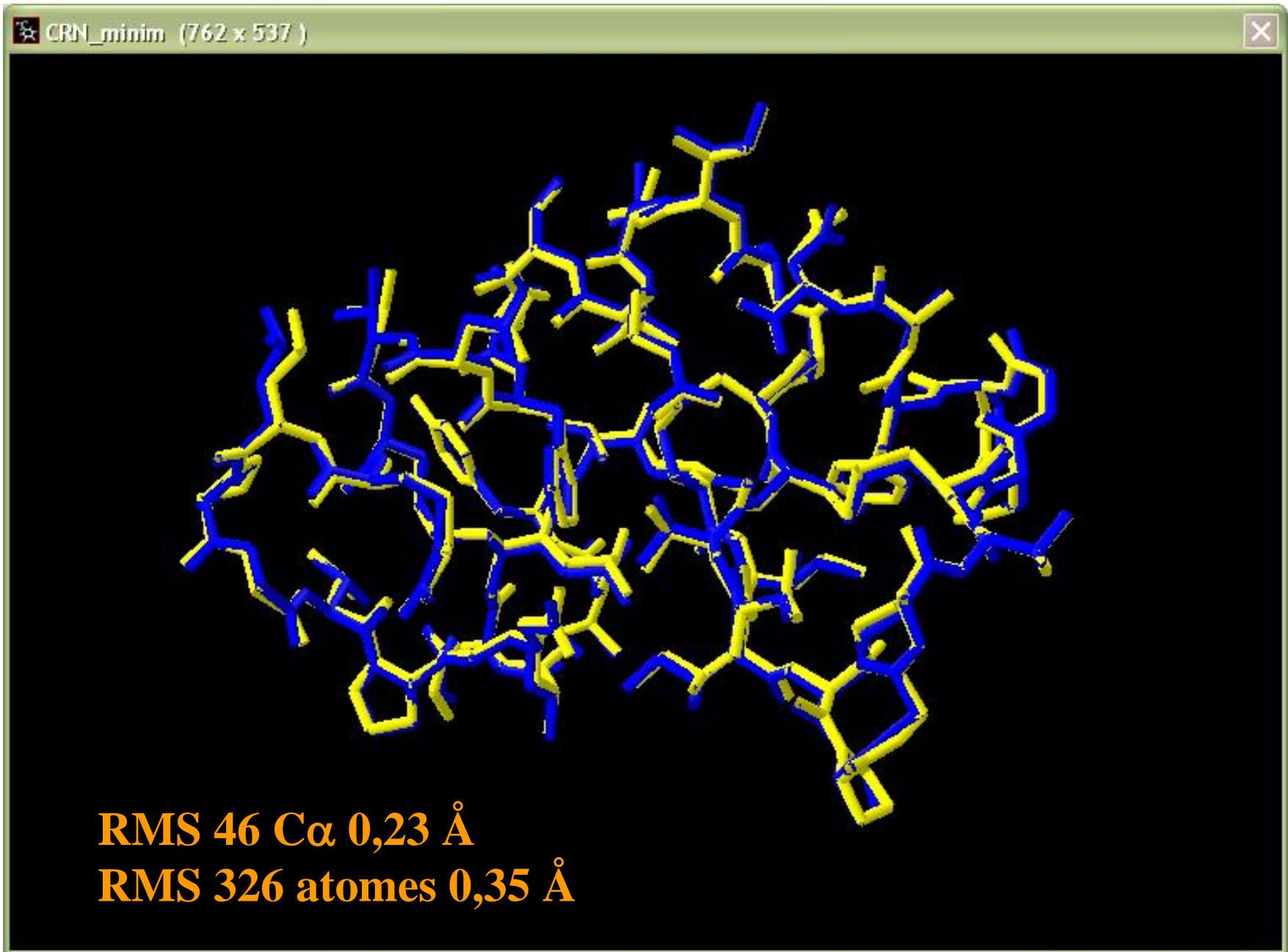
- ✓ Unité du RMSD => e.g. Ångstroms
- ✓ Structures identiques => RMSD = "0"
- ✓ Structures similaires => RMSD $C\alpha$ faible (1 – 3 Å)
- ✓ Structures distantes => RMSD $C\alpha$ > 3 Å
- ✓ Structures quelconques => RMSD $C\alpha$ >9 Å

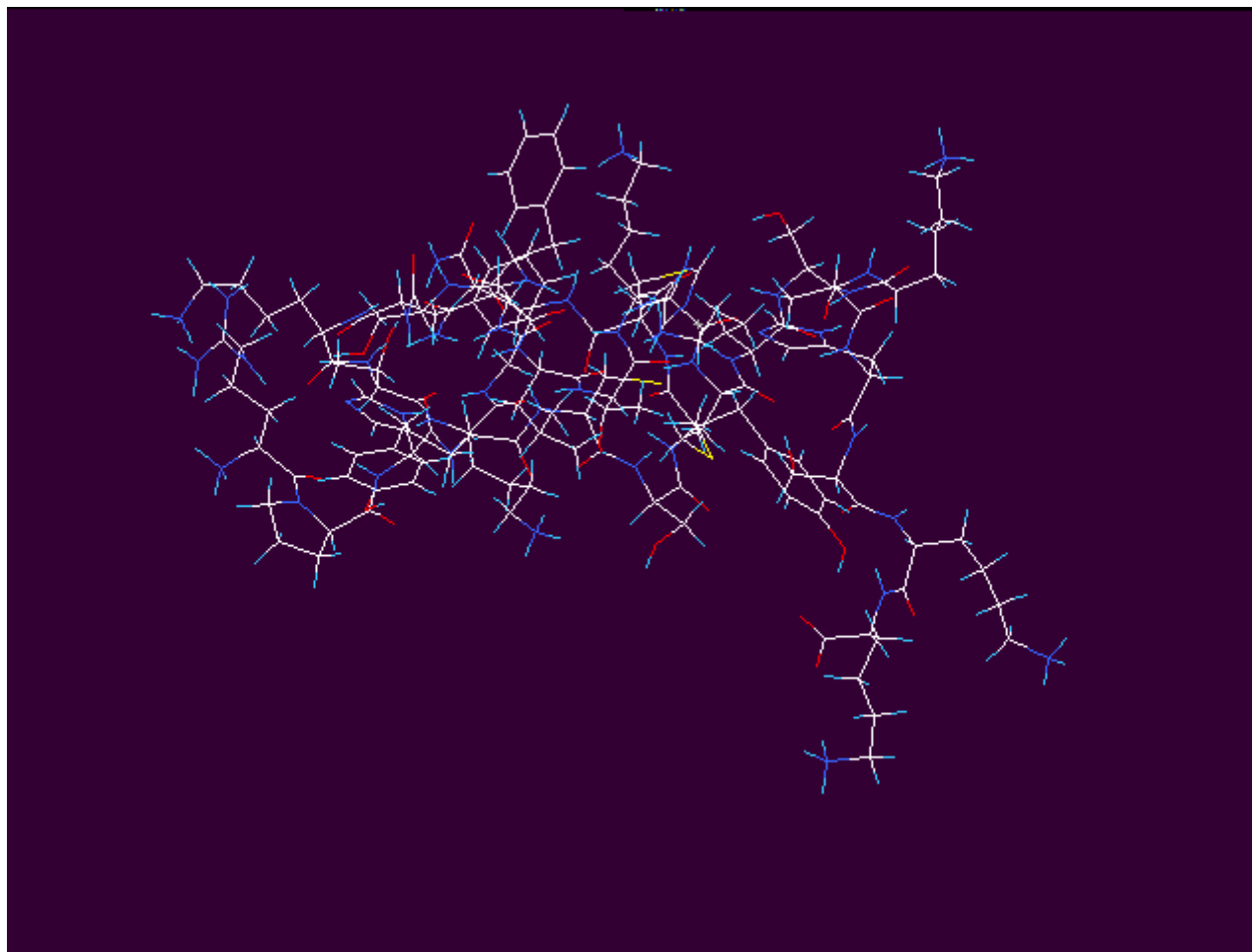


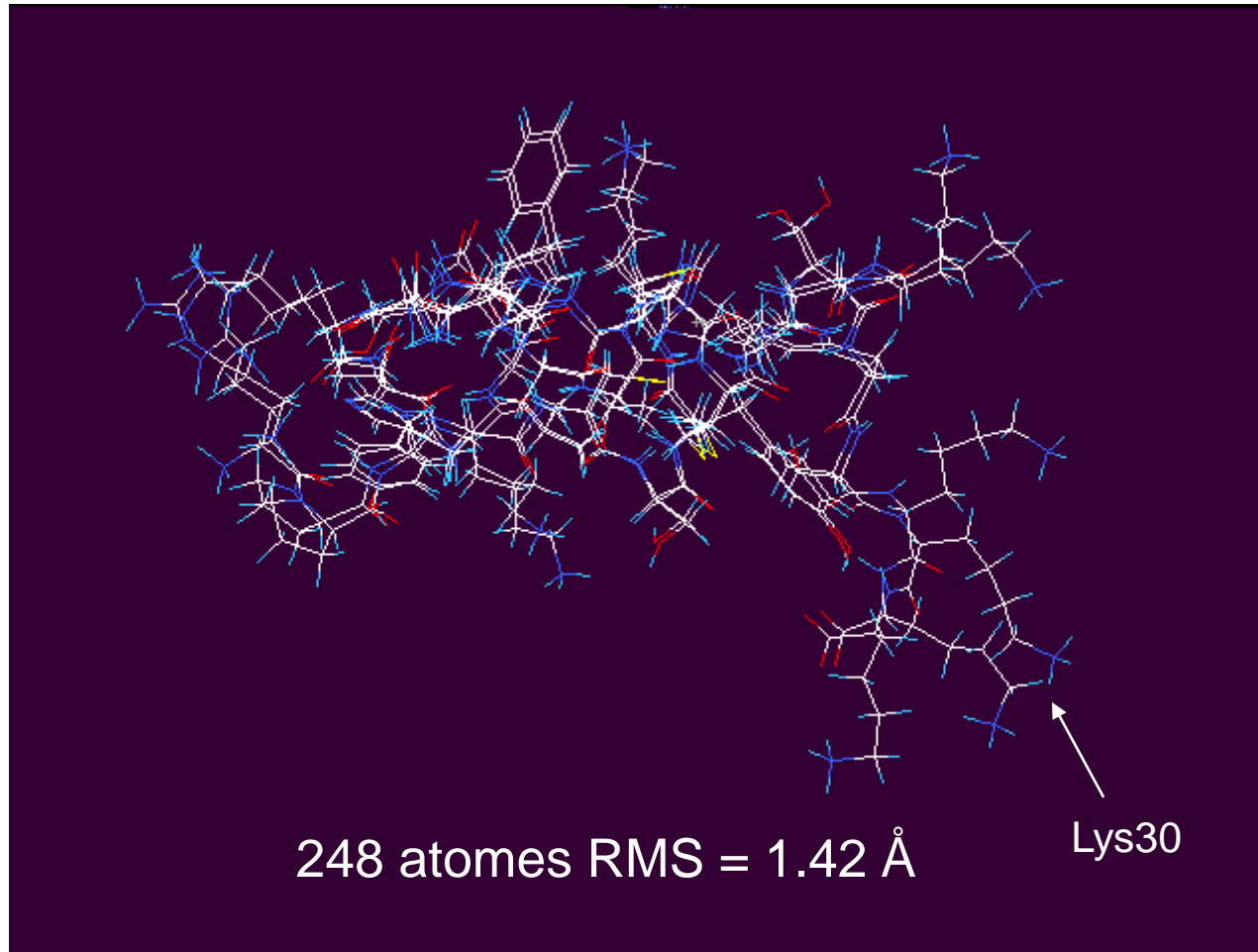


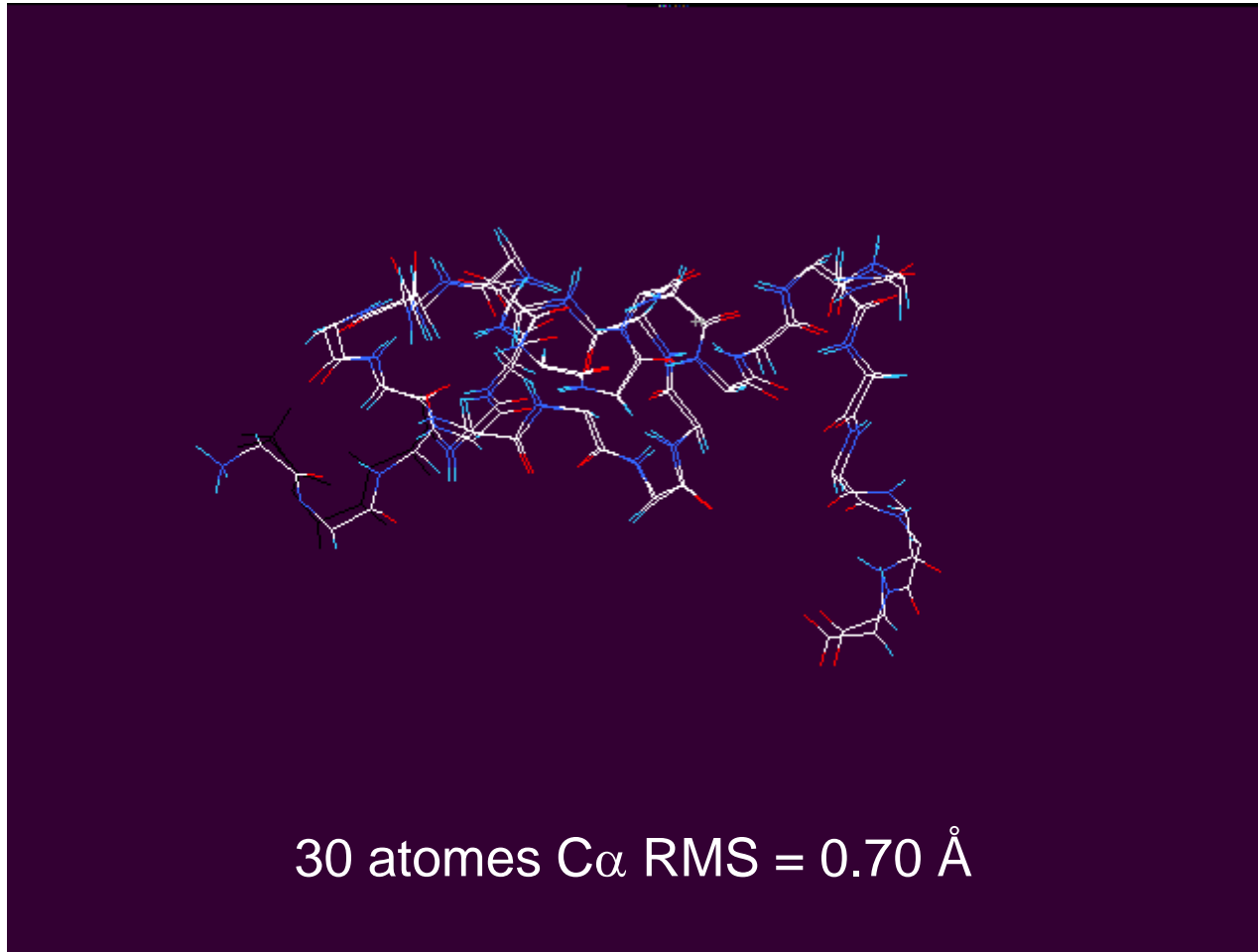












<http://www.cgl.ucsf.edu/home/meng/grpmt/structalign.html>

Nom	Adresses Internet
MultiProt	http://bioinfo3d.cs.tau.ac.il/MultiProt/
SuperPose	http://wishart.biology.ualberta.ca/SuperPose/
SALIGN	http://salilab.org/DBAli/?page=tools&action=f_salign
K2, K2SA	http://zlab.bu.edu/k2sa/
CE	http://cl.sdsc.edu/
Mammoth	http://ub.cbm.uam.es/software/mammoth.php
FATCAT	http://fatcat.burnham.org/fatcat-cgi/cgi/fatcat.pl?-func=pairwise
POSA	http://fatcat.burnham.org/POSA
GANGSTA+	http://agknapp.chemie.fu-berlin.de/gplus/index.php
TopMatch	http://topmatch.services.came.sbg.ac.at/
ASH	https://sysimm.org/ash_service/
C-alpha match	http://bioinfo3d.cs.tau.ac.il/c_alpha_match/
SSM	http://www.ebi.ac.uk/msd-srv/ssm/
ALADYN	http://aladyn.escience-lab.org/
NPS@	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_fit.html

- ✓ Les protéines ont des similarités structurales avec d'autres et parfois une origine évolutive commune (homologie).
- ✓ Créée par inspection elle des structures et à l'aide de méthodes automatiques.
- ✓ Description détaillée et complète des rapports structuraux et évolutifs entre toutes les protéines dont la structure est connue.

Scop Classification Statistics

SCOP: Structural Classification of Proteins. **1.75** release
 38221 PDB Entries (23 Feb 2009). 110800 Domains. 1 Literature Reference
 (excluding nucleic acids and theoretical models)

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902



Cyrus Chothia

Structural Classification of Proteins



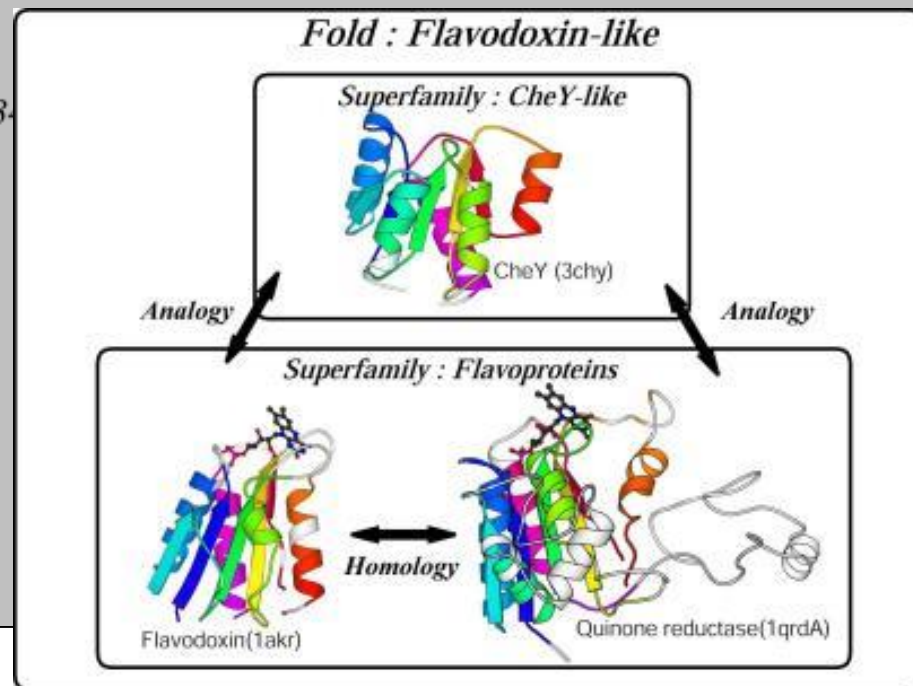
Protein: Triosephosphate isomerase from Human (*Homo sapiens*)

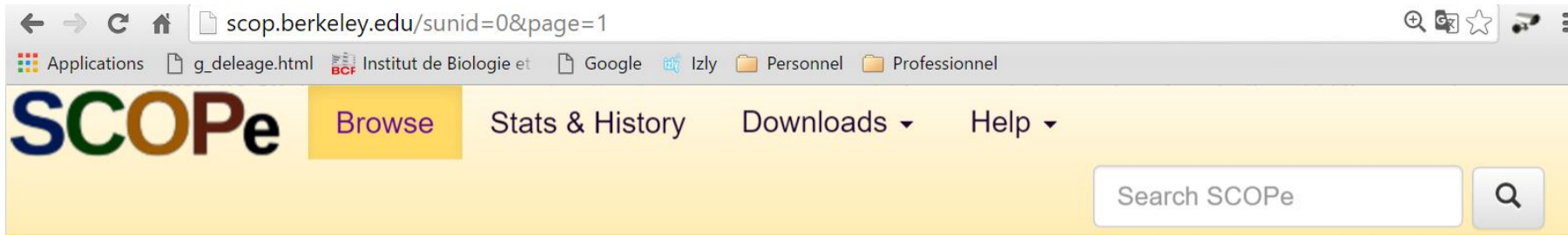
Lineage:

1. Root: [scop](#)
2. Class: [Alpha and beta proteins \(a/b\)](#)
Mainly parallel beta sheets (beta-alpha-beta units)
3. Fold: [TIM beta/alpha-barrel](#)
contains parallel beta-sheet barrel, closed; n=8, S=8; strand order 123- the first seven superfamilies have similar phosphate-binding sites
4. Superfamily: [Triosephosphate isomerase \(TIM\)](#)
5. Family: [Triosephosphate isomerase \(TIM\)](#)
6. Protein: Triosephosphate isomerase
7. Species: [Human \(*Homo sapiens*\)](#)

PDB Entry Domains:

1. [1hti](#)
complexed with pga
 1. [chain a](#)
 2. [chain b](#)

















scop.berkeley.edu/sunid=0&page=1

Applications g_deleage.html Institut de Biologie et Google Izly Personnel Professionnel

SCOPe Browse Stats & History Downloads ▾ Help ▾

Search SCOPe

Classes in SCOPe 2.06:

1.  a: All alpha proteins [46456] (289 folds)
2.  b: All beta proteins [48724] (177 folds)
3.  c: Alpha and beta proteins (a/b) [51349] (148 folds)
4.  d: Alpha and beta proteins (a+b) [53931] (385 folds)
5.  e: Multi-domain proteins (alpha and beta) [56572] (69 folds)
6.  f: Membrane and cell surface proteins and peptides [56835] (59 folds)
7.  g: Small proteins [56992] (94 folds)
8.  h: Coiled coil proteins [57942] (7 folds)
9.  i: Low resolution protein structures [58117] (25 folds)
10.  j: Peptides [58231] (133 folds)
11.  k: Designed proteins [58788] (44 folds)
12.  l: Artifacts [310555] (1 fold)


SCOPe: Structural Classification of Proteins — extended. Release 2.06 (updated 2016-08-25, stable release February 2016)

Copyright © 1994-2016 The **SCOP** and **SCOPe** authors








scope@compbio.berkeley.edu

Lineage for **Class a: All alpha proteins**



1. Root: [SCOPE 2.06](#)

2.  [Class a: All alpha proteins](#) [46456] (289 folds)



Folds:

1.  [a.1: Globin-like](#) [46457] (2 superfamilies)
core: 6 helices; folded leaf, partly opened
2.  [a.2: Long alpha-hairpin](#) [46556] (20 superfamilies)
2 helices; antiparallel hairpin, left-handed twist
3.  [a.3: Cytochrome c](#) [46625] (1 superfamily)
core: 3 helices; folded leaf, opened
4.  [a.4: DNA/RNA-binding 3-helical bundle](#) [46688] (14 superfamilies)
core: 3-helices; bundle, closed or partly opened, right-handed twist; up-and down
5.  [a.5: RuvA C-terminal domain-like](#) [46928] (9 superfamilies)
3 helices; bundle, right-handed twist
6.  [a.6: Putative DNA-binding domain](#) [46954] (1 superfamily)
core: 3 helices; architecture is similar to that of the "winged helix" fold but topology is different
7.  [a.7: Spectrin repeat-like](#) [46965] (16 superfamilies)
3 helices; bundle, closed, left-handed twist; up-and-down
8.  [a.8: immunoglobulin/albumin-binding domain-like](#) [46996] (11 superfamilies)
3 helices; bundle, closed, left-handed twist; up-and-down; mirror topology to the spectrin-like fold
9.  [a.9: Peripheral subunit-binding domain of 2-oxo acid dehydrogenase complex](#) [47004] (1 superfamily)

Lineage for Fold a.1: Globin-like

1. Root: SCOPe 2.06
2.  Class a: All alpha proteins [46456] (289 folds)
3.  Fold a.1: Globin-like [46457] (2 superfamilies)
core: 6 helices; folded leaf, partly opened

Superfamilies:

1.  a.1.1: Globin-like [46458] (5 families) *S*
2.  a.1.2: alpha-helical ferredoxin [46548] (3 families) *S*
contains two Fe4-S4 clusters

More info for Fold a.1: Globin-like

Timeline for Fold a.1: Globin-like:

- Fold a.1: Globin-like first appeared (with stable ids) in SCOP 1.55
- Fold a.1: Globin-like appears in SCOPe 2.05

SCOPe: Structural Classification of Proteins — extended. Release 2.06 (updated 2016-08-25, stable release February 2016)

Copyright © 1994-2016 The **SCOP** and **SCOPe** authors

scope@compbio.berkeley.edu

SCOPE




[Browse](#)

[Stats & History](#)






[Downloads](#) ▾

[Help](#) ▾

Lineage for Superfamily a.1.1: Globin-like





1. Root: [SCOPE 2.06](#)
2.  [Class a: All alpha proteins](#) [46456] (289 folds)
3.  [Fold a.1: Globin-like](#) [46457] (2 superfamilies)
core: 6 helices; folded leaf, partly opened
4.  [Superfamily a.1.1: Globin-like](#) [46458] (5 families) **S**

Families:











1.  [a.1.1.1: Truncated hemoglobin](#) [46459] (2 protein domains)
lack the first helix (A)
2.  [a.1.1.2: Globins](#) [46463] (27 protein domains)
Heme-binding protein
3.  [a.1.1.3: Phycocyanin-like phycobilisome proteins](#) [46532] (7 protein domains)
*oligomers of two different types of globin-like subunits containing two extra helices at the N-terminus
binds a bilin chromophore
automatically mapped to Pfam [PF00502](#)*
4.  [a.1.1.4: Nerve tissue mini-hemoglobin \(neural globin\)](#) [74660] (2 protein domains)
*lack the first helix but otherwise is more similar to conventional globins than the truncated ones
automatically mapped to Pfam [PF00042](#)*
5.  [a.1.1.0: automated matches](#) [191420] (1 protein)
not a true family






More info for [Superfamily a.1.1: Globin-like](#)

Lineage for **Family a.1.1.2: Globins**

1. Root: [SCOPe 2.06](#)
2.  Class [a: All alpha proteins](#) [46456] (289 folds)
3.  Fold [a.1: Globin-like](#) [46457] (2 superfamilies)
core: 6 helices; folded leaf, partly opened
4.  Superfamily [a.1.1: Globin-like](#) [46458] (5 families) *S*
5.  Family [a.1.1.2: Globins](#) [46463] (27 protein domains)
Heme-binding protein

Protein Domains:

1.  [Ascaris hemoglobin, domain 1](#) [46520] (1 species)
 Species [Pig roundworm \(Ascaris suum\) \[TaxId:6253\]](#) [46521] (1 PDB entry)
2.  [Bacterial dimeric hemoglobin](#) [46526] (1 species)
 Species [Vitreoscilla stercoraria \[TaxId:61\]](#) [46527] (7 PDB entries)
3.  [Chimeric hemoglobin beta-alpha](#) [46516] (1 species)
 Species [Synthetic, based on Homo sapiens sequence](#) [46517] (1 PDB entry)
4.  [Cytoglobin](#) [109626] (1 species)
 Species [Human \(Homo sapiens\) \[TaxId:9606\]](#) [109627] (8 PDB entries)
 Uniprot [Q8WWM9](#) 18-171
5.  [Dehaloperoxidase](#) [46530] (1 species)
 Species [Amphitrite ornata \[TaxId:129555\]](#) [46531] (31 PDB entries)

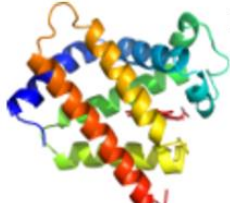
1. Root: SCOPe 2.06
2.  Class a: All alpha proteins [46456] (289 folds)
3.  Fold a.1: Globin-like [46457] (2 superfamilies)
core: 6 helices; folded leaf, partly opened
4.  Superfamily a.1.1: Globin-like [46458] (5 families) *S*
5.  Family a.1.1.2: Globins [46463] (27 protein domains)
Heme-binding protein
6.  Protein Cytoglobin [109626] (1 species)

Species:

 Human (Homo sapiens) [TaxId:9606] [109627] (8 PDB entries)
Uniprot Q8WWM9 18-171

1. Domains for 1umo:

1.  Domain d1umoa_ : 1umo A: [107959]
complexed with hem

2.  Domain d1umob_ : 1umo B: [107960]
complexed with hem

PDB entry 1umo

[View 1umo on RCSB PDB site](#)

Description: the crystal structure of cytoglobin: the fourth globin type discovered in man

Class: oxygen storage/transport

Keywords: oxygen storage/transport, cytoglobin, oxygen storage, histogloblin, hgb, transport, stap

Deposited on **2003-08-26**, released **2004-09-02**

The last revision prior to the SCOPe 2.06 freeze date was dated **2009-02-24**, with a file datestamp of **2009-03-01**.

Experiment type: XRAY

Resolution: 2.59 Å

R-factor: 0.21161

AEROSPACI score: 0.27 ([click here for full SPACI score report](#))

Chains and heterogens:

- Chain 'A':

Compound: cytoglobin

Species: *Homo sapiens* [[TaxId:9606](#)]

Database cross-references and differences (RAF-indexed):

- Uniprot [Q8WWM9](#)
 - engineered mutation (37)
 - engineered mutation (82)

Domains in SCOPe 2.06: [d1umoa_](#)

- Chain 'B':

Compound: cytoglobin

Species: *Homo sapiens* [[TaxId:9606](#)]

Database cross-references and differences (RAF-indexed):

- Uniprot [Q8WWM9](#)
 - engineered mutation (37)
 - engineered mutation (82)

Domains in SCOPe 2.06: [d1umob_](#)

- **Heterogens:** HEM, HOH

CATH

PROTEIN STRUCTURE CLASSIFICATION

[Home](#) | [Search](#) | [Documentation](#) | [Tools](#) | [Download](#)

[Home](#)

Welcome to CATH

CATH is a manually curated classification of protein domain structures. Each protein has been chopped into structural domains and assigned into homologous superfamilies (groups of domains that are related by evolution). This classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.

[Find out more about CATH >>](#)

New in CATH v3.4

CATH v3.4 is built from 104,238 PDB chains. We have added the following data since v3.3:

- 49 folds (total 1,282)
- 163 superfamilies (total 2,549)
- 1,311 sequence families (total 11,330)
- 24,232 domains (total 152,920)

[Download CATH data >>](#)


Janet Thornton



Christine Orengo

Using CATH

[Search](#)
[Browse](#)
[Download](#)
[Tutorials](#)
[Introduction to CATH](#)
[FAQ](#)

CATH Tools

[Find My Sequence](#)
[Find My Structure](#)
[Linking to CATH](#)

About CATH

[Release Statistics](#)
[Glossary](#)
[CATH Team](#)
[References](#)

CATH

<http://www.cathdb.info/>

✓ Classification hiérarchique des structures des domaines protéiques qui regroupe les protéines à quatre niveaux principaux, **Classe (C)**, **Architecture (A)**, **Topologie (T)** et **Superfamilles (h)** homologues.

- ✓ La **classe**, dérivée du contenu de structure secondaire, est assignée pour plus de 90% de structures de protéine automatiquement.
- ✓ L'**architecture** décrit l'orientation des structures secondaires indépendamment des connectivités, et est actuellement assignée element.
- ✓ Le niveau **Topologie** groupe les structures selon leurs raccordements topologiques et le nombre de structures secondaires.
- ✓ Les **superfamilles** homologues groupent des protéines avec des structures et des fonctions fortement semblables.

NB Les attributions des structures aux niveaux Topologie et Superfamilles impliquent des comparaisons de séquences et de structures.

Class

This is the highest level of classification and is derived from the gross secondary structure content. There are 4 classes:

1. Mainly Alpha
2. Mainly Beta
3. Alpha Beta
4. Few Secondary Structures.

The classification of protein domains into these 4 classes has been achieved for 90% of structures automatically,

- Percentage composition of secondary structures
- Percentage of secondary structures contacts
- Secondary structure alternation
- Percentage of parallel strands

This is discussed in detail in the paper referenced below.

References

[Analysis of domain structural class using an automated class assignment protocol.](#)

Michie AD, Orengo CA, Thornton JM

J Mol Biol **262** p168-85 (1996 Sep 20) 

Homologous Superfamily

This level groups together protein domains which are thought to share a common ancestor and can therefore be described as homologous. Similarities are identified either by high sequence identity or structure comparison using [SSAP](#). Structures are clustered into the same homologous superfamily if they satisfy two or more of the following criteria:

- Sequence identity $\geq 35\%$, overlap $\geq 60\%$ of larger structure equivalent to smaller.
- SSAP score ≥ 80.0 , sequence identity $\geq 20\%$, 60% of larger structure equivalent to smaller.
- SSAP score ≥ 70.0 , 60% of larger structure equivalent to smaller, and domains which have related functions, which is informed by the literature and Pfam protein family database.
- Significant similarity from HMM-sequence searches and HMM-HMM comparisons using SAM, HMMER (<http://hmmer.wustl.edu>) and PRC (<http://supfam.org/PRC>).

References

[The Pfam protein families database.](#)

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR

Nucleic Acids Res **32** pD138-41 (2004 Jan 1) [PubMed](#)

[Hidden Markov models for sequence analysis: extension and analysis of the basic method.](#)

Hughey R, Krogh A

Comput Appl Biosci **12** p95-107 (1996 Apr) [PubMed](#)

SSAP

Full name: Secondary Structure Alignment Program.

This secondary structure alignment program uses dynamic programming to align proteins by matching vectors between residues. For each residue in a protein a local structure environment is defined by a set of inter-atomic vectors; the method matches residues by comparing these structural environments. An important aspect of these environments is that because they are defined independantly for each residue, they are rotationally invariant which makes their comparison insensitive to the displacement of substructures. The method also allows other residue properties to be included in the comparison: including solvent accessible area and torsional angles (which correspond to degree of burial and secondary structure) improves the alignment of remote protein structures. An important advantage of the method is that it is completely automatic. This algorithm is used to cluster structurally similar proteins in the CATH system.

References

[A local alignment method for protein structure motifs.](#)

Orengo CA, Taylor WR

J Mol Biol 233 p488-97 (1993 Oct 5) [PubMed](#)

[Protein structure alignment.](#)

Taylor WR, Orengo CA

CATH Version 3.2

The v3.2.0 release of CATH has a number of improvements in the backend data and the frontend web pages.

Changes since v3.1:

- 20,330 newly assigned domains
- 87 new homologous superfamilies
- 26 new folds (topologies)

The table below summarises the number of clusters within each of the four classes in CATH.

Class	Architecture	Topology	Homologous Superfamily	S35 Family	S60 Family	S95 Family	S100 Family	Domains
1	5	310	682	2078	2689	3540	6685	23491
2	20	196	438	2062	2902	4468	7656	29992
3	14	512	956	4558	6473	8135	16346	58967
4	1	92	102	173	217	301	445	1765
Total	40	1110	2178	8871	12281	16444	31132	114215

CATH Tools

This page documents all the tools that can be used to query and analyse the CATH resource. ¹
near future.

- [!\[\]\(8c4dca64662d21542001ca0ed7eeb688_img.jpg\) BLAST](#) - search CATH using a FASTA sequence (or ID)
- [!\[\]\(3de35c640e7147a3fb61ee393128d2ae_img.jpg\) Cathedral Server](#) - search CATH using a PDB structure
- [!\[\]\(d1438aeefda19c86ae7477bf1fb30796_img.jpg\) SSAP Server](#) - perform a pairwise structure comparison

<http://www.cathdb.info/cgi-bin/search.pl>

[**https://xip.uclb.com/i/software/CATHEDRAL.html**](https://xip.uclb.com/i/software/CATHEDRAL.html)

<http://www.cathdb.info/cgi-bin/SsapServer.pl>

95 million protein domains classified into 6,119 superfamilies

Search by keywords, PDB code, GO term, etc


Search

Core classification files for the latest version of CATH-Plus (v4.2) are [now available to download](#). [Daily updates](#) of our very latest classifications [are also available](#).



3D Structure

Find out what 3D structure your protein adopts


 Find out more

 Go



Protein Evolution


Learn about a particular protein family and how it evolved

 Find out more



Protein Function

Investigate the function of your protein

 Find out more

 Go

CATH PDB code search

Search results for **1hti**

Found 2 related domains

PDBsum [1hti]

Domain	CATH code	Length	Image
1htiA0	3.20.20.90	248	
1htiB0	3.20.20.90	248	

Homologous Superfamily 3.20.20.90

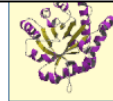
- Alpha Beta
- Barrel
- TIM Barrel
- Triose phosphate isomerase, FMN-dependent oxidoreductases and phosphate binding enzymes and Tryptophan biosynthesis enzymes

Displaying 1-20 of 24 sequence families

Click on the CATH code below to browse further

1 | 2 | Next

Code	Representative	Sequence family description	Image
<input type="checkbox"/> 3.20.20.90.1	1tpfA0	ISOMERASE (INTRAMOLECULAR OXIDOREDUCTASE)	
<input type="checkbox"/> 3.20.20.90.2	1hg3A0	ISOMERASE	
<input type="checkbox"/> 3.20.20.90.3	1gox00	OXIDOREDUCTASE (OXYGEN(A))	
<input type="checkbox"/> 3.20.20.90.4	1djnA1	OXIDOREDUCTASE	
<input type="checkbox"/> 3.20.20.90.5	1oya00	OXIDOREDUCTASE (FLAVOPROTEIN)	
<input type="checkbox"/> 3.20.20.90.6	1dorA1	OXIDOREDUCTASE	
<input type="checkbox"/> 3.20.20.90.7	1pii02	BIFUNCTIONAL (ISOMERASE AND SYNTHASE)	
<input type="checkbox"/> 3.20.20.90.8	1pii01	BIFUNCTIONAL (ISOMERASE AND SYNTHASE)	
<input type="checkbox"/> 3.20.20.90.9	1a5300	SYNTHASE	
<input type="checkbox"/> 3.20.20.90.10	1ubsA0	COMPLEX (LYASE/PEPTIDE)	
<input type="checkbox"/> 3.20.20.90.11	1nsj00	ISOMERASE	
<input type="checkbox"/> 3.20.20.90.12	1ak500	OXIDOREDUCTASE	
<input type="checkbox"/> 3.20.20.90.13	2tpsA0	THIAMIN BIOSYNTHESIS	
<input type="checkbox"/> 3.20.20.90.14	1rpxA0	3-EPIMERASE	
<input type="checkbox"/> 3.20.20.90.15	1dbtA0	LYASE	



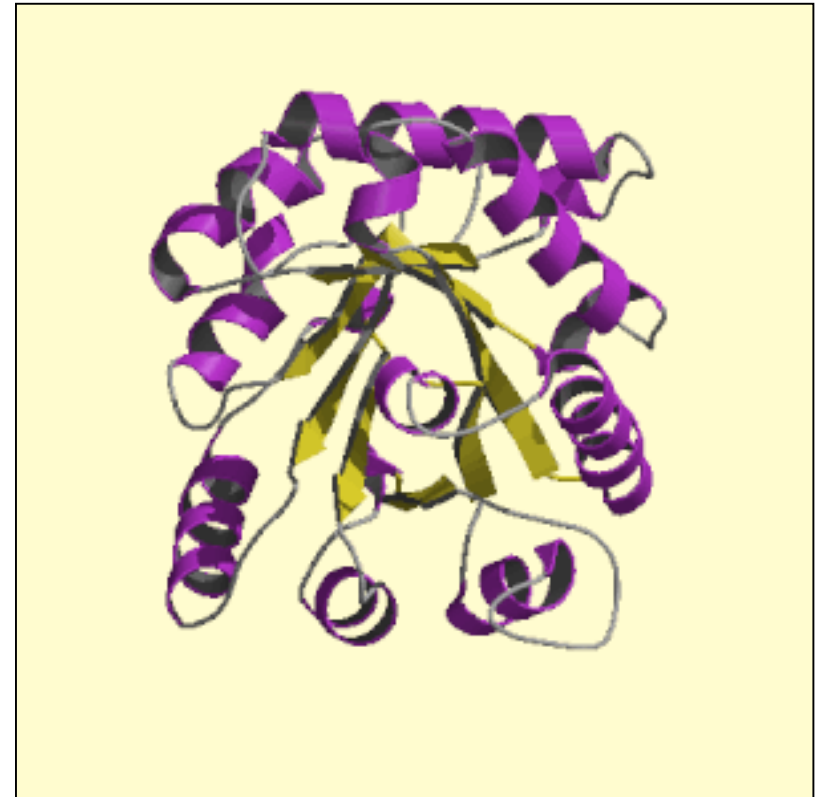
Homologous superfamily representative
1tpfA0

Go! PDB code
 CATH code
 General text

Goto...
SSAP
DHS
Gene3D
PDBsum

Navigation
Home
Top of heirarchy
Up one level
Next 20 entries

Help
Select a topic ▾



- FSSP : Classification basée sur une comparaison exhaustive des structures 3D à l'aide du programme DALI.
- HSSP : base (alignements, structure secondaire, profils, variabilité de séquences) des protéines homologues aux protéines de structure connues
- Mise-à-jour FSSP Sun Jun 16 17:22:08 BST 2002 (3242 sequence families representing 30624 protein structures)

FSSP: select structural neighbours of 1amk

Please cite: L. Holm and C. Sander (1996) Science 273(5275):595-60.

Select (check) structural neighbours to display

3D superimposition	Multiple alignment (wide)	Multiple alignment (narrow)	Multiple families (wide)
Multiple families (narrow)	Reset selection		

STRID2	Z	RMSD	LALI	LSEQ2	%IDE	PROTEIN
<input type="checkbox"/> 1amk	48.6	0.0	250	250	100	triose phosphate isomerase biological_unit
<input type="checkbox"/> 1qdsA	46.0	0.3	250	250	99	triosephosphate isomerase Mutant
<input type="checkbox"/> 1if2A	45.9	0.2	249	249	99	triosephosphate isomerase (tim) Mutant
<input type="checkbox"/> 1iigB	44.4	0.4	249	249	70	triosephosphate isomerase
<input type="checkbox"/> 1cilA	44.2	0.5	248	248	68	triosephosphate isomerase (tim) biological_unit
<input type="checkbox"/> 1iihB	44.2	0.5	249	249	70	triosephosphate isomerase
<input type="checkbox"/> 1tcdA	44.1	0.5	248	248	68	triosephosphate isomerase (tim) biological_unit
<input type="checkbox"/> 1kv5A	44.0	0.5	249	249	69	triosephosphate isomerase, glycosomal (tim) Mutant
<input type="checkbox"/> 4timB	43.0	0.5	244	249	70	Triosephosphate isomerase complex with 2-Phosphoglycera
<input type="checkbox"/> 1ag1T	42.9	0.5	244	249	70	triosephosphate isomerase biological_unit
<input type="checkbox"/> 5timB	42.8	0.4	244	249	70	Triosephosphate isomerase complex with sulfate
<input type="checkbox"/> 6timB	42.5	0.5	243	249	70	Triosephosphate isomerase complex with glycerol-3-phosp
<input type="checkbox"/> 1trdB	42.2	0.5	243	249	70	Triosephosphate isomerase 1
<input type="checkbox"/> 1tcdB	41.9	1.0	248	249	68	triosephosphate isomerase (tim) biological_unit
<input type="checkbox"/> 1iihA	41.8	1.0	249	249	70	triosephosphate isomerase
<input type="checkbox"/> 1cilB	41.8	1.0	248	249	68	triosephosphate isomerase (tim) biological_unit
<input type="checkbox"/> 1iigA	41.7	1.0	249	249	70	triosephosphate isomerase
<input type="checkbox"/> 5timA	40.5	1.0	244	249	70	Triosephosphate isomerase complex with sulfate
<input type="checkbox"/> 4timA	40.5	1.0	244	249	70	Triosephosphate isomerase complex with 2-Phosphoglycera
<input type="checkbox"/> 1tpfB	40.5	1.1	244	249	70	Triosephosphate isomerase
<input type="checkbox"/> 1tpe	40.4	1.1	244	249	70	Triosephosphate isomerase
<input type="checkbox"/> 6timA	40.3	1.1	244	249	70	Triosephosphate isomerase complex with glycerol-3-phosp
<input type="checkbox"/> 1tpfA	40.3	1.1	243	250	70	Triosephosphate isomerase
<input type="checkbox"/> 1ag1O	40.3	1.0	244	249	70	triosephosphate isomerase biological_unit
<input type="checkbox"/> 1tpdB	40.2	1.0	243	249	70	Triosephosphate isomerase

FSSP: search for '1hti'

Please cite: L. Holm and C. Sander (1996) Science 273(5275):595-60.

Searching for 1hti in Protein Index

2 matches were found

PDBid	Repre	Rmsd	Lali	Lseq	%ide	Compound
1htiA	1amk	1.5	235	248	50	Triosephosphate isomerase
1htiB	1amk	1.0	244	248	51	Triosephosphate isomerase

Click on a hyperlink to view the structural alignments.

return to [FSSP home page](#) / [Dali Domain Dictionary](#)

(C) L. Holm, EMBL-EBI, Hinxton, May 1996




Liisa Holm

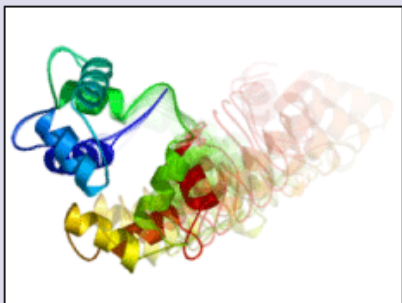
Eichier Édition Affichage Historique Marque-pages Outils ?

Database of Macromolecular Movements +

← → <http://molmovdb.mbb.yale.edu/molmovdb/> ☆

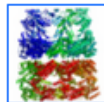


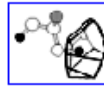
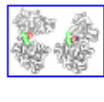
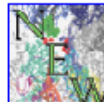
Les plus visités Débuter avec Firefox À la une

molmovdb.org 



Database of Macromolecular Movements with Associated Tools for Flexibility and Geometric Analysis

This describes the motions that occur in proteins and other macromolecules, particularly using movies. Associated with it are a variety of free software tools and servers for structural analysis. ([Citation info](#))

Services	More
 <p>Morph Server ↗</p>	 <p>All Resources</p>
 <p>Hinge Analysis ↗</p>	 <p>Geometry Programs ↗</p>
 <p>Motion Databases ↗</p>	 <p>New Stuff: Conformation Prediction ↗</p>

Search: Full-text

Copyright and Credits

Energétique et modélisation moléculaire

Méthodes quantiques

On distingue les noyaux et les électrons

Les interactions électrons-électrons et électrons-noyaux sont explicites.

Les interactions sont régies par les charges électroniques et nucléaires (énergie potentielle) et les mouvements électroniques.

Les interactions déterminent la distribution spatiale des électrons et des noyaux

Ainsi que leurs énergies.

Mécanique moléculaire

Noyaux et électrons sont modélisés en une seule particule

Particules sont sphériques (rayons atomiques expérimentaux ou théoriques) et possède une charge nette (théorique)

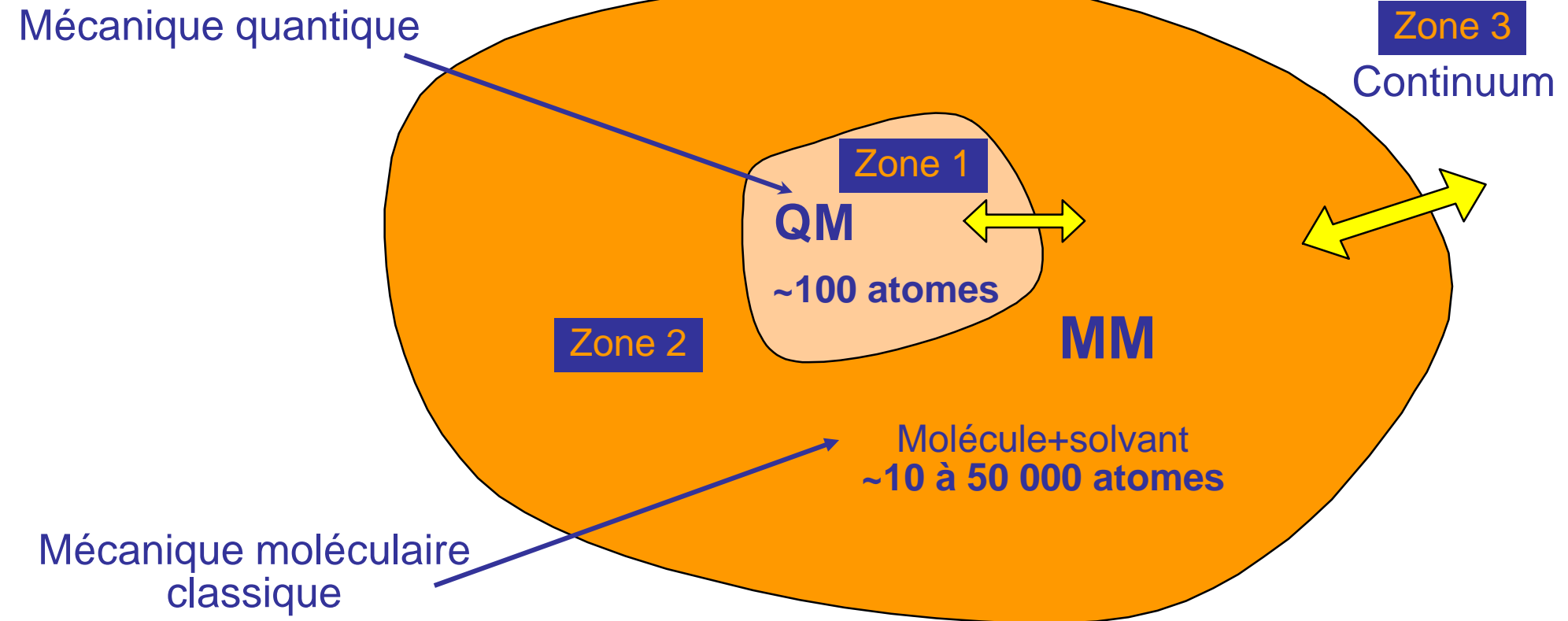
Les interactions sont modélisées par des ressorts et des potentiels classiques

Les interactions doivent être préassignées à un jeu spécifique de type d'atomes

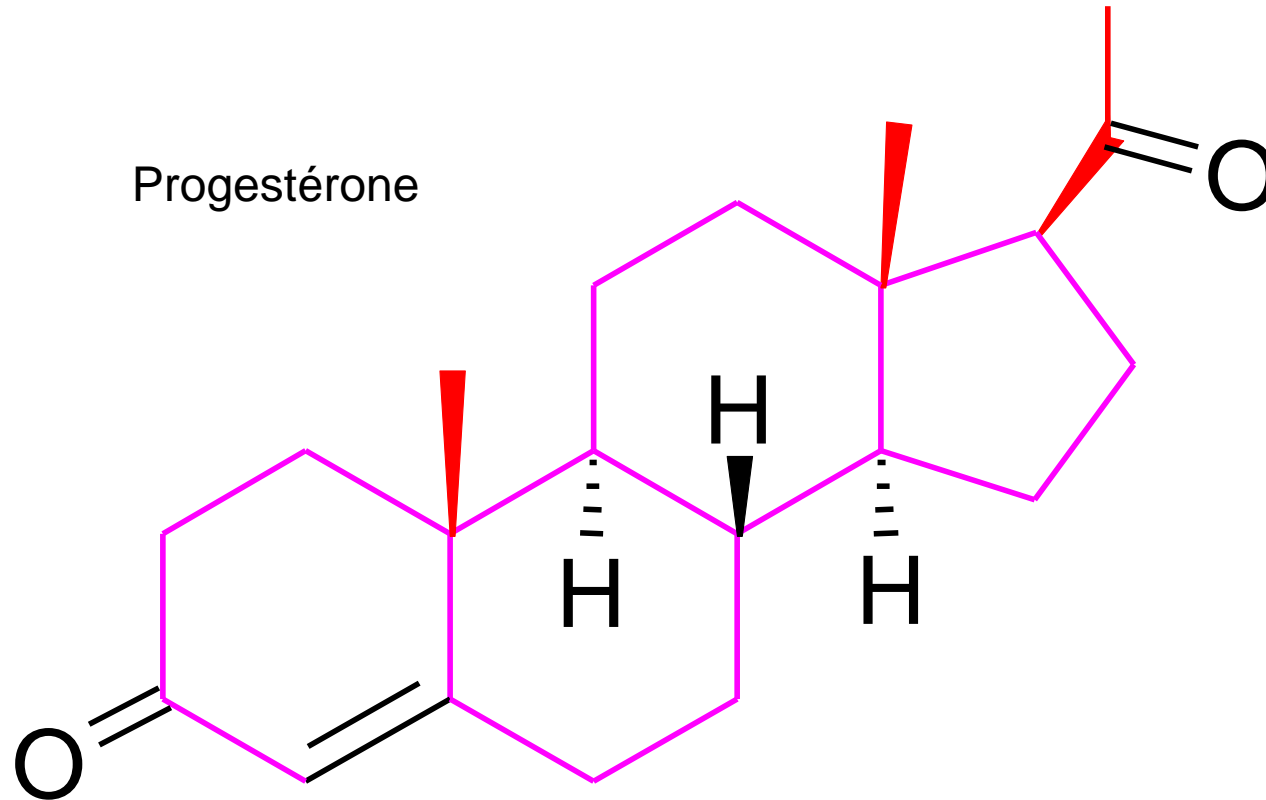
Les interactions déterminent la distribution spatiale des particules atomiques et leurs énergies

Méthodes hybrides

QMMM description en mécanique quantique autour du site actif et en mécanique moléculaire pour le reste de la molécule



Couplage de modèles : quantique, mécanique moléculaire, continuum



Une petite molécule est une molécule dont on peut énumérer toutes les conformations
≈ 500 atomes

Petite molécule associée à la notion de drogue/médicament

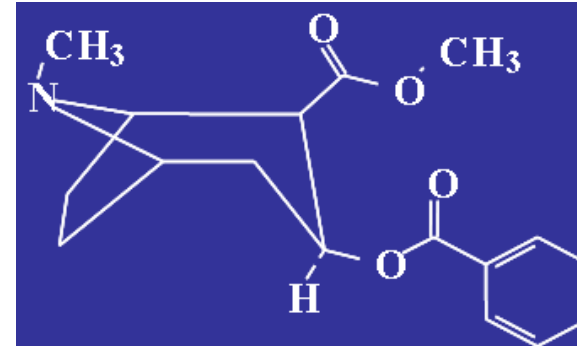
$\text{LogP} = \text{Log du coefficient de partition entre le 1-octanol et l'eau (logKO/W)}$

$\text{LogP élevé} \Rightarrow \text{hydrophobie élevée}$

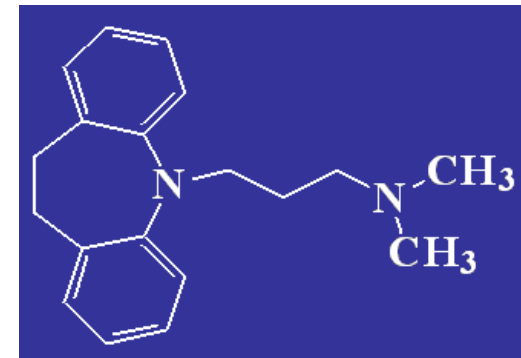
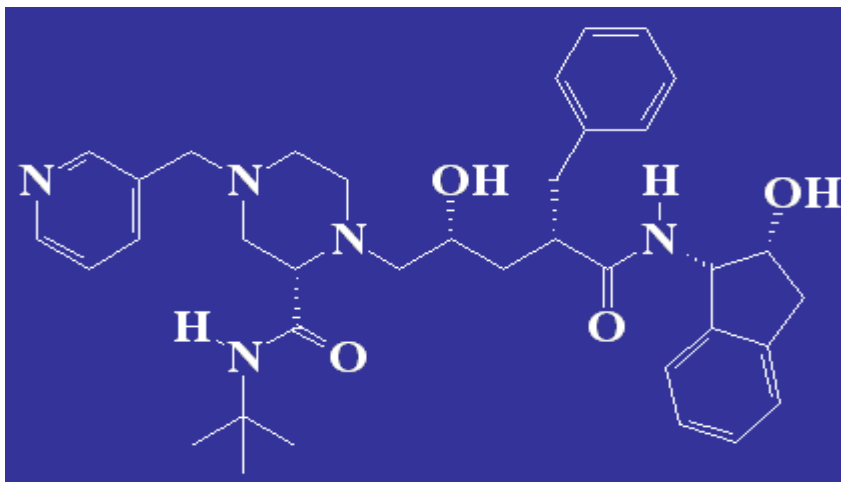
Morphine ($\text{logP} = 0,24$)



Cocaine ($\text{log P} = 2,72$)



Indinavir ($\text{log P} = 2,78$)



Imipramine ($\text{log P} = 4,49$)

Une « bonne » molécule en chimie thérapeutique :

- $MW \leq 500$ (opt = ~ 350)
- # accepteurs de liaison-H ≤ 10 (opt = ~ 5)
- # donneurs de liaison-H ≤ 5 (opt = ~ 2)
- $-2 < cLog P < 5$ (opt = ~ 3.0)
- # Flexibilité : rotules ≤ 5



Christopher Lipinski

Mécanique quantique

La mécanique quantique: l'énergie de la molécule est calculée à partir des interactions explicites entre les électrons et les noyaux.

méthodes *ab-initio*

méthodes semi-empiriques

Intérêt : précision des énergies calculées, modèle physique

Problèmes : temps de calcul prohibitif donc limité à 10-100 atomes

Gaussian

Mécanique moléculaire

La mécanique moléculaire est une méthode empirique permettant de reproduire raisonnablement des résultats expérimentaux à partir de modèles mathématiques simples des interactions. Ces modèles sont paramétrés pour les principaux types atomiques qui servent à décrire les molécules d'intérêt. L'ensemble fonctions mathématiques + paramètres est appelé champ de force*.

Intérêt : rapidité de calcul, taille non limitée

Problèmes : paramétrisation, précision.

- GROMOS (Berendsen & van Gunsteren, Groningen, Zurich) / Membranes & Protéines
- AMBER (UCSF, P. Kollman & Weiner) / ADN & Protéines
- JUMNA (IBPC, IBCP R. Lavery) ADN
- CHARMM (Harvard, Strasbourg, M. Karplus) Protéines/ADN/Membranes



Wilfred van Gunsteren



Martin Karplus

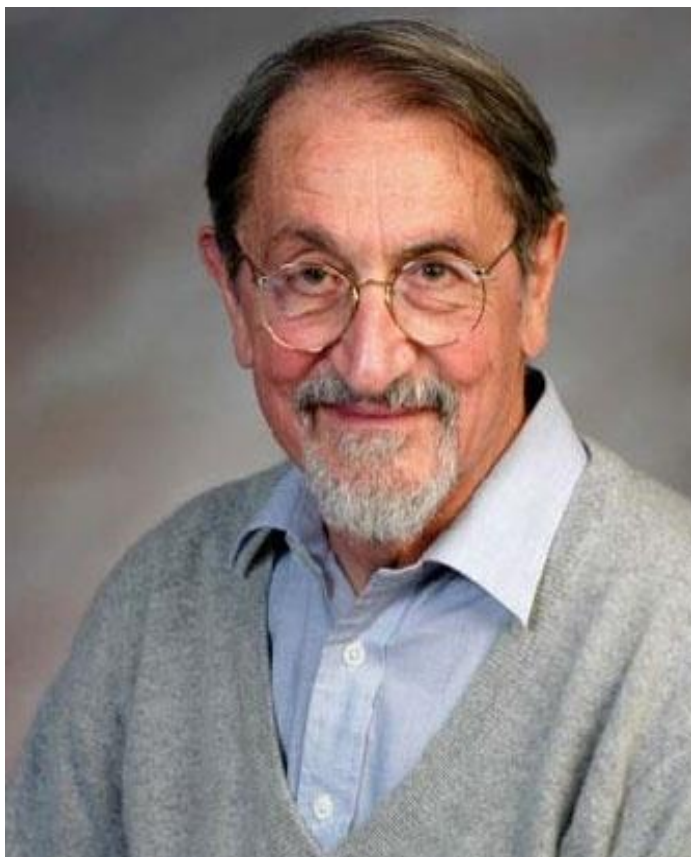


Peter Kollman
† 2001



Richard Lavery

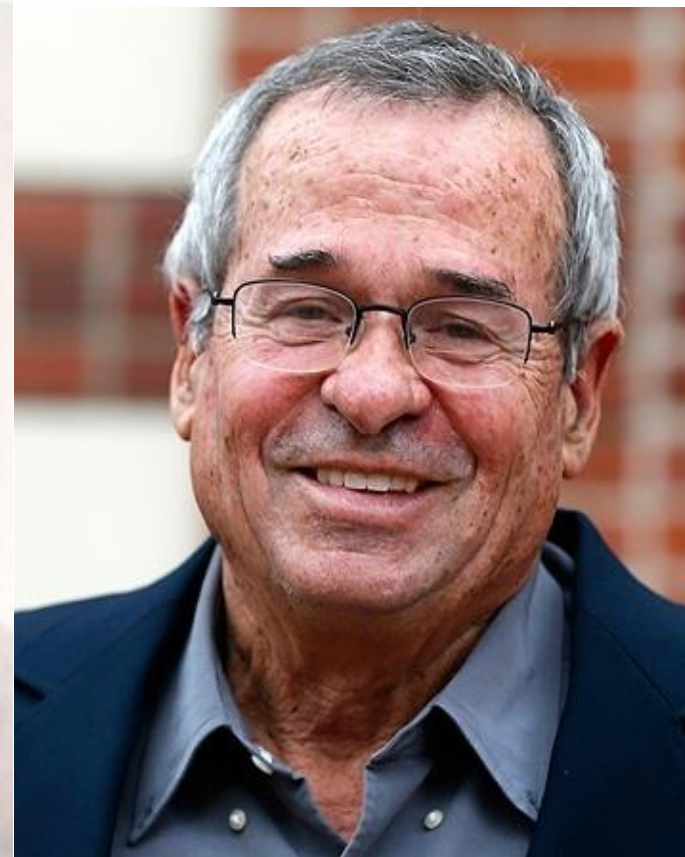
Prix NOBEL de chimie 2013 attribué à des spécialistes de la modélisation



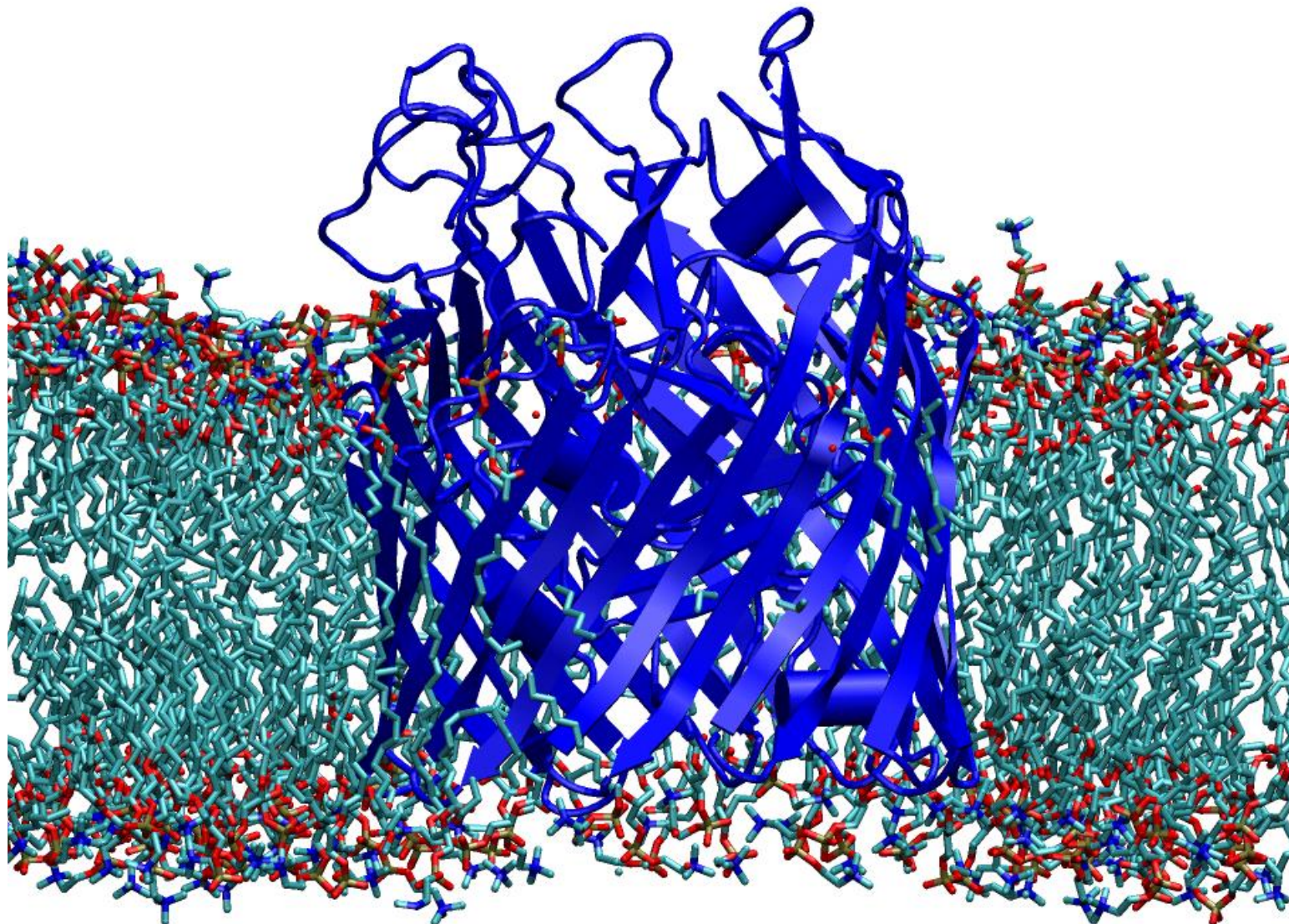
Martin Karplus



Michael Levitt

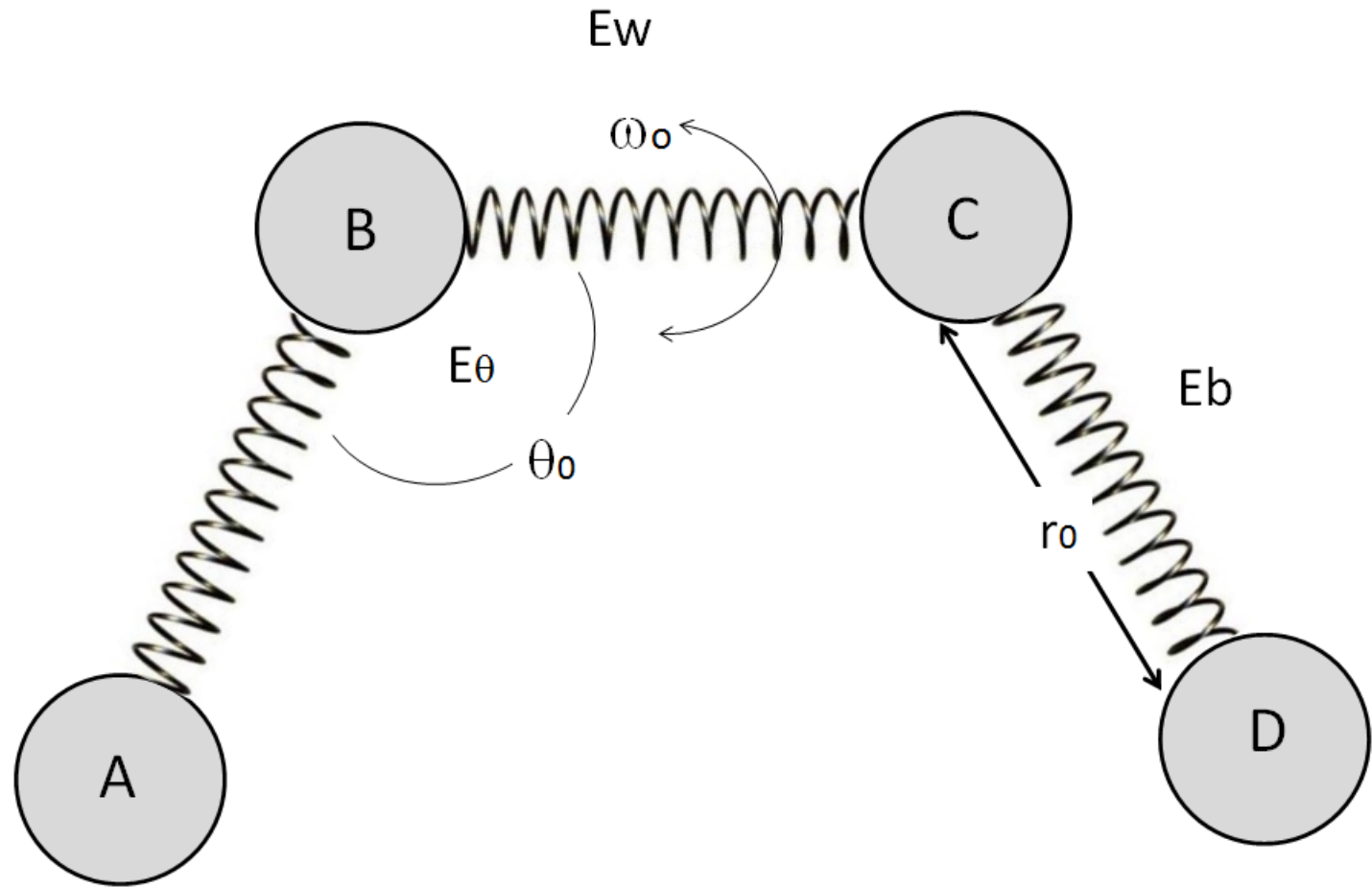


Arieh Warshel

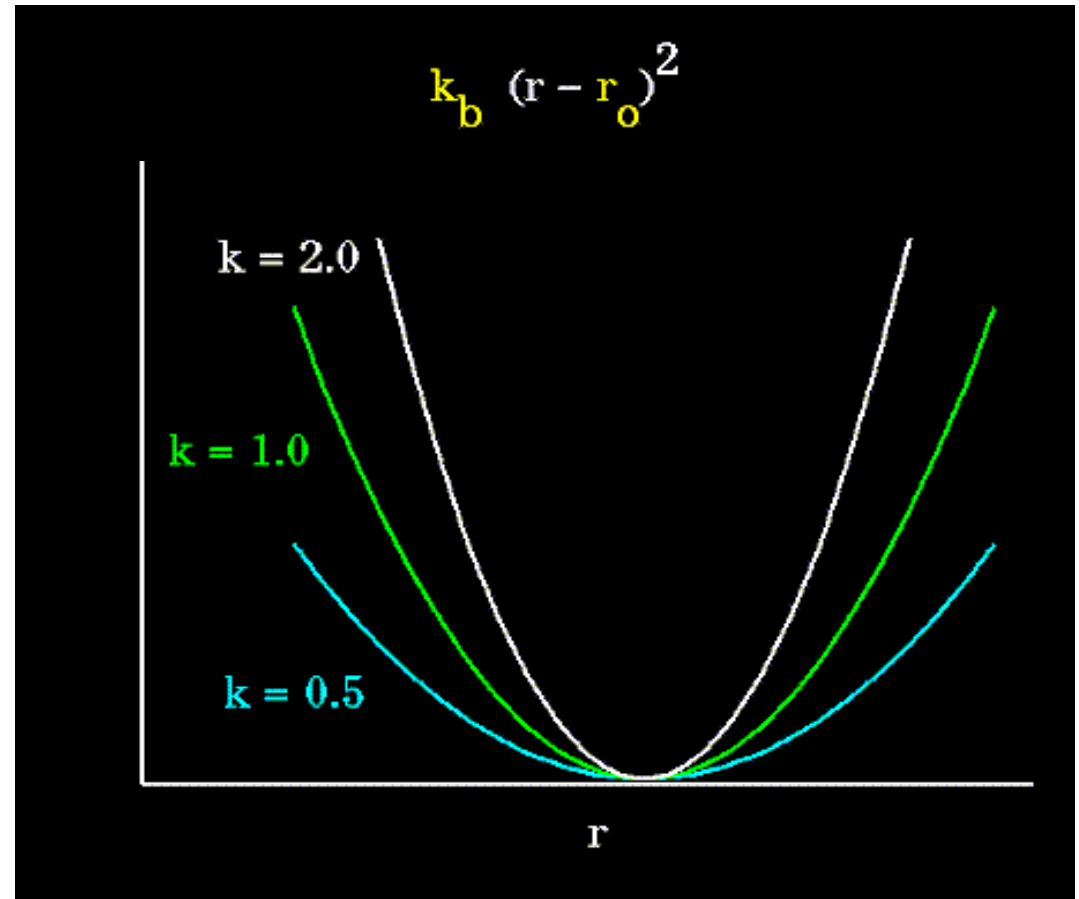
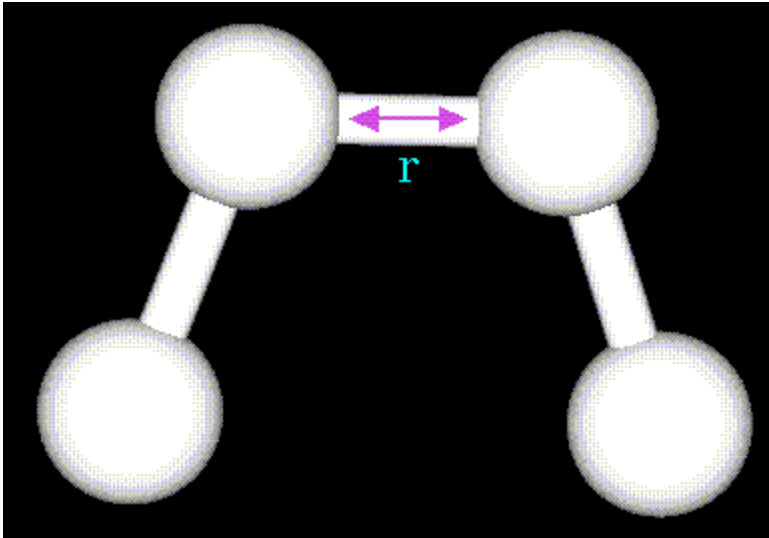


> 5000 atomes (jusqu'à 50k atomes)

Grosse molécule associée à la notion de cible



$$E = \sum_{\text{bonds}} k_b (r - r_0)^2$$

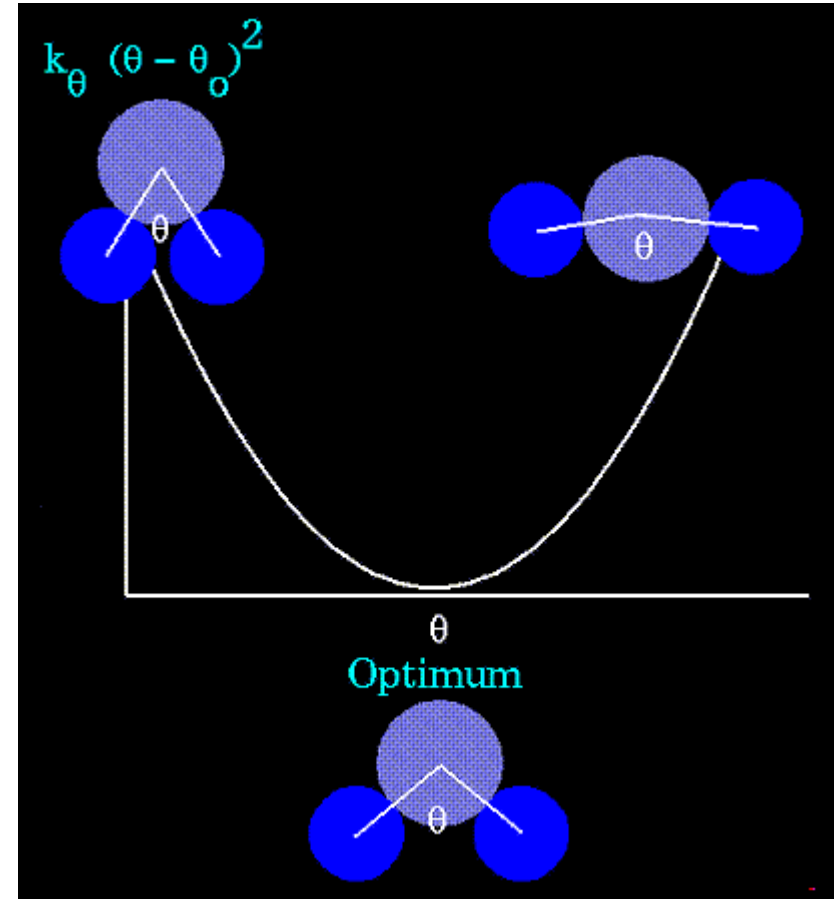
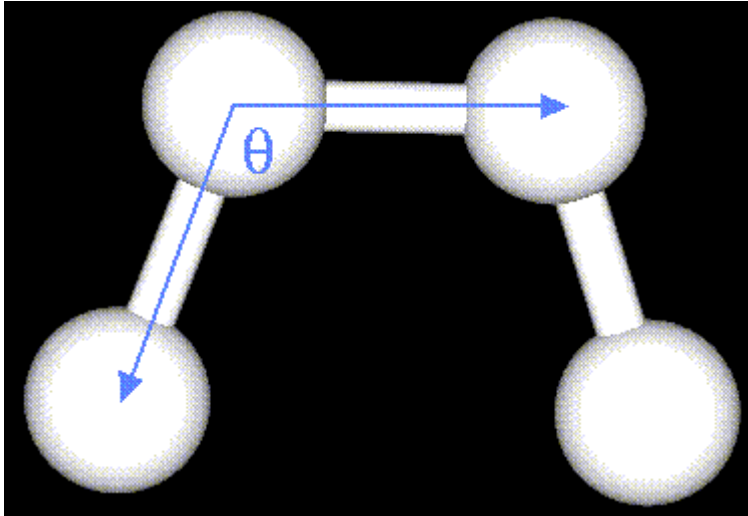


La constante k_b définit la « pente » de la parabole

C-C, $k_b = 310 \text{ kcal/mol.Å}^2$, $b_0 = 1.526 \text{ Å}$

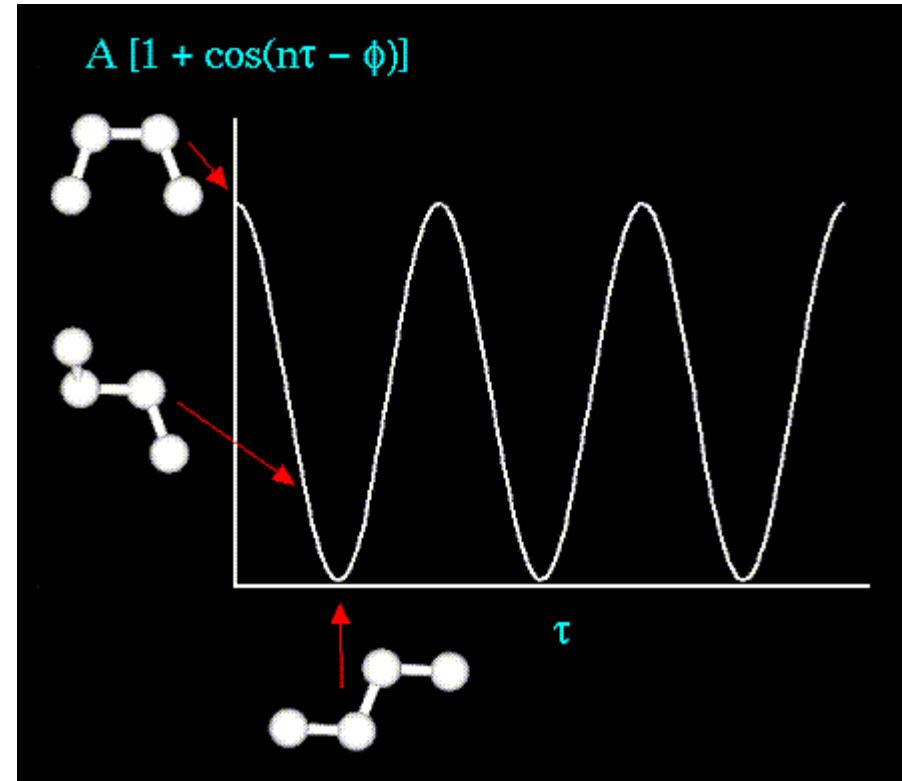
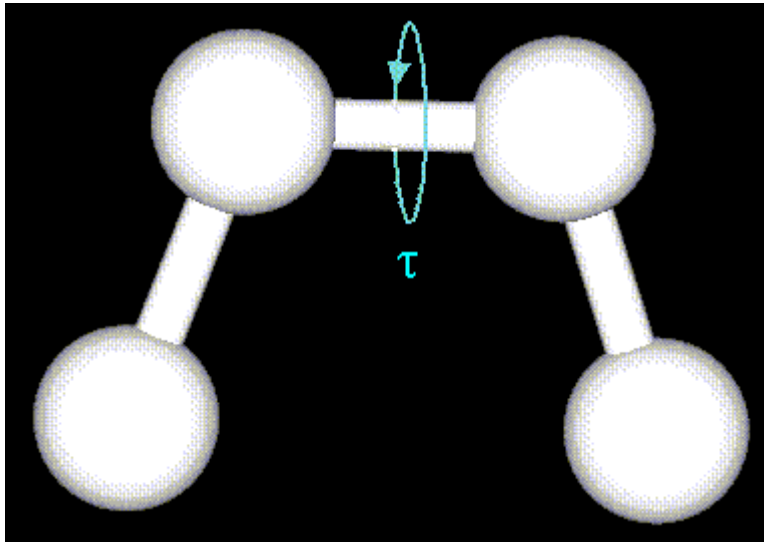
C=C, $k_b = 570 \text{ kcal/mol.Å}^2$, $b_0 = 1.229 \text{ Å}$

$$E = \sum_{\text{angles}} k_{\theta} (\theta - \theta_0)^2$$

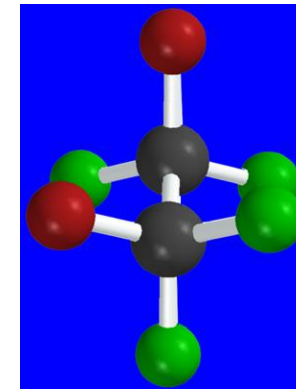
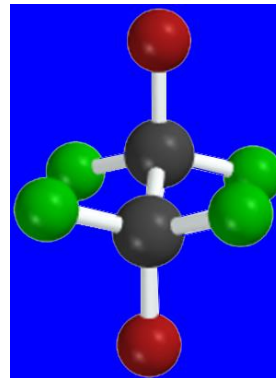
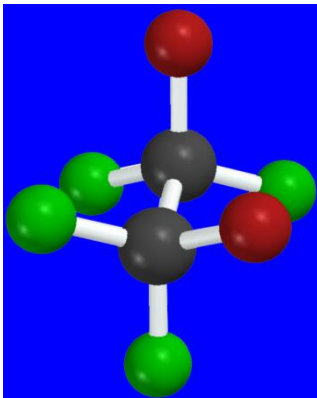
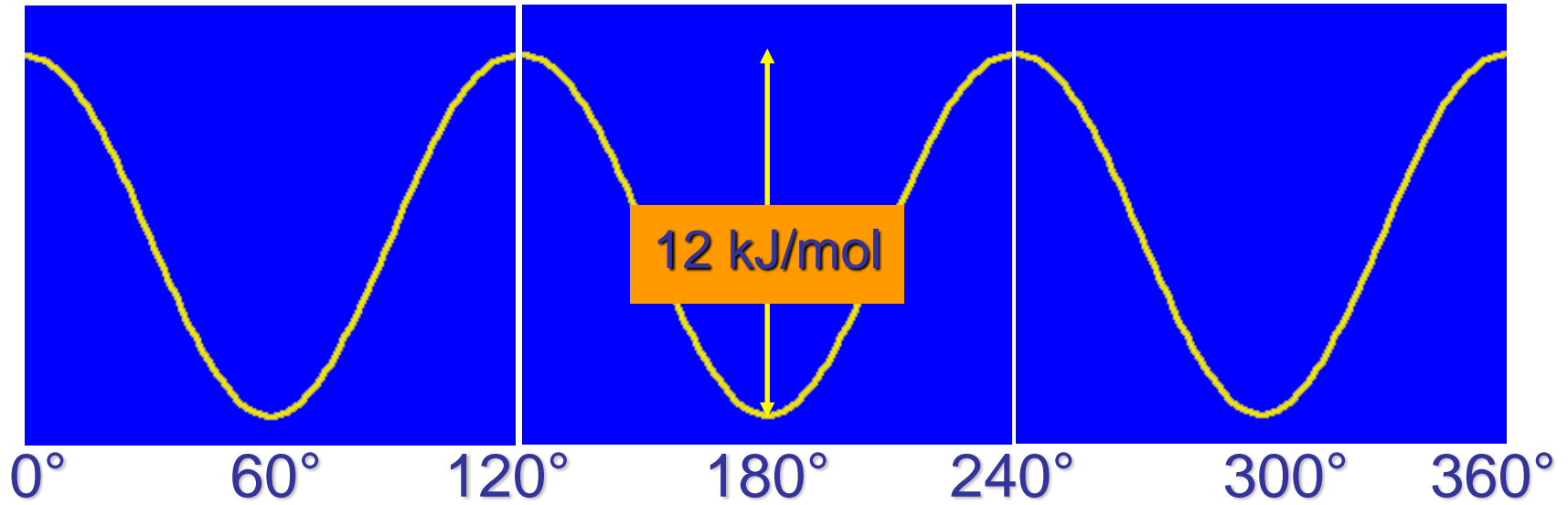
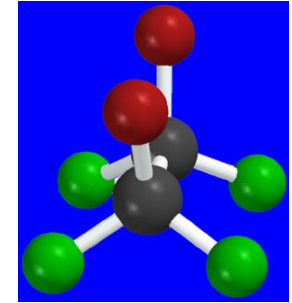
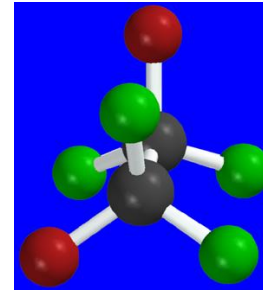
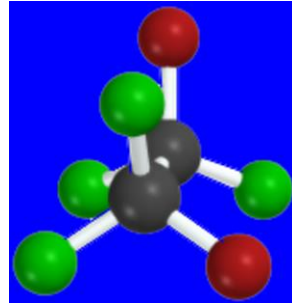
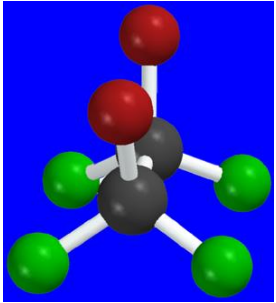


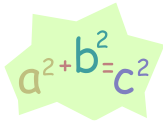
$$E = \sum A [1 + \cos(n\tau - \phi)]$$

torsions



La constante A définit l'amplitude des cycles





$$E_{\text{liaison}} = \sum k_b (r - r_o)^2$$

r_o = Longueur idéale de liaison (1.53 Å pour C -C)

r = Longueur réelle de la liaison

k_b constante de force (1000 kcal/mol.Å²)

$$+ E_{\text{angle}} = \sum k_{\theta} (\theta - \theta_o)^2$$

θ_o = Angle de valence idéal (109 °C -C -C)

θ = Angle de valence mesuré (sur la structure)

k_{θ} constante de force (500 kcal/mol.rad²)

$$+ E_{\text{angle}} = \sum k_{\delta} (1 + \cos(n\delta + \phi_o)) \quad n=1,2,3,4,6$$

δ = Angle de torsion

n = Périodicité

k_{δ} constante de force

$$+ E_{\text{impropre}} = \sum k_{\omega} (\omega - \omega_o)^2$$

ω_o = Angle de planarité idéale (180 °N -C α -CO)

ω = Angle de planarité mesurée

k_{ω} constante de force

E liaison > 0



Les interactions inter et intra-moléculaires sont de plusieurs types :

Interactions polaires

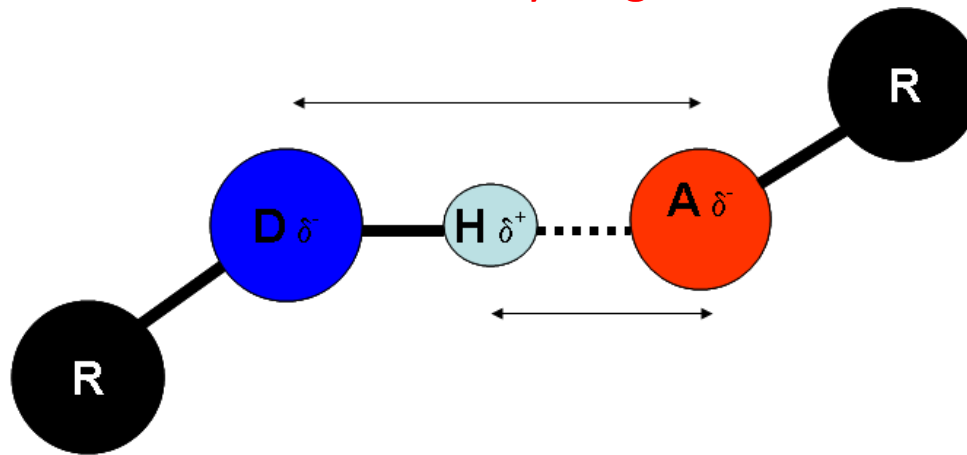
- liaisons hydrogènes
- interactions électrostatiques, pont salins
- dipôles-dipôles

Interactions hydrophobes

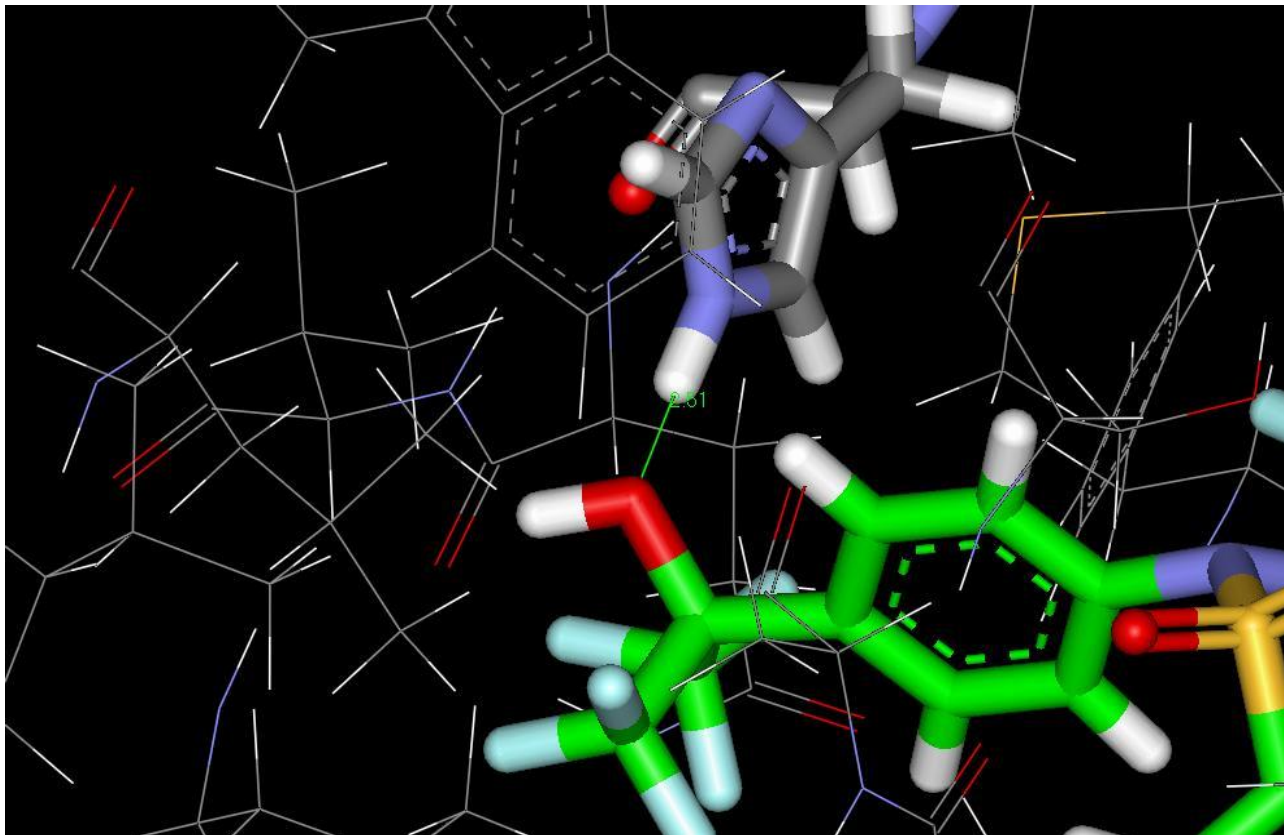
- aromatiques
- aliphatiques

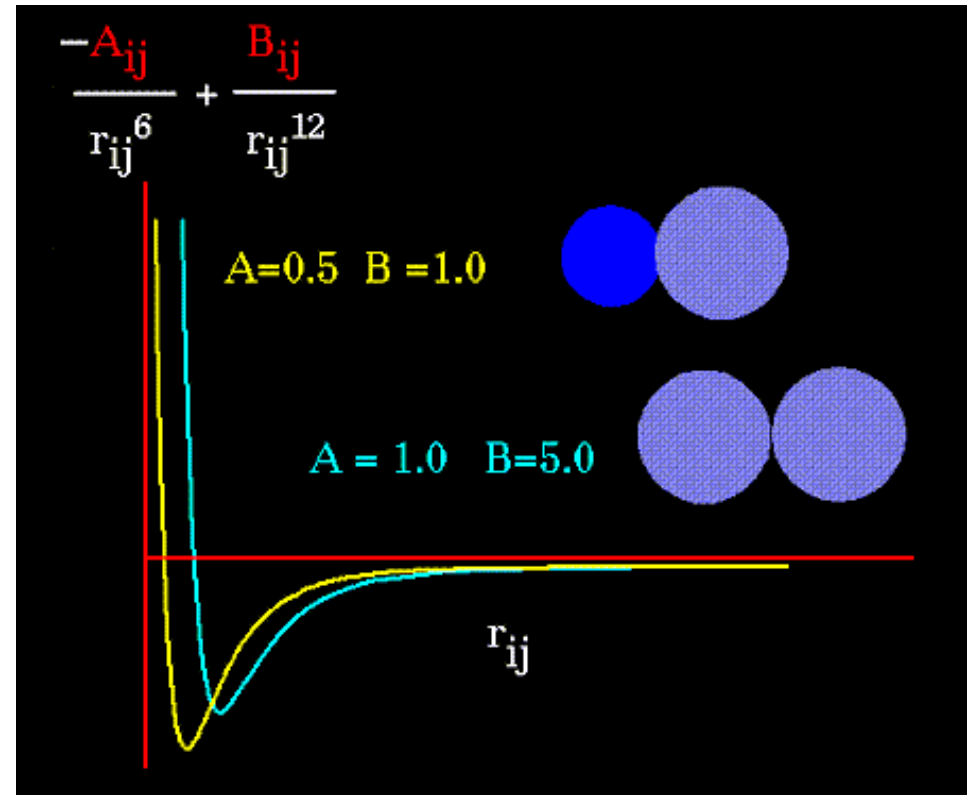
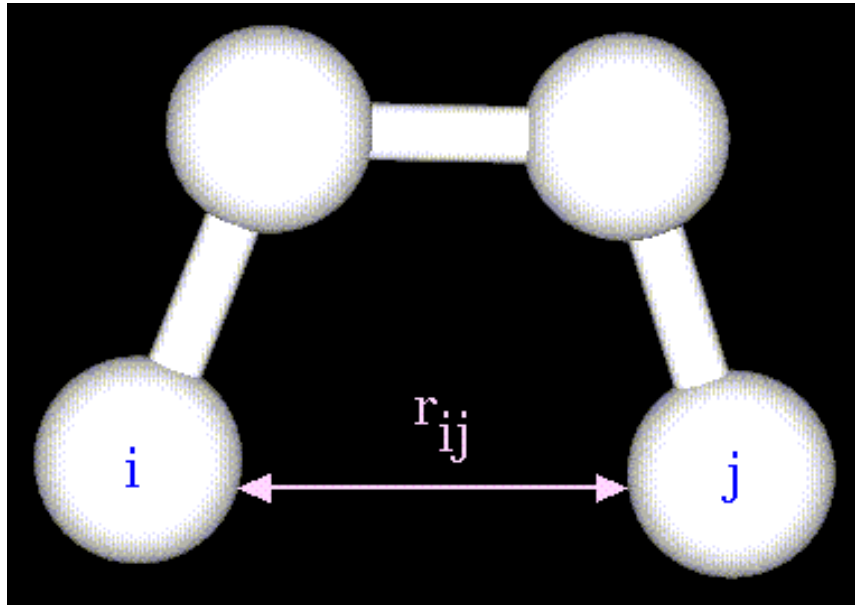
Liaisons hydrogènes

Liaison-H :
"Partage" un H



Energie ~ 5Kcal/mol

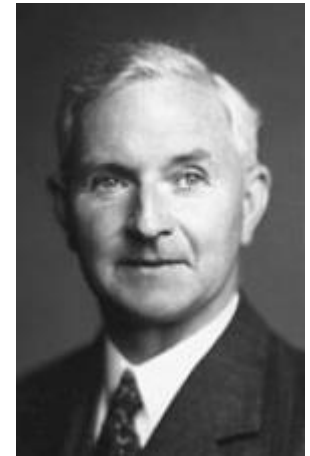
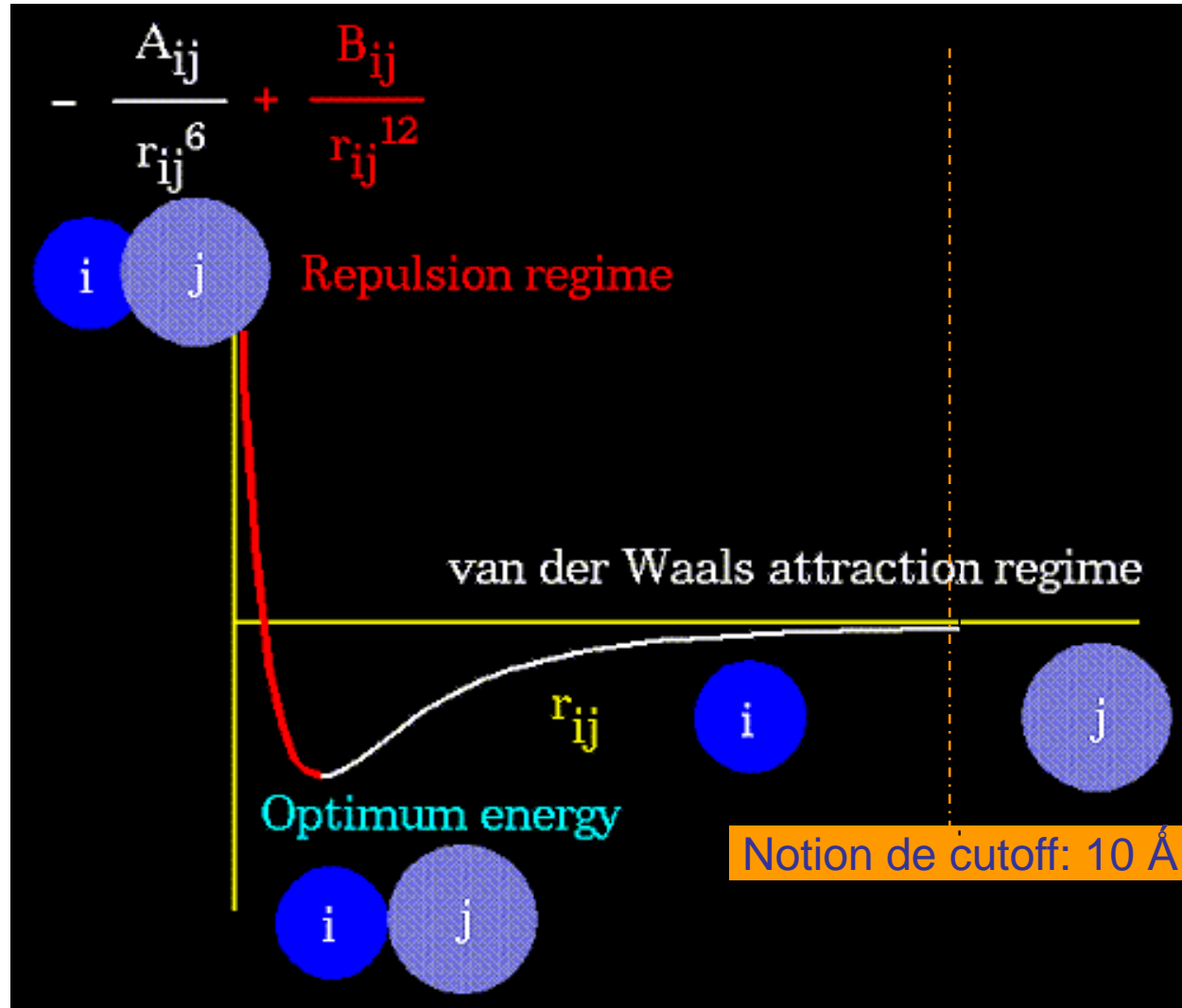




$$E = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$$

van der Waals term Electrostatic term

A (Attractif) et B (répulsif) influent sur la profondeur et la position du puits de potentiel



John Lennard-Jones
1894-1954

$$+ \sum 4\epsilon_{ij} \left[\left(\frac{r_0}{r_{ij}} \right)^{12} - \left(\frac{r_0}{r_{ij}} \right)^6 \right]$$

r_0 distance de Van der Waals

r_{ij} distance entre les atomes i et j

ϵ_{ij} , dépendance d'interaction entre les atomes (profondeur du puits)

$$+ \sum \frac{1}{4\pi \frac{\epsilon}{\epsilon_0}} \frac{q_i \cdot q_j}{r_{ij}}$$

D constante diélectrique (80 dans l'eau, 1 dans le vide)

q_i, q_j charges portées par les atomes i et j

ϵ est la permittivité dans le milieu

ϵ_0 est la permittivité dans le vide.

$$D = \frac{\epsilon}{\epsilon_0}$$

E liée

$$\begin{aligned}
 V = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2 \\
 & + \sum_{\text{dihedrals } n=1} \sum^N K_\phi^{(n)} [1 + \cos(n\phi - \delta)]
 \end{aligned}$$

E non liée

$$\begin{aligned}
 & + \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{i,j} \left(\frac{q_i q_j}{D r_{ij}} \right)
 \end{aligned}$$

- Parmall3X.pro CHARMM

	Atoms	kcal/A ²	ro
bonds HC	NH3	405.0000	1.04
bonds C	C	250.0000	1.38
bonds C	CT	187.0000	1.53
bonds C	N	403.0000	1.305
bonds C	NP	403.0000	1.305
bonds C	NR	403.0000	1.305
bonds C	NH1	403.0000	1.305
bonds C	NH2	403.0000	1.305
bonds C	NC2	403.0000	1.305
bonds C	O	595.0000	1.215
bonds C	OC	450.0000	1.22
bonds C	OH1	450.0000	1.38
bonds CT	CT	110.0000	1.53
bonds CT	N	261.0000	1.49
bonds CT	NH1	261.0000	1.49
bonds CT	NH2	261.0000	1.49
bonds CT	NH3	261.0000	1.49
bonds CT	NC2	261.0000	1.49
bonds CT	OH1	400.0000	1.42
bonds CT	S	450.0000	1.8100
bonds S	S	500.0000	2.0200

Exemple

	Atoms		kcal/Rad ²	Φ _o
angle H	NH1	H	30.000	107.500
angle H	NH1	C	35.3	120.0
angle H	NH1	CT	40.4	120.0
angle H	NH2	H	40.0000	120.0000
angle H	NH2	C	35.3	120.0
angle H	NH2	CT	40.4	120.0
angle H	OH1	CT	35.0000	108.0000
angle H	S	CT	35.0000	108.0000
angle H	OH1	C	35.0000	108.0000
angle H	OH2	H	120.000	104.5000
angle HC	NH3	HC	40.0000	109.5000
angle HC	NH3	CT	40.0000	109.5000
angle HC	NC2	HC	40.0000	120.0000
angle HC	NC2	C	35.3	120.0
angle HC	NC2	CT	40.4	107.5
angle HA	C	C	41.0	120.0
angle HA	C	NH1	41.0	120.0
angle HA	C	NH2	41.0	120.0
angle HA	C	NR	41.0	120.0
angle HA	C	O	41.0	120.0
angle HA	CT	HA	39.3	107.5
angle HA	CT	C	49.3	107.5
angle HA	CT	CT	30.000	108.000
angle HA	CT	N	51.5	107.5
angle HA	CT	NH1	51.5	107.5
angle HA	CT	NH3	51.5	107.5
angle HA	CT	NC2	51.5	107.5
angle HA	CT	OH1	50.0	107.5
angle HA	CT	S	40.000	107.500
angle C	C	C	60.0000	126.5000
angle C	C	CT	45.8	122.3
...				

L'énergie d'une protéine est la résultante :

D'une composante positive (E_{Liaison}) qui varie comme le nombre N d'atomes

$$E_b = k N / 2$$

D'une composante négative (sauf cas particulier) qui varie comme

$$E_{nb} = \frac{k' N (N-1)}{2}$$

Plus une protéine est de taille importante plus l'énergie doit être faible (et négative).

$$E \sim -K N \log(N)$$

Notion de cutoff: 10 Å

Res		bonds	angles	torsion	improper	NonBonds	elect.	TOTAL KJ/mol
HHT	1	0.000	6.183	7.540	0.000	0.00	-1.55	12.172
MET	1	0.299	3.103	1.277	0.290	-35.28	131.34	101.024
LYSH	2	0.261	9.799	10.282	2.089	-24.65	32.88	30.660
PRO	3	0.258	11.687	37.094	2.302	-28.21	-29.99	-6.862
VAL	4	0.305	5.176	11.376	228.355	-47.18	-2.75	195.281
THR	5	0.283	3.200	5.572	0.616	-30.08	-27.92	-48.330
LEU	6	0.167	3.060	3.789	240.623	-39.63	-2.49	205.515
TYR	7	0.467	10.807	18.036	6.466	-39.22	-47.55	-50.993
ASP	8	0.238	4.188	3.346	1.454	-54.17	-6.30	-51.249
VAL	9	0.103	4.751	2.441	230.448	-38.26	-9.33	190.156
ALA	10	0.082	4.335	0.672	0.195	-28.65	-11.06	-34.432
GLU	11	0.213	4.572	4.609	0.197	-26.50	-6.97	-23.874
TYR	12	0.431	3.458	6.223	4.268	-42.00	-44.99	-72.610
ALA	13	0.176	5.643	1.328	0.769	-20.30	23.20	10.808
GLY	14	0.083	3.459	2.917	0.013	-9.35	26.72	23.840
VAL	15	0.250	9.572	8.090	214.946	-25.75	-0.88	206.225
SER	16	0.262	5.665	7.694	0.548	-19.41	-16.16	-21.393
...								
ALA	32	0.166	7.809	0.784	2.478	-23.14	-8.96	-20.859
LYSH	33	0.350	4.704	4.953	0.948	-22.34	-10.01	-21.391
THR	34	0.221	5.256	7.807	0.919	-28.96	-22.55	-37.299
ARG	35	0.419	9.431	2.995	2.925	-50.00	-260.86	-295.087
GLU	36	0.339	5.161	6.122	1.350	-39.73	-13.73	-40.486
LYSH	37	0.230	10.246	3.236	3.897	-40.54	-13.07	-35.997
VAL	38	0.320	3.649	4.050	233.542	-37.90	-10.52	193.138
GLU	39	0.218	2.884	3.221	0.122	-40.42	-10.84	-44.824
ALA	40	0.052	4.074	4.491	1.059	-24.66	-15.56	-30.536
ALA	41	0.129	2.730	2.246	1.437	-32.63	-12.09	-38.180
MET	42	0.085	5.116	3.731	0.520	-49.87	-7.36	-47.775
ALA	43	0.247	3.484	2.975	0.291	-20.67	-2.03	-15.700
GLU	44	0.234	4.382	1.743	0.786	-32.80	-12.62	-38.272
LEU	45	0.295	9.293	14.002	242.358	-42.22	-5.22	218.512
ASN	46	0.239	12.827	5.434	1.988	-26.66	-192.49	-198.667
TYR	47	0.354	8.413	1.867	5.100	-52.38	-42.04	-78.690
ILE	48	0.169	11.573	14.236	5.867	-30.01	21.57	23.412
PRO	49	0.253	16.682	19.383	1.805	-24.03	-35.18	-21.088
ASN	50	0.121	7.159	6.759	1.811	-17.91	-174.76	-176.812
ARG	51	0.286	11.947	12.838	0.459	-28.32	-189.62	-192.406
OXT	51	0.000	0.000	0.000	0.000	-2.27	2.83	0.558
TOTAL		13	70	336	2385	-1592	-1923	-411



Exemple d'une protéine (vide):

1BTA (inhibiteur de ribonucléase): 89 aa (915 atomes)

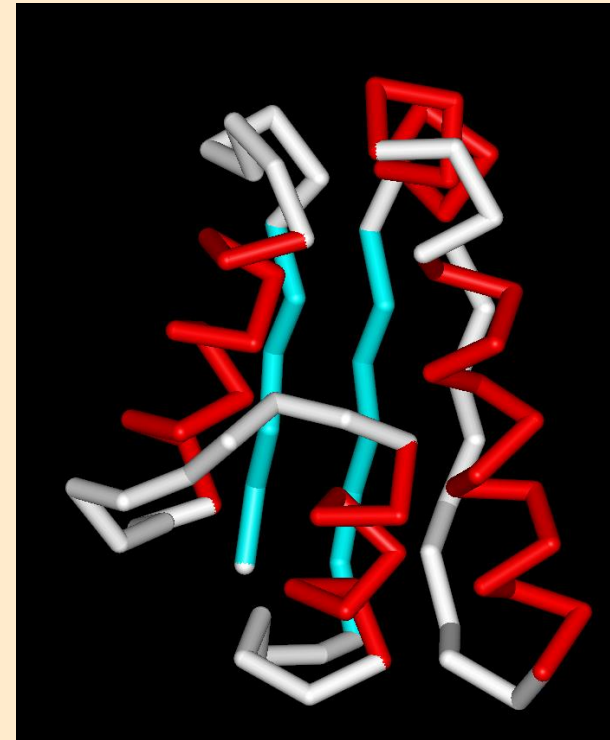
Champ de force GROMOS

Conformation 1



$E=911$ kJ/mol

Conformation 2

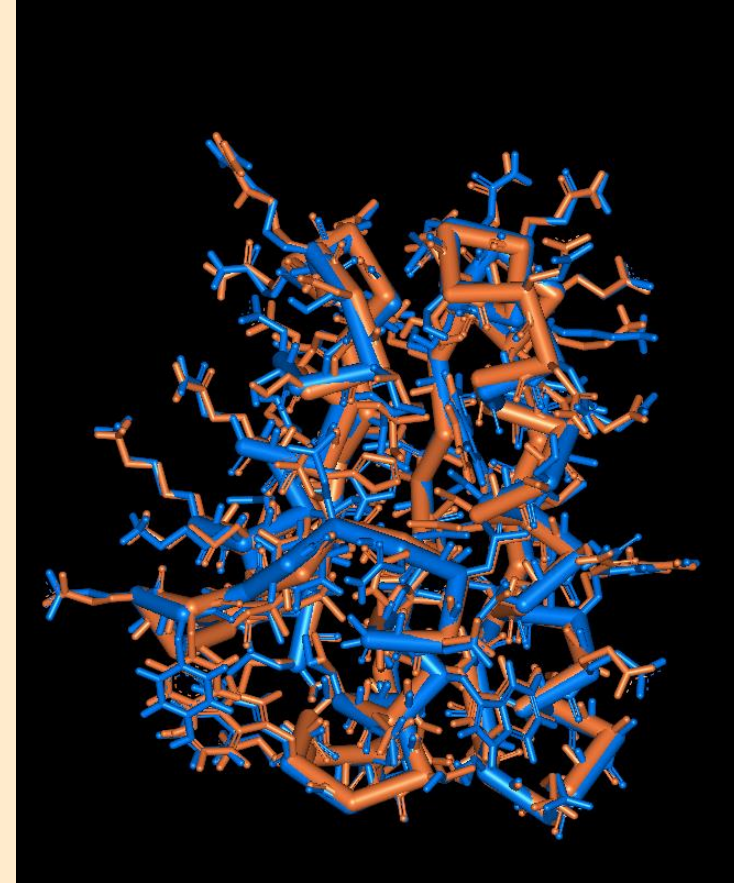
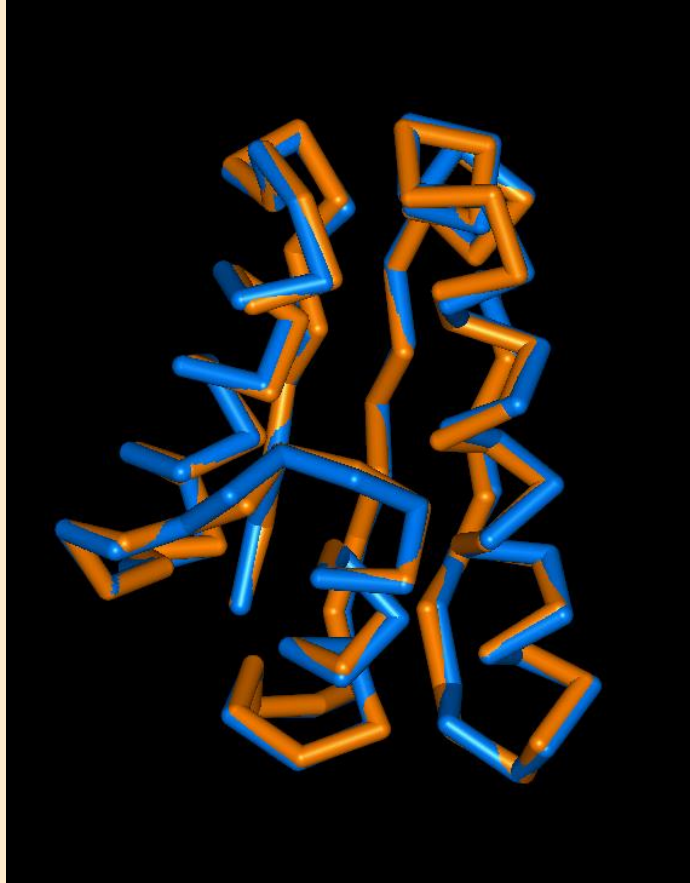


$E=-4998$ kJ/mol

Exemple d'une protéine (vide):

1BTA (inhibiteur de ribonucléase): 89 aa (915 atomes)

Champ de force GROMOS



Les protéines se trouvent en général (sauf les protéines membranaires) dans un solvant aqueux. Des interactions existent entre la protéine et les molécules d'eau de façon à contribuer à leur stabilité. (solvant **explicite** ou **implicite**)

Solvant explicite

Interactions non liées

Ajout de molécules de solvant (eau) autour de la molécule à considérer

Exemple 1BTA placée dans une boîte d'eau

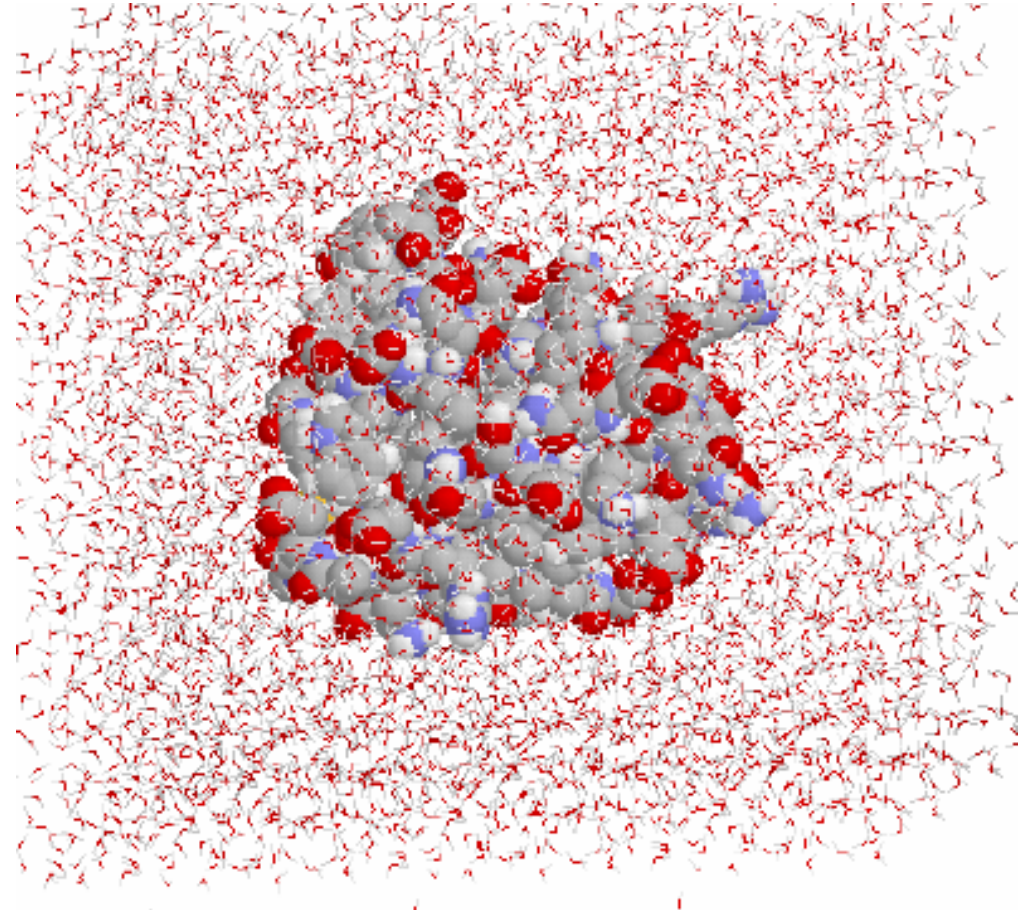
18114 atomes

12394 liaisons covalentes

7087 angles de valence

483 angles dièdres

328 058 954 interactions non liées



Calcul très précis sans approximation (sauf le cut-off) mais très coûteux

Plusieurs centaines de millions de termes à calculer

Solvant implicite

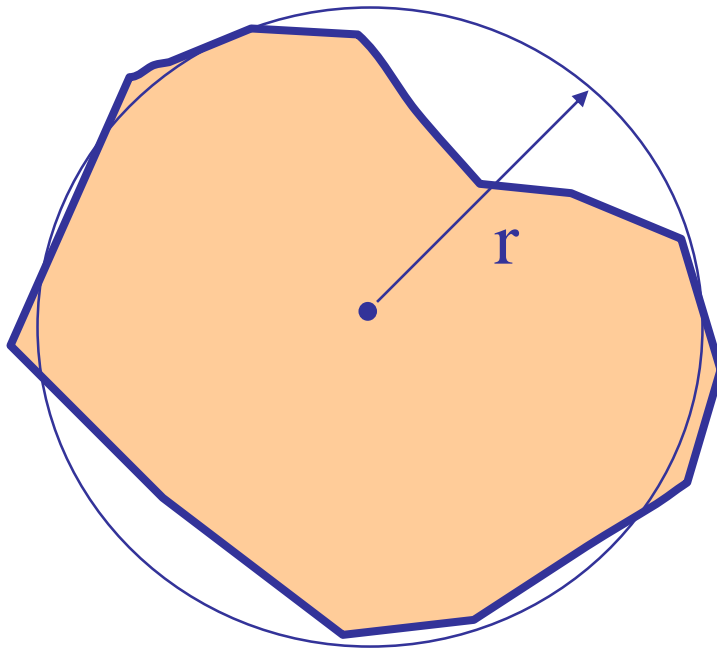
On traite l'influence du solvant par un terme électrostatique qui mime l'effet d'écrantage des charges des molécules d'eau.

Modification de la constante diélectrique du milieu.

Un protéine n'est ni du vide ($\epsilon=1$) ni de l'eau ($\epsilon=80$)

$$\epsilon = 4 r$$

r rayon d'une sphère équivalente centrée sur le centre de masse



Solvant implicite (dans le vide)

928 liaisons covalentes

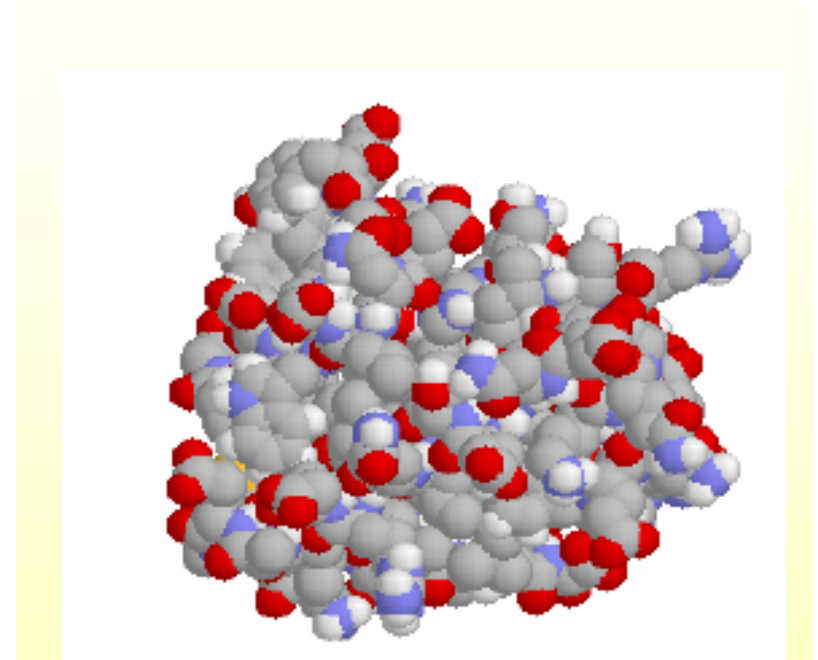
1354 angles de valences

483 angles diédres

830 780 interactions non liées

Interactions non liées

Calcul peu précis mais moins coûteux



Seulement 850 000 termes à calculer !!

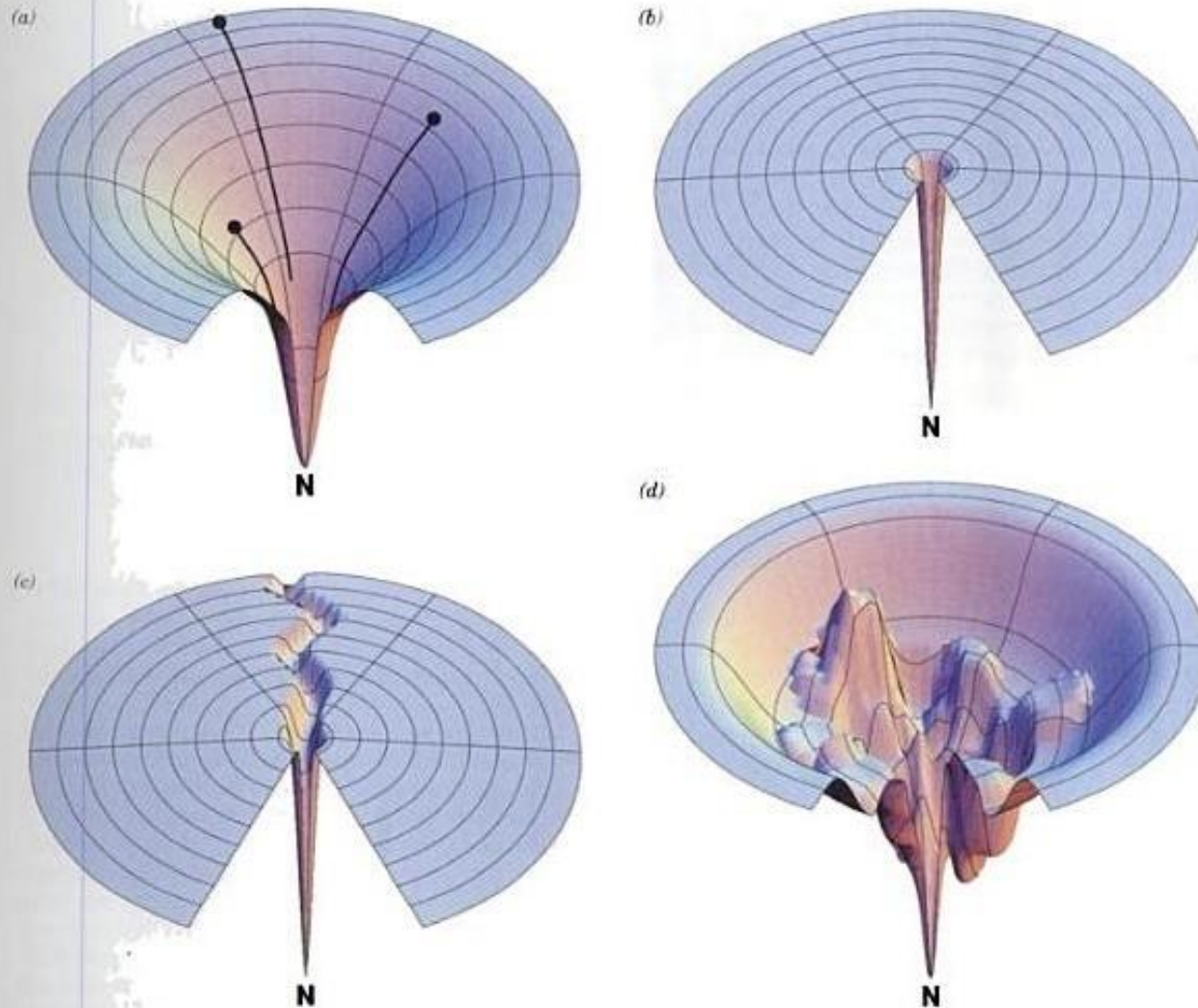
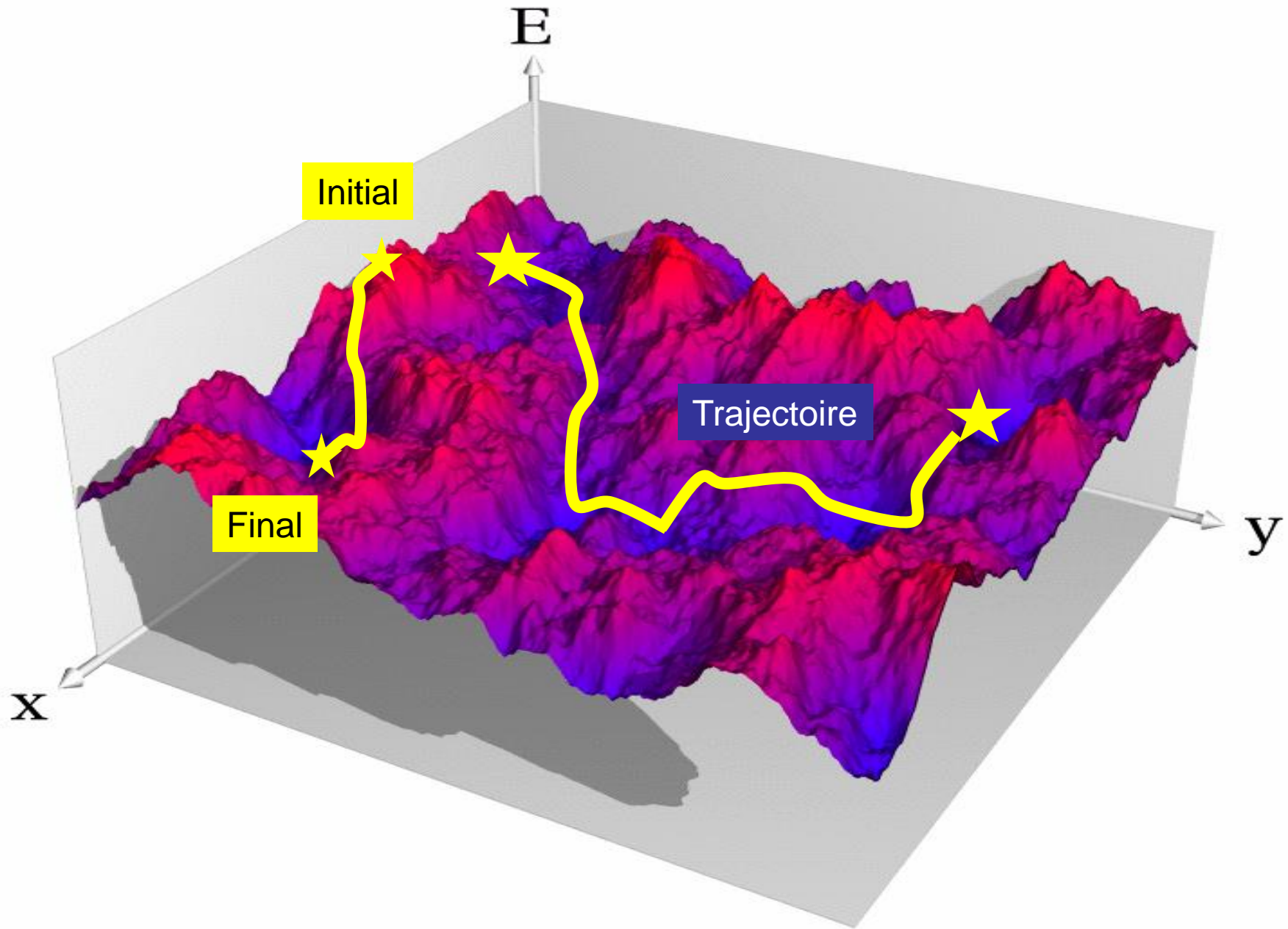


FIGURE 9-11 Entonnoirs de repliement. (a) Paysage en entonnoir idéalisé. À mesure que la chaîne polypeptidique établit un nombre croissant de contacts intracaténaux, son énergie libre interne (sa hauteur au dessus de l'état natif, N) décroît ainsi que sa liberté conformationnelle (la largeur de l'entonnoir). Des polypeptides de conformations différentes (points noirs) suivent différentes voies (lignes noires) pour atteindre l'état natif. (b) Le paysage de Levinthal en « terrain de golf » dans lequel la chaîne polypeptidique doit trouver son état natif (le trou) par recherche

aléatoire, c'est-à-dire sur une surface énergétique plane. (c) Le paysage de repliement classique où la chaîne se déplace au hasard sur une surface énergétique plane jusqu'à ce qu'elle rencontre une gorge qui conduit à l'état natif. (d) Surface énergétique accidentée présentant des minima locaux dans lesquels un polypeptide en cours de repliement peut être piégé momentanément. On pense que les entonnoirs de repliement des protéines naturelles ont cette topographie là. [Avec la permission de Ken Dill, University of California at San Francisco.]



Minimisation http://uel.unisciel.fr/chimie/modelisation/modelisation_ch03/co/modelisation_ch03.html

Les algorithmes de minimisation d'énergie mesurent l'énergie le long de la surface de façon incrémentale pour déterminer des directions qui mènent à un minimum.

Algorithmes incapables de changer de vallée => Minima locaux

1ère dérivée

Steepest Descent (plus grande pente)

Suit le gradient de la fonction énergie à chaque pas en utilisant le calcul de la pente.

Très efficace en début de cycle.

Peut conduire à des oscillations

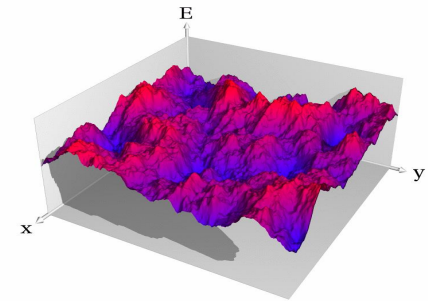
Mauvaise convergence à l'approche du minimum (recherche d'extremum)

Conjugate Gradients

Tient compte des gradients calculés aux étapes précédentes pour éviter les oscillations.

Bonne convergence.

Peut rencontrer des problèmes quand les conformations initiales sont très distordues



2^e dérivée Newton-Raphson

Utilise la dérivée seconde Cherche uniquement les minima

On calcule d'abord l'énergie initiale E_0

Chaque atome de coordonnées x,y,z est déplacé d'un facteur de déplacement dx,dy,dz ,
La nouvelle énergie E_1 est calculée.

La dérivée première de l'énergie potentielle (ou encore gradient d'énergie) est :

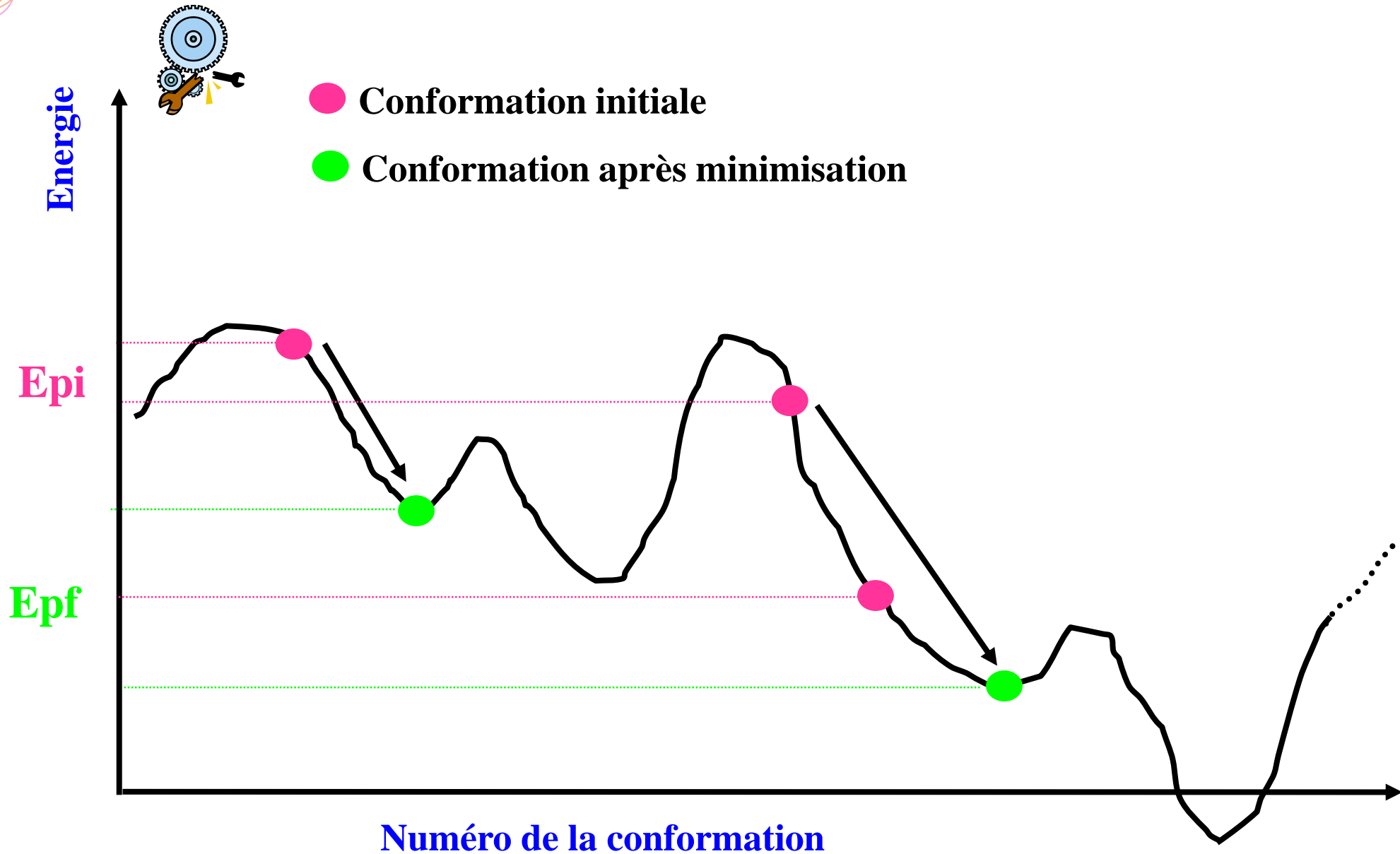
$$\text{Grad}(E) = dE/dxyz = (e_1 - e_0)/dxyz$$

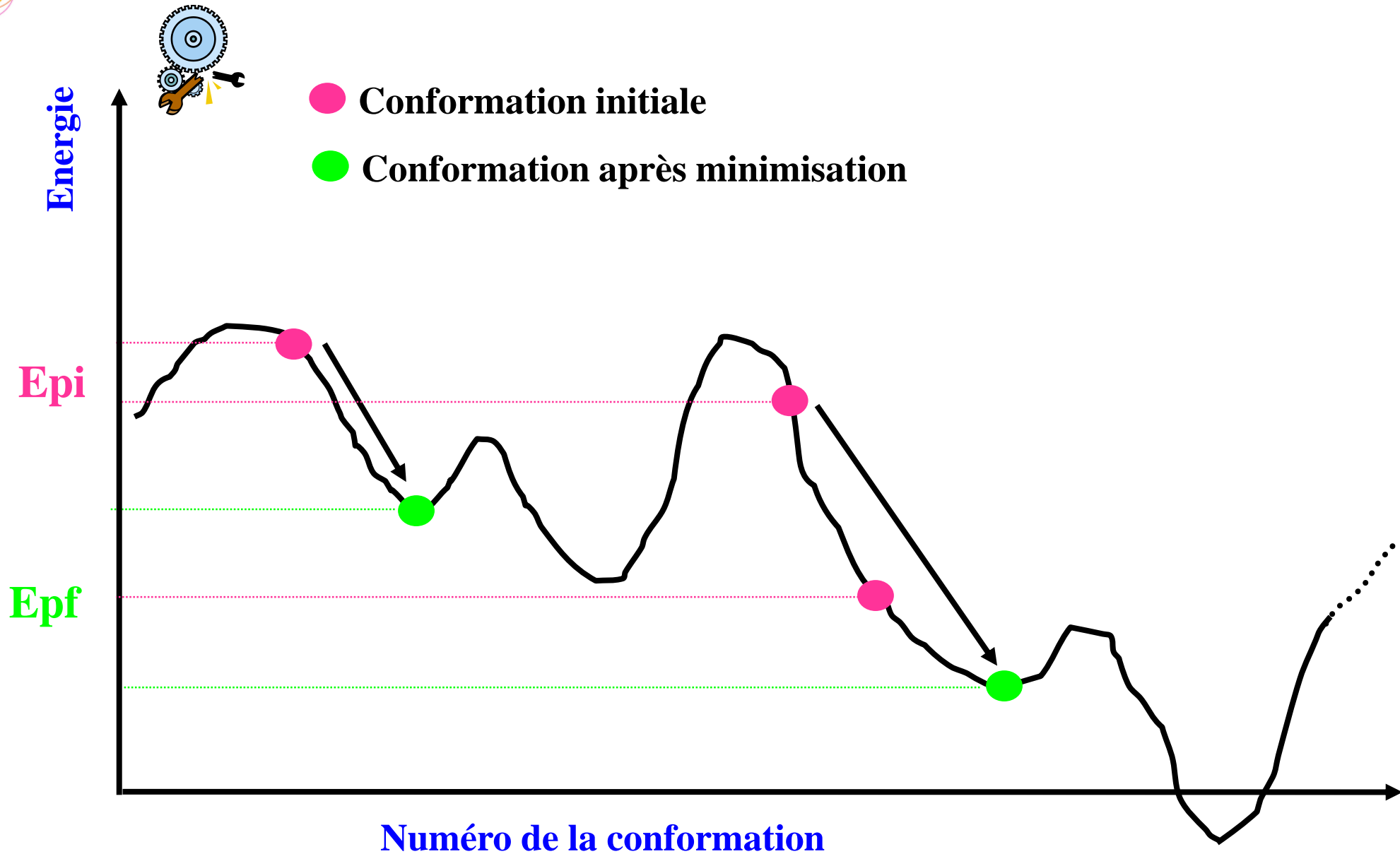
Si le gradient est <0 , l'énergie diminue avec $dxyz$.

Si le gradient est >0 , l'énergie augmente avec $dxyz$.

Chaque atome est déplacé sur une distance dépendant de $dE/dxyz$. Cet algorithme suivra donc la direction imposée par les forces interatomiques dominantes et consiste à rechercher la direction de la plus grande pente au cours de laquelle l'énergie décroît le plus rapidement. La direction suivie sera celle indiquée par l'opposé au gradient d'énergie, c'est-à-dire dans la direction où l'énergie diminue le plus vite.

Cette méthode est rapide dans les premiers cycles mais converge très lentement en fin de cycle.





- **Avantages**

- Convergence assez rapide
- Energie chute très rapidement (10^7 à $100 \text{ kcal.mol}^{-1}$) par effet du champ de force
- Atteinte rapide du minimum le plus proche

- **Inconvénients**

- Dépend de la conformation initiale
- Pas de réarrangements très importants => Exploration conformationnelle faible
- Barrières d'énergie infranchies
- Dépendance du champ de force
- Toujours associer valeur d'énergie avec champ de force et taille de la molécule

- **A utiliser en début et fin d'optimisation par dynamique moléculaire**

Energy Minimisation Preferences [X]

Dc Steps of

Dc Steps of

Dc Steps of

Bonds Non-bonded Cutoff Å

Angles Electrostatic

Torsions Show Energy Report

Improper

Stop when delta E between two steps is below kJ/mol

Stop when Force acting on any atom is below

Lock non-selected residues Lock / Constrain is for Carbon Alpha only

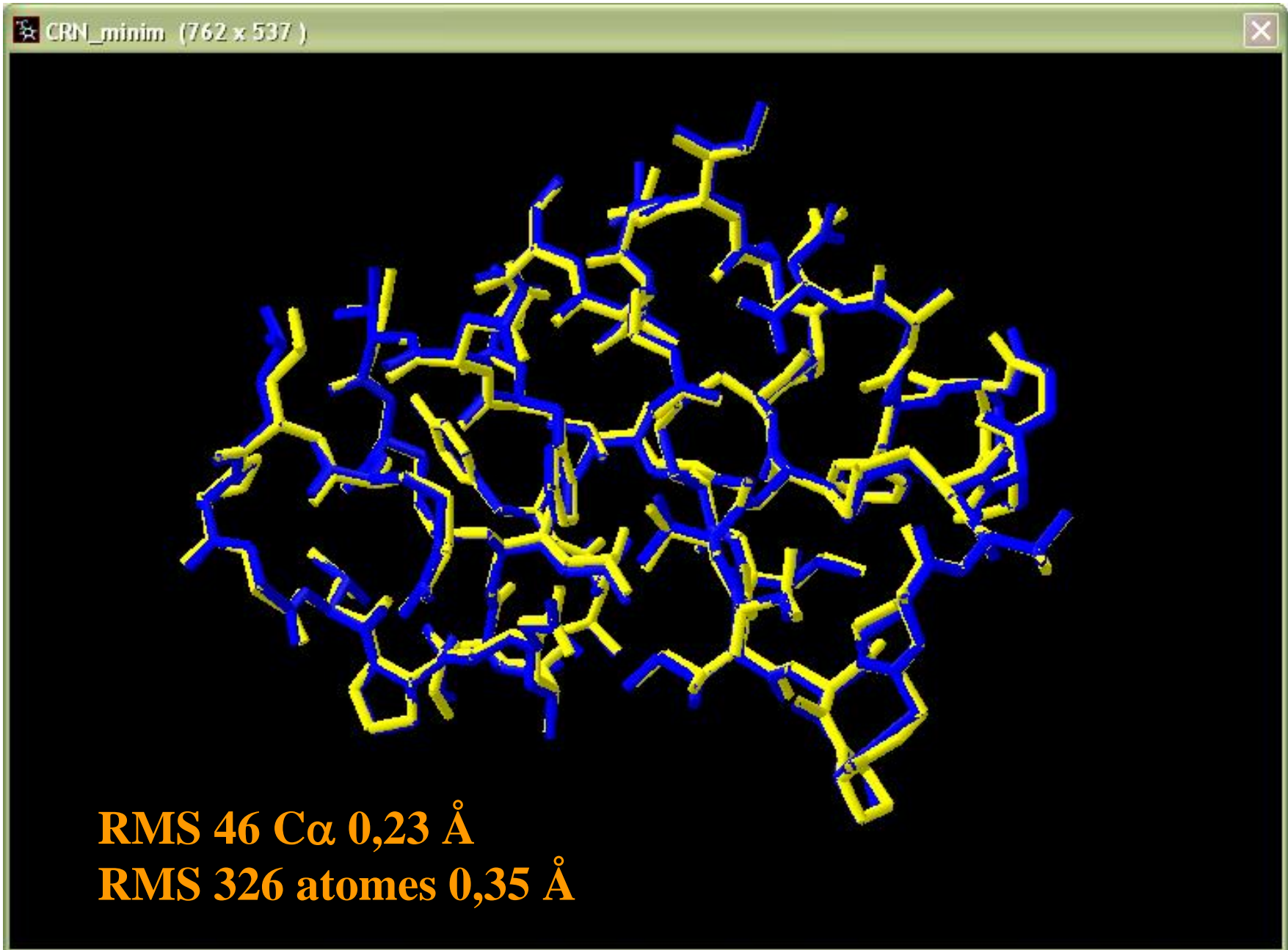
Use an harmonic constraint:

(selected residues)

(non selected residues)

Res		bonds	angles	torsion	improper	NonBond	elect.	TOTAL KJ/mol
HHT	1	0.003	0.152	2.788	0.000	0.00	4.53	7.477
MET	1	0.979	2.977	0.757	0.178	-33.11	124.46	96.242
LYSH	2	1.088	9.382	4.108	2.147	-25.68	31.09	22.128
PRO	3	0.207	14.520	27.634	3.960	-28.97	-32.12	-14.767
VAL	4	1.203	8.194	7.434	5.932	-40.47	-3.11	-20.818
THR	5	0.464	1.507	5.877	0.437	-32.13	-30.55	-54.393
LEU	6	0.689	6.280	2.668	2.970	-34.04	-8.04	-29.469
TYR	7	0.907	4.759	12.335	2.702	-38.21	-58.15	-75.653
ASP	8	0.404	2.088	1.671	1.189	-52.34	-11.73	-58.711
VAL	9	0.813	5.135	0.824	4.355	-42.37	-11.12	-42.361
ALA	10	0.373	1.326	2.180	0.613	-24.22	-13.33	-33.061
GLU	11	0.468	2.251	1.744	0.459	-26.42	-10.34	-31.833
TYR	12	0.577	2.120	5.859	1.155	-41.70	-48.03	-80.021
ALA	13	0.758	4.263	0.486	1.543	-20.28	18.98	5.750
GLY	14	0.729	2.297	1.941	0.095	-8.73	22.05	18.378
VAL	15	0.831	5.751	1.813	4.194	-22.95	-4.70	-15.054
SER	16	1.192	6.260	5.301	3.455	-19.11	-36.07	-38.974
...								
ALA	32	0.894	7.498	1.612	1.717	-21.22	-19.90	-29.408
LYSH	33	0.552	3.056	2.472	0.853	-23.57	-12.40	-29.039
THR	34	0.484	3.084	7.255	1.295	-33.56	-32.11	-53.548
ARG	35	1.640	6.546	2.840	3.317	-50.80	-271.59	-308.047
GLU	36	0.292	2.676	3.370	0.824	-38.40	-19.01	-50.244
LYSH	37	0.495	7.235	3.639	2.618	-40.80	-17.47	-44.287
VAL	38	1.044	5.797	2.800	4.646	-34.18	-13.22	-33.111
GLU	39	0.696	0.890	5.005	0.911	-40.21	-13.49	-46.194
ALA	40	0.377	1.820	2.371	1.212	-22.39	-19.69	-36.303
ALA	41	0.521	1.623	2.106	1.719	-33.71	-15.49	-43.227
MET	42	0.483	2.622	3.026	0.756	-47.88	-14.68	-55.675
ALA	43	0.435	1.718	3.971	0.978	-20.93	-6.02	-19.847
GLU	44	0.491	2.741	2.291	0.295	-33.04	-15.39	-42.613
LEU	45	0.566	11.583	7.521	4.467	-42.77	-5.75	-24.388
ASN	46	1.179	12.806	4.930	2.653	-28.46	-199.58	-206.468
TYR	47	0.591	6.808	1.987	2.743	-47.39	-46.61	-81.872
ILE	48	1.721	11.693	5.809	6.928	-32.71	20.13	13.565
PRO	49	0.404	14.541	17.348	1.810	-25.82	-34.77	-26.480
ASN	50	1.420	7.162	3.482	3.268	-13.85	-189.05	-187.566
ARG	51	1.878	6.650	8.697	1.980	-26.35	-199.45	-206.590
OXT	51	0.000	0.000	0.000	0.000	-2.33	3.70	1.375
Total		43	283	231	114	-1559	-2297	-3183





Exemple d'une protéine (vide):

1BTA (inhibiteur de ribonucléase): 89 aa (915 atomes)

Champ de force GROMOS

Conformation 1 (PDB)



$E=911$ kJ/mol

Conformation 2
(minimisation d'énergie)



$E=-4998$ kJ/mol

Problème du voyageur de commerce:

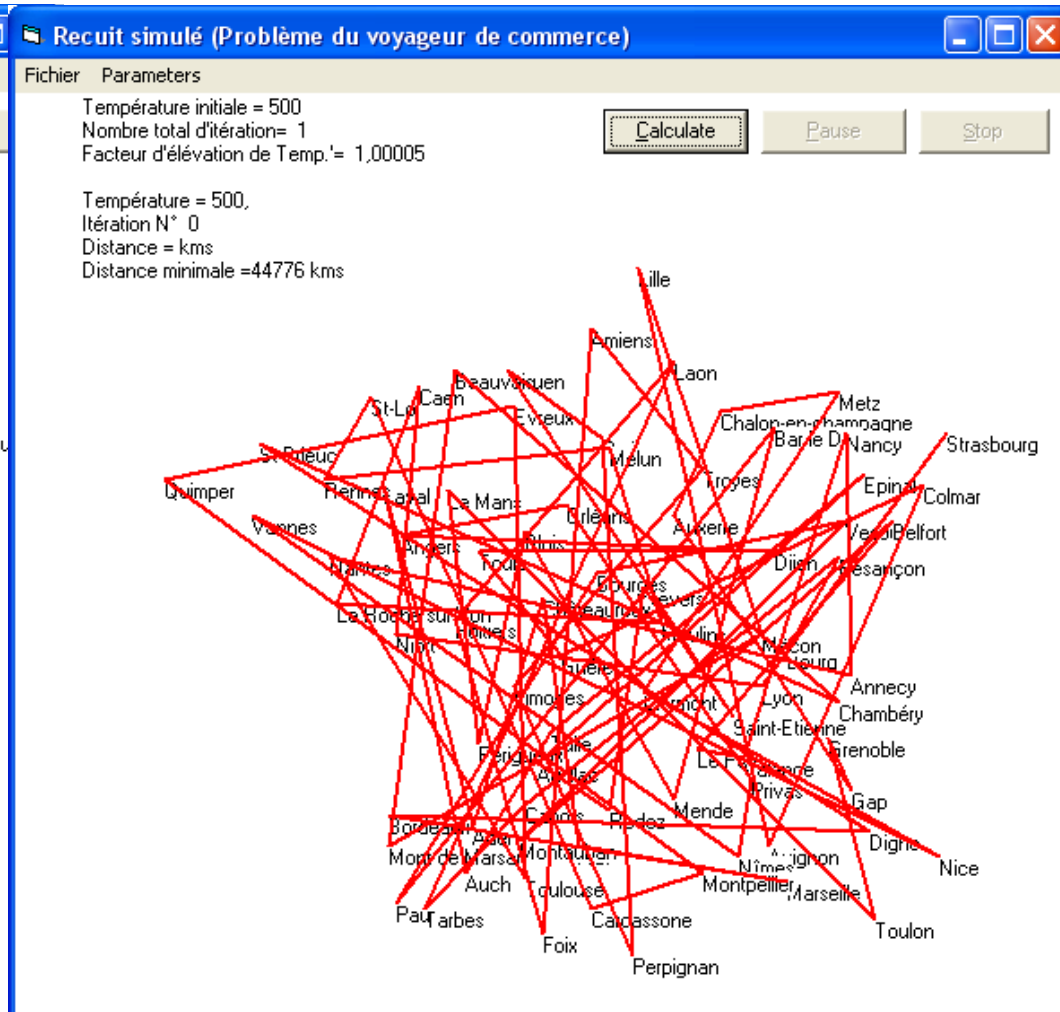
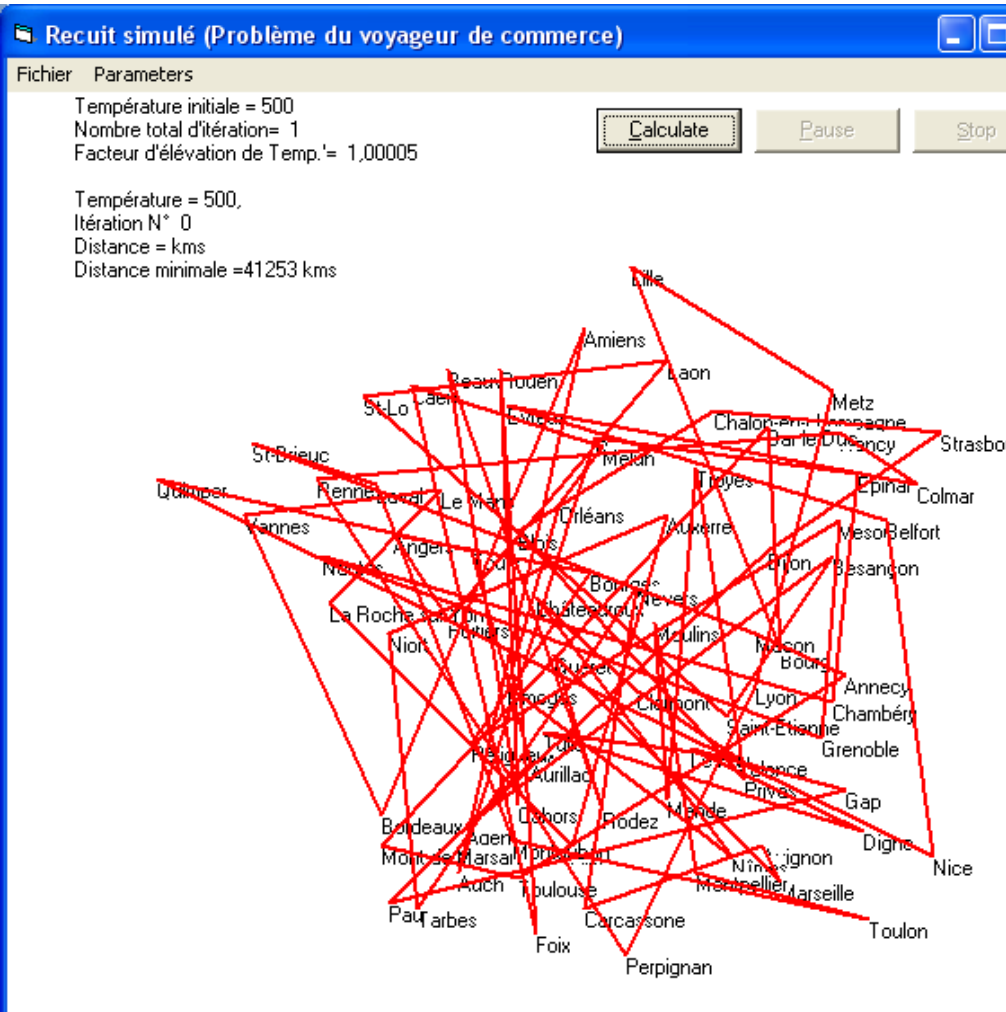
Comment visiter le maximum de villes en France en faisant le minimum de distance et sans revenir en arrière?

$$N = (n-1)!/2$$

Nb de villes	Possibilités	Temps de calcul (1 μ s/trajet)
5	12	12 microsecondes
10	181 440	0,18 seconde
15	43 milliards	12 heures
20	60 E+15	1928 ans
25	310 E+21	9,8 milliards d'années (!)

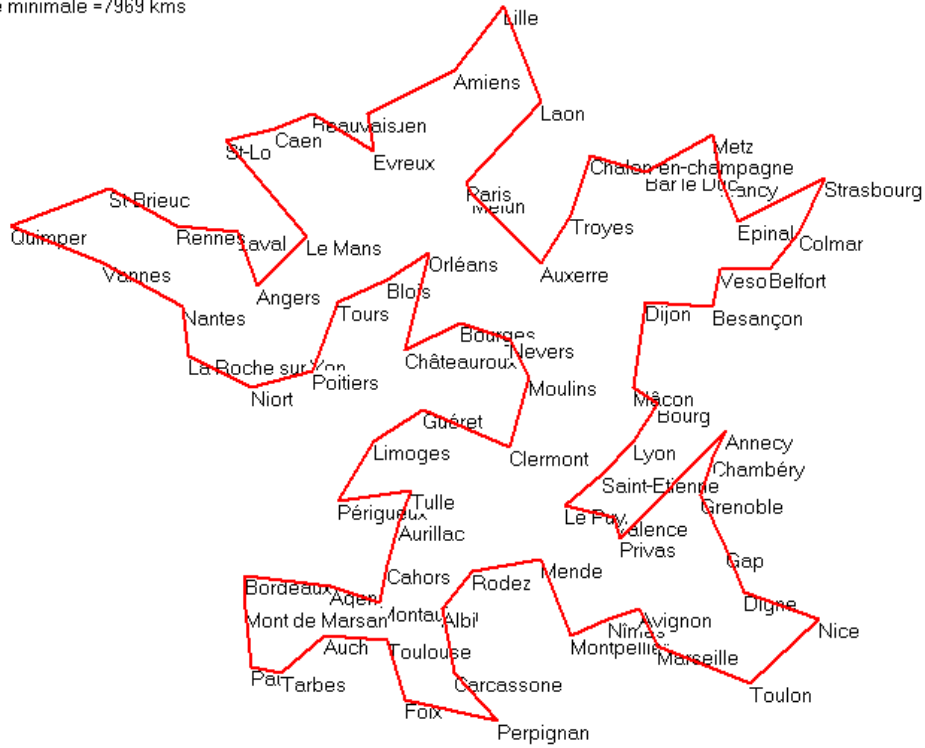
Ici, la fonction à minimiser F est donc la longueur du chemin. L'exploration aléatoire de l'espace d'état consiste à choisir aléatoirement deux villes et à intervertir leur place dans le chemin courant. On retient la solution si elle est meilleure sinon on prend une probabilité d'acceptation = $-\Delta F/T$

Exemple



Température initiale = 300
 Nombre total d'itération= 2000
 Facteur d'élévation de Temp.'= 1,002

Température = 300,
 Itération N° 0
 Distance = kms
 Distance minimale =7969 kms



Température initiale = 300
 Nombre total d'itération= 2000
 Facteur d'élévation de Temp.'= 1,002

Température = 300,
 Itération N° 0
 Distance = kms
 Distance minimale =7904 kms



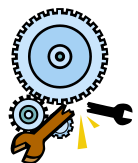
<http://www.lps.ens.fr/~weisbuch/livre/b9.html>

Température = 300,
Itération N° 0
Distance = kms
Distance minimale = 7841 kms

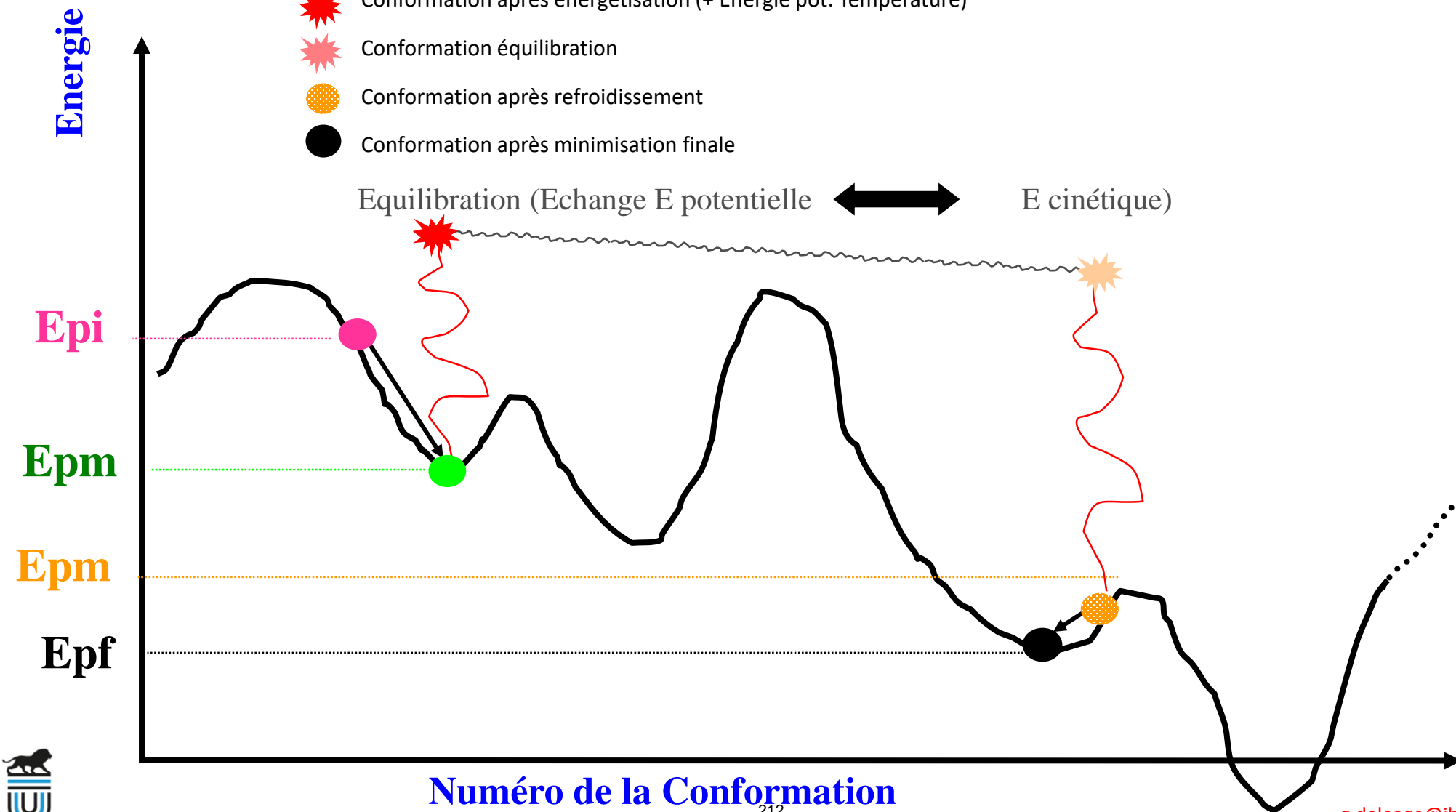
https://interstices.info/jcms/c_43811/le-recuit-simule

http://perso.ibcp.fr/gilbert.deleage/Cours/software/recuit_simule.exe

http://uel.unisciel.fr/chimie/modelisation/modelisation_ch05/co/modelisation_ch05.html



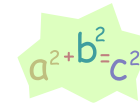
- Conformation initiale
- Conformation après minimisation
- ★ Conformation après énergétisation (+ Energie pot. Température)
- ★ Conformation équilibration
- Conformation après refroidissement
- Conformation après minimisation finale



● Objectifs

- Franchir des barrières énergétiques
- Fourniture d'énergie potentielle

$$E_{cin} = \sum \frac{1}{2} m v_i^2 = (3N - 6) \frac{kT}{2} = Fd$$



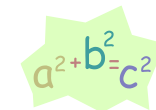
● Mathématiquement

- Trajectoire des atomes déterminée par intégration numérique des équations du mouvement de Newton (nécessite un pas d'intégration ~1fs)

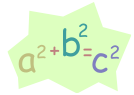
$$m_i \frac{d^2 r_i(t)}{dt^2} = m_i a_i = F_i$$

avec

$$F_i = - \frac{\partial E_p(r_1, \dots, r_i, \dots, r_n)}{\partial r_i}$$



- Algorithme de Verlet. Intégration sur $\Delta t = 1$ fs



$$r(t + \Delta t) = r(t) + \frac{dr(t)}{dt} \Delta t + \frac{d^2 r_i(t) \Delta t^2}{dt^2 \cdot 2} + \dots = r(t) + v_i(t) \Delta t + a_i(t) \frac{\Delta t^2}{2}$$

$$r(t - \Delta t) = r(t) - v_i(t) \Delta t + a_i(t) \frac{\Delta t^2}{2}$$

$$r(t + \Delta t) = 2r(t) - r_i(t - \Delta t) + a_i(t) \Delta t^2$$

Nouvelle position indépendante de la vitesse

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t}$$

mais V Calculable à tout instant t

- **Avantages**

- Réarrangements très importants
- Exploration conformationnelle importante
- Barrières d'énergie franchies

- **Inconvénients**

- Convergence lente voire nulle!
- Température sans grande signification physique
- Paramétrages délicats
 - Modification du champ de force pendant la dynamique
 - K_b , K_ϕ , K_w à augmenter pour "protéger la chimie de la molécule"
- Minimum minimorum pas forcément atteint

- **Compromis entre exploration forte - intégrité structurale**



- **t=0: r(0) structure initiale (minimisée), v(0) aléatoires**

- **Immersion dans une boîte de solvant**

- **Montée en température**

- Incréments de 20K (100 pas d'incrément) de Δt

- **Equilibration (300K)**

- Ajustement des vitesses pour conserver $T = \text{cte}$

- Echange entre E_{cin} et E_{pot}

- A l'équilibre, $E_{\text{tot}} = E_{\text{pot}} + E_{\text{cin}} = \text{Cte}$

- E_{cin} permet de franchir des barrières énergétiques de l'ordre de $kT/2$ par degré de liberté

- **Production de conformations**

- Trajectoires échantillonnées et stockées

- Vitesses sont stockées pour une analyse statistique

- **Refroidissement**

```

heat  { ps=100  T=( 0 , 300)  }
equil { ps=100  T=300        }
prod  { ps=500  T=300        }
cool  { ps=100  T=( 300 , 0)  }
  
```

- **Objectif => Calculer l'énergie de la protéine (X-PLOR)**

$E_{\text{Totale}} = E_{\text{Empirique}} + E_{\text{Contrainte}}$

$E_{\text{Contrainte}} = E_{\text{Noe}} + E_{\text{Dièdres}}$

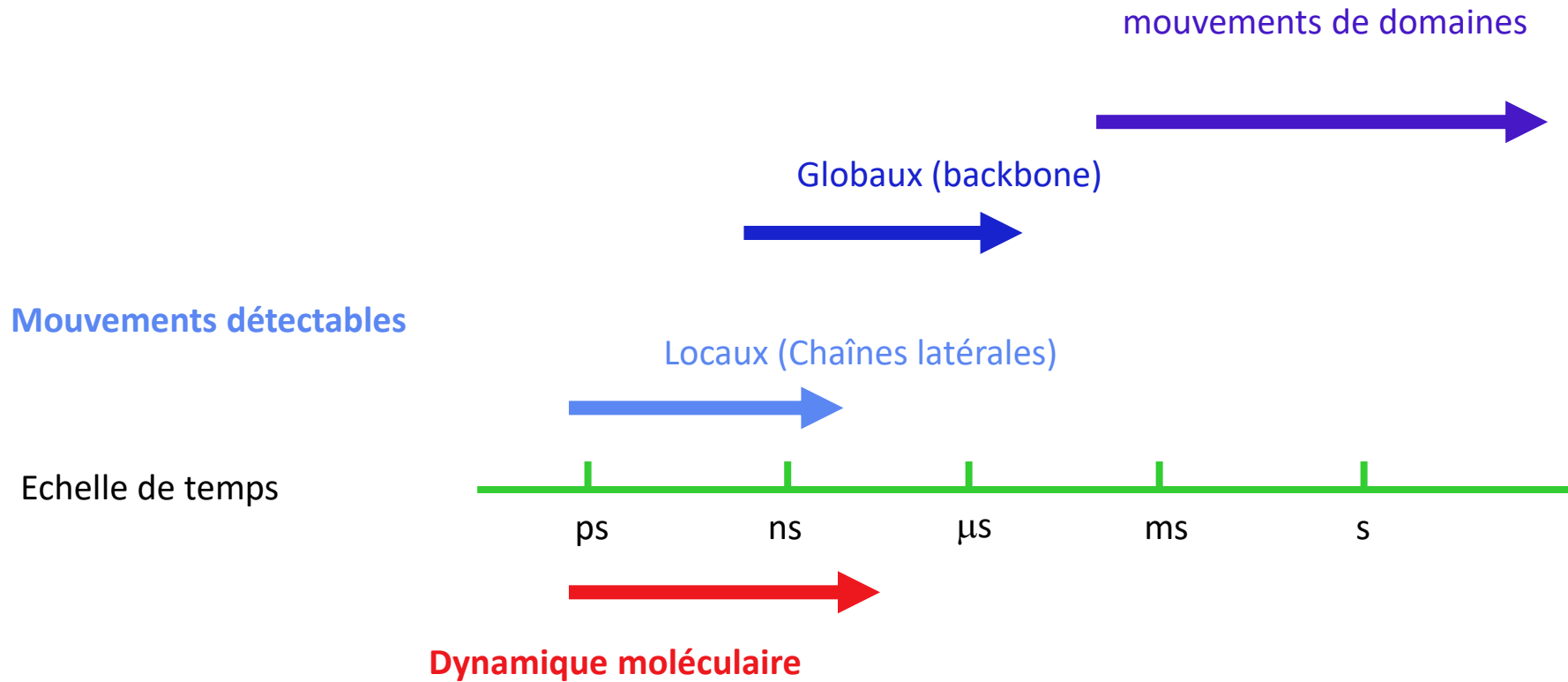
$E_{\text{Empirique}} = \text{Energie interne} + \text{Energie non liée}$

- **RMN**

- Fortes contraintes expérimentales de distances et angles dièdres
- Dynamique forte
- Forte température d'emblée (1000 ou 2000 K)
- Génération de nombreuses structures
- Champ de force adapté
 - Renforcement des constantes k_b et k_F)
 - Conservation de la planéité de la liaison peptidique
- Exploration forte de l'espace des conformations

- **Optimisation de structures sans contraintes**

- Pas de contraintes expérimentales
- Choix d'un champ de force réaliste
- Immersion du modèle dans un solvant
- Dynamique douce
- Incréments progressifs de température (10 ou 20 K)
- Exploration faible de l'espace des conformations



Modélisation moléculaire



La modélisation intervient aussi dans les déterminations expérimentales

- ✓ Seulement on détecte pour 20% des protéines un homologue évident dans la PDB or pour ~ 70% des protéines \exists une structure empreinte permettant de modéliser
- ✓ Prédire les fonctions de protéines ayant des faibles degrés de similarité
- ✓ Identifier des protéines qui ont un nouveau repliement "new folds". Une chaîne est considérée comme repliement déjà connu ssi:

1. z-score ≥ 4.0

$$z = \frac{x - \mu}{\sigma}$$

x = valeur mesurée

μ = moyenne de la population

σ = déviation standard

2. RMSD $< 3.0\text{\AA}$
3. Nombre de positions alignées $\geq 70\%$ de la longueur de la chaîne.

- ✓ Estimation du nombre total de folds ~ 4,000
- ✓ Une base de données de 1100 folds couvre 90% des familles de protéines

Govindarajan S., Recabarren R., & Goldstein R.A. 1999 Proteins: Structure, Function, and Genetics 35:408-414

✓ Modélisation par homologie

✓ Standard (Id>30%)

- ✓ 2 protéines qui ont plus de 30% d'identité de séquences ont 80% de leurs C α superposables avec un écart quadratique moyen de 1 Å (RMSD=1Å)

✓ A faible taux d'identité (Id<30%)

- ✓ 2 protéines qui ont une topologie « probablement » identique en dépit de l'absence de similarité importante (arguments expérimentaux ou de structures secondaires).

✓ Threading (Identification de repliement)

- ✓ Une séquence est testée sur une librairie de repliements pour déterminer sa compatibilité structure-séquence probable, méthode d'alignement séquence-structures tridimensionnelles.

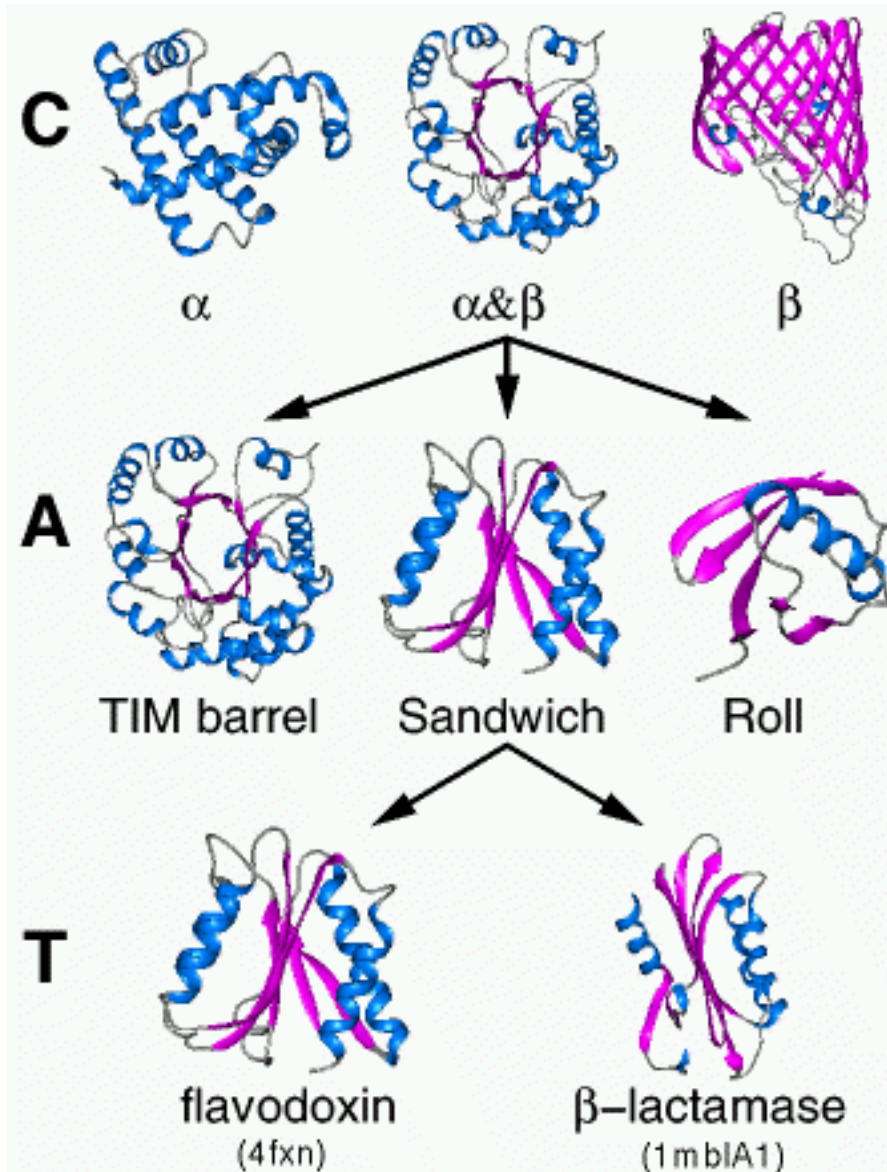
✓ Alphabets structuraux (légo moléculaire)

- ✓ Assemblage de morceaux et reconstitution de structures à partir de fragments.

✓ *Ab initio* (en progression dans CASP, folding@home)

- ✓ Structure directement déduite de la séquence à partir de règles empiriques.

Les méthodes de montage ou enfilage des repliements (threading)

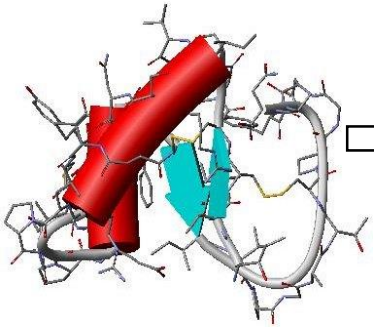


Le nombre de repliements différents estimé est de plusieurs ordres de grandeur en dessous du nombre de familles de séquences différentes (dégénérescence structure-séquence).

Les méthodes de threading (Smith et al., 1997) visent donc à calculer l'adéquation entre une séquence de structure inconnue et une structure 3D connue (ou parties de celle-ci). La séquence est « enfilée » sur chaque repliement d'une librairie de coeurs représentatifs de la PDB (~1500 coeurs). La compatibilité de la séquence avec un repliement particulier peut se mesurer grâce à un potentiel empirique : potentiels statistiques de distances entre résidus, précalculés sur la PDB (Sippl, 1990) ou scores de paires de résidus étant donnée la structure secondaire où se trouvent ceux-ci.

Les techniques de "threading" ("enfilage") sont basées sur la méthode suivante:

1. une librairie de repliements protéiques non redondants ("coeurs") est constituée à partir de la banque de données structurale des protéines (PDB Brookhaven). Chaque repliement est une chaîne tridimensionnelle, la séquence étant complètement oubliée.
2. La séquence « test » est ensuite ajustée de manière optimale à chaque repliement de la librairie (insertions-délétions permises dans les boucles), l' "énergie" de chaque ajustement ("montage") est calculée en sommant les "interactions" des résidus deux à deux. Les repliements de la librairie sont ensuite rangés par ordre d'énergie, le plus probable pour la séquence étant celui d'énergie la plus basse.



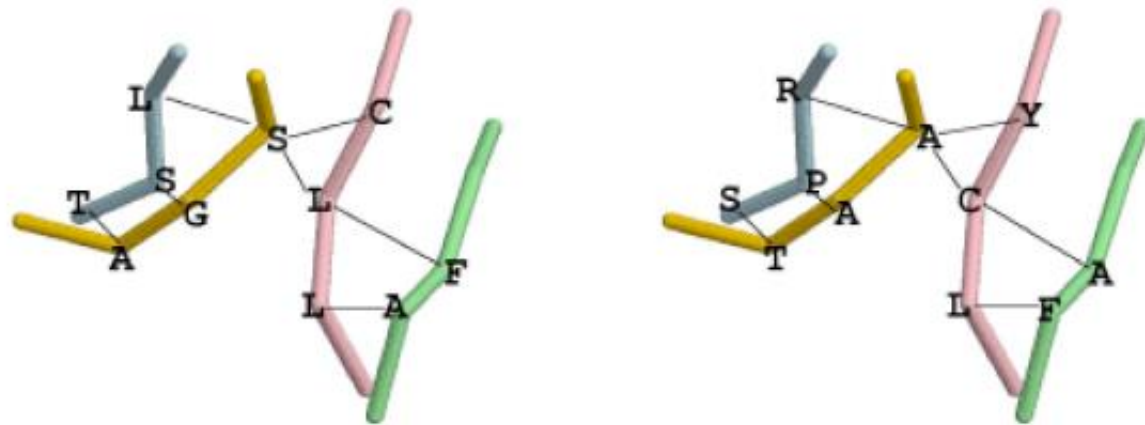
C	C	H	H	C	C	C	E	E	C	C	C	C	H: Helix, E: extended C: Coil
e	e	e	b	b	e	b	b	b	b	e	e	e	e exposé b buried ou enfoui
Ce	Ce	He	Hb	Cb	Ce	Cb	Cb	Eb	Cb	Ce	Ce	Ce	
M	F	T	V	N	E	H	I	D	R	L	Y	A	



Optimisation du score de correspondance par glissement et insertions



T A A G S P R T S L L L A F A L L C Y
T A A G S P R T S L L L A F A L L C Y





Séquence :

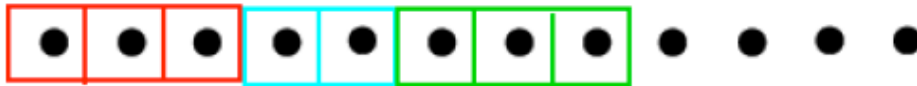
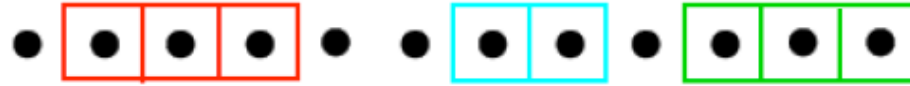
Taille : $L_S = 12$



Cœur :

Nbre d'éléments : $M = 3$

Taille : $L_C = 8$



Degrés de liberté :

$$\tilde{n} = L_S - L_C + 1 = 5$$



Nombre d'alignements : $\binom{M+\tilde{n}-1}{M}$

$$\binom{7}{3} = 35$$



Avec $M=15$ et $\tilde{n}=100$ $\binom{114}{15}$

$$\sim 2.10^{18}$$

3DPSSM : <http://www.sbg.bio.ic.ac.uk/~3dpssm/>

bioinbgu : <http://www.cs.bgu.ac.il/~bioinbgu/form.html>

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

FROST : <http://www-mig.jouy.inra.fr/frost/>

Un alphabet structural est une série (ou librairie) de **petits prototypes** qui approximent **chaque partie** des structures protéiques.

Ils sont composés d'un **nombre limité** d'éléments structuraux **récurrents** des structures protéiques.

Les associations entre ces **"lettres" structurales** sont gouvernées par des règles logiques et forment des mots (de structures protéiques).

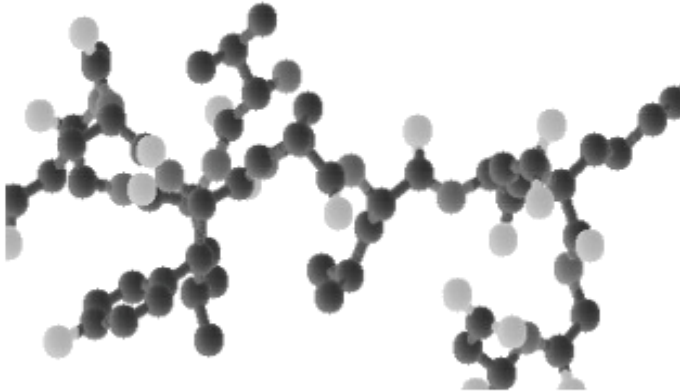
Un alphabet structural n'a pas d'*a priori* vis-à-vis des structures secondaires, *i.e.* **ce n'est pas une catégorisation des boucles.**

Idée: Reconstruire les structures 3D à partir des blocs

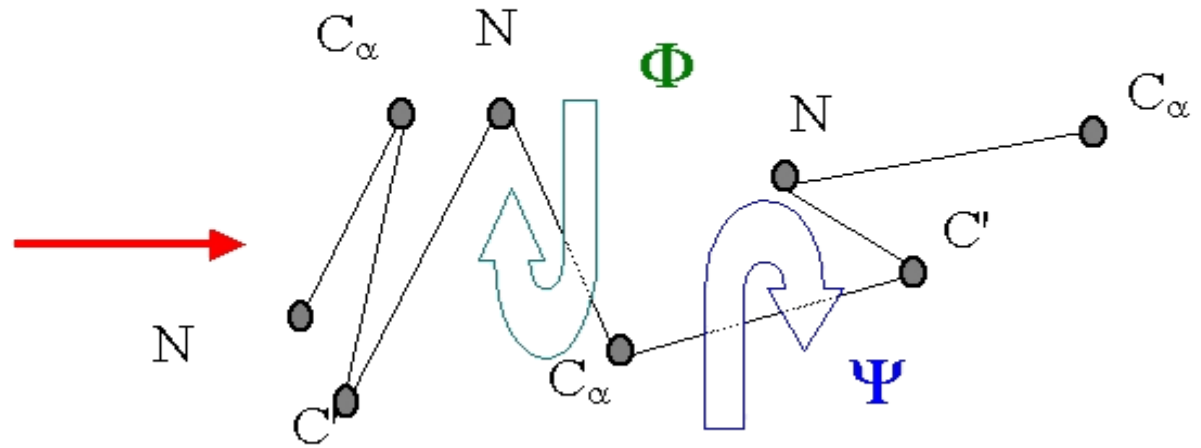
de Brevern A.G., Camproux A.C., Hazout S., Etchebest C., and Tuffery P. (2001), *Protein structural alphabets: beyond the secondary structure description*, Recent Adv. In Prot. Eng., 1:319-331 .

Information 3D.

3D



dihedral angles



5 résidus \Rightarrow 8 angles dièdres (ϕ , ψ)

Combien de blocs?

Nombre de
prototypes

Approximation de
la structure 3D
locale

Prédiction

Important

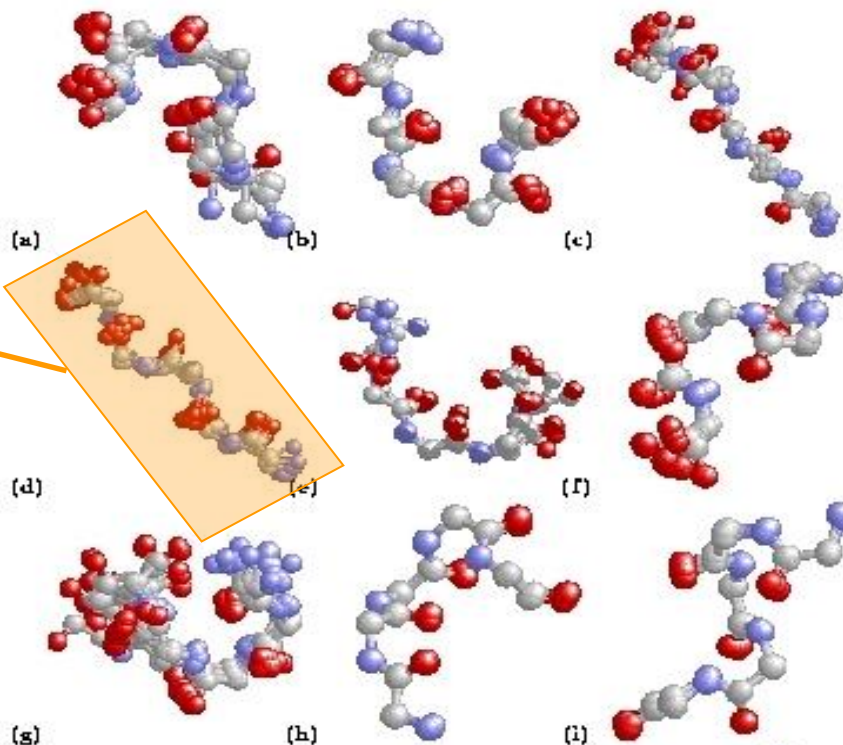
Equipe	Année	Nombre	Longueur
Unger	1989 / 93	103 / 83	6
Pretrelski	1992	113	8
Schuchhardt	1996	100	9

Limité

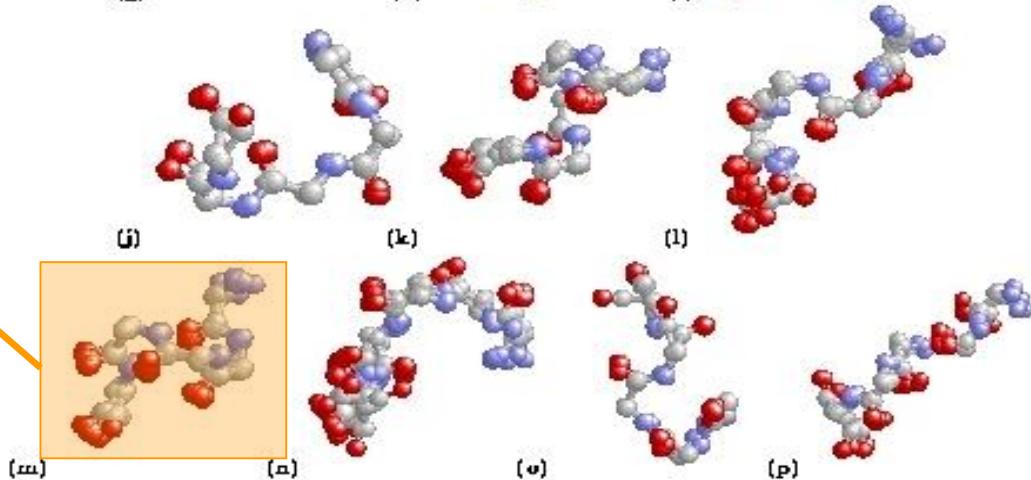
Equipe	Année	Nombre	Longueur
Rooman	1990	4	4, 5, 6 & 7
Fetrow	1993 / 97	6	7
Bystroff	1998 / 2000	13 / 16	5 to 17 (8)
Camproux	1999 / 2004	12 / 27	4
de Brevern	2000	16	5
Hunter	2003	28	7

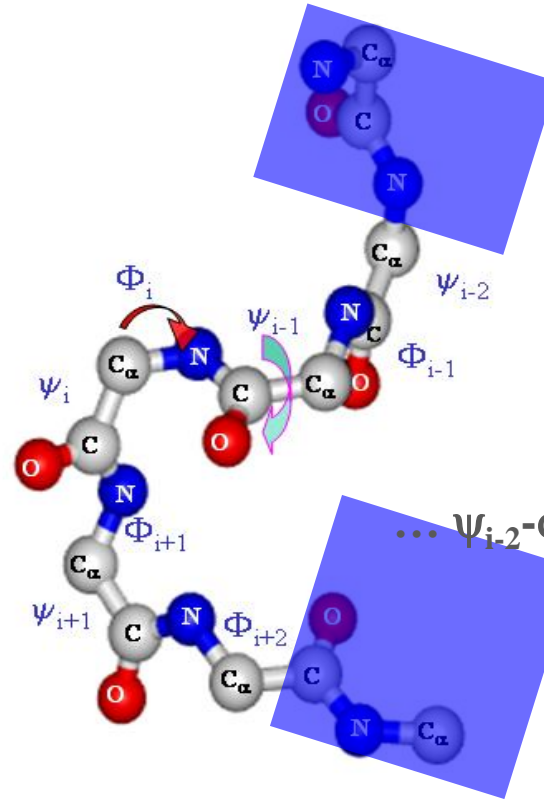
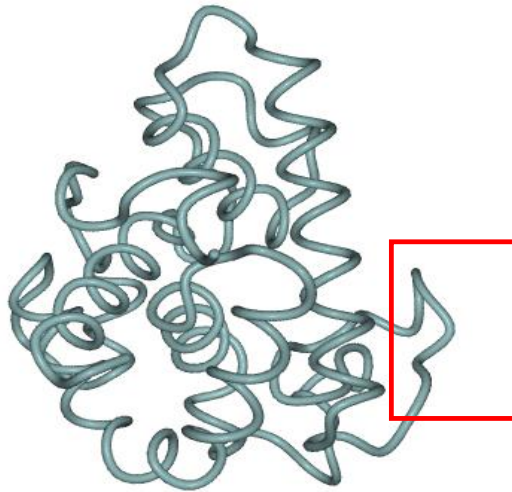
Les Blocs Protéiques (A. De Brevern)

BP *d*
Brin étendu



BP *m*
Hélice






fragment *f*
(5 aa, 8 angles dièdres)

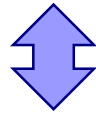
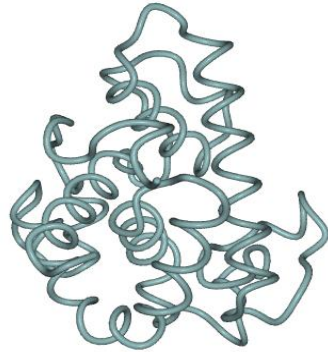
... ψ_{i-2}-φ_{i-1}-ψ_{i-1}-φ_i-ψ_i-φ_{i+1}-ψ_{i+1}-φ_{i+2} ...

fragment f
(5 aa, 8 angles)

	ψ_{i-2}	ϕ_{i-1}	ψ_{i-1}	ϕ_i	ψ_i	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}
PB <i>a</i>	41.14	75.53	13.92	-99.80	131.88	-96.27	122.08	-99.68
PB <i>b</i>	108.24	-90.12	119.54	-92.21	-18.06	-128.93	147.04	-99.90
PB <i>c</i>	-11.61	-105.66	94.81	-106.09	133.56	-106.93	135.97	-100.63
PB <i>d</i>	141.98	-112.79	132.20	-114.79	140.11	-111.05	139.54	-103.16
PB <i>e</i>	133.25	-112.37	137.64	-108.13	133.00	-87.30	120.54	77.40
PB <i>f</i>	116.40	-105.53	129.32	-96.68	140.72	-74.19	-26.65	-94.51
PB <i>g</i>	0.40	-81.83	4.91	-100.59	85.50	-71.65	130.78	84.98
PB <i>h</i>	119.14	-102.58	130.83	-67.91	121.55	76.25	-2.95	-90.88
 PB <i>i</i>	130.68	-56.92	119.26	77.85	10.42	-99.43	141.40	-98.01
PB <i>j</i>	114.32	-121.47	118.14	82.88	-150.05	-83.81	23.35	-85.82
PB <i>k</i>	117.16	-95.41	140.40	-59.35	-29.23	-72.39	-25.08	-76.16
PB <i>l</i>	139.20	-55.96	-32.70	-68.51	-26.09	-74.44	-22.60	-71.74
PB <i>m</i>	-39.62	-64.73	-39.52	-65.54	-38.88	-66.89	-37.76	-70.19
PB <i>n</i>	-35.34	-65.03	-38.12	-66.34	-29.51	-89.10	-2.91	77.90
PB <i>o</i>	-45.29	-67.44	-27.72	-87.27	5.13	77.49	30.71	-93.23
PB <i>p</i>	-27.09	-86.14	0.30	59.85	21.51	-96.30	132.67	-92.91

Calcul de distance => 16 scores

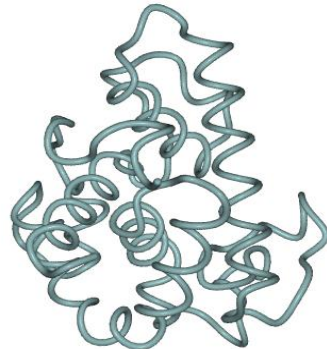
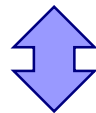
Le plus proche



Codage réversible de la structure 3D de la Myoglobin
en 1 mot

>153L

```
ZZmnopfklpccebjafklmmnopabecjklmmmmmmmm  
mmmmmmmmmmmmmmmmnopa fklmmmmmmmmnooolapgeh  
iafkopagcjkopafklmccehj fklmklmmmmmmmmmmmm  
mmmmmmmmmbcfklmmmmmmmmmmnomklmnb fklmmgoia  
hilmmmmmmmmmmmmmmmmnoZZ
```



BPs : fréquence → PB *d* (19%) et PB *m* (30%), ensuite > 6%

transition → nombre limité de transition d'un BP à un autre

approximation → 0.41 Å (médiane = 0.26 Å)

forte discrimination entre BPs (le second plus proche est ... loin)

Permet la comparaison des structures 3D des protéines avec des outils de séquences

de Brevern A.G. (2005), *New assessment of a structural alphabet*, In *Silico Biology*, 5, 26.

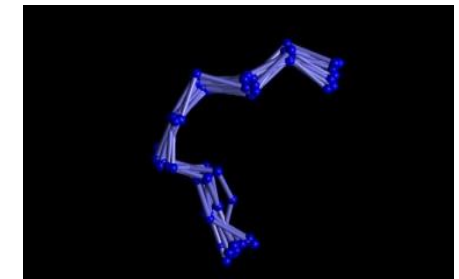
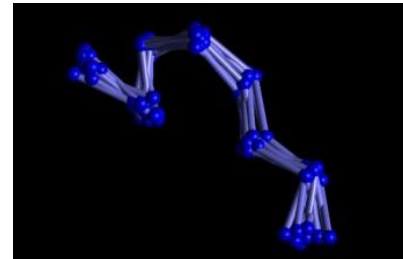
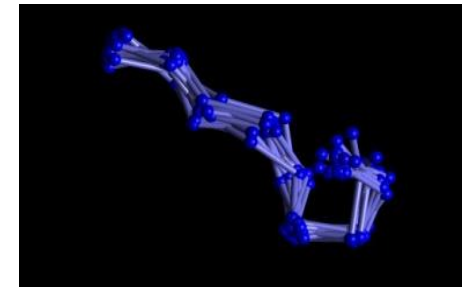
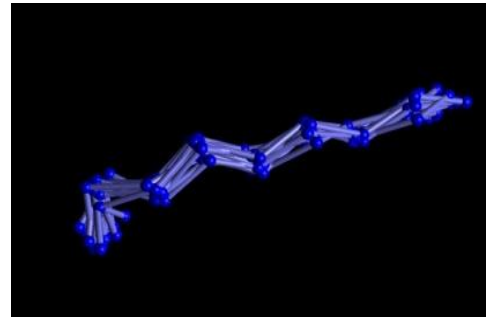
1 Mot Structural (MS) = série de 5 PBs consécutifs (les + fréquents)

1MS définit en fait une forme de grammaire, de successions très fréquentes.

5 C α \rightarrow 9 C α

72 MSs les + fréquents \rightarrow 0.9 Å

mmmnp
mmnopa
mnopac
opacd

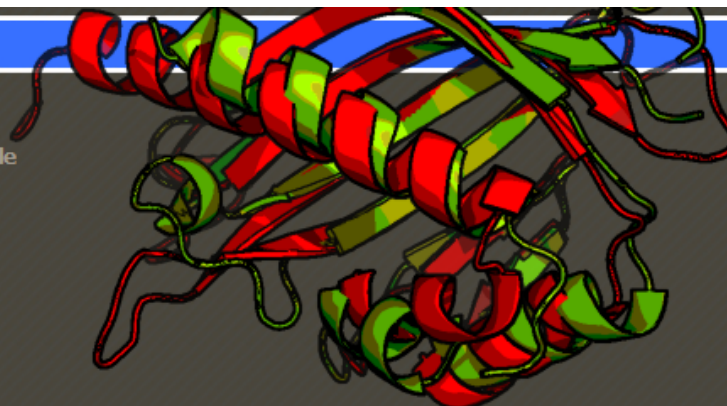


Recouvrement permet d'avoir des chemins
(création d'un graphe)

mulPBA

[HOME](#) [ABOUT](#) [EXAMPLE](#) [CONTACT](#)

Welcome to MulPBA webservice. A powerful method for multiple structural alignment based on a structural alphabet.



http://www.dsimb.inserm.fr/dsimb_tools/mulpba/

Introduction

mulPBA is a tool for comparison of protein structures based on similarity in the local backbone conformation. The local backbone conformation is defined as pentapeptide dihedrals, using Protein Blocks (PBs)[de Brevern et al. 2000, Joseph et al. 2010]. The protein structures represented as PB sequences, are aligned by dynamic programming scored by a PB substitution matrix [Joseph et al. 2011]. A progressive alignment strategy similar to CLUSTALW was adopted for multiple PB sequence alignment. Highly similar stretches identified by the pairwise alignments are given higher weights during the progressive alignment. Encoding the information on backbone conformation as PB sequences, enables 'sequence-like' structural alignment, which is minimally influenced by structural flexibility (see examples). The residue equivalences from PB based alignments are identified to obtain a three dimensional fit of the structures, followed by an iterative refinement of the structural superposition.

Compare and align multiple protein structure - PDB data

Use this form to analyse chain from Protein Data Bank files. In order to make multiple protein analysis a minimum of 3 protein are required. The files can also be uploaded see bellow.

Protein Data Bank uploader

1N1BA
1N20A
1N23A
1N1ZA

Address of the results page:

http://www.dsimb.inserm.fr/dsimb_tools/mulpba/TMP/mulpba.20130219145135_92882/results.html

Your analysis ID is : 20130219145135_92882

Summary

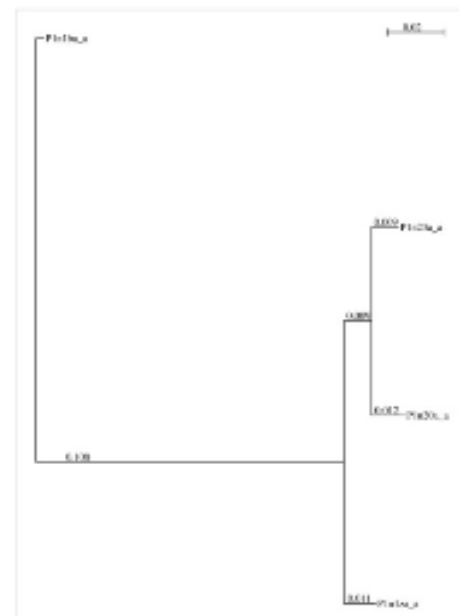
Color code for Quality of alignment



Score	7.550972
Nrms	98.449615
GDT score	98.158913
RMSD of core	0.349511
N 3.5A	519.500000
30% gap	98

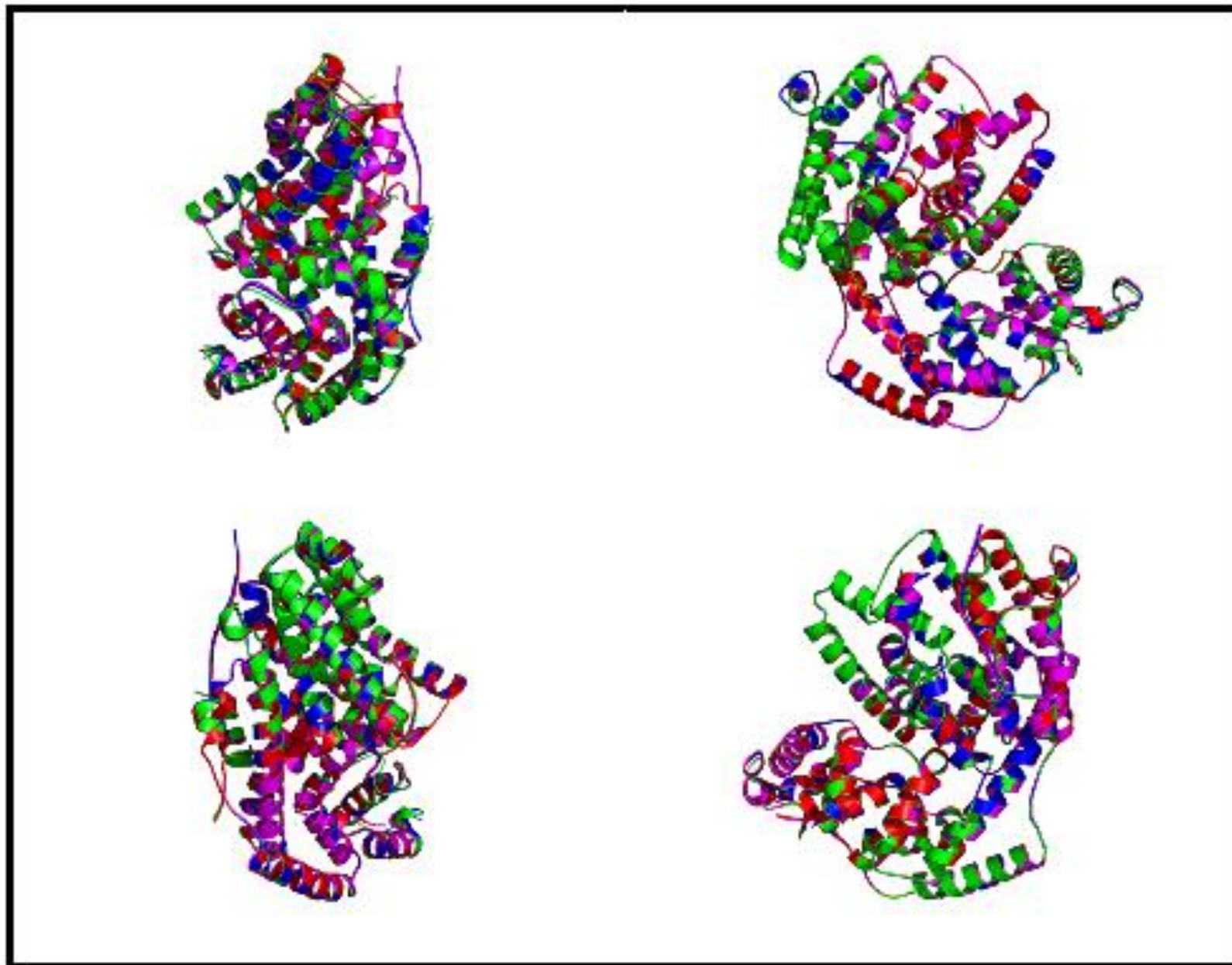


Dendrogram



Alignment

Position			64
1n1b a	-----	LWDSNYIQSLNTPYTEERHLDRKAELIVQVRILLKEKMEPVQQLELIHDLKYLGLSDFQDEI	KEIL
PB	-----	z z f k l m m m p c c f k l b f k l m m m m m m m m m m m m m m m m p c c d f k l m m m m m m m m n o p r m m m m m m m m m m m m m m m	
Position			54
1n20 a	IRRSGNYQPALWDSNYIQSLNTPYTEERHLDRKAELIVQVRILLKEKMEPVQQLELIHDLKYLGLSDFQDEI	KEIL	
PB		z z c e h i a c d f b d f k l m m m p c c f k l b f k l m m m m m m m m m m m m m m m m g c c d f k l m m m m m m m m n o p r m m m m m m m m m m m m m m m	
Position			55
1n23 a	-RRSGNYQPALWDSNYIQSLNTPYTEERHLDRKAELIVQVRILLKEKMEPVQQLELIHDLKYLGLSDFQDEI	KEIL	
PB		- z z e h i a d d f b d f k l m m m p c c f k l b f k l m p g c d f k l m m m m m m m m n o p r m m m m m m m m m m m m m m m	
Position			54
1n1z a	IRRSGNYQPALWDSNYIQSLNTPYTEERHLDRKAELIVQVRILLKEKMEPVQQLELIHDLKYLGLSDFQDEI	KEIL	
PB		z z c e h i a d d f b d f k l m m m p c c f k l b f k l m g c c d f k l m m m m m m m m n o p r m m m m m m m m m m m m m m m	



Robetta: full-chain protein structure prediction server - Mozilla Firefox

Robetta: full-chain protein structure predicti... +

<http://rosetta.bakerlab.org/>

Full-chain Protein Structure Prediction Server

Simons et al. (1997) J Mol Biol. 268:209-225

Model 1 **Target – T0513**

2.66 Å over 62 residues

0.84 Å over 39 residues

de novo prediction by Robetta in CASP-8

REGISTRATION
[[Register / Update](#)] [[Login](#)]

DOCUMENTATION
[[Docs / FAQs](#)]

SERVICES
Domain Parsing & 3-D Modeling
[[Queue](#)] [[Submit](#)]

Interface Alanine Scanning
[[Queue](#)] [[Submit](#)]

Fragment Libraries
[[Queue](#)] [[Submit](#)]

DNA Interface Residue Scanning
[[Queue](#)] [[Submit](#)]

RELATED SITES
[RosettaBackrub Server](#)
[RosettaAntibody Server](#)
[RosettaDesign Server](#)
[RosettaDock Server](#)
[Rosetta Commons](#)
[Foldit](#)
[Rosetta@home](#)
[Human Proteome Folding Project](#)

Robetta hardware generously provided by

Robetta uses the [Rosetta](#) software package ([licensing information](#))



David Baker

Robetta Server - Mozilla Firefox

Eichier Édition Affichage Historique Marque-pages Outils ?

Robetta Server +

http://robetta.bakerlab.org/submit.jsp

www.bakerlab.org

ROBETTA BETA

Full-chain Protein Structure Prediction Server

Structure Prediction Fragment Libraries Alanine Scanning DNA Interface Scan

[Queue] [Submit] [Queue] [Submit] [Queue] [Submit] [Queue] [Submit]

[Register / Update] [Docs / FAQs] [Login]

Submit a job to the Server

Required

Prediction Type: Ginzu : Domain Prediction
3-D Model : Full Prediction (available after Ginzu completes)

[Registered Username:](#) or [Registered Email Address:](#)

Target Name:

Paste [Fasta](#) [TRANSLATE DNA TO AA](#)

or Upload [Fasta:](#)

Do not warn me if my sequence matches one already submitted

Note: please do not submit known PDB sequences or CASP targets intentionally

Optional

[Reply Email:](#)

For De Novo Models

[Exclude Homologs:](#) (for testing)

[Rosetta NMR](#) (click links below for input format)

[Chemical Shifts:](#)

[NOE Constraints:](#)

[Dipolar Constraints:](#)



Structure Prediction Fragment Libraries Alanine Scanning DNA Interface Scan

[Queue] [Submit] [Queue] [Submit] [Queue] [Submit] [Queue] [Submit]
 [Register / Update] [Docs / FAQs] [Login]

Robetta Queue

19 Job(s) Queued Status: Sort By:

Username: ID: Target: Notes: Host:

Results Per Page:

Page 1 of 17

Jobs may be removed one week after they complete to conserve disk space.

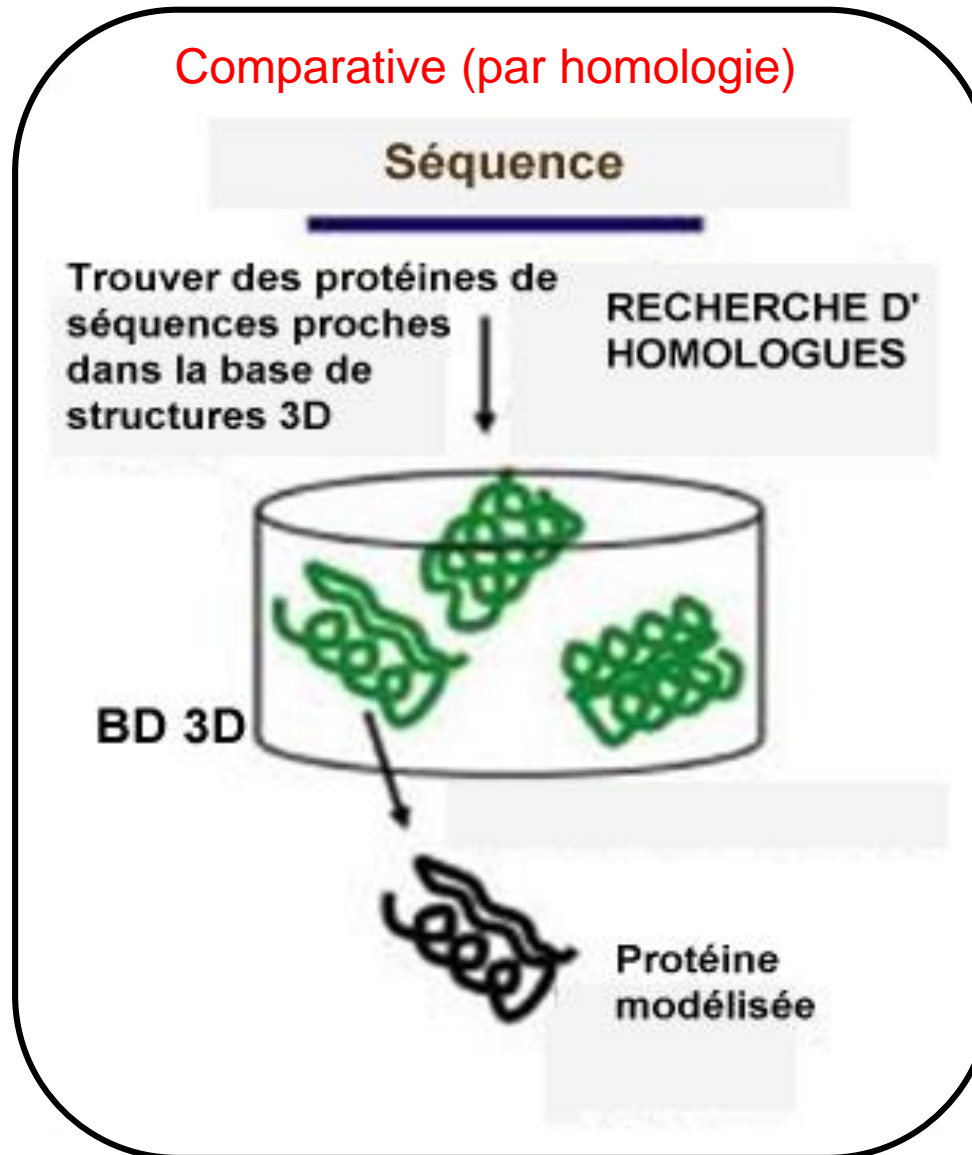
To conserve disk space, please remove your jobs after they complete by following these [instructions](#).

*ETC is the Estimated Time of Completion in days. These values may increase when higher priority users submit jobs.

1 2 3 4 5 6 7 8 9 10 11 ...

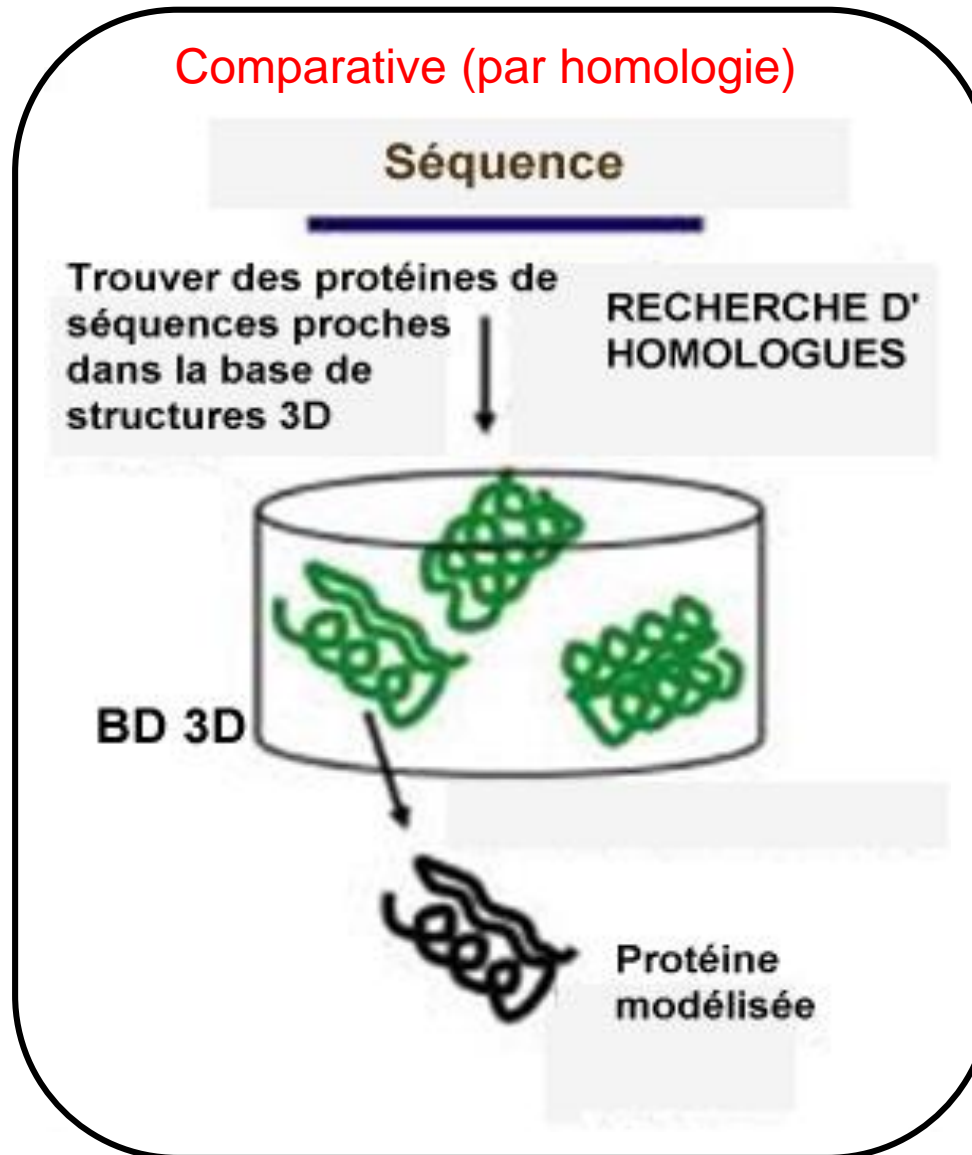
ID	Status (*ETC)	Method	Username	Target	Length	Domain Prediction	Host
21768	Active (1)	Ginzu	roshnipanda	mouseahr230-421	192	--	vit.x.x
21767	Active (0)	Ginzu	Tanya	interleukin35	372	--	141.250.x.x
21766	Complete	Ginzu	Ecilalnir	Medium.AC.US	364	1:blast:1f88A:364	generio-nat1.x.x
21765	Active (15)	Full	uguruzuner	harzianum	182	1:blast:1xvxA:182	r74-192-215-111.bostonta02.clstb.tl.dh.x.x
21764	Active (15)	Full	Olivetree	SNAP25a	206	1:blast:1sfcC:107 2:blast:1sfcD:99	h0e04cd77ec0.dhcp-vit.x.x
21763	Active (14)	Full	Olivetree	Snap25b	206	1:blast:1sfcC:112 2:blast:1sfcD:94	h0e04cd77ec0.dhcp-vit.x.x
21762	Complete	Ginzu	rgv77	VDAC	283	1:blast:2k4tA:283	132.248.x.x
21761	Complete	Ginzu	ribotide	D90 YR1	648	1:blast:1zvzA:648	bbennett.PHRM.x.x
21760	Active (14)	Full	odehir	BALB-delta5	396	1:blast:3ir2A:198 2:blast:2jyWA:198	p1138-ipbf405sasajima.aichi.on.x.x
21759	Complete	Ginzu	conqhv	CA-SP1	245	1:blast:3h4eA:245	137.216.x.x
21758	Complete	Ginzu	Breeze	CT	273	1:msa:173 2:msa:100	ps15410powers.opmc.x.x
21757	Active (13)	Full	rgonzalez	alpha2	650	1:outpref:75:a 2:pfam:PF08399:130:a/b 3:hsearch:3ibsA:200 4:psiblast:3li8A:181 5:outpref:64:a	148.247.x.x
21756	Active (11)	Full	Dglab	HisRS domain	528	1:psiblast:2j3iA:528	borcim103.x.x
21755	Active (9)	Full	Adelef	cTerminal 3522	120	1:pfam:PF06990:120:a/b	200-89-68-78.x.x
21754	Complete	Ginzu	Brub	hgrf	75	1:outpref:75	n105057.x.x

Comparative (par homologie)



2 protéines homologues (même ancêtre commun),
détectées par l'identité de séquences,
partage le même repliement (structure 3D proche)

Comparative (par homologie)



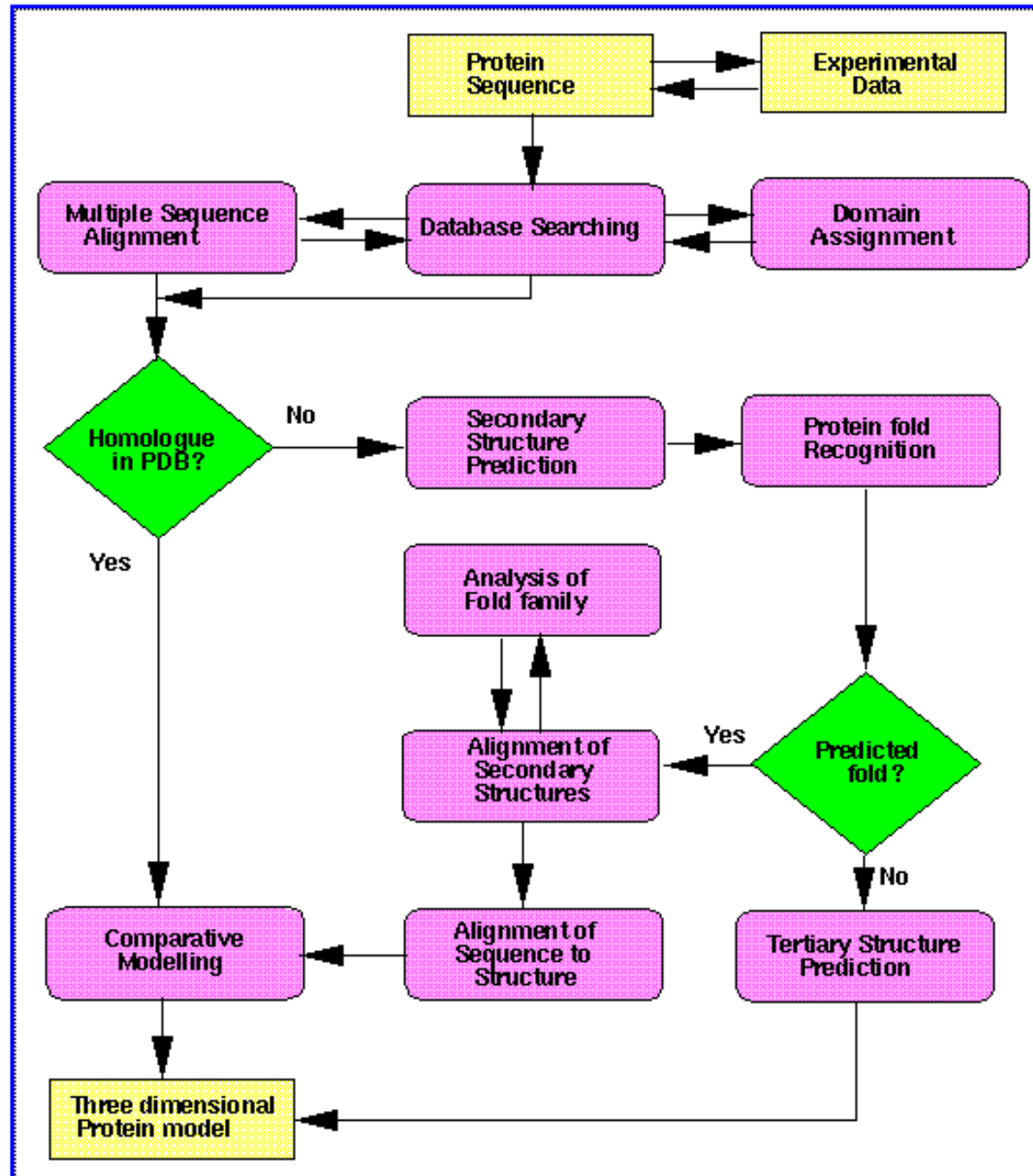
2 protéines homologues (même ancêtre commun),
détectées par l'identité de séquences,
partage le même repliement (structure 3D proche)

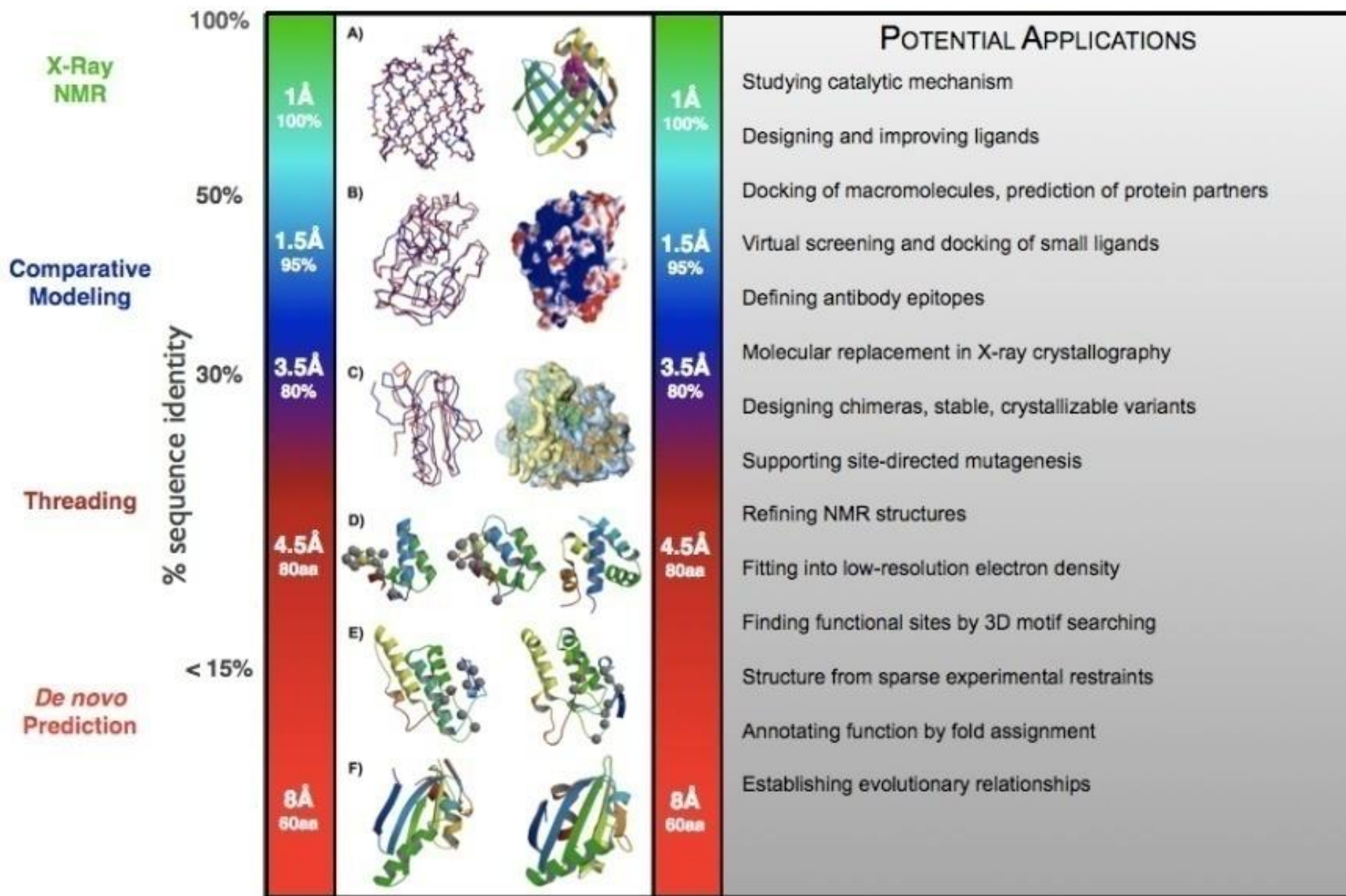
- **Modélisation homologique**

- 2 protéines qui ont plus de 30% d'identité de séquences ont 80% de leurs Ca superposables avec un écart quadratique moyen de 1 Å (RMSD=1Å)
- =>topologie architecture identique
- Automatique si > 60% d'identité
- elle si > 30 et < 60% similarité
- Non raisonnable si < 30% similarité

- **Modélisation par analogie**

- 2 protéines qui ont la même fonction et une topologie « probablement » identique en dépit de l'absence de similarité importante (arguments expérimentaux ou de structures secondaires).
- Exemple du motif A de la glycoprotéine P
- Exemple des motifs structuraux de la protéine gp41 du VIH



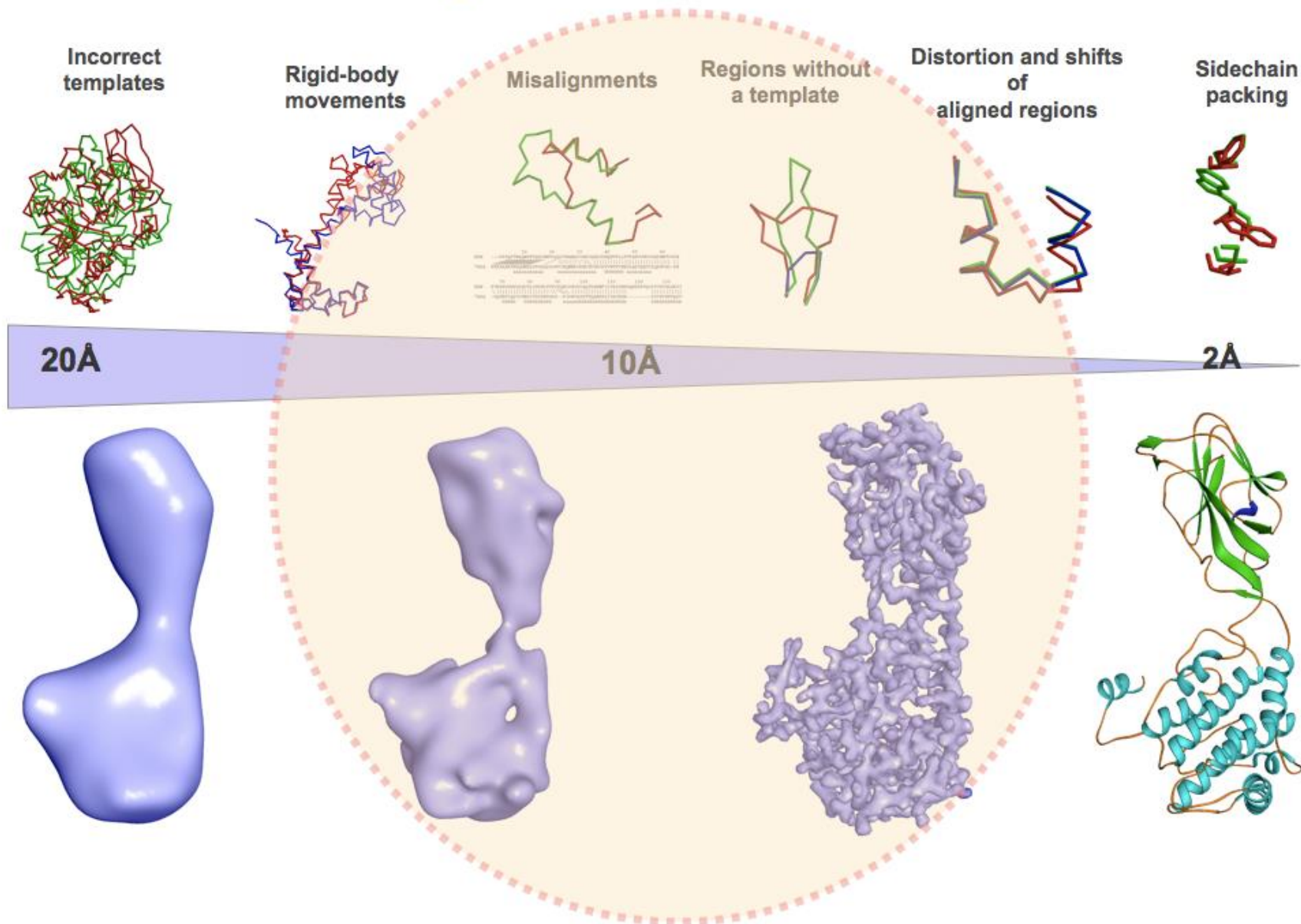


- ✓ **15% < IS < 25% : Threading**
- ✓ **25% < IS < 35%** : Conception de mutants ou de variants pour cristallogénèse -
Modèle par homologie à environ 3.5 Å - 80% d'exactitude stérique
- ✓ **35% < IS < 50%** : Remplacement moléculaire en cristallographie - Définition d'épitopes - Criblage virtuel - Docking de petites molécules -
Modèle par homologie à environ 1.5 Å - 95% d'exactitude stérique
- ✓ **50% < IS < 90%** : Docking de macromolécules - Prédiction de partenaires protéiques - Conception et amélioration de ligands -
Modèle par homologie à environ 1 Å - 95-100% d'exactitude stérique



Un modèle est faux par définition l'important est qu'il soit utile

Errors in comparative models vs. resolution



- L'alignement de 2 séquences aléatoires de longueur 300 peut produire jusqu'à ~20% identités.

- En dessous de 20% d'identité un alignement peut paraître plausible mais ne peut pas être validé par les méthodes conventionnelles. Twilight zone

PERCENT
IDENTITY

100

ALIGNMENT
METHODS

90

80

Automatic pairwise
methods

70

60

50

Consensus methods

40

30

Profile methods

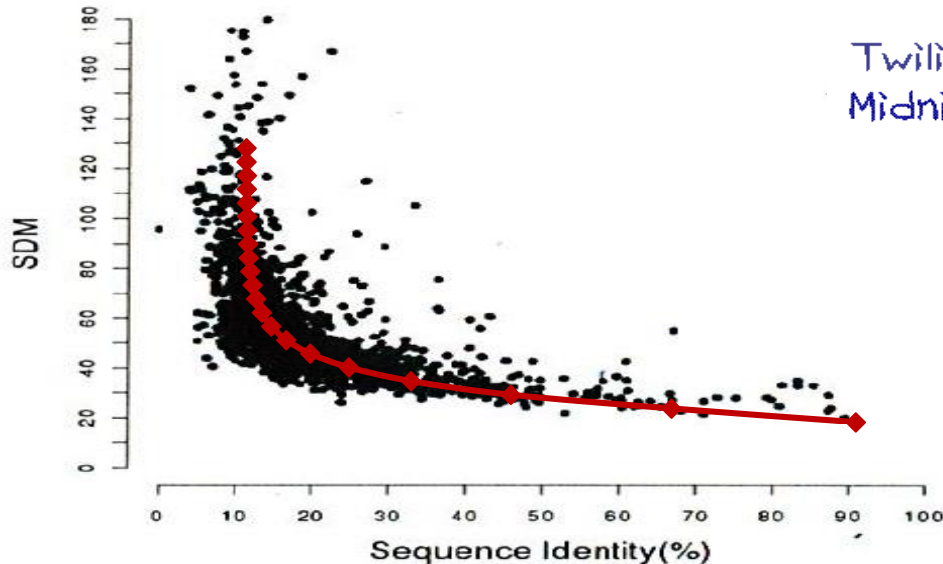
20

Twilight zone
Midnight zone

10

Structure prediction

0



$$S = 12,3 + 98,4e^{(-0.95\Delta)}$$

$$a^2 + b^2 = c^2$$

Alignement
des
Séquences
33 identités
(21% id)

Alignement
des
structures 3D
24 identités
(15% Id)

```

ANTHEPROT 2000 V6.0 by G. Deléage (g.deleage@ibcp.fr), IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE
Alignment : D:\Users\Gilbert\Documents\Enseignement\Bioinformatique\2010-2011\1HBG_1GDI.mus using Muscle
Matrix
Amino acids identity 100% >= 75 <= 50 < 50
Gap opening penalty 10 Gap extension penalty .05 Number of perfect matches 33

      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
1 1HBG:A|PD G-LSAAQRQVIAATWKDIAGADNGAGVGKKCLIKFLSAHPQMAAVFGF-SGASD-----PGVAALGAKVL 63
2 1GDI:A|PD GALTESQAALVKSSWEEFN--NIPKHTHRFFILVLEIAPAAKDLFSFLKGTSEVPQNNPELQAHAGKVF 68
3Consensus * * : * :: :*::: . * . :: :* . * * :*: .*::: * : * .*:::

      80      90      100     110     120     130     140
      |      |      |      |      |      |      |
1 1HBG:A|PD AQIGVAVSHLGDEGKMV--AQMKAVGVRHKGYGKHKIKAQYFEP LGASLLSAMEHRIGGKMNAAAKDAWA 131
2 1GDI:A|PD KLVYEAAIQLEVTGVVVTDATLKNLGSVHVSKG---VADAHFPVVKEAILKTIKEVVGAKWSEELNSAWT 135
3Consensus : * . :* * :* * :* :* * . * : :* : :*:::: .*: . : :*:::

      150     160     170     180     190     200     210
      |      |      |      |      |      |      |
1 1HBG:A|PD AAYADISGALISGLQS--
2 1GDI:A|PD IAYDELAIVIKKEMDDAA
3Consensus ** ::: .: . :.
    
```

```

ANTHEPROT 2000 V6.0 by G. Deléage (g.deleage@ibcp.fr), IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE
Alignment : D:\ANTHEPRO\tempali.aln using Clustalw
Matrix BLOSUM
Amino acids identity 100% >= 75 <= 50 < 50
Gap opening penalty 10 Gap extension penalty .05 Number of perfect matches 24

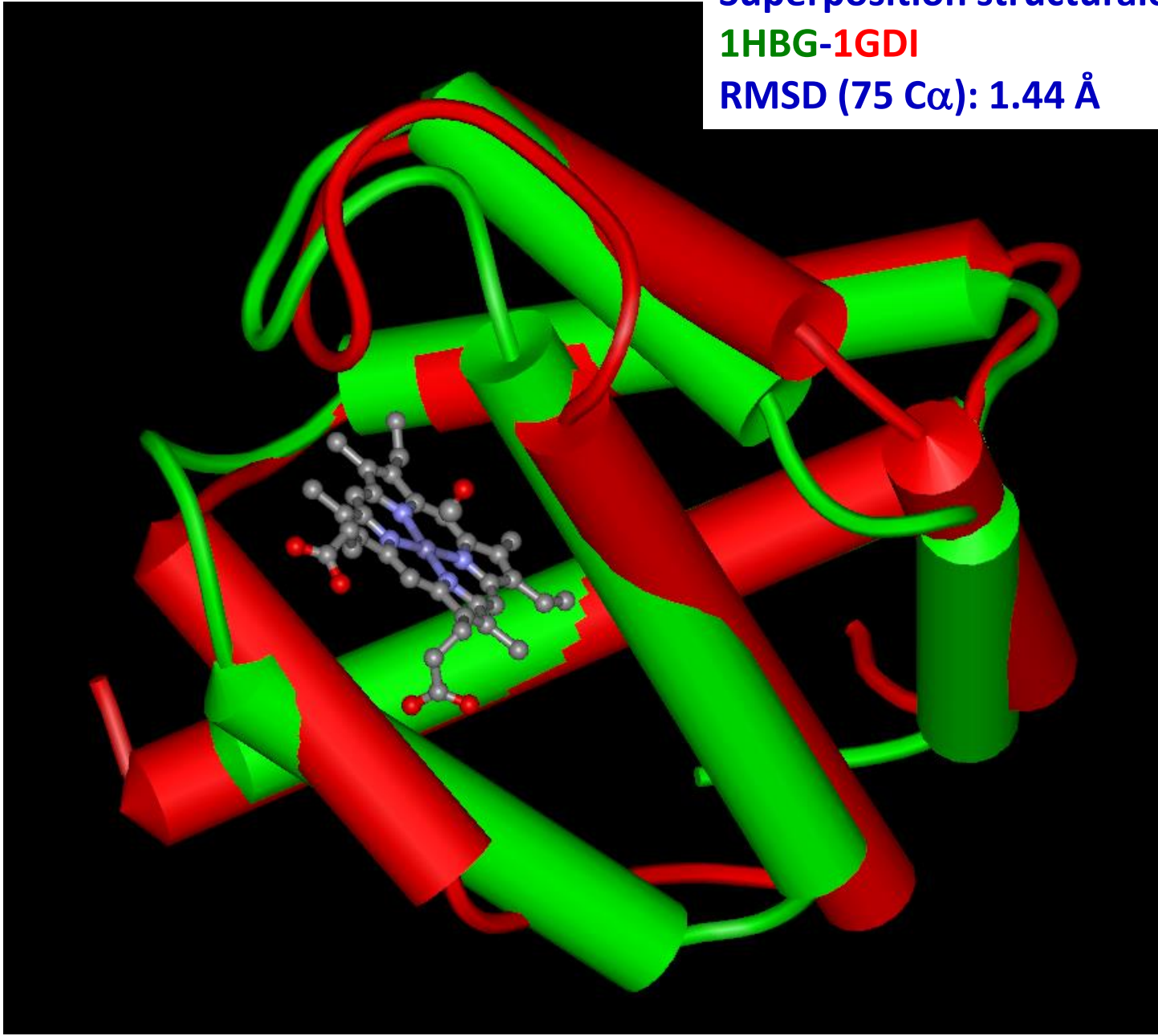
      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
1 1HBG_A|PD -GLSAAQRQVIAATWKDIAGADNGAGVGKKCLIKFLSAHPQMAAVFG--FSGASDP----GVAALGAKVL 63
2 1GDI_A|PD GALTESQAALVKSSWEEFN--ANIPKHTHRFFILVLEIAPAAKDLFSFLKGTSEVPQNNPELQAHAGKVF 68
3Consensus .*: * : : :*::: * . :: :* . * * :*: .*: . . . * : * .*:::

      80      90      100     110     120     130     140
      |      |      |      |      |      |      |
1 1HBG_A|PD AQIGV-AVSHLGDEGKMVAQMKAVGVRHKGY-GNKHKAQYFEP LGASLLSAMEHRIGGKMNAAAKDAWA 131
2 1GDI_A|PD KLVYEAAIQLEVTG--VVVTD-ATLKNLGSVHVSKGVDADAHFPVVKEAILKTIKEVVGAKWSEELNSAWT 135
3Consensus : *:. :* . * . . . * : : * : :*:::: .*: . : :*:::

      150     160     170     180     190     200     210
      |      |      |      |      |      |      |
1 1HBG_A|PD AAYADISGALISGLQS--
2 1GDI_A|PD IAYDELAIVIKKEMDDAA
3Consensus ** ::: .: . :.
    
```



Superposition structurale
1HBG-1GDI
RMSD (75 C α): 1.44 Å



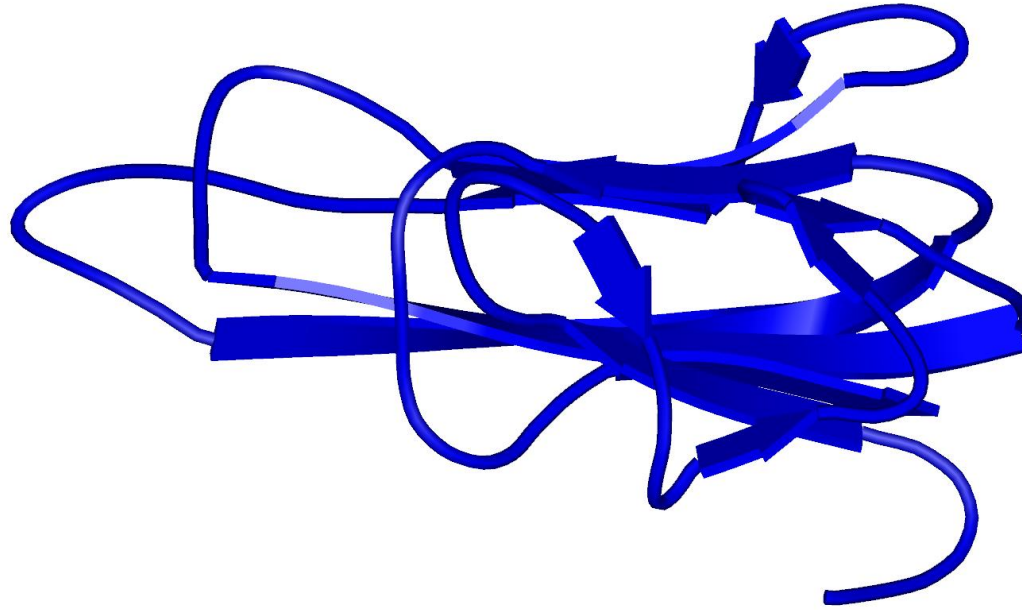
Alignement 2STV-1TNFA

ANTHEPROT 1.2.62 by G. Deléage (g.deleage@ibcp.fr), IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE
 Alignment : C:\anthepto\ALITEMP.aln using Clustalw
 Matrix GONNET
 Amino acids identity 100% >= 75 <= 50 < 50
 Gap opening penalty 10 Gap extension penalty .05 Number of perfect matches 25 Identity : 13.369 %

	10	20	30	40	50	
1 PDB3S4GA	TMRAVKRMINTHLEHKRFALIN	SGNTNATAGTVQN-	LSNGI	IQGDDINQR		49
2 PDB1TNFA	-----	RTPSDKPV	AHVVANPQAEGQLQW-	LNRR		27
3Consensus		. : . : . . *	. * * : : *	: * : *	: * : *	
	60	70	80	90	100	
1 PDB3S4GA	SGDQVRIVSHKLHVRGTAITV	SQTFRFI	WFRDNMNR	GTTPTV	LEVLTAN	99
2 PDB1TNFA	AN--	-ALLANGVEL	RDNQLVVPSEGLYLIYSQV	LFKGQGC	PSTHVL	74
3Consensus	::	: : : : *	: : * . . : : : *	. : * * *	: *	
	110	120	130	140	150	
1 PDB3S4GA	FMSQYNPITLQQKRFTILK	DVTLNCSLTGESIK--	DR	IINLPGQLV	NYNG	147
2 PDB1TNFA	ISRIAVSYQTKVNL	LSAIKSPCQRETPEGA	EAKPWYEP	IYLG	GVFQLEKG	124
3Consensus	:	. : : : *	. : * . * . * * *	: *	: *	
	160	170	180	190	200	
1 PDB3S4GA	ATAVAASNGP	GAIFMLQIGDSL	VGLWDSSYEAVYTDA			
2 PDB1TNFA	DRLSAEIN	RPDYLLFAES	GQVYFGI	IAL-----		
3Consensus	*	* * * . : : : *	: * : . * :			

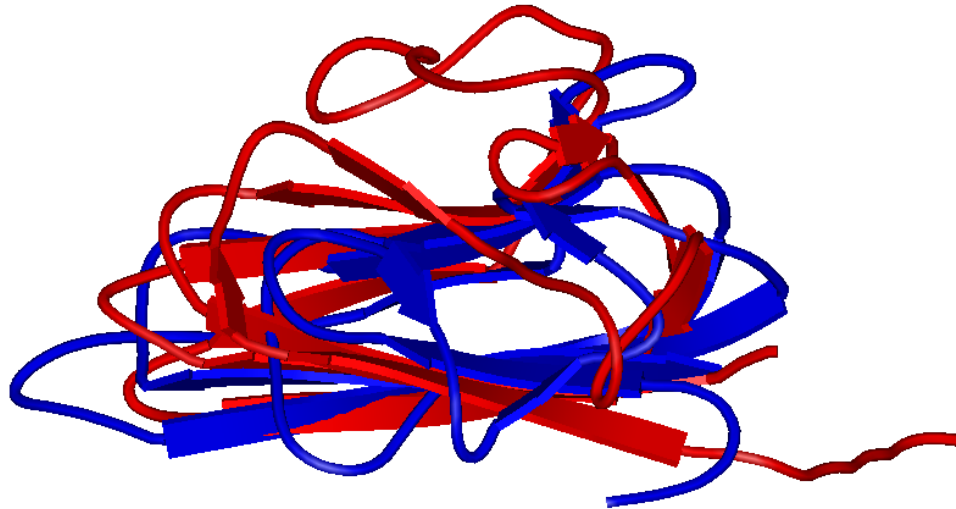


1TNFA



2STV





Rmsd $C\alpha$ = 2.9 Å

Structure-based sequence alignment: sequence identity: 2.82 % pdb1tnf 152 residues -
> pdb2stv 184 residues matching Ca: 58 (38.16% / 31.52%) rms deviation: 1.179500 min. length: 8

```

6          16          26          36          46          56
RTPS-----DKPVAHVVANPQAEGLQWLNRRANALL-----ANGVELRDNQLVVPSEGLYLIYSQ
          **  *****          * * aa  B *          *****
          **  *****          B aa  * *          *          *****
TMRAVKRMINTHLEHKRFALINSGNTNATAGTVQNLSNGIIQGDINQRSGDQ-----VRIV--SHKLH
12          22          32          42          52          62          72

          66          76          86          96
VLFKGQGCPSTHVLLTHTISRIAVSYQTK-----V-NLLSAIKSPCQ
*****          *****          *          *          *****
*****          *****          *          *          *****
VRGTAITVSQT---FRFIWRDN---MNRGTTPTVLEVLNTANFMSQYNPITLQQKRFTILLKDVTLNCSL
          82          92          102          112          122          132

          106          116          126          136          146          1
RETPEGAEAKPWYEPIYLGGVFQLEKGD-----RLSAEINRPDY---LLFAESGQVFGIAL
          ****          **  **  *          ***  *          *          *****
          ****          **  **  *          ***  *          *          *****
TGE-----SIKDRIINLPG-QLVNYNGATAVAASNGPGAIFMLQIGDSLVGLWDS-----SYEAVY
          142          152          162          172          182          1

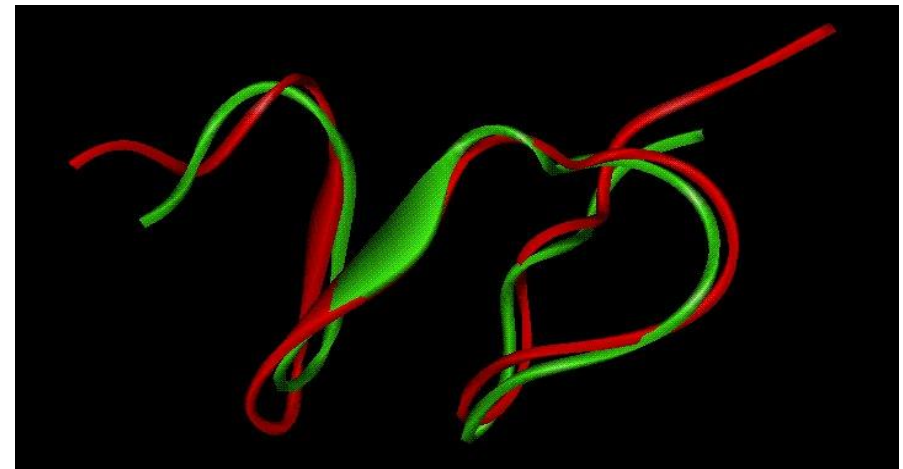
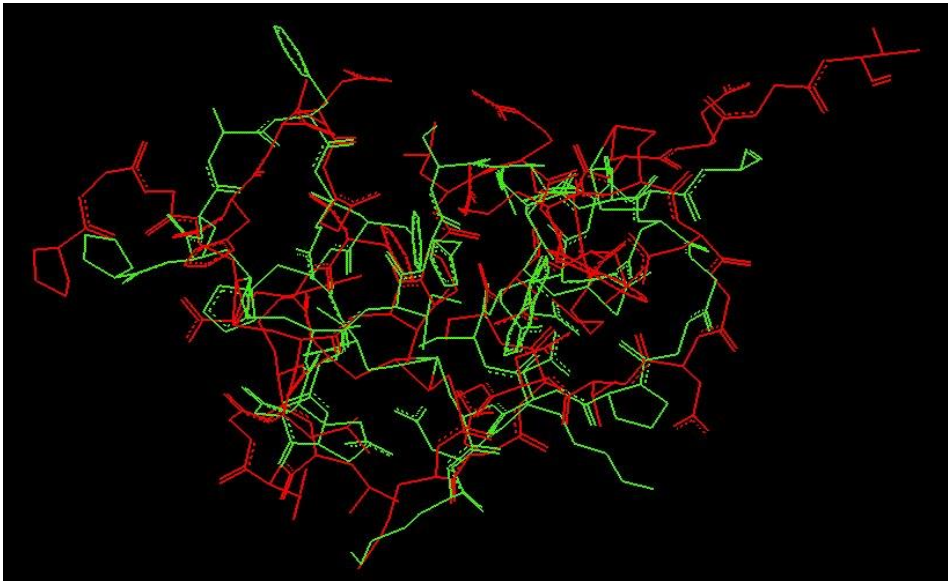
```

1AJJ : LDL receptor

1CR8 : Low Density Lipoprotein Receptor Related Protein

	10	20	30	40	
				
1.pdb1ajj.ent	P--CSAFEFHC--LSGECIHSSWRCDGGPDCKDKSDEENCA--				37
2.pdb1cr8.ent	PGGCHTDEFQCRLDGLCIPLRWRCDDTDCMDSSEDEKSCGEV				42
Homology	* * : **:* *.* ** ***** .** *.***:.*				

50 % d'identité de séquence



Ecart quadratique moyen (Cα des résidus conservés) : 1,6 Å

- **Choix de l'empreinte**

BLAST détecte pour 20% des protéines un homologue évident dans la PDB or pour ~ 70% des protéines \exists une structure empreinte permettant de modéliser

- Recherche de protéines homologues de structure 3D connue (choix)
- Alignement entre la séquence et la (es) structure(s) (problème des ponts disulfures)
- Choix des insertions et délétions à effectuer (remplacements équivalents)

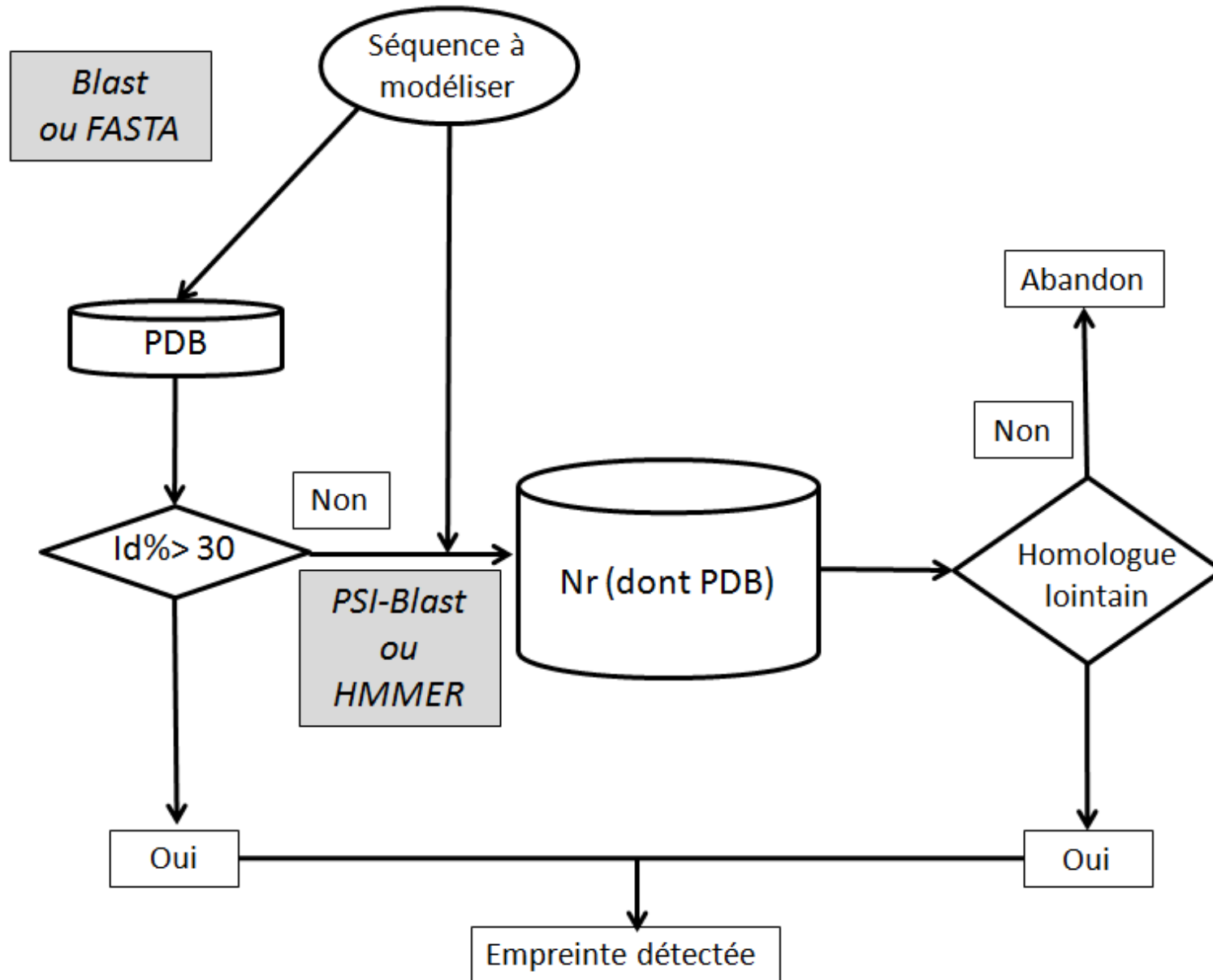
- **Construction**

- Approche substitutive et ou progressive (SPDB viewer...)
 - Substitution des chaînes latérales avec conservation des orientations des CL
 - Réorientation elle des chaînes latérales si «clashes »
 - Recherche de boucles dans les régions avec insertions
- Approche globale ou géométrique (modeller ou geno3D)

- **Optimisation - validation**

- Minimisation d'énergie
- Dynamique moléculaire
- Validations biologiques
- Choix d'expériences à partir d'hypothèses

Détection et validation d'empreinte



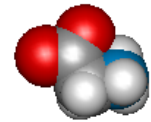
Structure 3D

150
...AKL**GMS**-VHKLPWSTERILMNFSEK

Protéine d'intérêt

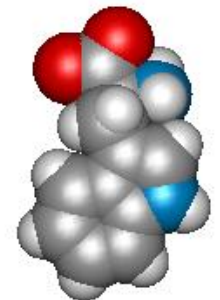
...ARL**WMST**IRRI**PMGST**-RILMQGTEK

Gly



?

Trp



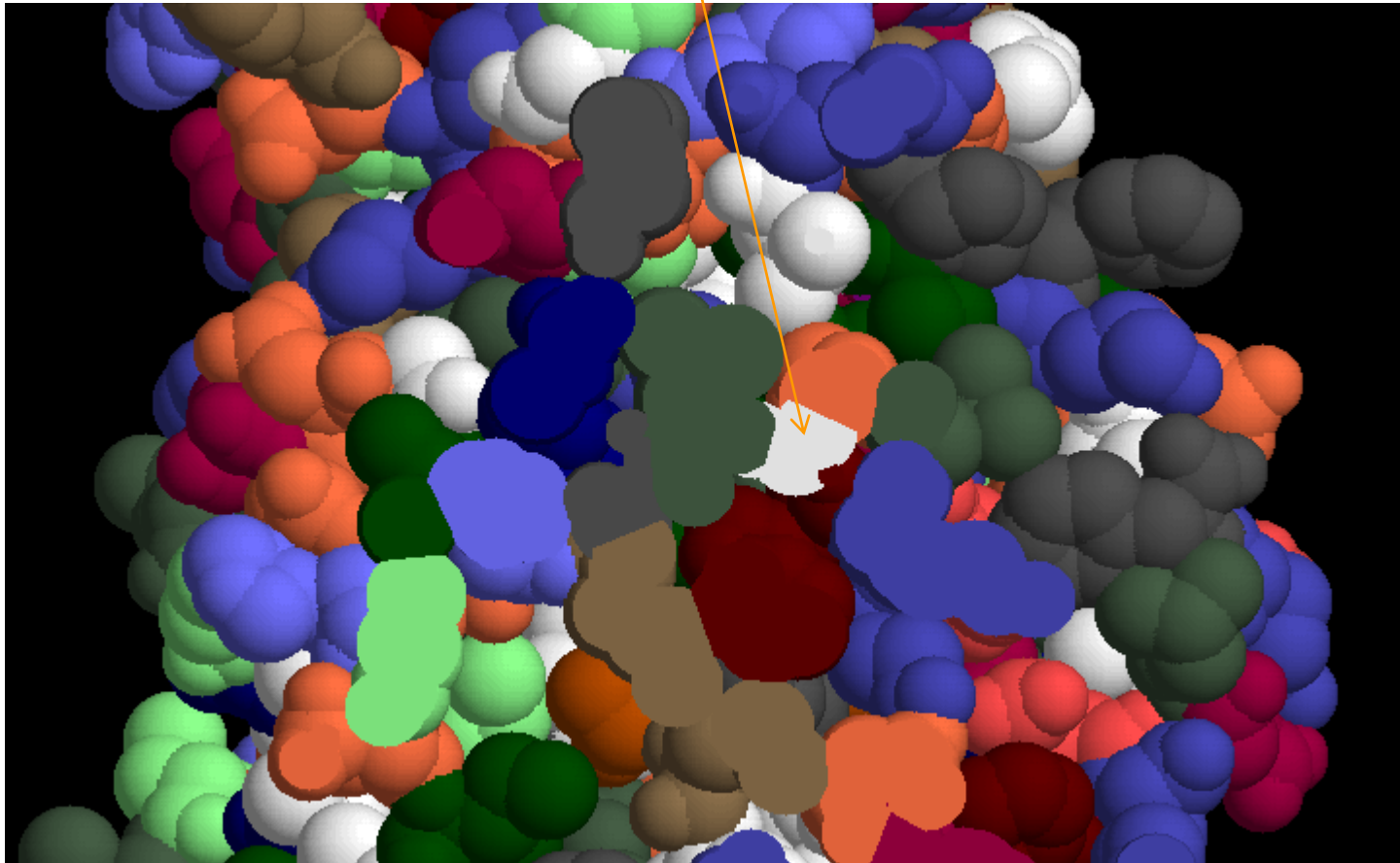
150

Structure 3D

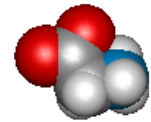
...AKLGMS - VHKLPWSTERILMNFSEK

Protéine d'intérêt

...ARLWMSTIRRI PMGST - RILMQGTEK

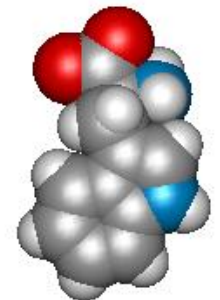


Gly



?

Trp

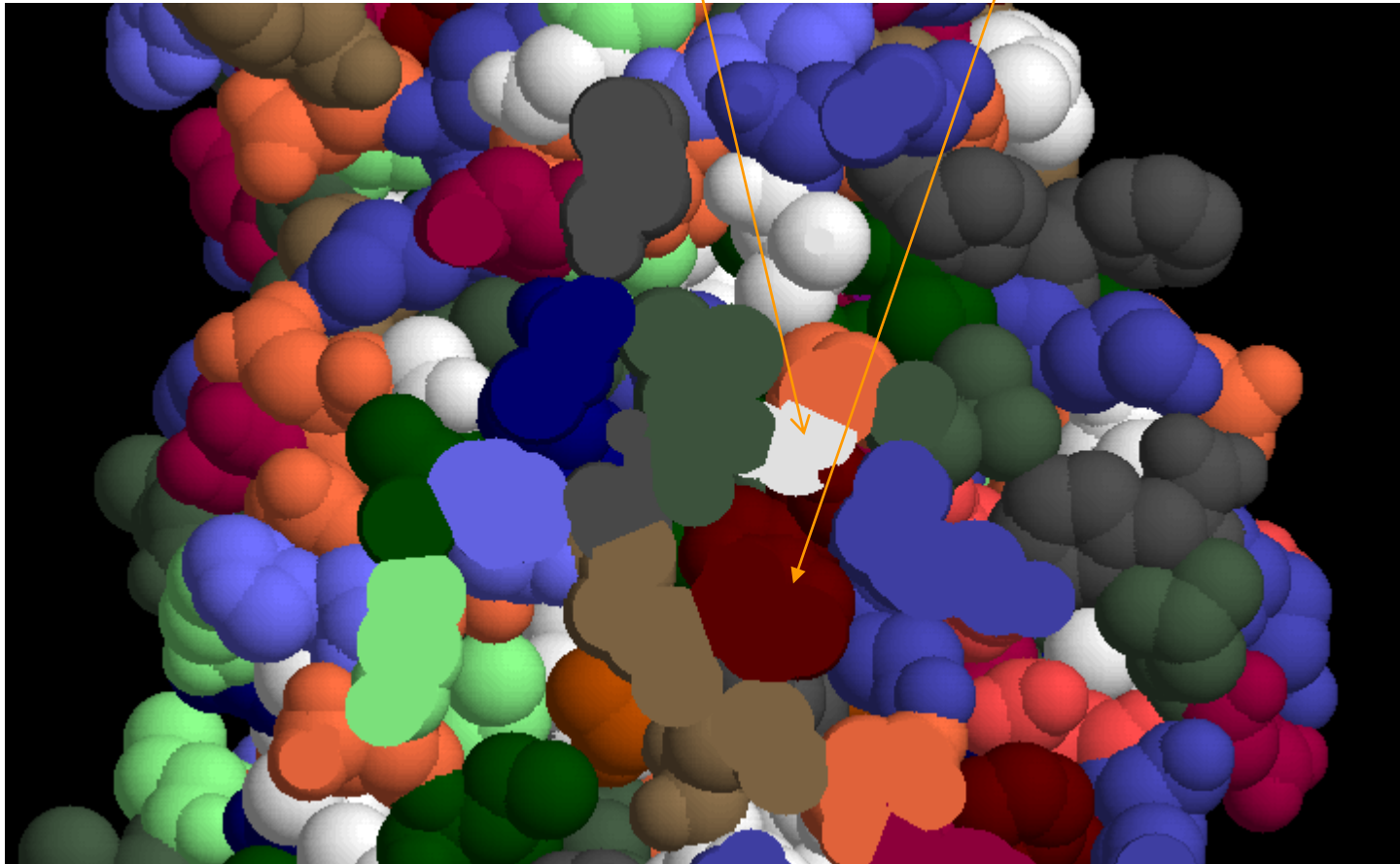


Structure 3D

...AKLGMS - VHKLP MWSTERILMNFSEK

Protéine d'intérêt

...ARLWMSTIRRI PMGST - RILMQGTEK



Construction

TD 1	25/09 14H30-17H45 Distanciel Master de Bioinformatique S. Megy
	29/09 14H00-17H15 Distanciel Master de Biochimie P. Gouet
TD2	21/10 8H-11H15

Mutation des acides aminés

- **Logiciel graphique**

Avantages

- Utilisable avec une logique de mutation (connue du biologiste)
- Instantané et facilité à mettre en œuvre (sur un PC)
- Apport de la connaissance du biologiste

Inconvénients

- Qualité dépend de l'expertise du modélisateur
- Difficulté d'imposer des contraintes externes
- Choix des conformations (rotamères)
- Boucles?
- Optimisation découplée de la construction
- Une seule empreinte utilisable.
- Démo approche mutation avec Swiss PDB viewer

TD 1 25/09 14H30-17H45 Distanciel Master de Bioinformatique S. Megy

<https://univ-lyon1.webex.com/meet/simon.megy>

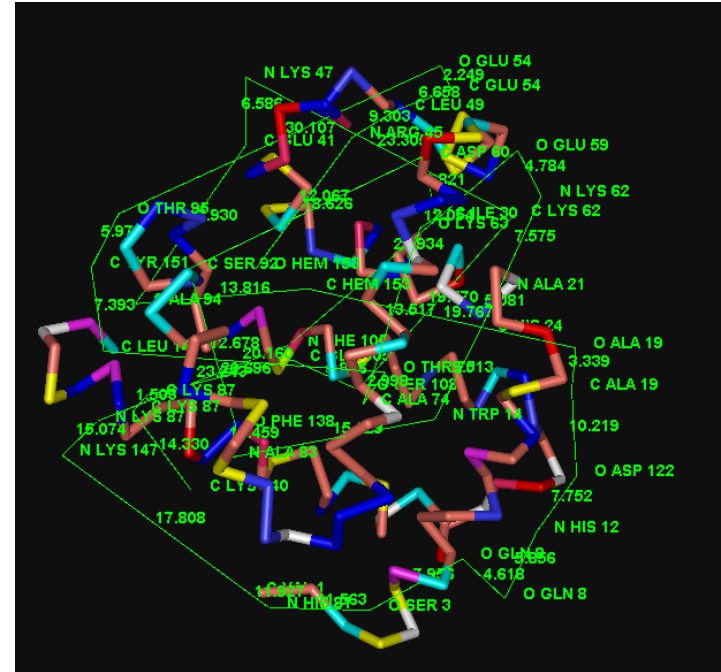
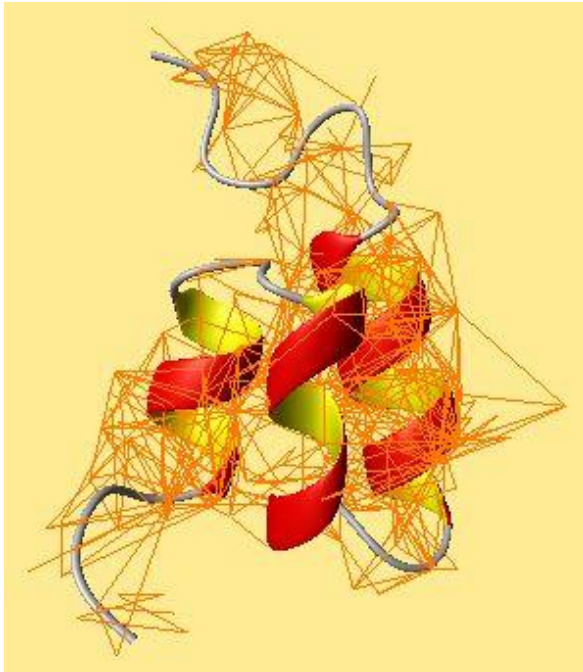
29/09 14H00-17H15 Distanciel Master de Biochimie P. Gouet

<https://univ-lyon1.webex.com/univ-lyon1/j.php?MTID=m10b1876c55c162fafc4d61aead473aec>

TD2 21/10 8H-11H15

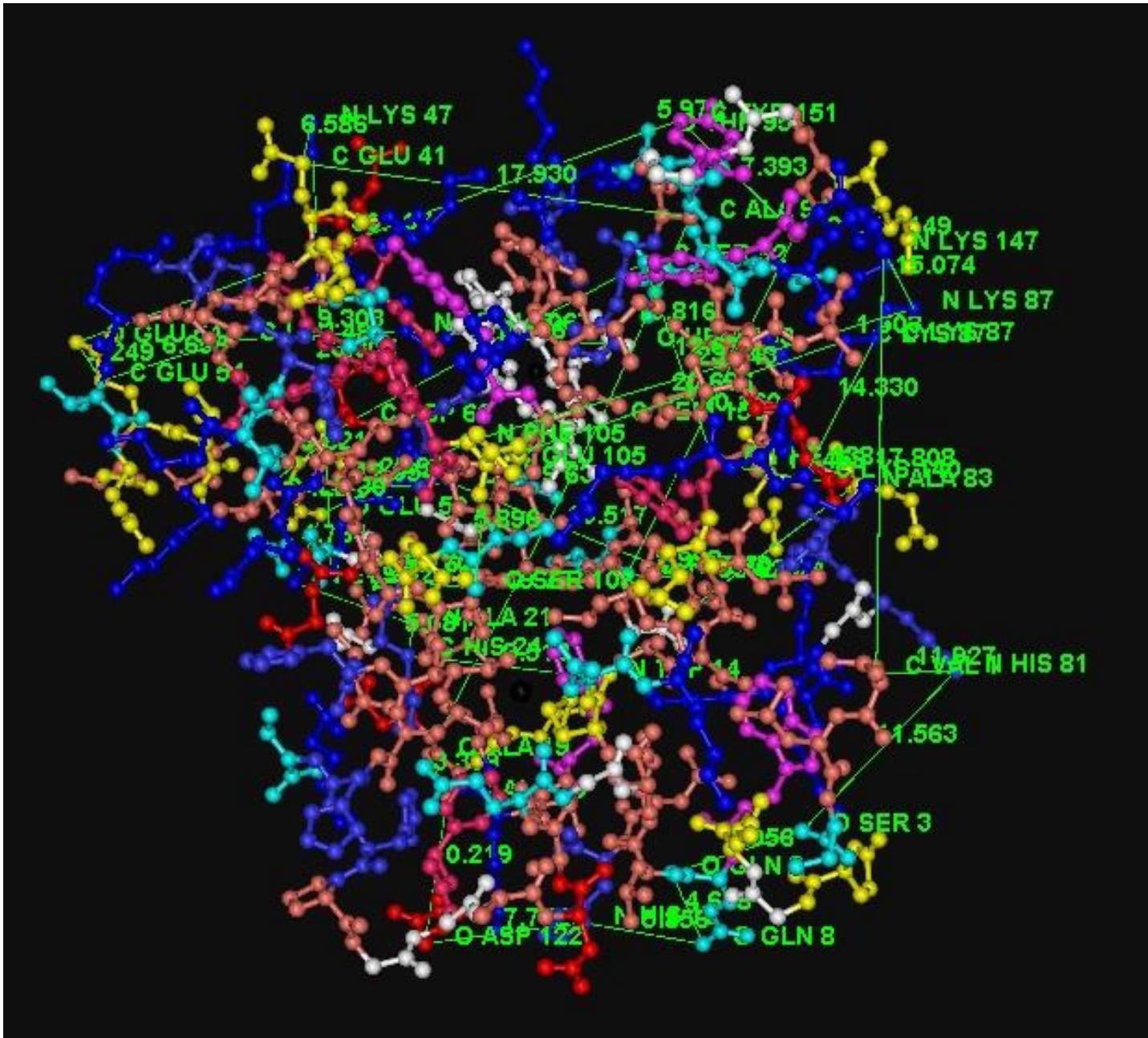
Principe :

- 1) A partir de l'alignement les atomes en commun entre l'empreinte(s) et la protéine source sont identifiés
- 2) Des distances sont mesurées sur le(s) empreinte(s) entre les atomes communs (y compris sur la chaîne latérale)



- 3) Transformation des distances en contraintes ($C = D_{\text{moy}} \pm \sigma$)
- 4) Imposition des contraintes mesurées avec celles de la chimie
- 5) Optimisation par dynamique moléculaire

Construction globale par « distance geometry »



Modeller



```

A I L V G S M P R R D G M E R K D L L K A N V K I F K C Q G A
V E V C P V D C F Y E G P N F L V I H P D E C I D C A L C E P
S A C K P E C P V N I I Q G S - - Y A I D A D S G I D G S
C - - I A C G A C K P E C P V N I I Q G S - - I Y A I D A D S
    
```



Andrej Sali

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

About MODELLER

MODELLER News

Download & Installation

Release Notes

Data file downloads

Registration

Accelrys licensing

Discussion Forum

Subscribe

Browse archives

Search archives

About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is [available for download](#) for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are [commercially available](#). There are also many other [resources and people using Modeller](#) in graphical or web interfaces or other frameworks.

1. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2006.

<http://www.salilab.org/modeller/>

M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.

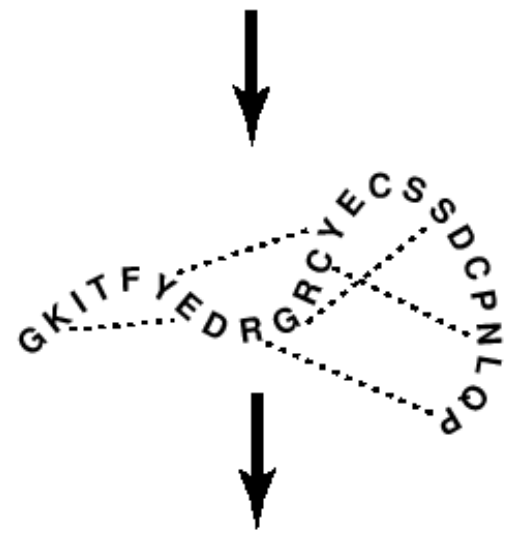
<http://salilab.org/modeller/tutorial/>



1. ALIGN SEQUENCE WITH STRUCTURES:

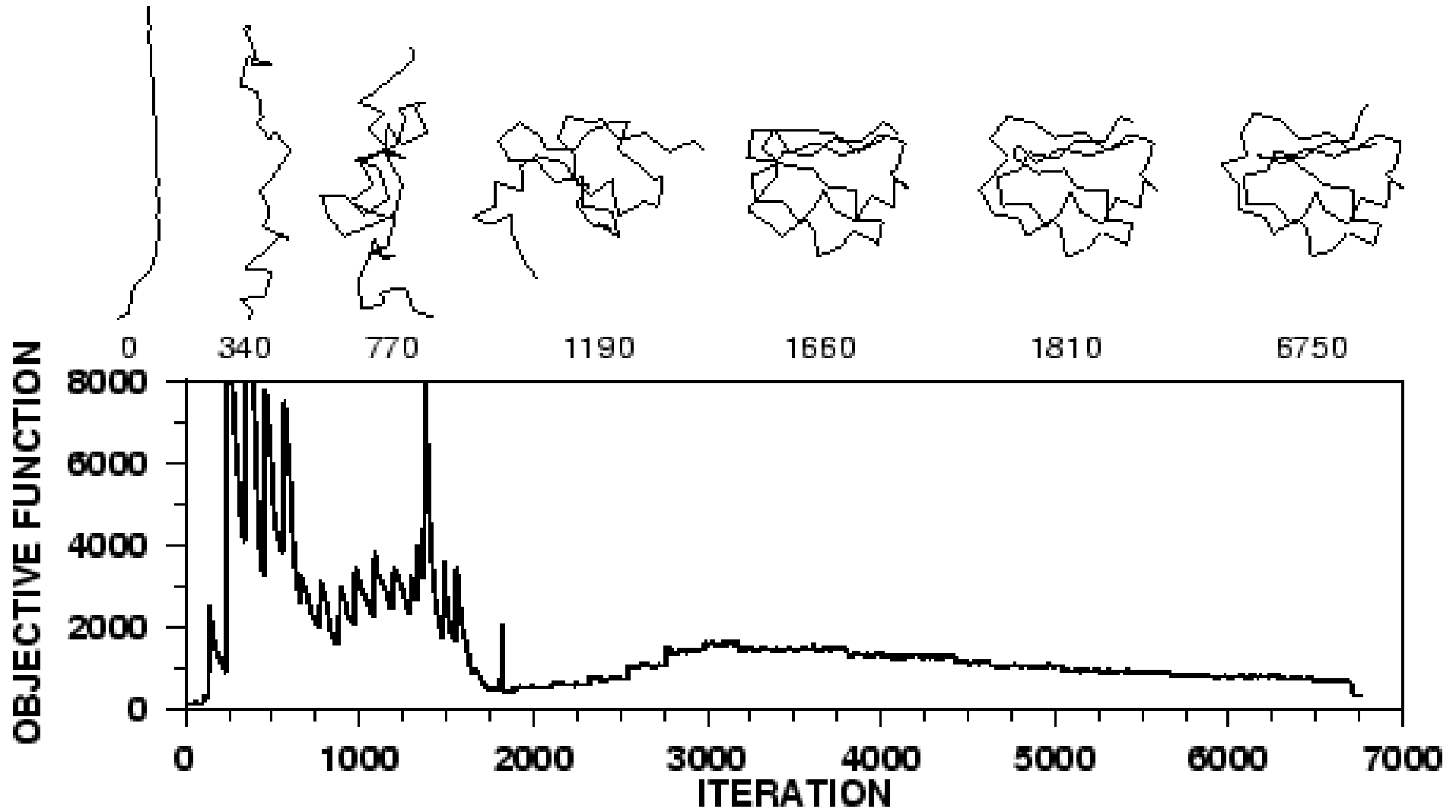
3D	GRISFFEDAGF-GHCYECSSDC-NI
3D	GKITFYEDRGFQHCYECSSDC-NI
SEQ	GKITFYEDRG---RCYECSSDCPNI

2. EXTRACT SPATIAL RESTRAINTS:



3. SATISFY SPATIAL RESTRAINTS:





Avantages

Prise en compte de plusieurs empreintes

Utilisable à faible taux d'identité

Utilisation possible de contraintes expérimentales

- RMN (attribution partielle)
- transfert de fluorescence (FRET)
- mutations connues
- Structures secondaires (prédiction ou RMN)
- Pas de problèmes dans les boucles
- Indice de convergence (superposition des solutions)

Inconvénients

- Sous détermination
- Pas de solution unique
- Temps de construction
- Qualité de la fonction objectif

Un cas concret d'utilité de modèle

ANTHEPROT V4.0 by G. Deléage (g.deleage@ibcp.fr)
 IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE
 Matrix BLOSUM

Amino acids identity : : 100
 >= 50
 < 50

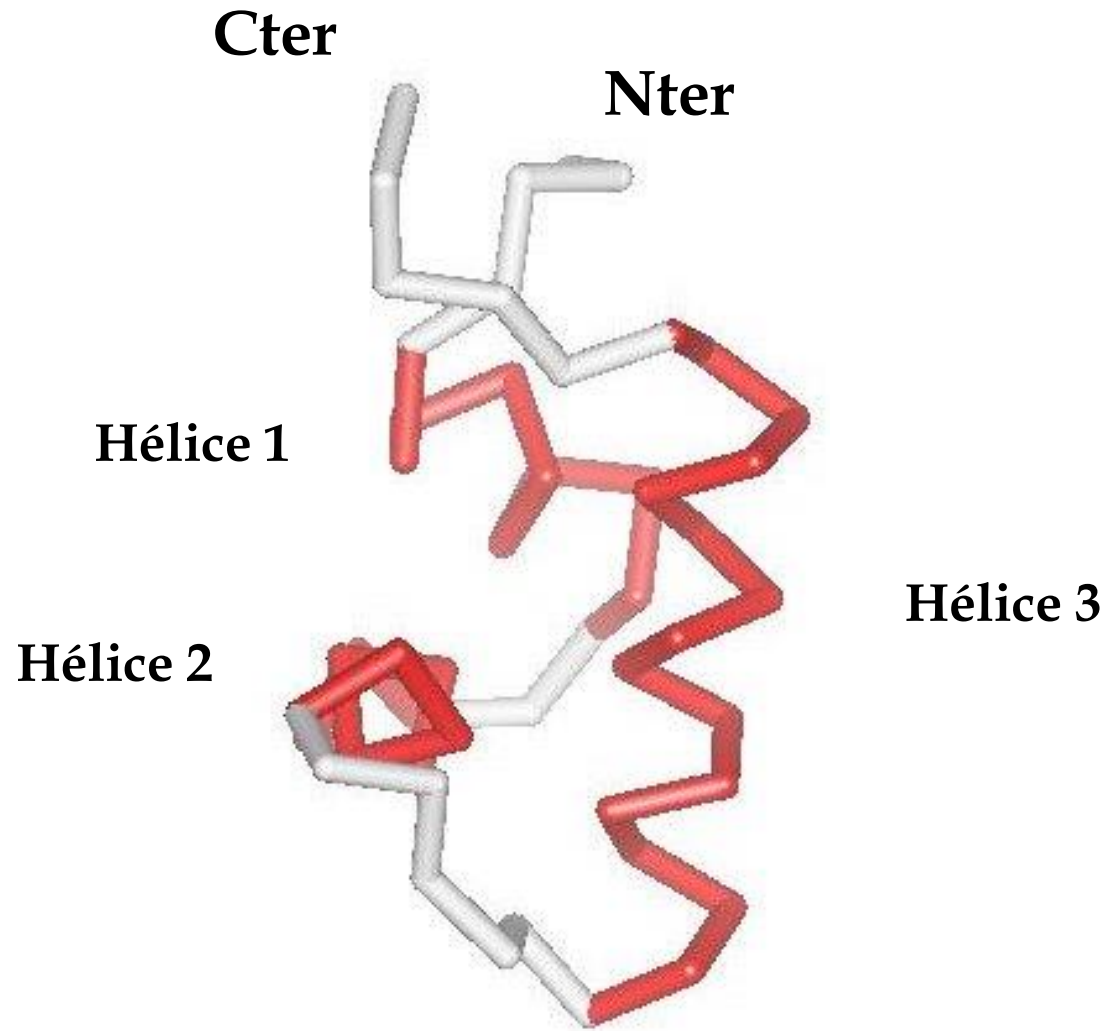
Gap opening penalty 10 Gap extension penalty .05
 Number of perfect matches 23

	10	20	30	40	50
	Hélice 1		Hélice 2		Hélice 3
1 MOLE	██████████		██████████		██████████
2 FRUR_AVE	---		---		---
	M K P V T L Y D V A E Y A G V S Y Q T V S R V V N Q A S H --- V S A K T R E K V E A A M A E L N Y I P N R				51
	---		M K L D E I A R L A G V S R T T A S Y V I N G K A K Q Y R V S D K T V E K V M A V V R E H N Y H P N -		50

Hélice 1 L6 - G14
 Hélice 2 Y17 - V24
 Hélice 3 K33 - E44



Une seule insertion de 3 acides aminés à effectuer



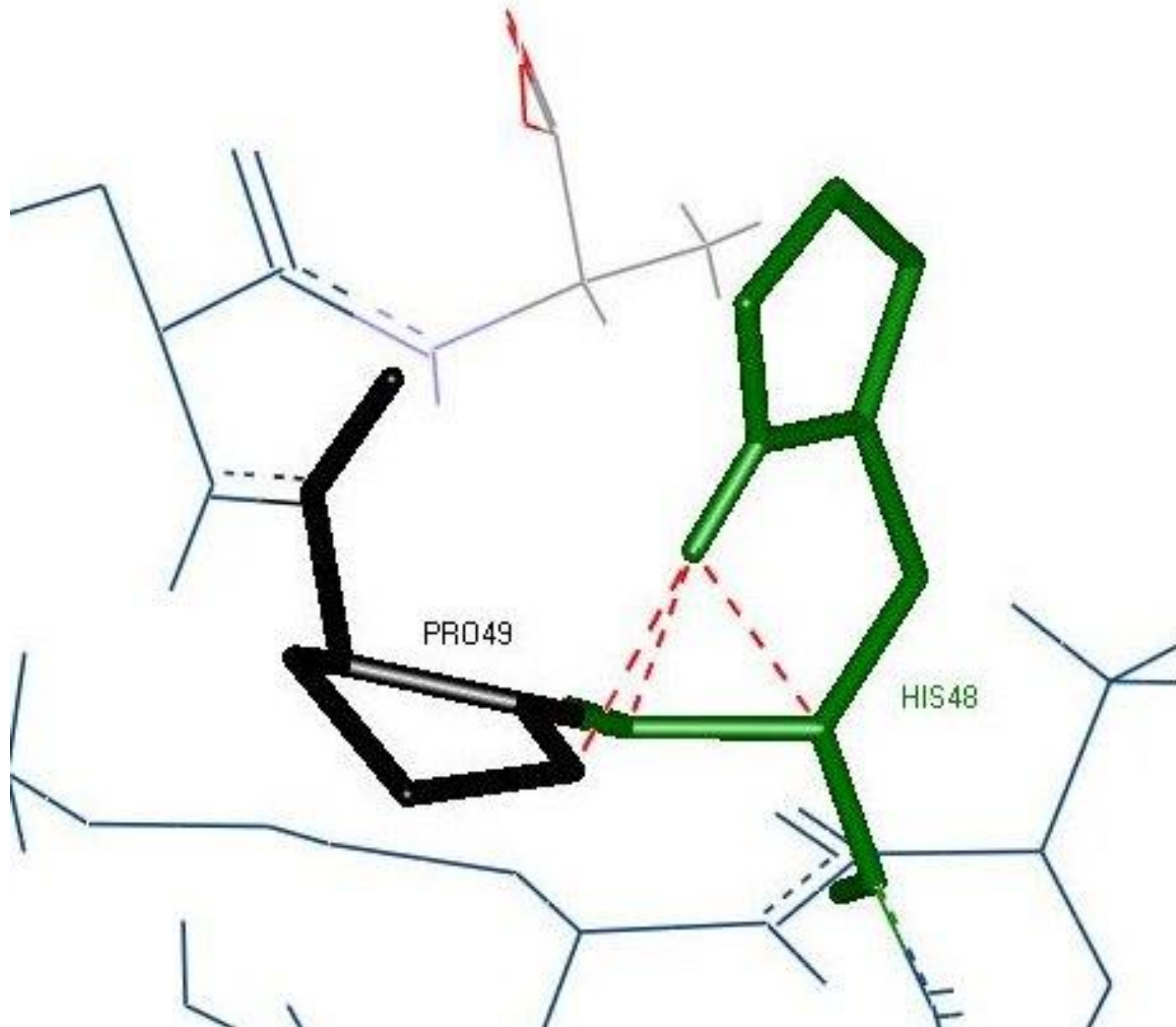
Pas de ponts disulfures

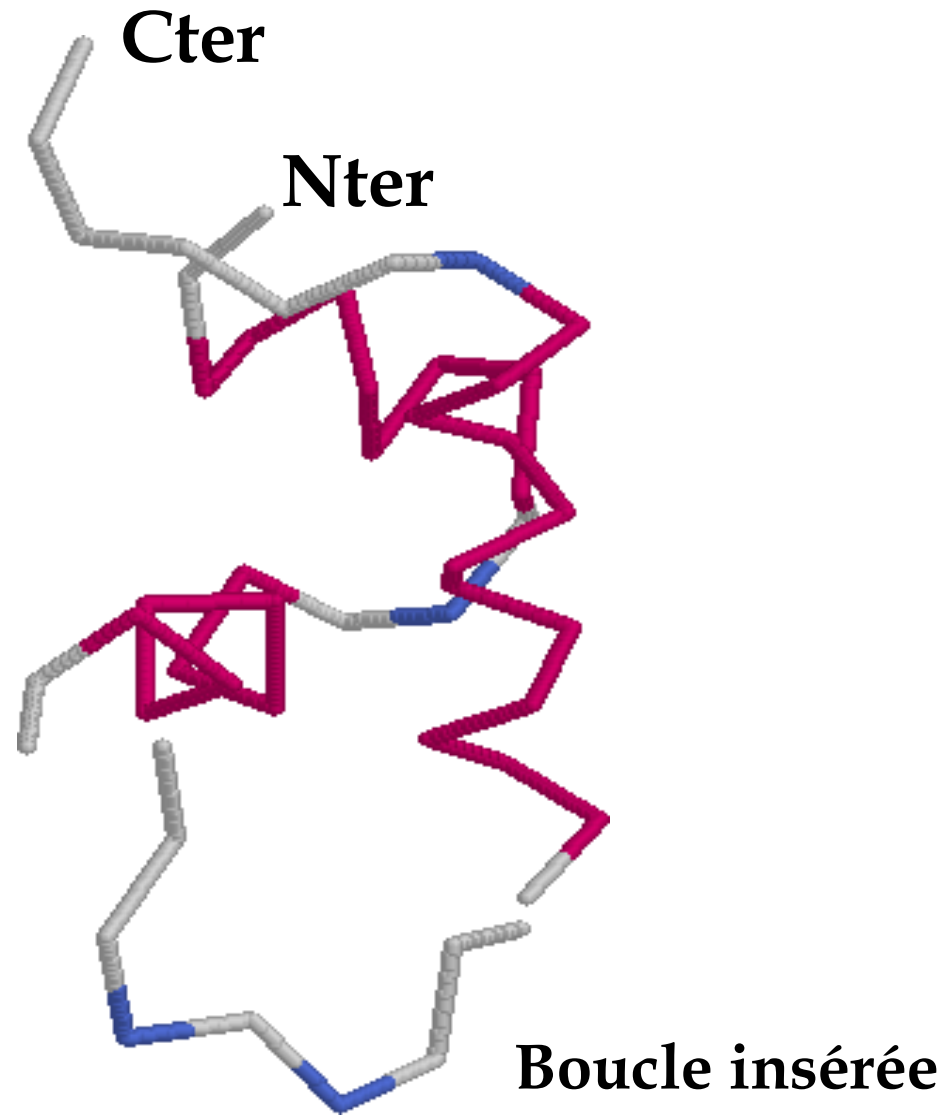
● Construction

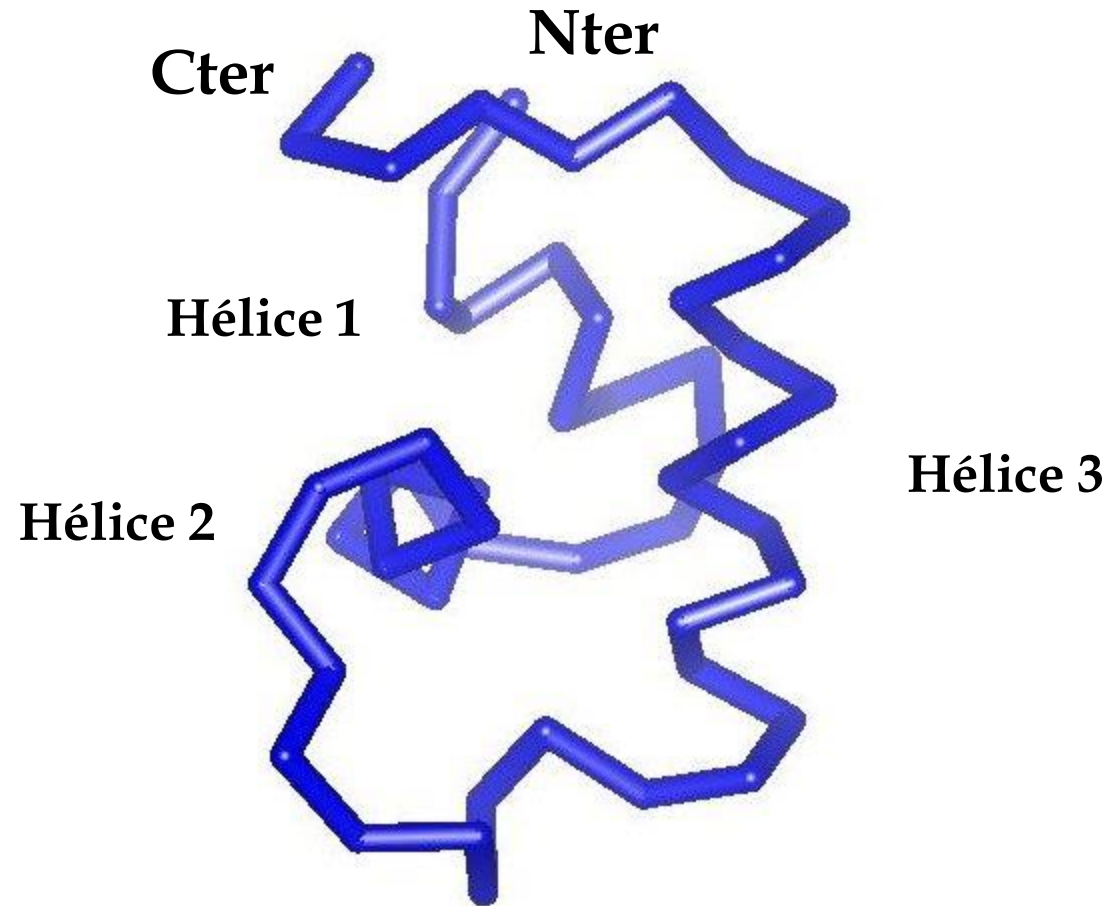
- Substitution des chaînes latérales avec conservation des orientations des CL
- Réorientation elle des chaînes latérales si «clashes »
- Recherche de boucles dans les régions avec insertions

● Optimisation - validation

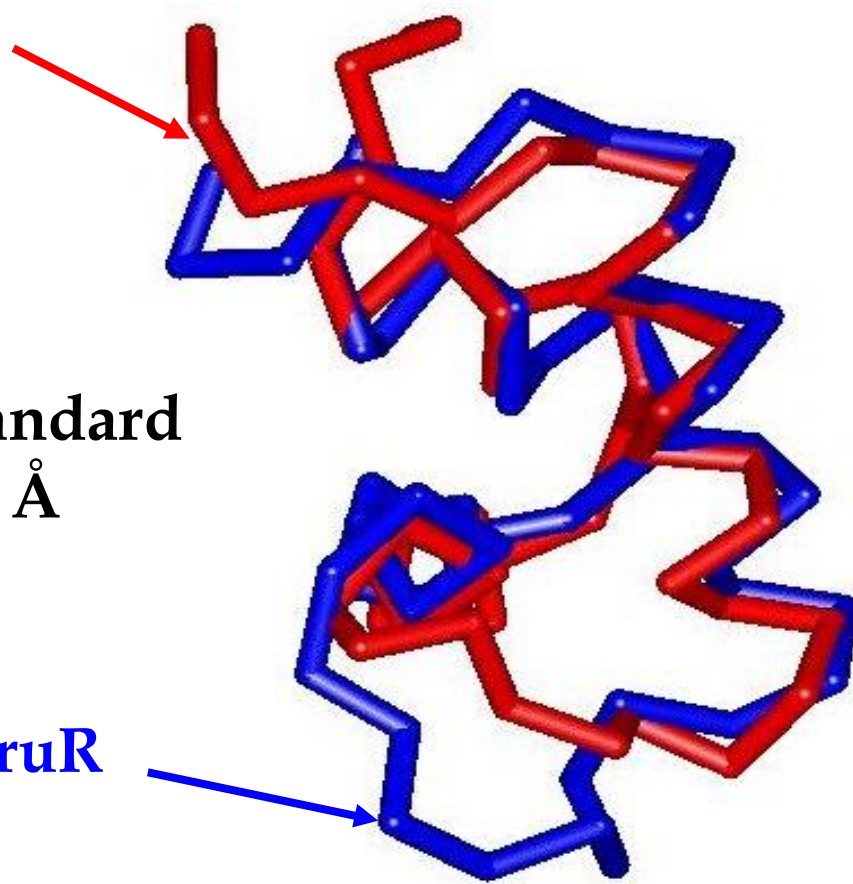
- Minimisation d'énergie
- Dynamique moléculaire
- Validations biologiques
- Choix d'expériences à partir d'hypothèses







Empreinte Lacl



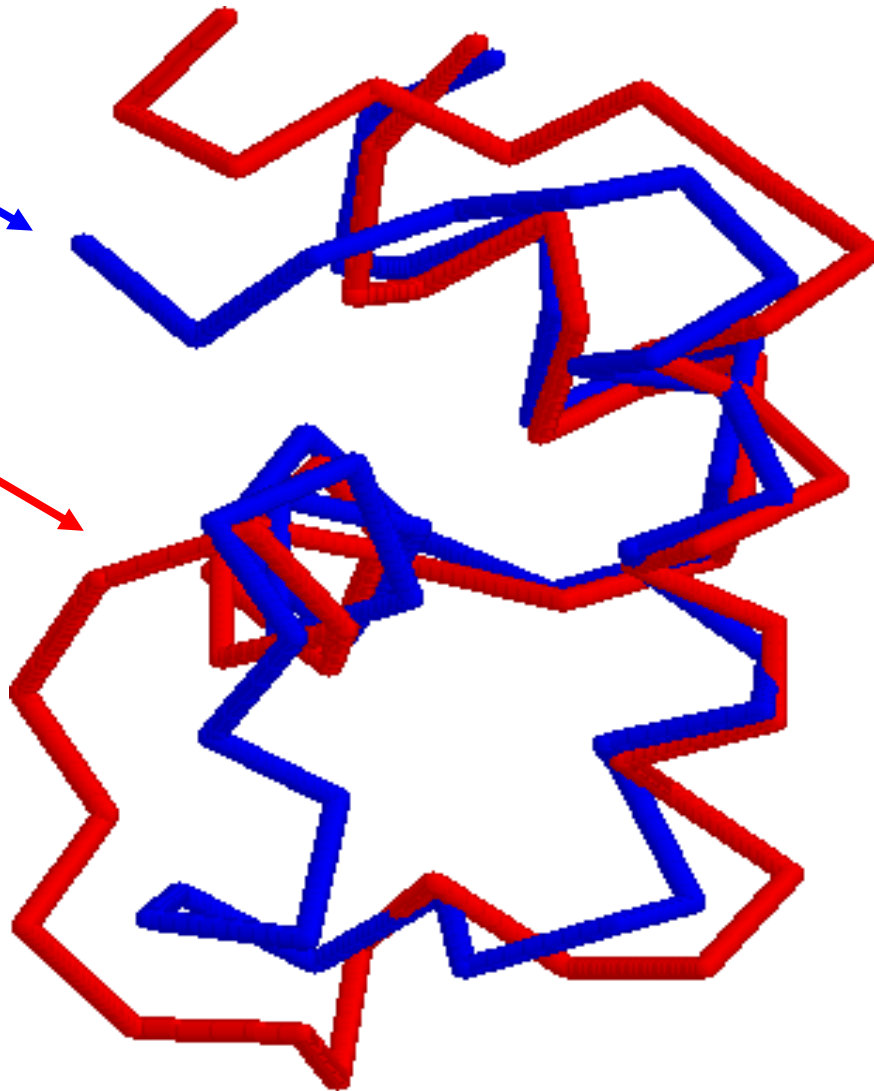
Déviatiion standard
 $C\alpha = 2,2 \text{ \AA}$

Modèle FruR

Structure RMN

Modèle homologique

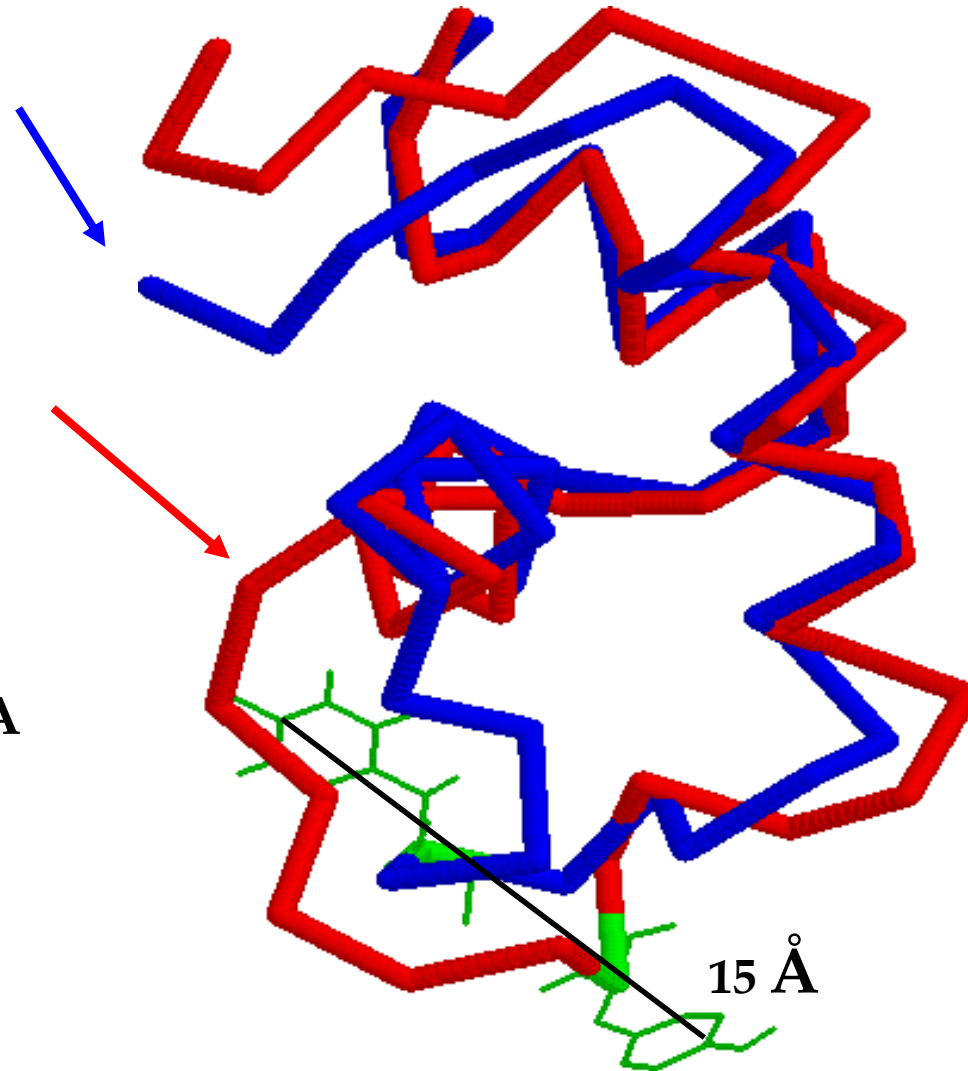
Déviati on standard
 $C\alpha = 2,9 \text{ \AA}$

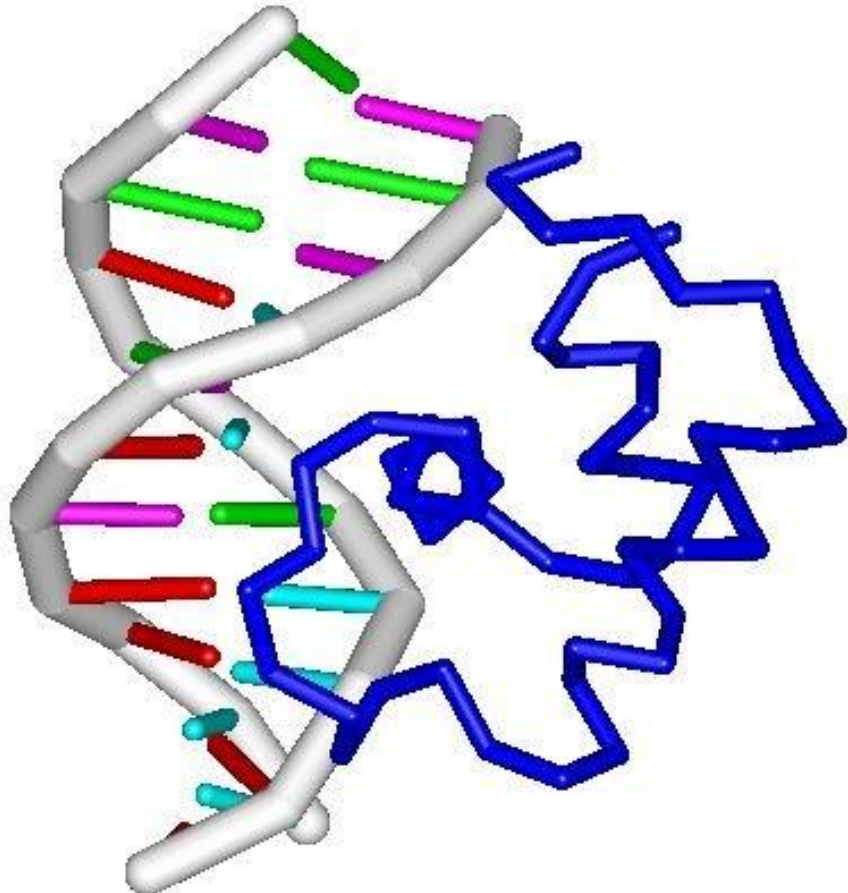


Structure RMN

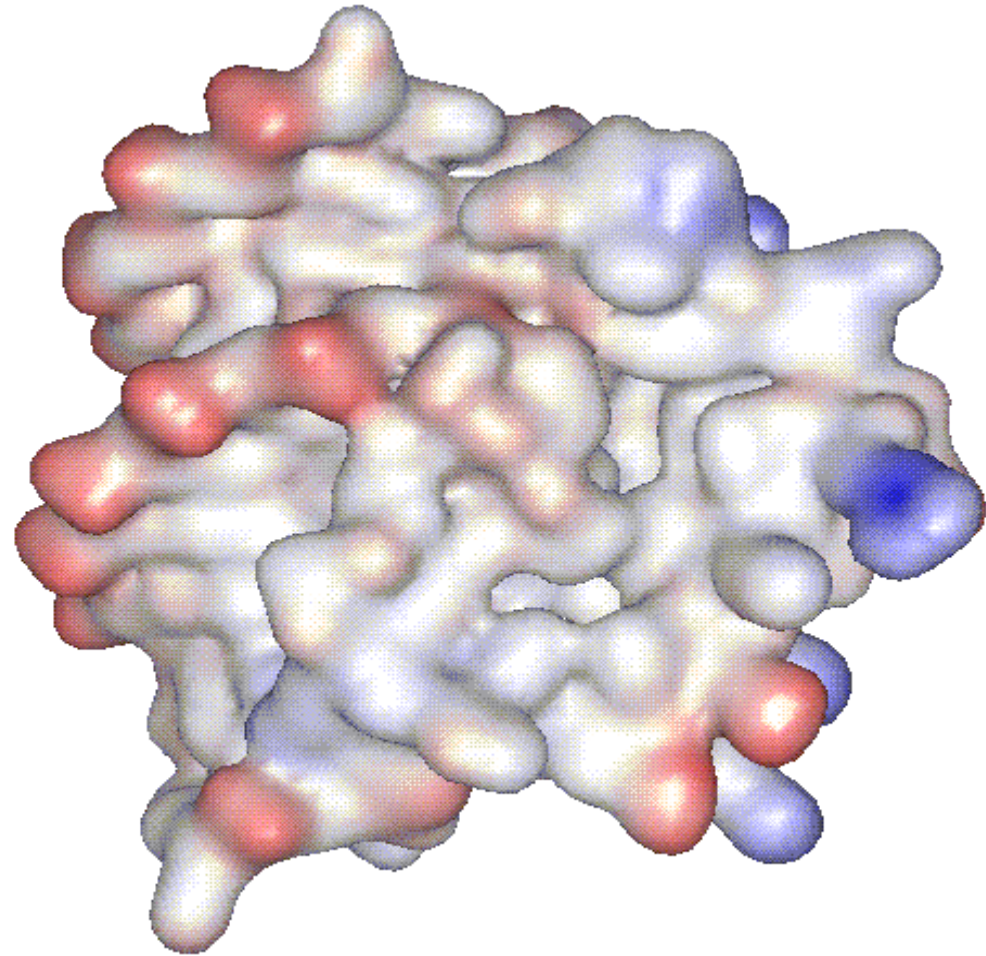
Modèle homologique

Distance TYR28 CZ 15A

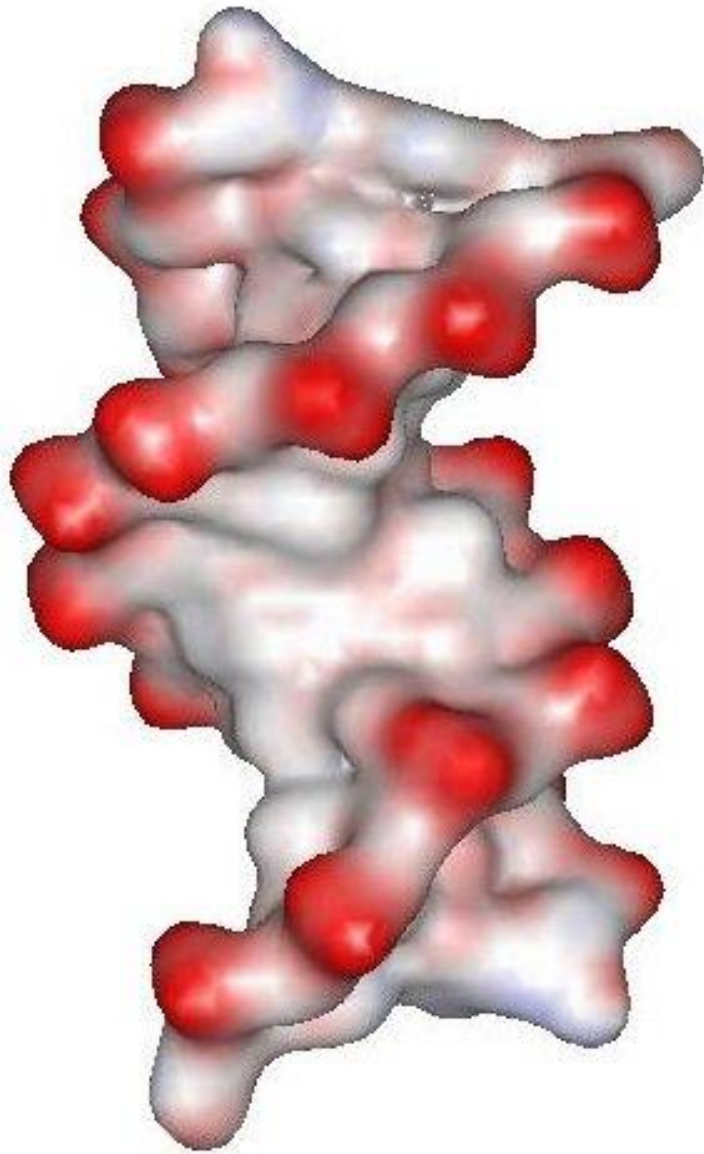




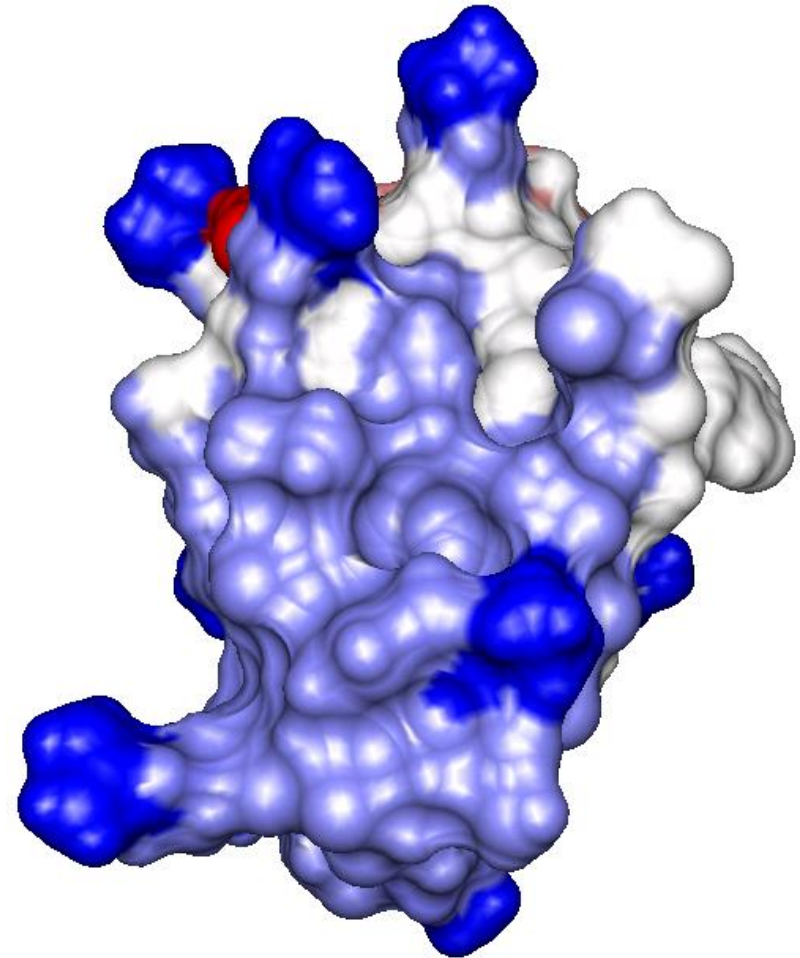
ADN de LacI répresseur



Rouge : Potentiel négatif
Bleu : Potentiel positif



Surface « vue » par le répresseur



Surface « vue » par l'ADN

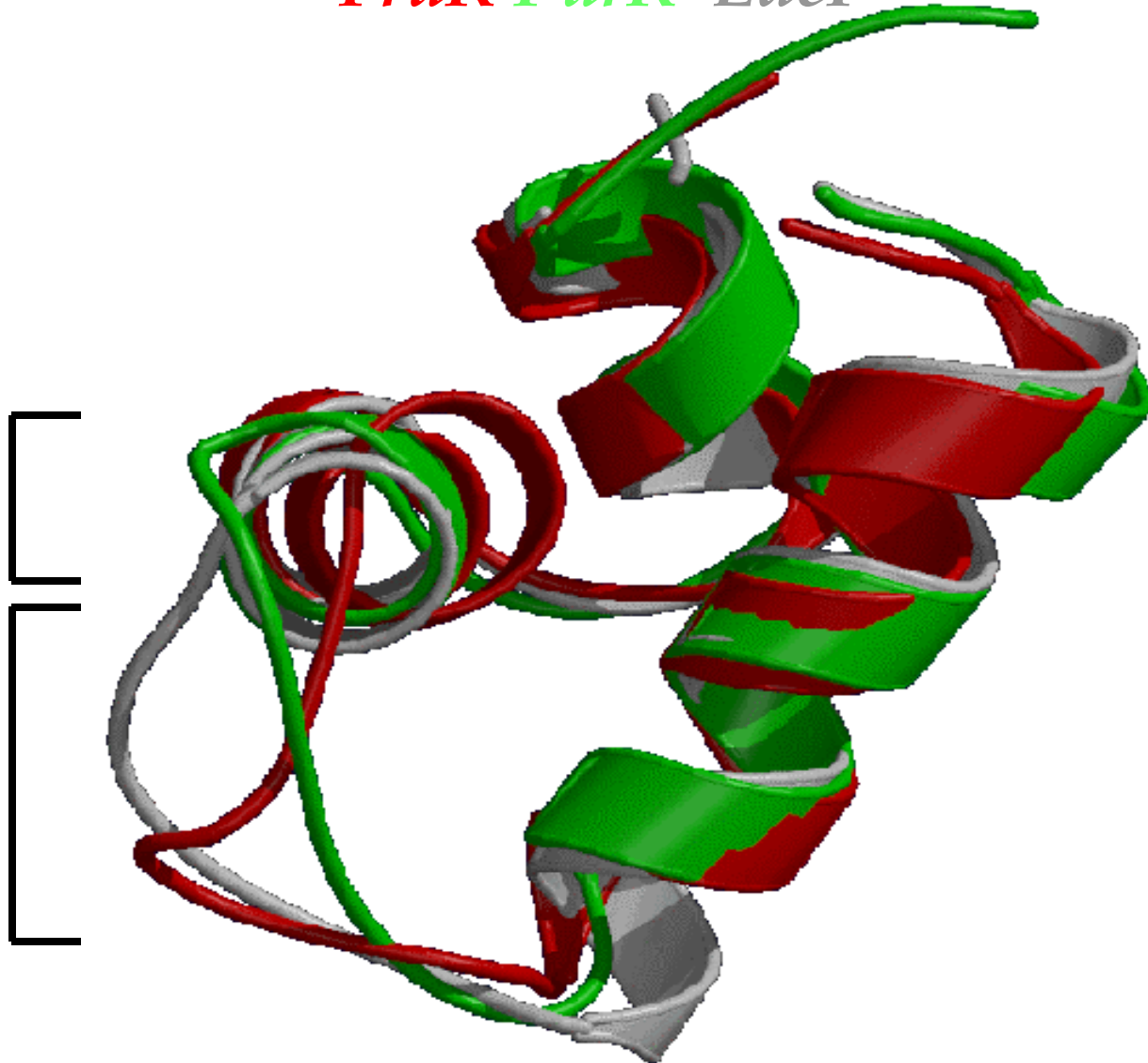
FruR *PurR* *LacI*

Fixation à l'ADN

Maximum de différence



Spécificité de
reconnaissance



- **Répresseur du DBD de LacI isolé**
 - Définition insuffisante pour « analyser les liaisons H » intramoléculaires
- **Complexe Répresseur du DBD de LacI avec l'ADN**
 - Mise en évidence des liaisons H inter moléculaires
- **Comparaison des séquences de tous les répresseurs**
 - Identifications des résidus => liaisons H intermoléculaires
 - Alignements des séquences disponibles
 - Analyse des liaisons H dans le répresseur FruR

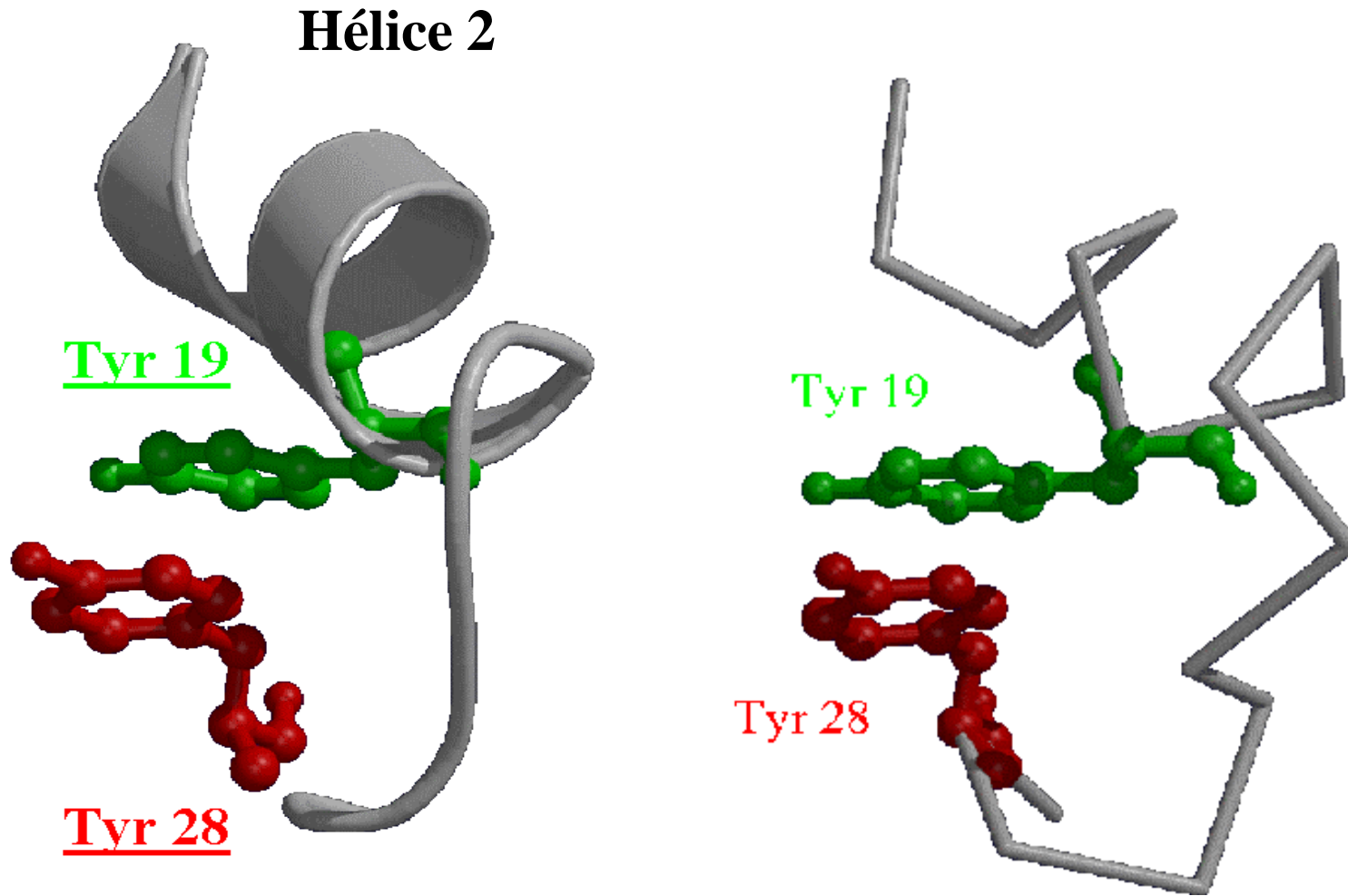
ANTHEPROT V4.0 by G. Deléage (g.deleage@ibcp.fr)
 IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE
 Date :11-24-1998 at :09:06:49

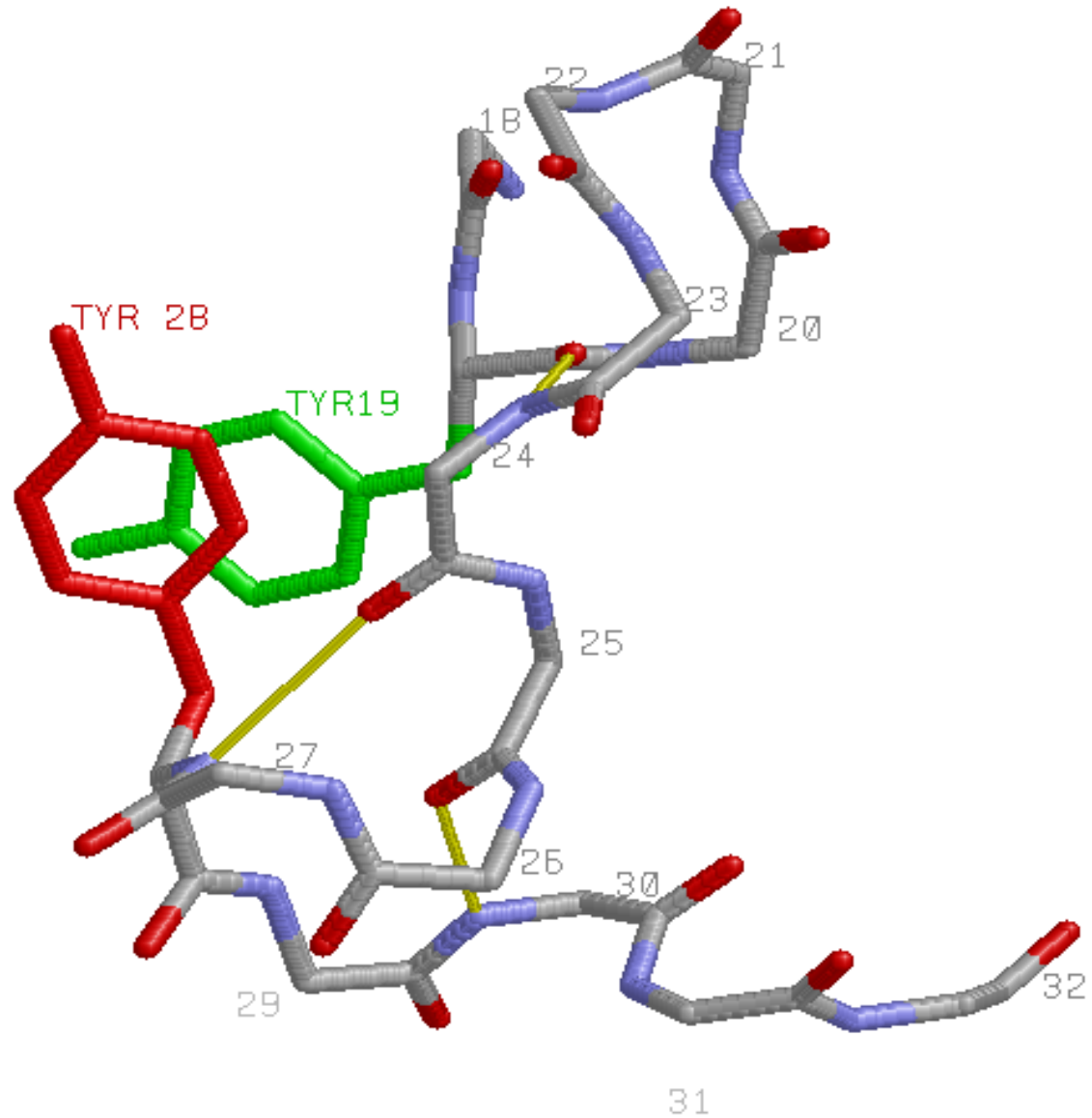
Matrix

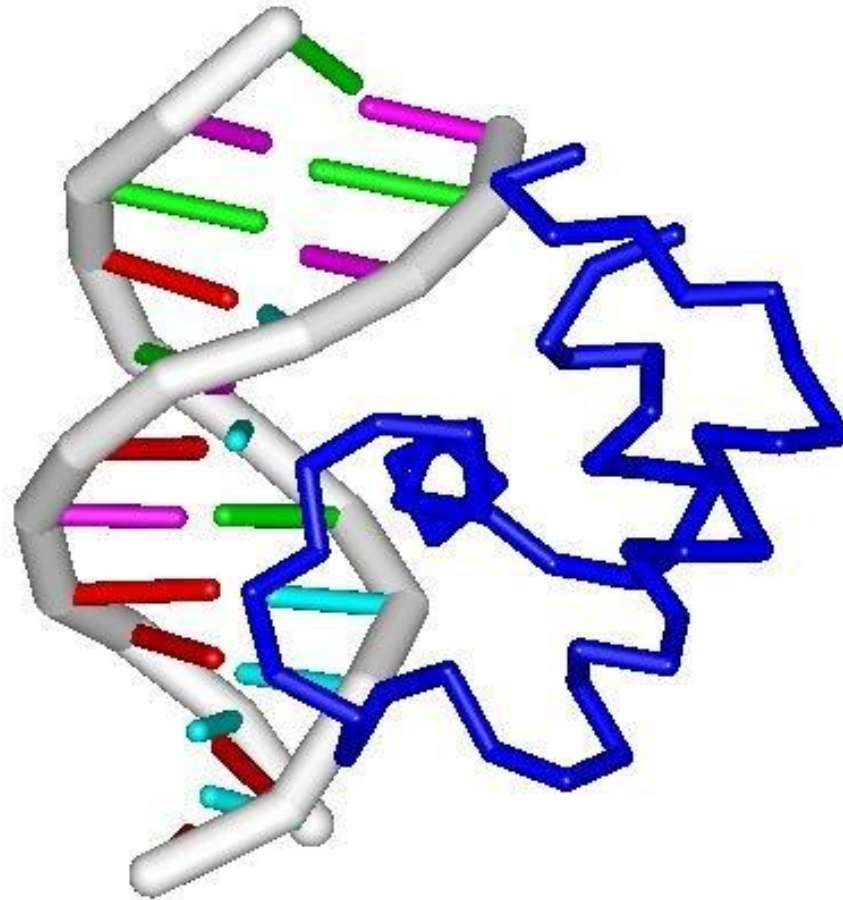
Amino acids identity : : 100
 >= 75
 >= 50
 < 50

Gap opening penalty 0 Gap extension penalty 0 Number of perfect matches 37

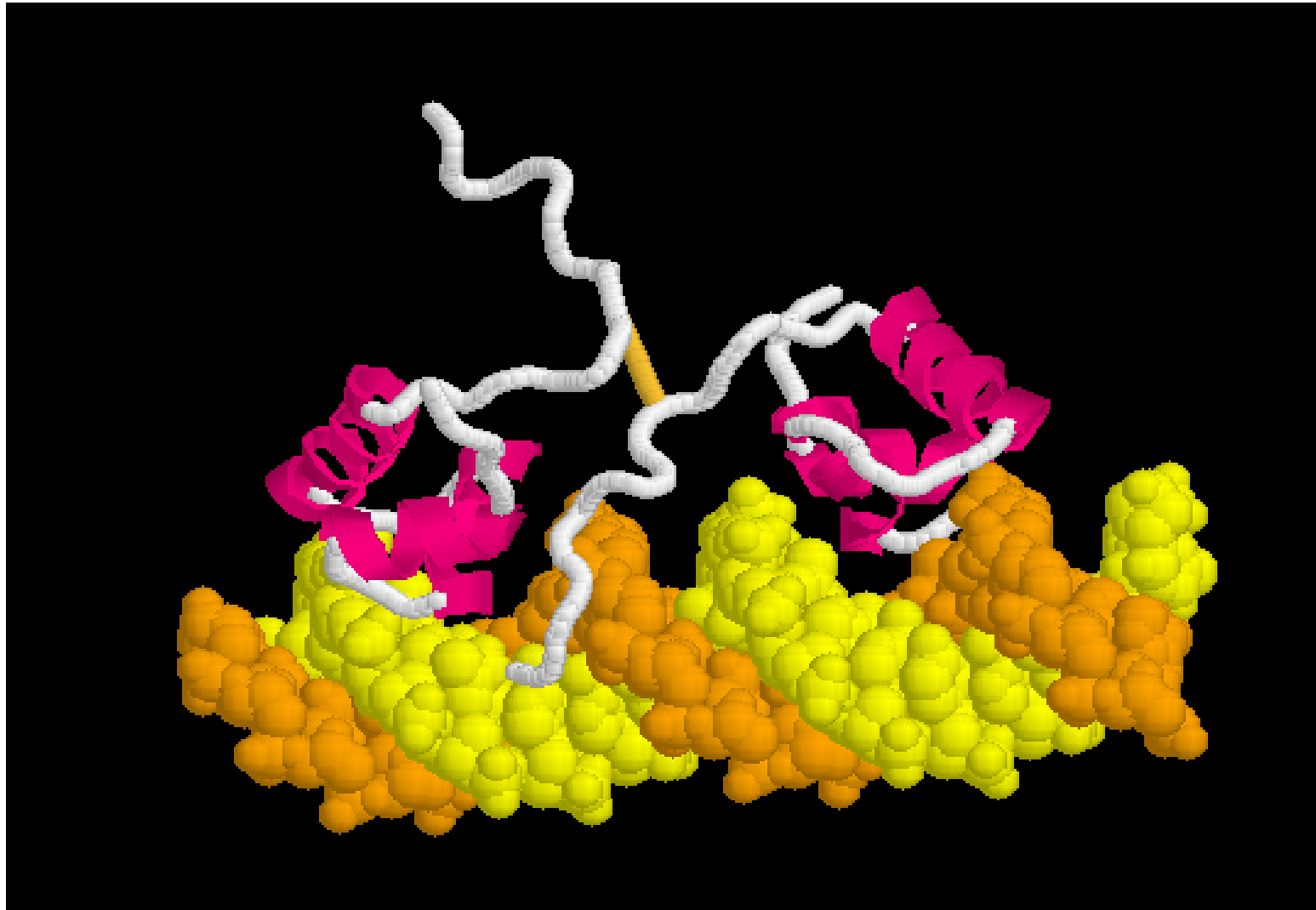
	10	20	30	40	50	60	70
LACI_ECOLI	MKPVTLYDVAEYAGVSYQTVSRVNV	---	QASHVSAKTREKVEAAMAELNYIPNRVAQQLAGKQSL	LLIGVA			
	HHHHHHHH	HHHHHHHH	HHHHHHHHHHHHHHHH	hhhhhhhhh			
PURR_ECOLI	---	ATIKDVAKRANVSTTTVSHVIN	---	KTRFVAEETRNAVWAAIKELHYSPSAVARSLKVNHTKSIGLL			
	HHHHHHHH	HHHHHHHHHH	HHHHHHHHHHHHHHHH	HHHHHHHHHH			
FRUR_ECOLI	---	MKLDEIARLAGVSRRTASVINGKAKQYRVSDKTVEKVMVVREHNYHPNAVAAGLRAGRTRSIGLV					
	HHHHHHHH	HHHHHHHHHH	HHHHHHHHHHHHHHHH	hhhhhhhhh			
	↑	↑	↑	↑			







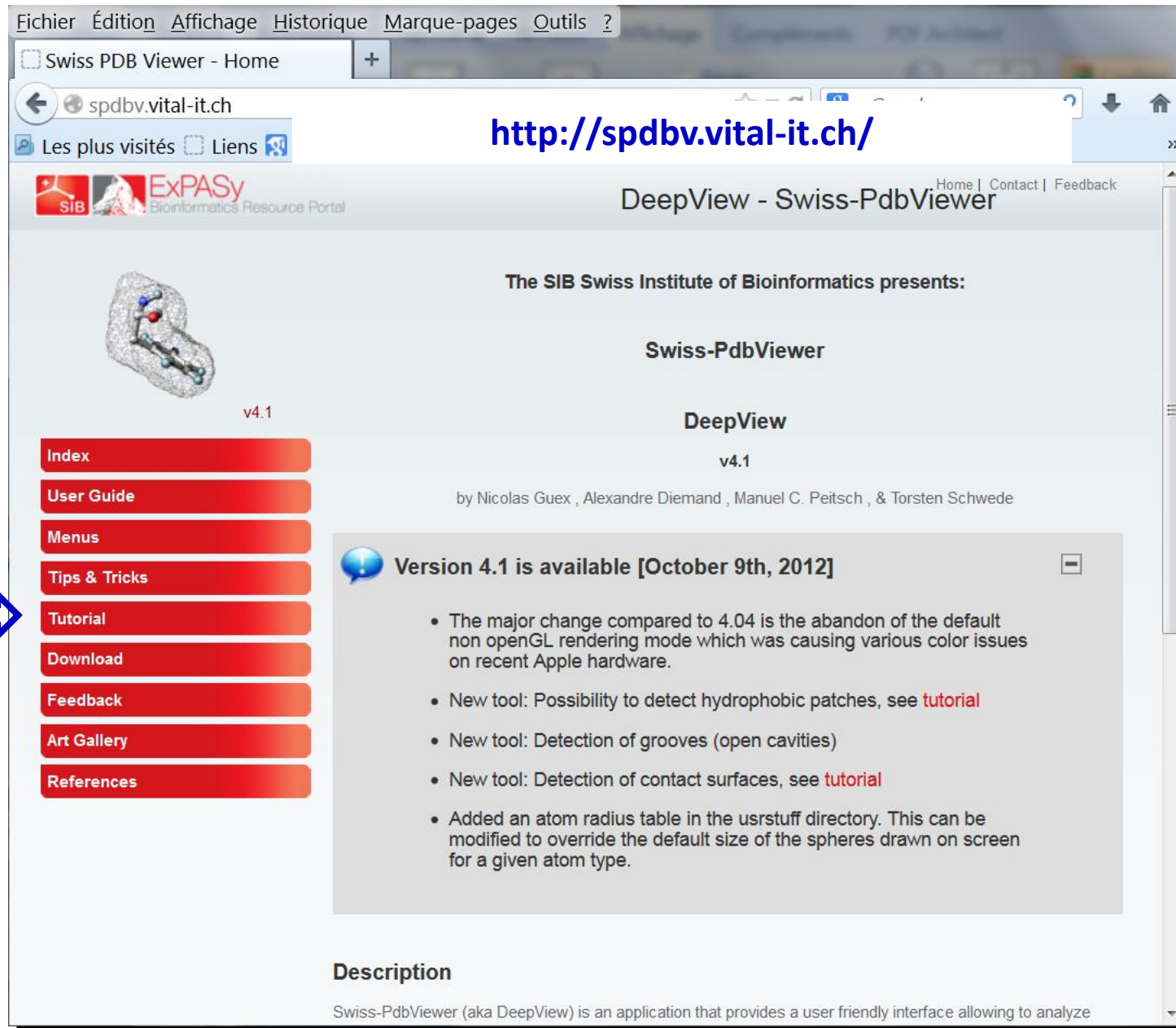
Les résidus donnant des liaisons H intra-moléculaires dans FruR sont conservés dans l'alignement et correspondent aux résidus donnant les liaisons H inter-moléculaires dans le complexe LacI-DNA d'où l'hypothèse d'une bascule des liaisons H de l'intra vers l'intermoléculaire



Structure du complexe 1efa

...

Modélisation avec Swiss-Pdb-Viewer



Fichier Édition Affichage Historique Marque-pages Outils ?

Swiss PDB Viewer - Home

spdbv.vital-it.ch

http://spdbv.vital-it.ch/

Les plus visités Liens

SIB ExPASy Bioinformatics Resource Portal

DeepView - Swiss-PdbViewer

Home | Contact | Feedback

The SIB Swiss Institute of Bioinformatics presents:

Swiss-PdbViewer

DeepView

v4.1

by Nicolas Guex , Alexandre Diemand , Manuel C. Peitsch , & Torsten Schwede

Version 4.1 is available [October 9th, 2012]

- The major change compared to 4.04 is the abandon of the default non OpenGL rendering mode which was causing various color issues on recent Apple hardware.
- New tool: Possibility to detect hydrophobic patches, see [tutorial](#)
- New tool: Detection of grooves (open cavities)
- New tool: Detection of contact surfaces, see [tutorial](#)
- Added an atom radius table in the usrstuff directory. This can be modified to override the default size of the spheres drawn on screen for a given atom type.

Description

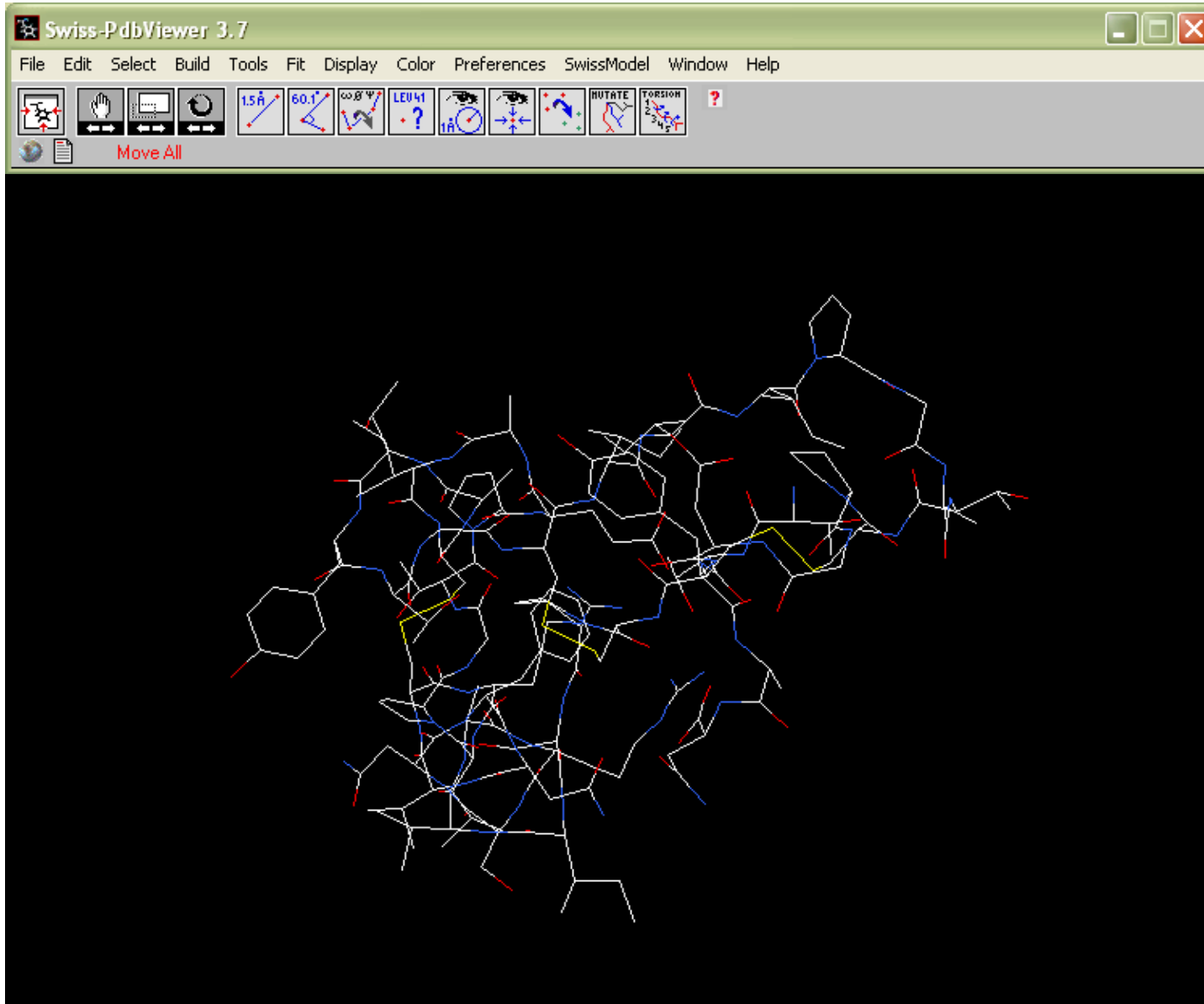
Swiss-PdbViewer (aka DeepView) is an application that provides a user friendly interface allowing to analyze



Nicolas Guex



- **Charger 3 fichiers de CRN**
- **Masquer les structures 2 et 3 (layers infos)**



- Menu Build loop
- Cliquer sur L 18 et Ala 24

Déviatiun angle en ° de la fermeture idéale

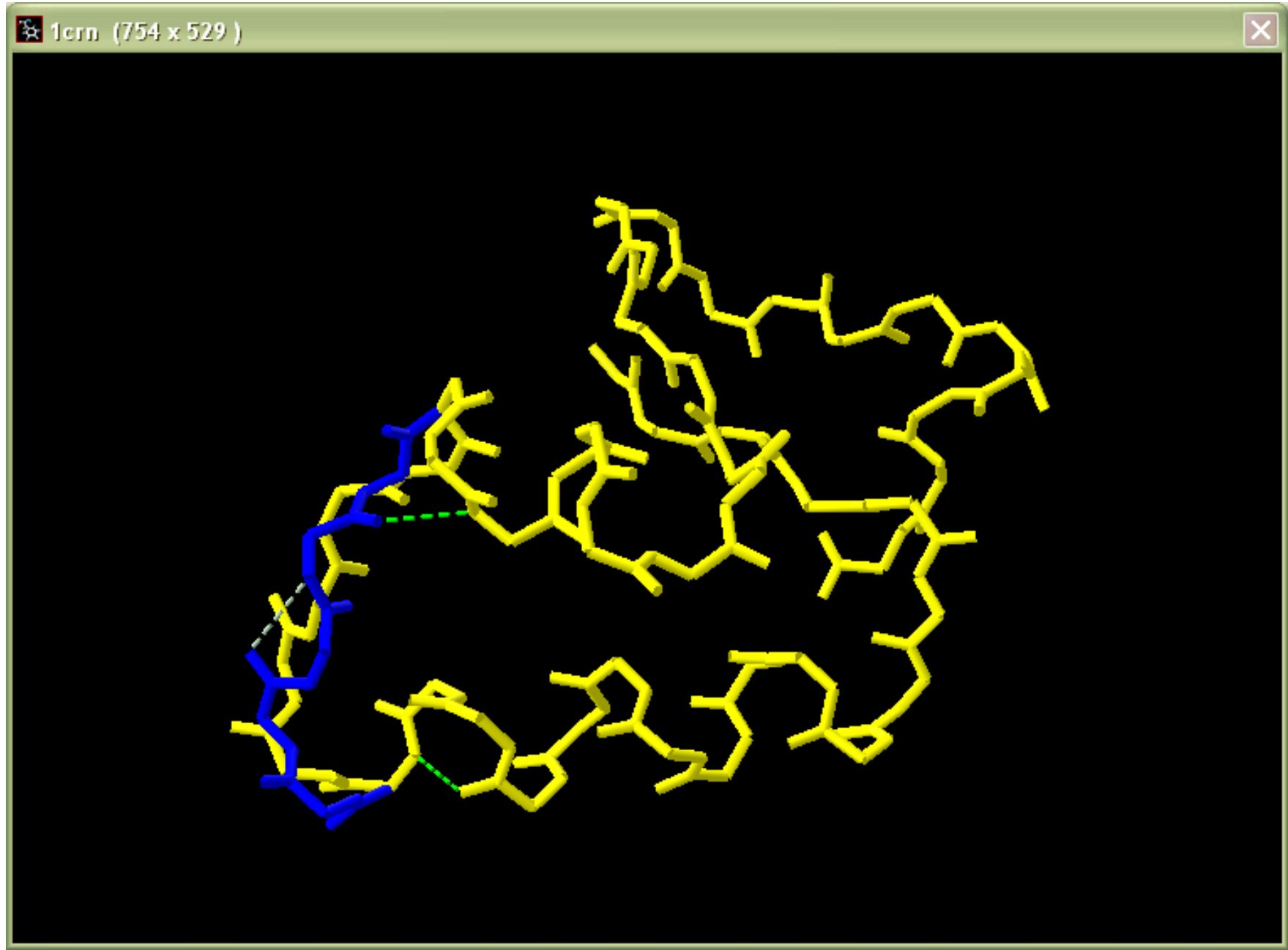
Energies (Force field ou champ moyen PP)

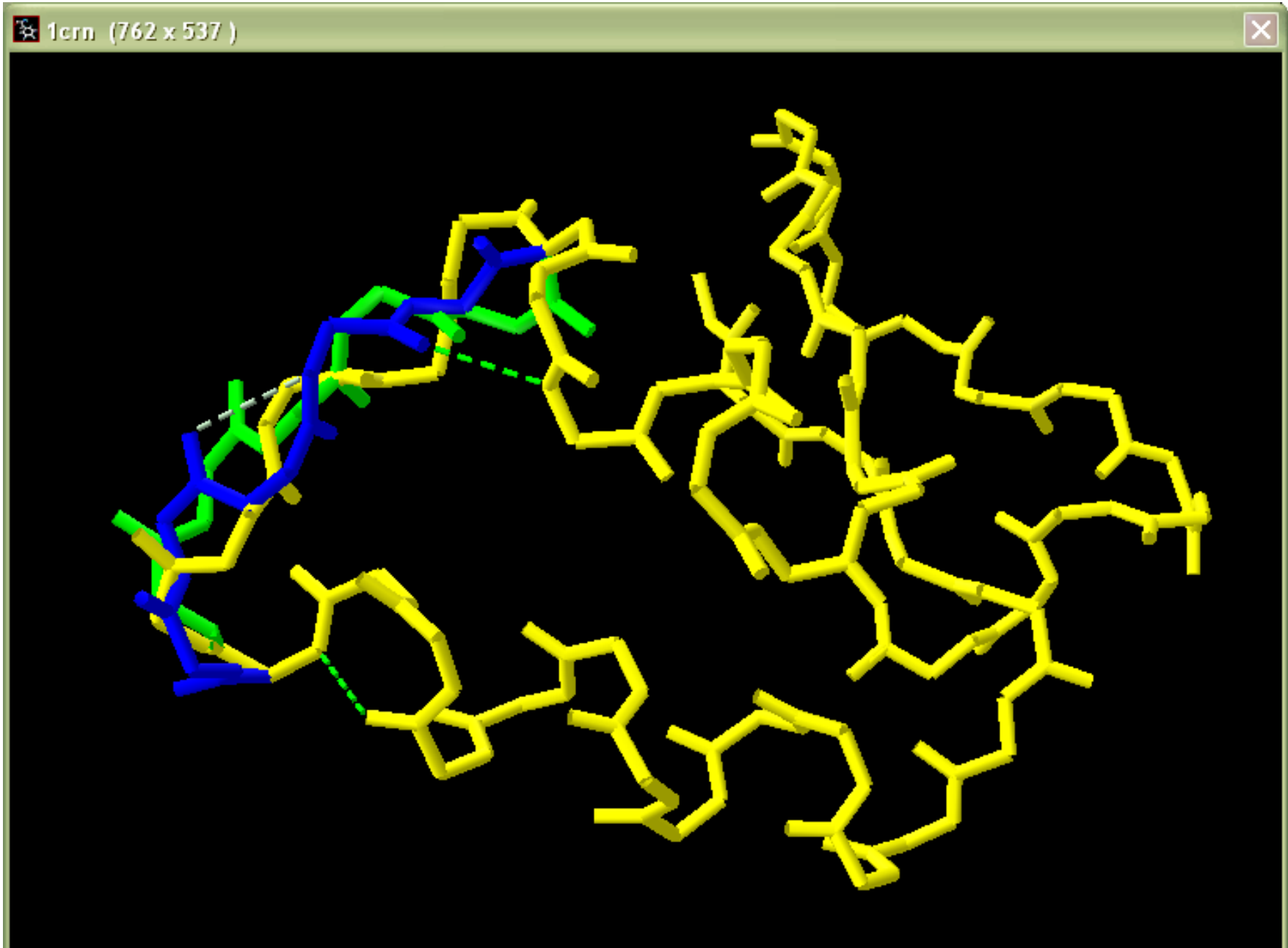
Déviatiun D (Å) de la fermeture idéale

C:\megacompil\Biochimie\spdt

clash score:0
PP:-10.56
FF:3329.4

C-N+	CA-C-N+	C-N+-CA+
0.04	-1.49	-7.07
0.06	-5.58	-8.02
-0.11	48.12	15.41
-0.11	48.59	12.73
0.36	22.29	-11.69
0.33	30.37	-1.97
0.47	4.69	-23.20
0.36	30.46	-10.55
0.02	30.43	-16.21
0.06	21.71	-10.97
0.25	30.82	-24.76
-0.27	9.26	20.27
-0.47	33.30	16.88
-0.15	43.50	26.43
-0.27	15.97	-2.21
0.16	-7.00	-28.18
-0.12	-6.97	-24.20
0.20	7.35	-24.71





Alignement

C:\megacompi\Biochimie\spdbv3_7\spdbv\temp...

```

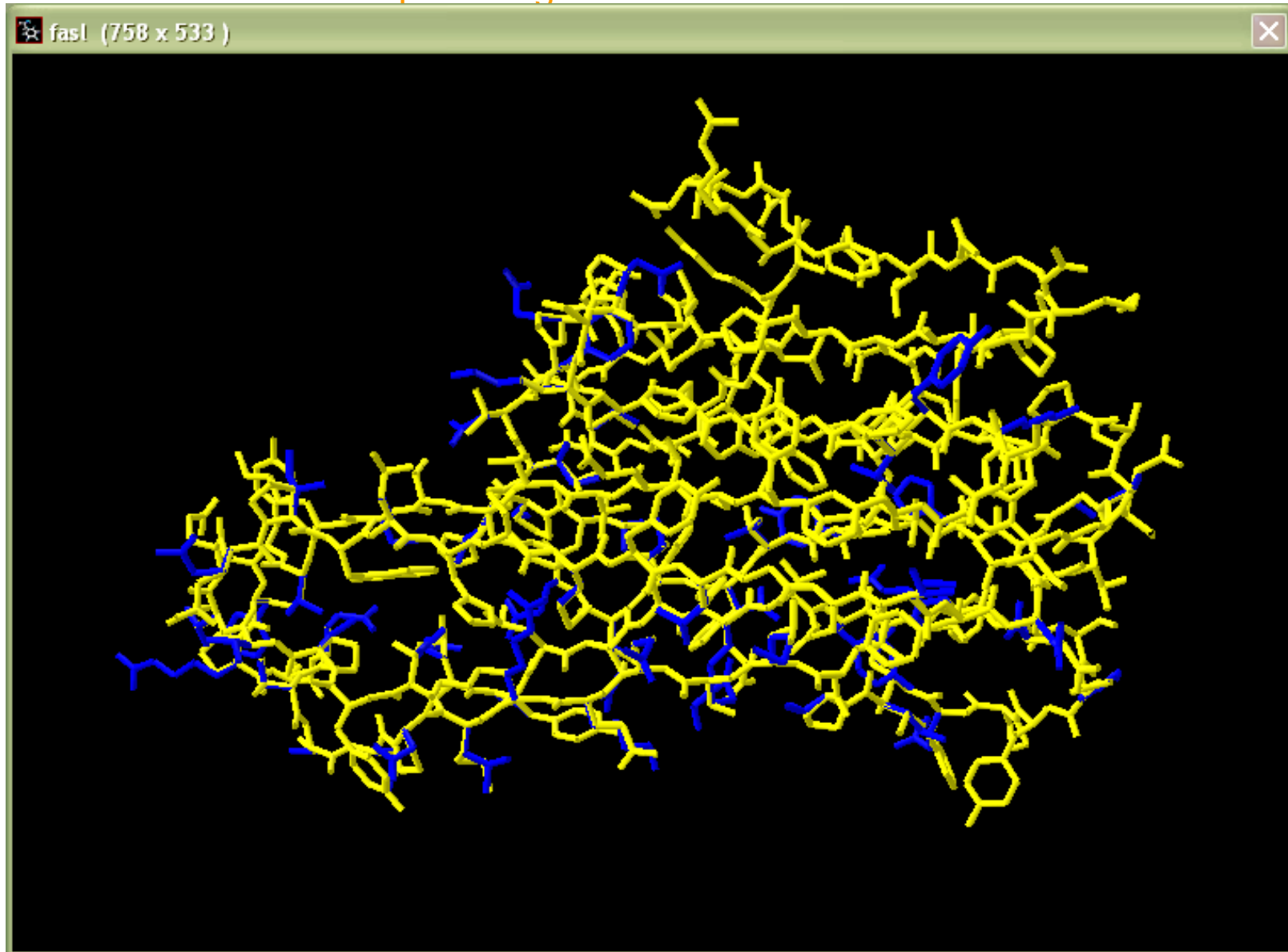
PGTPE          clash score:1  bad  G->X:0
. . . (-3)     bad Phi/Psi:2  bad  X->P:1
HLEHK          PP:-10.46      weak X->P:2
FF:707252.2    access:0.00      rms:0.00
  
```

2STV	10	HLEHK	2.50
2STV	92	VLNTA	2.50
1PYP	113	NNPID	3.00
4PTI	35	GCRAX	1.50
1EST	61	NQNNC	2.50
5ABP	130	KESAV	1.80
8ADH	140	GTSTF	2.40
8ADH	212	AGAAR	2.40
2TBV A	68	TEVGR	2.90
2TBV A	227	GAVVI	2.90
4RHV 1	211	VNEHD	3.00
4RHV 3	155	LQSTI	3.00
3RPZ A	188	GVAHG	1.90
1HVP A	25	GADDT	9.99
2SGA	116	SSGIV	1.50
3PGK	296	VTDKE	2.50
3PGK	296	VTDKE	2.50

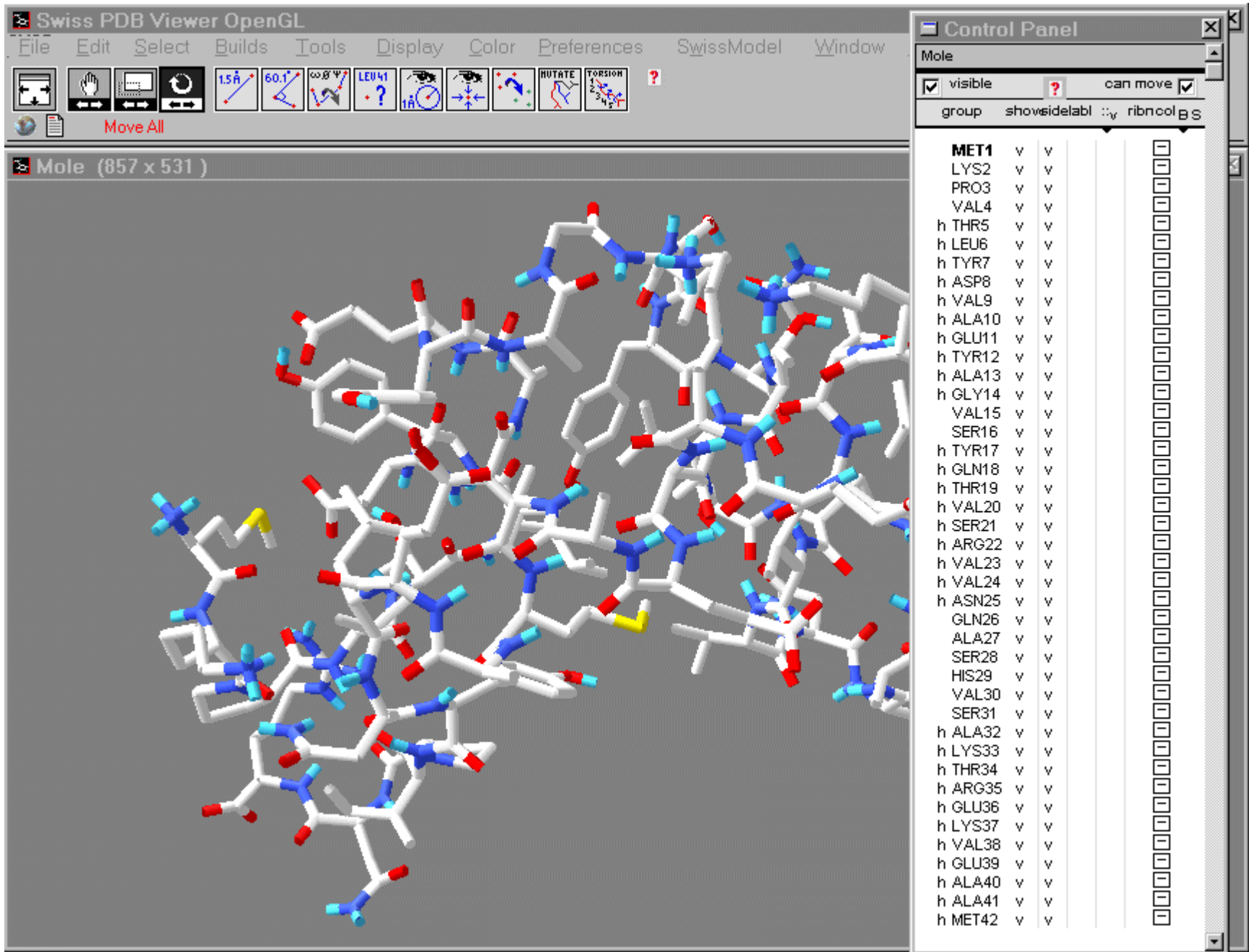
- **Un outil « automatique » de modélisation**
 - **Charger une structure (*.pdb) comme empreinte**
 - **Afficher la fenêtre de contrôle (menu windows)**
 - **Charger la séquence à modéliser**
 - **Choisir les régions à superposer**
 - **Demander une superposition itérative**
- **Observer la structure générée**
- **Réorienter les chaînes latérales**
- **Minimiser l'énergie**

Modélisation avec Swiss-Pdb-Viewer

Charger le fichier de séquence (Fasl.txt) load raw sequence
Load PDB file 1TNR-A puis magic fit



The screenshot displays the Swiss PDB Viewer OpenGL application window. The menu bar includes File, Edit, Select, Builds, Tools, Display, Color, Preferences, SwissModel, Window, and Help. A toolbar below the menu contains various icons for file operations and viewing options. The 'File' menu is open, showing options such as 'Open PDB File...' (highlighted with a blue circle and the number '1'), 'Open mmCIF File...', 'Open MOL File...', 'Open Text File...', 'Import...', 'Open Surface', 'Open Electrostatic Potential', 'Open Electron Density Map', 'Close', 'Close All Layers', 'Discard...', 'Save In Original Orientation', 'Save', 'Save Remote Job', 'Exit', and a list of recently opened files. Simultaneously, a file explorer window is open, showing the 'Execut' directory with a list of files including 'Bank', 'Dicro', 'Hpf', 'Hsf', '3sgb.pdb', '9abp.pdb', 'Crn.pdb', 'dimere.pdb', 'essai.pdb', 'fit_bad.pdb', 'Fit_hpf.pdb', 'fpe.pdb', 'frur.pdb', 'Gen_tmp.pdb', 'modele.pdb', 'modele1.pdb', 'modele2.pdb', 'modele3.pdb', 'Mole.pdb', 'Tempfile.pdb', and 'tme2.pdb'. The file explorer also shows the 'Nom' and 'Type' fields, with 'Type' set to 'Pdb Files (*.pdb *.ent)' and a 'Lecture seule' checkbox.



The screenshot displays the Swiss PDB Viewer OpenGL interface. The main window shows a 3D ball-and-stick model of a protein chain. The control panel on the right lists residues from MET1 to MET42, each with checkboxes for visibility and mobility.

group	showsidelabl	ribn	col	B	S
MET1	v	v			<input type="checkbox"/>
LYS2	v	v			<input type="checkbox"/>
PRO3	v	v			<input type="checkbox"/>
VAL4	v	v			<input type="checkbox"/>
h THR5	v	v			<input type="checkbox"/>
h LEU6	v	v			<input type="checkbox"/>
h TYR7	v	v			<input type="checkbox"/>
h ASP8	v	v			<input type="checkbox"/>
h VAL9	v	v			<input type="checkbox"/>
h ALA10	v	v			<input type="checkbox"/>
h GLU11	v	v			<input type="checkbox"/>
h TYR12	v	v			<input type="checkbox"/>
h ALA13	v	v			<input type="checkbox"/>
h GLY14	v	v			<input type="checkbox"/>
VAL15	v	v			<input type="checkbox"/>
SER16	v	v			<input type="checkbox"/>
h TYR17	v	v			<input type="checkbox"/>
h GLN18	v	v			<input type="checkbox"/>
h THR19	v	v			<input type="checkbox"/>
h VAL20	v	v			<input type="checkbox"/>
h SER21	v	v			<input type="checkbox"/>
h ARG22	v	v			<input type="checkbox"/>
h VAL23	v	v			<input type="checkbox"/>
h VAL24	v	v			<input type="checkbox"/>
h ASN25	v	v			<input type="checkbox"/>
GLN26	v	v			<input type="checkbox"/>
ALA27	v	v			<input type="checkbox"/>
SER28	v	v			<input type="checkbox"/>
HIS29	v	v			<input type="checkbox"/>
VAL30	v	v			<input type="checkbox"/>
SER31	v	v			<input type="checkbox"/>
h ALA32	v	v			<input type="checkbox"/>
h LYS33	v	v			<input type="checkbox"/>
h THR34	v	v			<input type="checkbox"/>
h ARG35	v	v			<input type="checkbox"/>
h GLU36	v	v			<input type="checkbox"/>
h LYS37	v	v			<input type="checkbox"/>
h VAL38	v	v			<input type="checkbox"/>
h GLU39	v	v			<input type="checkbox"/>
h ALA40	v	v			<input type="checkbox"/>
h ALA41	v	v			<input type="checkbox"/>
h MET42	v	v			<input type="checkbox"/>

Swiss PDB Viewer OpenGL

File Edit Select Builds Tools Display Color Preferences SwissModel Window Help

Compute H-bonds
 Compute Molecular Surface
 Compute Electrostatic Potential

Compute Energy (Threading)
 Compute Energy (Force Field)
 Energy Minimisation
 Fix Selected Sidechains

Fit molecules (from selection)
 Improve Fit...
 Magic Fit
 Iterative Magic Fit
 Generate Structural Alignment
 Calculate RMS...
 Set Layer Std Dev. into B-factors

Apply Transformation on Current Layer
 Build Crystallographic Symmetry
 Translate Layer along Unit Cell

Fix Atoms Nomenclature

Load Raw Sequence to Model...
 Load FoldFit Alignment...
 Save FoldFit Alignment...

Ignore Selected AA during modelling
 Use Selected AA during modelling
 Draw Residues to Ignore as TM

Lock Selected Residues of Model
 Unlock Selected Residues of Model

Homo Multimer Mode
 Build Preliminary Model
 Save Optimize Model Job

Update Threading Display Automatically
 Update Threading Display Now
 Auto Color by Threading Energy

Find Appropriate ExPdb Templates...
 Submit Modelling Request...

2

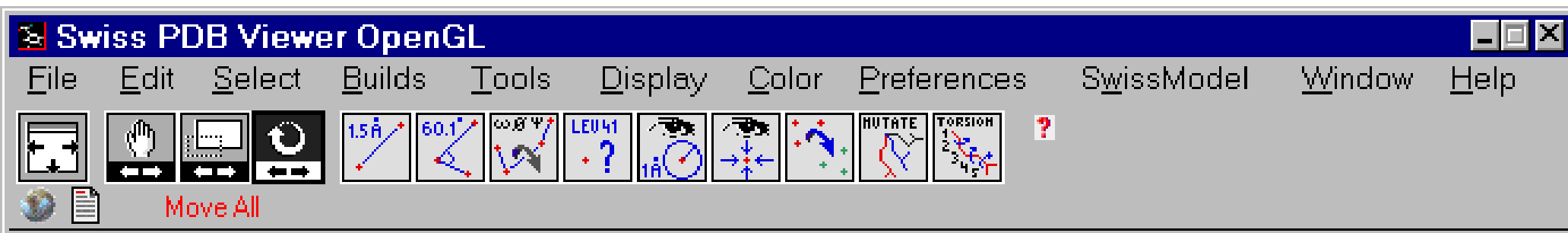
3

Align

Modele < M T V T I Y D V A R E A R V S M A T V S R V V N G N Q N V K P E T R D K V N E V I K K L N Y R P N A >

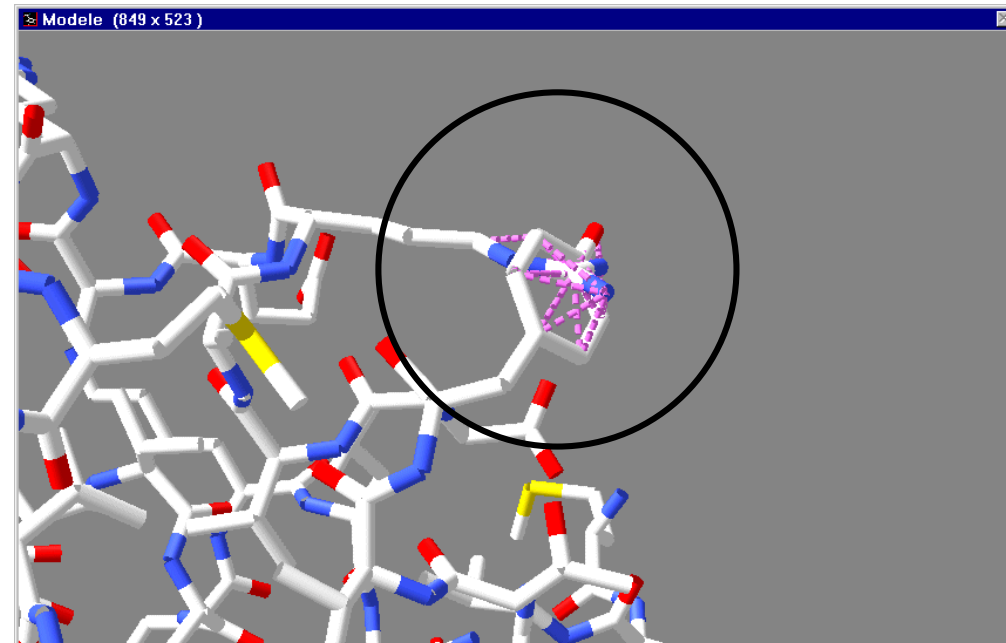
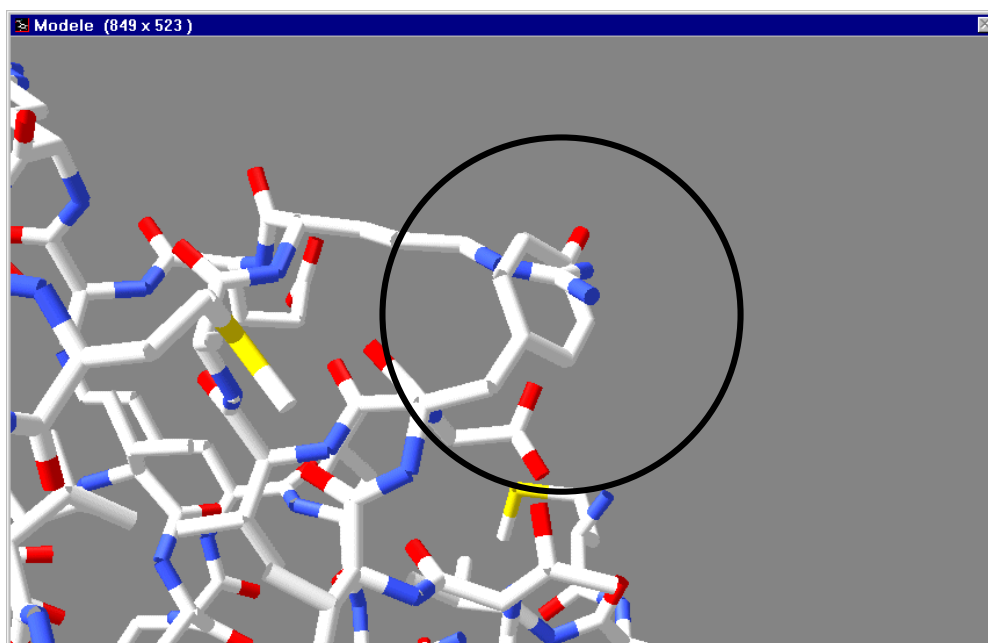
Mole MKP V T L Y D V A E Y A G V S Y Q T V S R V V N Q A S H V S A K T R E K V E A A M A E L N Y I P N R >

Mole: LEU45



Construction automatique

Mauvais contacts

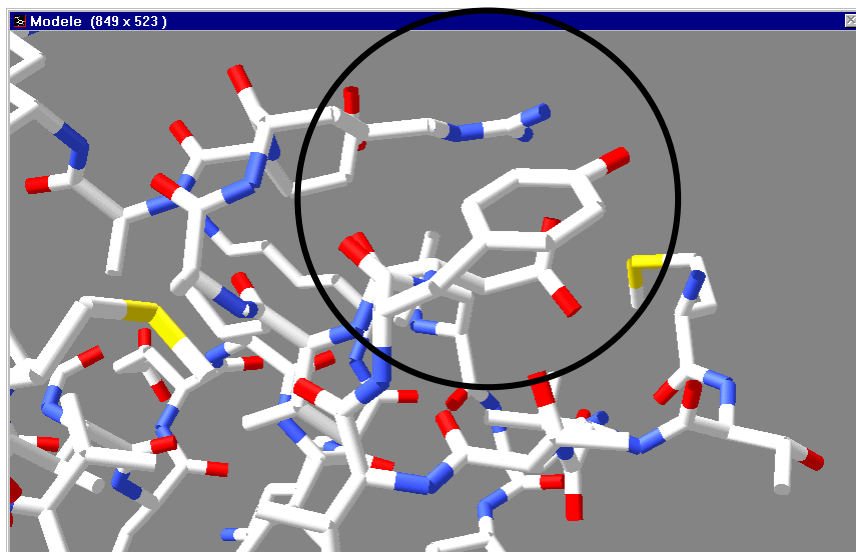


$$\Delta G_{\text{Tyr}_6} = 99999864 \text{ kJ/mole}$$

$$\Delta G_{\text{Arg}_{10}} = 99999624 \text{ kJ/mole}$$



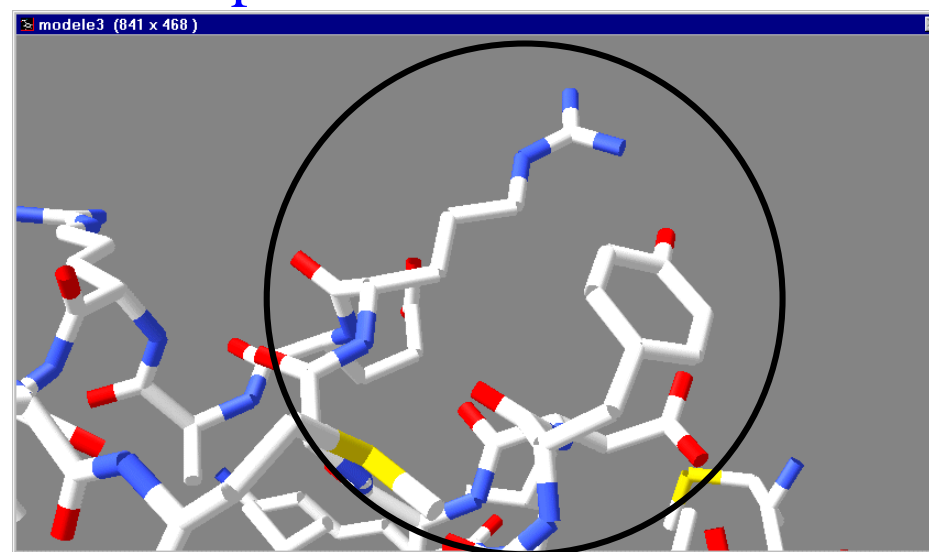
Réorientation elle



$$\Delta G_{\text{Tyr}_6} = +138 \text{ kJ/mole}$$

$$\Delta G_{\text{Arg}_{10}} = -191 \text{ kJ/mole}$$

Après minimisation

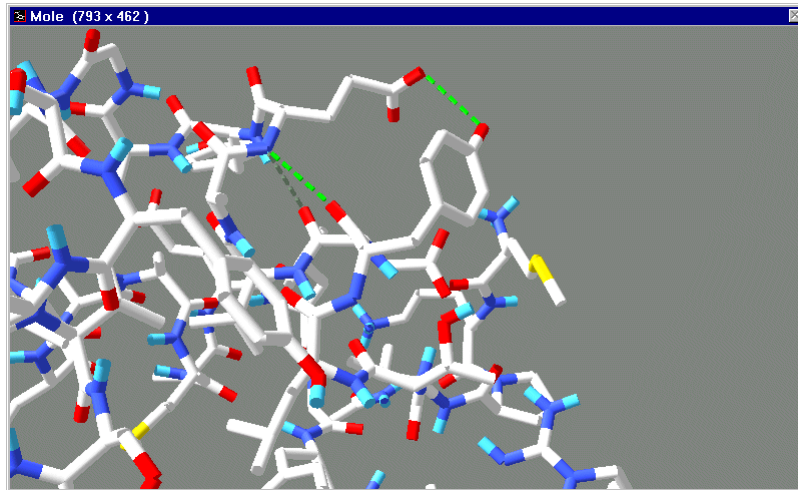


$$\Delta G_{\text{Tyr}_6} = -57 \text{ kJ/mole}$$

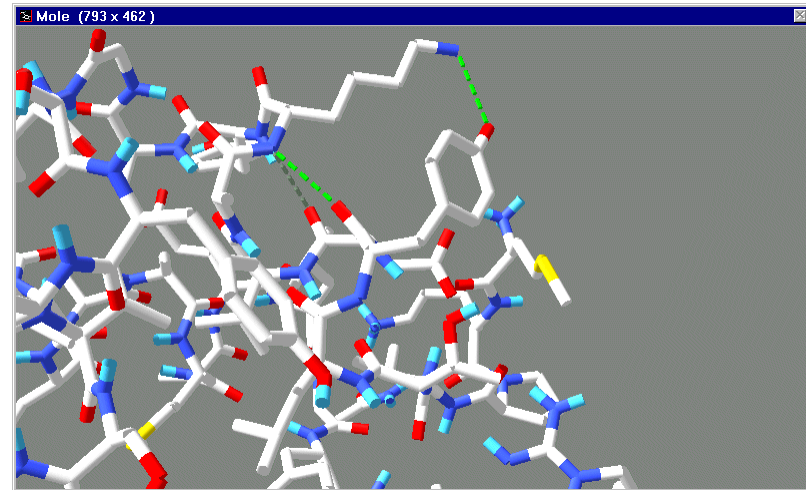
$$\Delta G_{\text{Arg}_{10}} = -273 \text{ kJ/mole}$$



Mutation automatique aa par aa

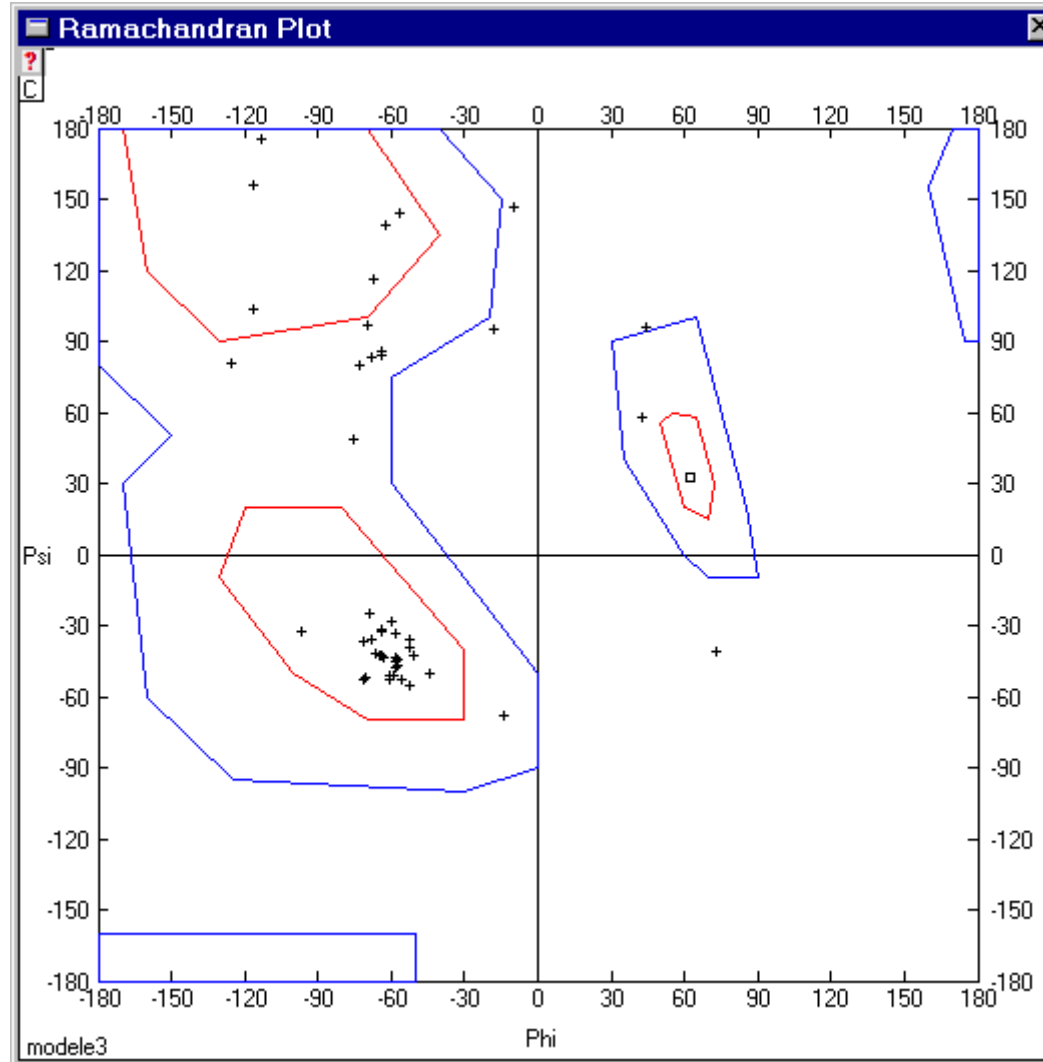


Rotamère optimal



Score = (4 x NbClash with backbone N CA and C atoms)
 + (3 x NbClash with backbone O atoms) + (2 x
 NbClash with sidechains atoms) - **NbHbonds** - 4 x Nb SSbonds

Total (kJ/mol) = bonds + angles + torsion + impropres + non liées + Electrostatique
 -3484 kJ/mol = 41,2 + 333 + 217 + 85 + (- 1445) + (- 2715)



Modélisation avec Geno3D


<http://geno3d-pbil.ibcp.fr>

GENO3D : AUTOMATIC MODELING OF PROTEINS THREE-DIMENSIONAL STRUCTURE - Mozilla F...

Eichier Édition Affichage Historique Marque-pages Outils ?

http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?p

GENO3D : AUTOMATIC MODELING OF ...



Pôle BioInformatique Lyonnais

Geno3D

Geno3D is the IBCP contribution to PBIL in Lyon, France

[HOME] [GENO3D] [HELP] [REFERENCES] [NEWS] [NPS@] [SuMo] [PBIL]

Tuesday, April 8th 2008: more options for PSI-BLAST search ([see news](#))
 Tuesday, April 27th 2010: A filesystem corruption on 24th and 27th causes Geno3D service downtime.

GENO3D Release 2 : AUTOMATIC MODELING OF PROTEINS THREE-DIMENSIONAL STRUCTURE

[Abstract] [[GENO3D help](#)] [Original server]

Database :

Sequence 1 name (optional) :

Paste protein 1 sequence below : [help](#)

```
VTTFFVALYDYESRTETDLSFKKGERLQIVNMTEGDWLLAHSLSSTGQTGYIPSNYVAPSD
SIQAEWYFGKITRRESERLLLNENPRGTFILVRESETTKGA YCLSVSDFDNAKGLNVKH
YKIRKLDSGGFYITSRTOFNSLQQLVA YYSKHADGLCHRLTTVCPTSKP
```

Terminé

Copier/coller de la séquence à modéliser

Combet, C., Jambon, M., Deléage, G. & Geourjon, C., Geno3D an automated protein modelling Web server, *Bioinformatics*, 2002, 18, 213-214

GENO3D : Geno3D2 START TO RUN MODEL - Mozilla Firefox

http://geno3d-pbil.ibcp.fr/cgi-bin/d3_geno3d2.pl

P B I L . i b c p . f r **Pôle BioInformatique Lyonnais**
Geno3D
 Geno3D is the IBCP contribution to PBIL in Lyon, France

[HOME] [GENO3D] [HELP] [REFERENCES] [NEWS] [NPS@] [SuMo] [PBIL]

Tuesday, April 8th 2008: more options for PSI-BLAST search ([see news](#))
 Tuesday, April 27th 2010: A filesystem corruption on 24th and 27th causes Geno3D service downtime.

Job **GENO3D2** (ID: 23227) is running on **GENO3D** server (started on 20101011-103340).
 Results will be shown below. **Please wait and don't go back.**

Run GENO3D2

FIRST STEP :
 Select template(s) to use for each chain in one or more pdb target :

PSI-BLAST run 3 for UNK_232270_0 □

TEMPLATE	E	FIRST	LAST	ID	ALIGNEMENT	COMMENT	NPSA link
<input type="checkbox"/> pdb1fmkA -0	1.000000e-53	1	168	100.000000	see alignment	PHOSPHOTRANSFERASE 24-JAN-97 1FMK CROSS_PDB	NPSA
<input type="checkbox"/> pdb1y57A -0	1.000000e-53	1	168	100.000000	see alignment	TRANSFERASE 02-DEC-04 1Y57 CROSS_PDB	NPSA
<input type="checkbox"/> pdb2ptkA -0	2.000000e-53	1	168	97.000000	see alignment	TYROSINE-PROTEIN KINASE 17-JUN-97 2PTK CROSS_PDB	NPSA
<input type="checkbox"/> pdb1kswA -0	5.000000e-53	2	168	100.000000	see alignment	TRANSFERASE 14-JAN-02 1KSW CROSS_PDB	NPSA
<input type="checkbox"/> pdb2srcA -0	5.000000e-53	2	168	100.000000	see alignment	TYROSINE-PROTEIN KINASE 29-DEC-98 2SRC CROSS_PDB	NPSA
<input type="checkbox"/> pdb2h8hA -0	2.000000e-52	3	168	100.000000	see alignment	TRANSFERASE 07-JUN-06 2H8H CROSS_PDB	NPSA
<input type="checkbox"/> pdb1qcfA -0	3.000000e-48	3	168	53.000000	see alignment	TYROSINE KINASE 04-MAY-99 1QCF CROSS_PDB	NPSA
<input type="checkbox"/> pdb1ad5A -0	3.000000e-48	3	168	53.000000	see alignment	TYROSINE-PROTEIN KINASE 20-FEB-97 1AD5 CROSS_PDB	NPSA

Terminé

Psi-Blast sur nr database ?

GENO3D : Geno3D2 START TO RUN MODEL - Mozilla Firefox

http://geno3d-pbil.ibcp.fr/cgi-bin/d3_geno3d2.pl

Pôle BioInformatique Lyonnais
Geno3D
Geno3D is the IBCP contribution to PBIL in Lyon, France

[HOME] [GENO3D] [HELP] [REFERENCES] [NEWS] [NPS@] [SuMo] [PBIL]

Tuesday, April 8th 2008: more options for PSI-BLAST search ([see news](#))
Tuesday, April 27th 2010: A filesystem corruption on 24th and 27th causes Geno3D service downtime.

Job **GENO3D2** (ID: 23227) is running on **GENO3D** server (started on 20101011-103340).
Results will be shown below. **Please wait and don't go back.**

Run GENO3D2

FIRST STEP :
Select template(s) to use for each chain in one or more pdb target :

PSI-BLAST run 3 for UNK_232270_0

TEMPLATE	E	FIRST	LAST	ID	ALIGNEMENT	COMMENT	NPSA link
<input checked="" type="checkbox"/> pdb1fmkA-0	1.000000e-53	1	168	100.000000	see alignment	PHOSPHOTRANSFERASE 24-JAN-97 1FMK CROSS_PDB	NPSA
<input checked="" type="checkbox"/> pdb1y57A-0	1.000000e-53	1	168	100.000000	see alignment	TRANSFERASE 02-DEC-04 1Y57 CROSS_PDB	NPSA
<input checked="" type="checkbox"/> pdb2ptkA-0	2.000000e-53	1	168	97.000000	see alignment	TYROSINE-PROTEIN KINASE 17-JUN-97 2PTK CROSS_PDB	NPSA
<input checked="" type="checkbox"/> pdb1kswA-0	5.000000e-53	2	168	100.000000	see alignment	TRANSFERASE 14-JAN-02 1KSW CROSS_PDB	NPSA
<input type="checkbox"/> pdb2srcA-0	5.000000e-53	2	168	100.000000	see alignment	TYROSINE-PROTEIN KINASE 29-DEC-98 2SRC CROSS_PDB	NPSA
<input type="checkbox"/> pdb2h8hA-0	2.000000e-52	3	168	100.000000	see alignment	TRANSFERASE 07-JUN-06 2H8H CROSS_PDB	NPSA
<input type="checkbox"/> pdb1qcfA-0	3.000000e-48	3	168	53.000000	see alignment	TYROSINE KINASE 04-MAY-99 1QCF CROSS_PDB	NPSA
<input type="checkbox"/> pdb1ad5A-0	3.000000e-48	3	168	53.000000	see alignment	TYROSINE-PROTEIN KINASE 20-FEB-97 1AD5 CROSS_PDB	NPSA

Terminé

Choix des empreintes (limité à 4 via le Web)

GENO3D : Geno3D2 START TO RUN MODEL - Mozilla Firefox

http://geno3d-pbil.ibcp.fr/cgi-bin/d3_geno3d2.pl

```
VTTFVALYDY ESRTETLSF KKGERLQIVN NTEGDWLLAH SLSTGQTGYI
PSNYVAPSDS IQAEEWYFGK ITRRESERLL LNAENPRGTF LVRESETTKG
AYCLSVSDFD NAKGLNVKHY KIRKLDGGF YITSRTQFNS LQLLVAYYSK
HADGLCHRLT TVCPTSKP
```

- This chain was modelled using 4 templates :

Template	Alignment	Secondary information (Sov)	Identity
pdb1fmkA_0	ali antheprot ali clustalw	--	100.0%
pdb1kswA_0	ali antheprot ali clustalw	--	100.0%
pdb1y57A_0	ali antheprot ali clustalw	--	100.0%
pdb2ptkA_0	ali antheprot ali clustalw	--	97.6%

- Template at each amino acid position :

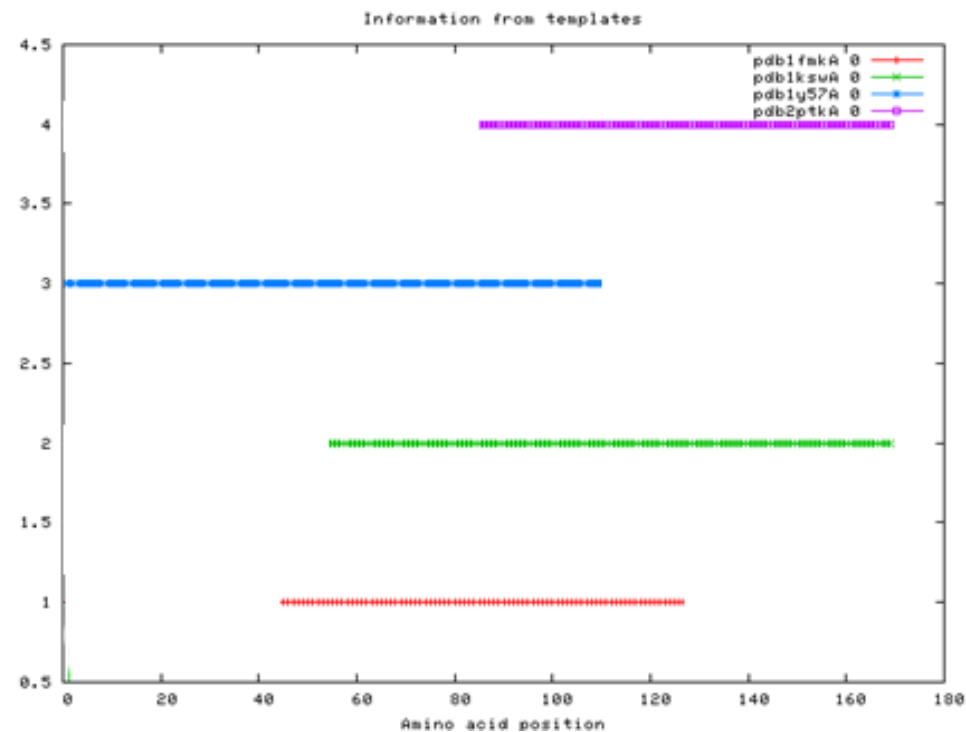
Information from templates

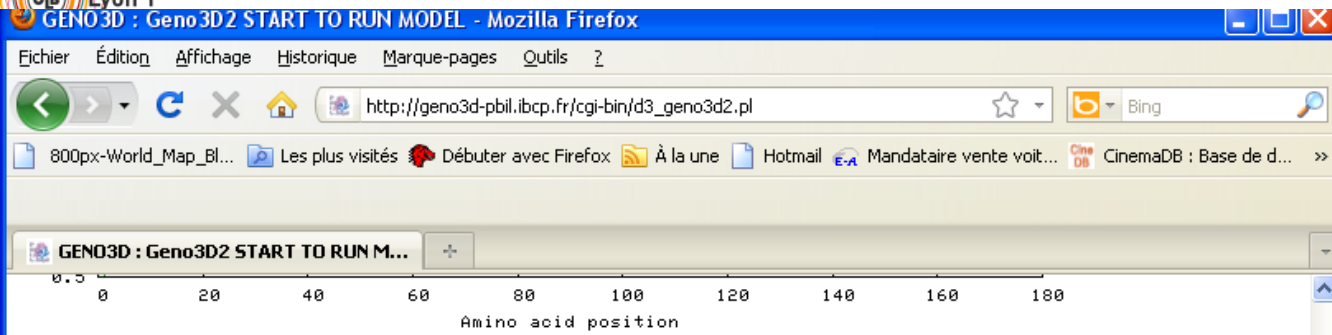
Information from templates

Terminé

Résumé du choix des empreintes

- %id
- Sov (si %id <30)
- Taux de couverture





- Deviation between templates on this chain (Angstrom) :

	1fmkA	1kswA	1y57A	2ptkA
1fmkA	0.00	0.66	2.23	1.05
1kswA	0.66	0.00	2.09	1.38
1y57A	2.23	2.09	0.00	2.77
2ptkA	1.05	1.38	2.77	0.00
. mean deviation :	1.698975			

1 Paramètres et soumission du job
attention à entrer un e-mail correct

SECOND STEP :
Enter your e-mail address :

2 Soumission du job

THIRD STEP :
Choose number of model to generate :

EXPERT OPTIONS :

Inter/intra restraints ratio :

Distance restraints cut off (Angstrom):

Margin in distance restraints (Angstrom):

Margin in angle restraints (degree) :

Maximal number of distance restraints :

Save full template in superposition : No Yes

LAUNCH Geno3D

User : public@193.51.160.224. Last modification time : Mon Oct 11 10:42:40 2010. Current time : Mon Oct 11 10:42:40 2010
This service is supported by 'Ministère de la recherche', 'Programme Bioinformatique inter-EPST', 'CNRS (IMABIO, COMI, GENOME) and Région Rhône-Alpes (Programme EMERGENCE)'. [Comments.](#)

Terminé

Pôle Bioinformatique Lyonnais
Geno3D is the IBCP contribution to FBIL in Lyon, France

[HOME] [GENOSD] [HELP] [REFERENCES] [NEWS] [NPS@] [SuMo] [FBIL]

Tuesday, April 8th 2008: more options for PSI-BLAST search ([see news](#))
Tuesday, April 27th 2010: A filesystem corruption on 24th and 27th causes Geno3D service downtime.

MODELING JOB 2010101110481629449 RUNNING !

Job ID : 2010101110481629449
You will receive an e-mail (g.deleage@ibcp.fr) when the molecular modelling will be finished ...

User : public@193.51.160.224. Last modification time : Mon Oct 11 10:48:20 2010. Current time : Mon Oct 11 10:48:20 2010 This service is supported by 'Ministère de la recherche', 'Programme Bioinformatique inter-EPST', 'CNRS (IMABIO, COMI, GENOME) and Région Rhône-Alpes (Programme EMERGENCE)'. [Comments.](#)

de PBIL WWW prod admin <pbiladm@ibcp.fr> ☆
 sujet **Geno3D2 Web Server : Started** 10:48
 pour g.deleage@ibcp.fr ☆

MOLECULAR MODELLING OF SEQUENCE 1 : UNK_232270_0
 COMMENT :
 SELECTED TEMPLATE(S) :

TEMPLATE	FSTAA	LSTAA	EXPECT	IDENT	COMMENT
pdb1fmkA-0	1	168	1.000000e-53	100	PHOSPHOTRANSFERASE 24-JAN-97 1FMK CROSS_PDB
pdb1y57A-0	1	168	1.000000e-53	100	TRANSFERASE 02-DEC-04 1Y57 CROSS_PDB
pdb2ptkA-0	1	168	2.000000e-53	97	TYROSINE-PROTEIN KINASE 17-JUN-97 2PTK CROSS_PDB
pdb1kswA-0	2	168	5.000000e-53	100	TRANSFERASE 14-JAN-02 1KSW CROSS_PDB

Mail de démarrage d'exécution
10H48

de PBIL WWW prod admin <pbiladm@ibcp.fr> ☆
 sujet **Geno3D2 Web Server : Ended**
 pour g.deleage@ibcp.fr ☆

MOLECULAR MODELLING OF SEQUENCE 1 : UNK_232270_0
 COMMENT :
 SELECTED TEMPLATE(S) :

TEMPLATE	FSTAA	LSTAA	EXPECT	IDENT	COMMENT
pdb1fmkA-0	1	168	1.000000e-53	100	PHOSPHOTRANSFERASE 24-JAN-97 1FMK CROSS_PDB
pdb1y57A-0	1	168	1.000000e-53	100	TRANSFERASE 02-DEC-04 1Y57 CROSS_PDB
pdb2ptkA-0	1	168	2.000000e-53	97	TYROSINE-PROTEIN KINASE 17-JUN-97 2PTK CROSS_PDB
pdb1kswA-0	2	168	5.000000e-53	100	TRANSFERASE 14-JAN-02 1KSW CROSS_PDB

Mail de terminaison
10H54

 JOB IDENTIFIER : 2010101110422427079
 NUMBER OF SEQUENCE TO MODEL : 1
 EMAIL ADDRESS : g.deleage@ibcp.fr
 NUMBER OF MODEL TO GENERATE : 3
 END-----

THE MODELING PROCESS HAS ENDED ON Mon Oct 11 10:54:03 CEST 2010

RESULTS ARE AVAILABLE FOR SEVEN DAYS at URL :

http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/tmp/MODELING/2010101110481629449/home.html

Lien sur les résultats

RESULTS OF MODELING 2010101110422427079

MOLECULAR 3D MODELS :

[.model_1.pdb](#)
[.model_2.pdb](#)
[.model_3.pdb](#)

Lien sur les modèles

MODEL ANALYSIS :

COMPLETE MODEL ANALYSIS (energy, ramachadran, ...) [report.txt](#)

BY CHAIN ANALYSIS (deviation, restraints, comparison to template(s)...):

- Chain 1 : [report_1.txt](#)

TEMPLATES VALIDATION : [valid.txt](#)

ARCHIVE FOR LOCAL ANALYSIS : [archive.tar.gz](#)

Les fichiers sont conservés 7 jours sur le serveur et possibilité de rapatrier l'archive complète des résultats et des fichiers

User : public@193.51.160.224 Last modification time : Mon Oct 11 10:53:49 2010 Current time : Mon Oct 11 11:11:13



Echier Édition Affichage Historique Marque-pages Outils ?

http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/tmp/MODELING/2010101

- Models energy (kcal/mol) :

```
. model 1 : -7127.33
. model 2 : -6952.99
. model 3 : -7029.19
. model 4 : -6932.40
. model 5 : 16589.70
. model 6 : -6851.69
. model 7 : -6922.69
. model 8 : -7159.96
. model 9 : 9671.65
. model 10 : -6973.96
```

Models with PROCHECK (Roman A. Laskowski) :

```
enerously disallowed
2.6% 0.7%
g postscript
all residue types : jpeg postscript
jpeg postscript
jpeg postscript
eg postscript
Model 2 75.5% 20.5% 2.6% 1.3%
. ramachandran plot : jpeg postscript
. ramachandran plots for all residue types : jpeg postscript
. main-chain parameters : jpeg postscript
. side-chain parameters : jpeg postscript
. residue properties : jpeg postscript
Model 3 69.5% 25.2% 2.0% 3.3%
. ramachandran plot : jpeg postscript
```

- Structural agreement between models (RMSD in angstrom) :

	Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Mod 6	Mod 7	Mod 8	Mod 9	Mod 10
Model1	0.00	0.89	1.13	0.99	0.95	0.96	1.05	0.76	0.97	0.87
Model2	0.89	0.00	1.25	1.01	1.10	0.94	1.01	0.88	0.90	0.96
Model3	1.13	1.25	0.00	1.18	1.15	1.14	1.32	1.13	1.18	1.18
Model4	0.99	1.01	1.18	0.00	1.05	1.01	1.11	0.98	1.04	0.99
Model5	0.95	1.10	1.15	1.05	0.00	1.05	1.15	0.93	1.06	1.03
Model6	0.96	0.94	1.14	1.01	1.05	0.00	1.09	0.93	0.89	0.86
Model7	1.05	1.01	1.32	1.11	1.15	1.09	0.00	1.08	1.09	0.94
Model8	0.76	0.88	1.13	0.98	0.93	0.93	1.08	0.00	0.89	0.91
Model9	0.97	0.90	1.18	1.04	1.06	0.89	1.09	0.89	0.00	0.91
Model 10	0.87	0.96	1.18	0.99	1.03	0.86	0.94	0.91	0.91	0.00

```
. mean deviation : 1.019661
. superimposed pdb file : super.pdb
. local deviation by residue : jpeg text
```

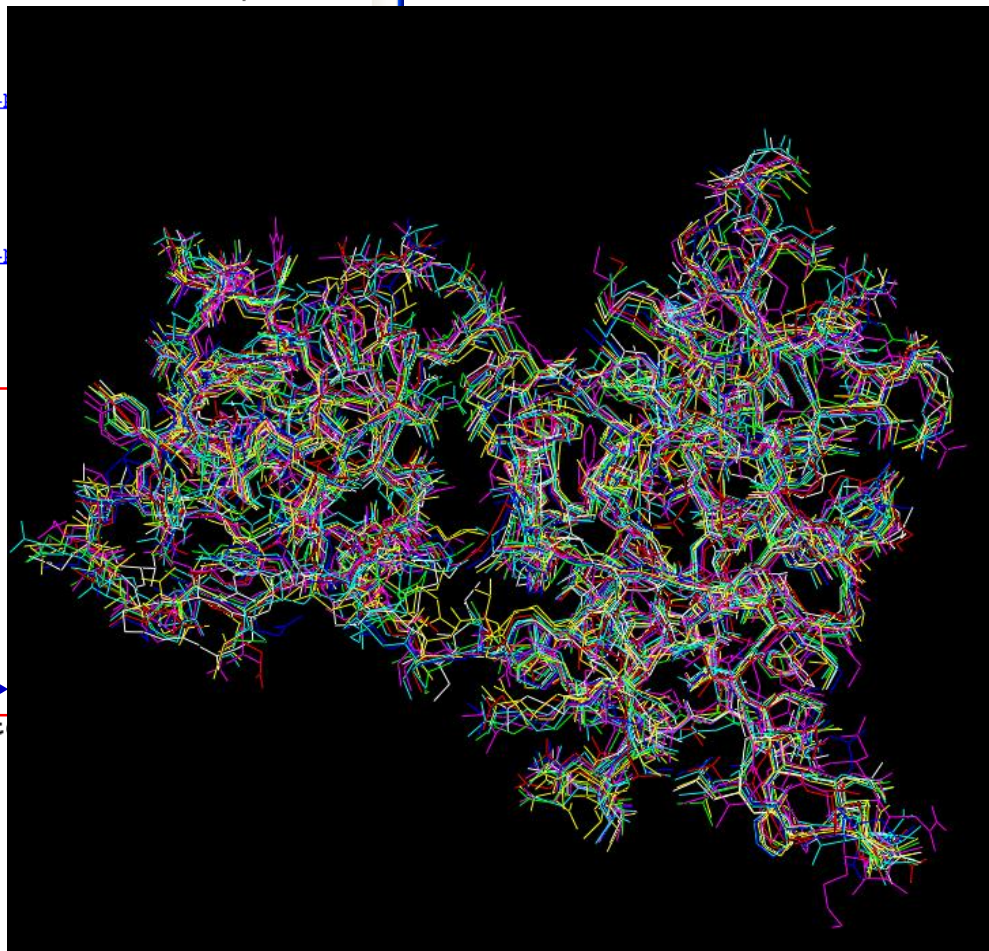
- Number of intrachain restraints deduced from template

```
. 6298 on MChain atoms (CA,C,N)
. 8959 on other atoms (CB,CE,OG,...)
. restrains repartition along sequence : jpeg text
```

- Number of violation of these intrachain restraints :

	MChain	Other
Model 1	181(2.90%)	161(1.80%)
Model 2	187(3.00%)	176(2.00%)
Model 3	195(3.10%)	188(2.10%)

. violation repartition along sequence : jpeg text



Geno3d result analysis - Mozilla Firefox

http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/tmp/MODELING/2010101

- Models energy (kcal/mol) :

```

.model 1 : -7127.33
.model 2 : -6952.99
.model 3 : -7029.19
.model 4 : -6932.40
.model 5 : 16589.70
.model 6 : -6851.69
.model 7 : -6922.69
.model 8 : -7159.96
.model 9 : 9671.65
.model 10 : -6973.96

```

Models with PROCHECK (Roman A. Laskowski) :

```

generously disallowed
2.6% 0.7%
g postscript
all residue types : jpeg pc
jpeg postscript
jpeg postscript
eg postscript
Model 2 75.5% 20.5% 2.6% 1.3%
.ramachandran plot : jpeg postscript
.ramachandran plots for all residue types : jpeg pc
.main-chain parameters : jpeg postscript
.side-chain parameters : jpeg postscript
.residue properties : jpeg postscript
Model 3 69.5% 25.2% 2.0% 3.3%
.ramachandran plot : jpeg postscript

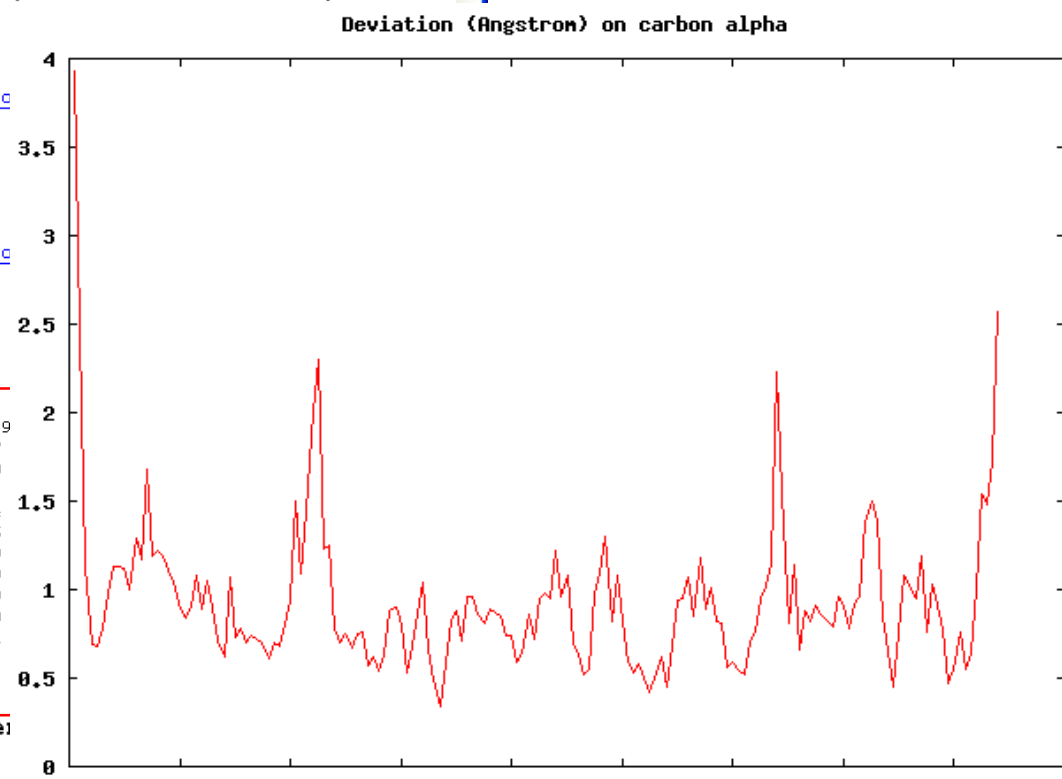
```

- Structural agreement between models (RMSD in angstrom) :

	Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Mod 6	Mod 7	Mod 8	Mod 9
Model1	0.00	0.89	1.13	0.99	0.95	0.96	1.05	0.76	0.97
Model2	0.89	0.00	1.25	1.01	1.10	0.94	1.01	0.88	0.90
Model3	1.13	1.25	0.00	1.18	1.15	1.14	1.32	1.13	1.18
Model4	0.99	1.01	1.18	0.00	1.05	1.01	1.11	0.98	1.04
Model5	0.95	1.10	1.15	1.05	0.00	1.05	1.15	0.93	1.06
Model6	0.96	0.94	1.14	1.01	1.05	0.00	1.09	0.93	0.89
Model7	1.05	1.01	1.32	1.11	1.15	1.09	0.00	1.08	1.09
Model8	0.76	0.88	1.13	0.98	0.93	0.93	1.08	0.00	0.89
Model9	0.97	0.90	1.18	1.04	1.06	0.89	1.09	0.89	0.00
Model 10	0.87	0.96	1.18	0.99	1.03	0.86	0.94	0.91	0.91

mean deviation : 1.019661
superimposed pdb file : [super.pdb](#)
local deviation by residue : [jpeg text](#)

- Number of intrachain restraints deduced from te



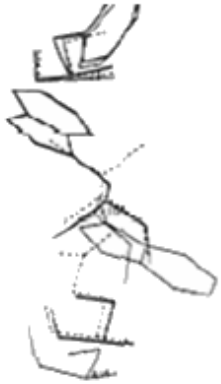
C:\Documents and Settings\Administrateur\Mes documents\Enseignement\Bioinformatique\2011100710354421384\ home_for_local_use.html

NUMBER OF VIOLATION OF CRYS PROTOCOL

	MChain	Other
Model 1	181(2.90%)	161(1.80%)
Model 2	187(3.00%)	176(2.00%)
Model 3	195(3.10%)	188(2.10%)

. violation repartition along sequence : [jpeg text](#)

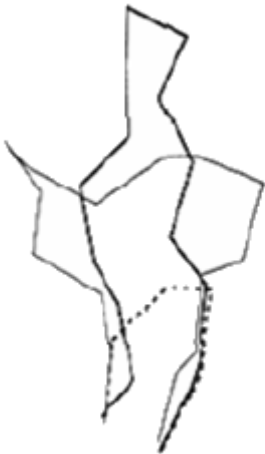




1) Erreur de positions de chaînes latérales



2) Distorsions et décalages



3) Absence de coordonnées dans le support (insertion et création de boucles)

4) Mauvais alignement



```
EDN ---KPPQFTWAQWFETQHINMTSQOCTNAMQ
7KSA KETAAAKFERQHMDSSSTAASSSNYCQMMK
      aaaaaaaaaa          aaaaaaa
```

5) Mauvais support



- ✓ Comparaison de différentes propriétés du modèle par rapport à la distribution statistique de ces mêmes propriétés dans une base de données: accessibilité, packing, formation du cœur hydrophobe, distribution spatiale des résidus chargés, volumes atomiques, liaison hydrogène .../...
- ✓ Estimation du repliement: Utilisation des potentiels de force moyenne et de profils 3D: Verify3D, Prosa, Harmony, Anolea.

Name	Type ^a	World Wide Web address ^b	Reference ^c
Model evaluation			
ANOLEA	S	www.fundp.ac.be/pub/ANOLEA.html	113
AQUA	P	www-nmr.chem.ruu.nl/users/rull/aqua.html	98
BIOTECH ^d	S	biotech.embl-ebi.ac.uk:8400/	73, 96
ERRAT	S	www.doe-mbi.ucla.edu/errata_server.html	40
PROCHECK	P	www.biochem.ucl.ac.uk/~roman/procheck/procheck.html	96
ProCeryon ^e	P	www.proceryon.com/	(d)
ProsaII ^e	P	www.came.sbg.ac.at	169
PROVE	S	www.ucmb.ulb.ac.be/UCMB/PROVE	134
SQUID	P	www.yorvic.york.ac.uk/~oldfield/squid	121
VERIFY3D	S	www.doe-mbi.ucla.edu/verify3d.html	105
WHATCHECK	P	www.sander.embl-heidelberg.de/whatcheck/	73

^aS, server; P, program.

^bSome of the sites are mirrored on additional computers.

^c(a) MolSoft Inc., San Diego. (b) Molecular Simulations Inc., San Diego. (c) Tripos Inc., St Louis. (d) ProCeryon Biosciences Inc. New York.

^dThe BIOTECH server uses PROCHECK and WHATCHECK for structure evaluation.

^eProCyon is a new software package that includes PeeP, PROFIT and PROSUP, a structure comparison program.

Librairie d'outils <http://zhanglab.ccmb.med.umich.edu/services/>

Logiciels/serveur de modélisation

Nom	Adresses Internet
Geno3D	http://geno3d-pbil.ibcp.fr
Modeller	http://salilab.org/modeller/
Swiss Model	http://swissmodel.expasy.org/
QUARK	http://zhanglab.ccmb.med.umich.edu/QUARK/
ROSETTA	http://rosetta.bakerlab.org/
MMM-tree Server	http://manaslu.aecom.yu.edu/MMMTree/
Prédiction de stabilité de mutants	http://zhanglab.ccmb.med.umich.edu/STRUM/

Logiciels/serveur de validation de modèles

Nom	Adresse Internet
ANOLEA	http://protein.bio.puc.cl/cardex/servers/anolea/
AQUA	http://nmr.cmbi.ru.nl/~jd/aqua/
ERRAT	http://nihserver.mbi.ucla.edu/ERRATv2/
PROCHECK	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/
ProsaII	https://prosa.services.came.sbg.ac.at/prosa.php
Verify 3D	http://nihserver.mbi.ucla.edu/Verify_3D/
Whatcheck	http://swift.cmbi.ru.nl/gv/whatcheck/

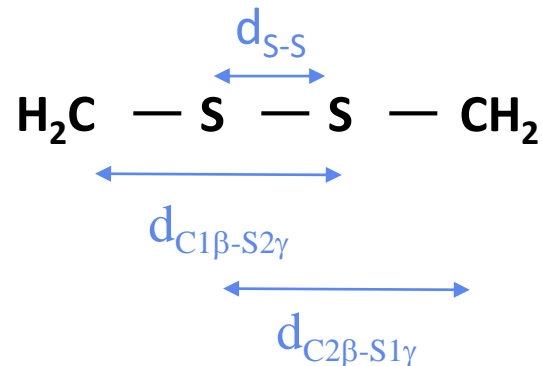
La protéine à modéliser à la même topologie, le même cœur structural, que la ou les empreintes utilisées

- Taux de similarité suffisant (supérieur à 30%)

ET / OU

- Preuves expérimentales : Activité biologique, Mutagénèse, Modifications chimiques , CD, Fluorescence
- Prédiction de structures secondaires (SOV automatique dans geno3D)

- Un outil « automatique » de modélisation sur le Web
- Prise en compte aisée des gaps et insertions
- Visualiser les zones peu fiables
- Possibilité d'utiliser plusieurs empreintes
- Générer un jeu de structures afin de pouvoir faire une étude statistique
- Possibilité de modéliser des protéines multi-domaines
- Utilisable même si le taux de similarité avec la protéine empreinte est faible (nécessité d'arguments biologiques)
- Possibilité de ne pas contraindre certaines régions ou au contraire introduire des contraintes externes (ponts SS)



$$d_{\text{S-S}} = 2,1 \text{ \AA}$$

$$d_{\text{C1}\beta\text{-S2}\gamma} = d_{\text{C2}\beta\text{-S1}\gamma} = 3,1 \text{ \AA}$$

- Générer à partir de la ou des empreintes des contraintes de distances et d'angles dièdres (Phi, Psi, Chi-1 et Oméga)
- Ces contraintes sont appliquées à la protéine que l'on souhaite modéliser
- La stratégie de modélisation est comparable à celle utilisée dans le cas de la modélisation sous contraintes RMN

- Générer une superposition des structures empreintes
- Calcul d'une déviation standard pour chacun des atomes communs
- A partir de ces données statistiques génération des contraintes en tenant compte des insertions, des gaps, et de la conservation des acides aminés et des atomes
- Run de modélisation moléculaire sous contraintes avec le logiciel X-PLOR 3.85

- Paramètres internes à **analyse_mod**

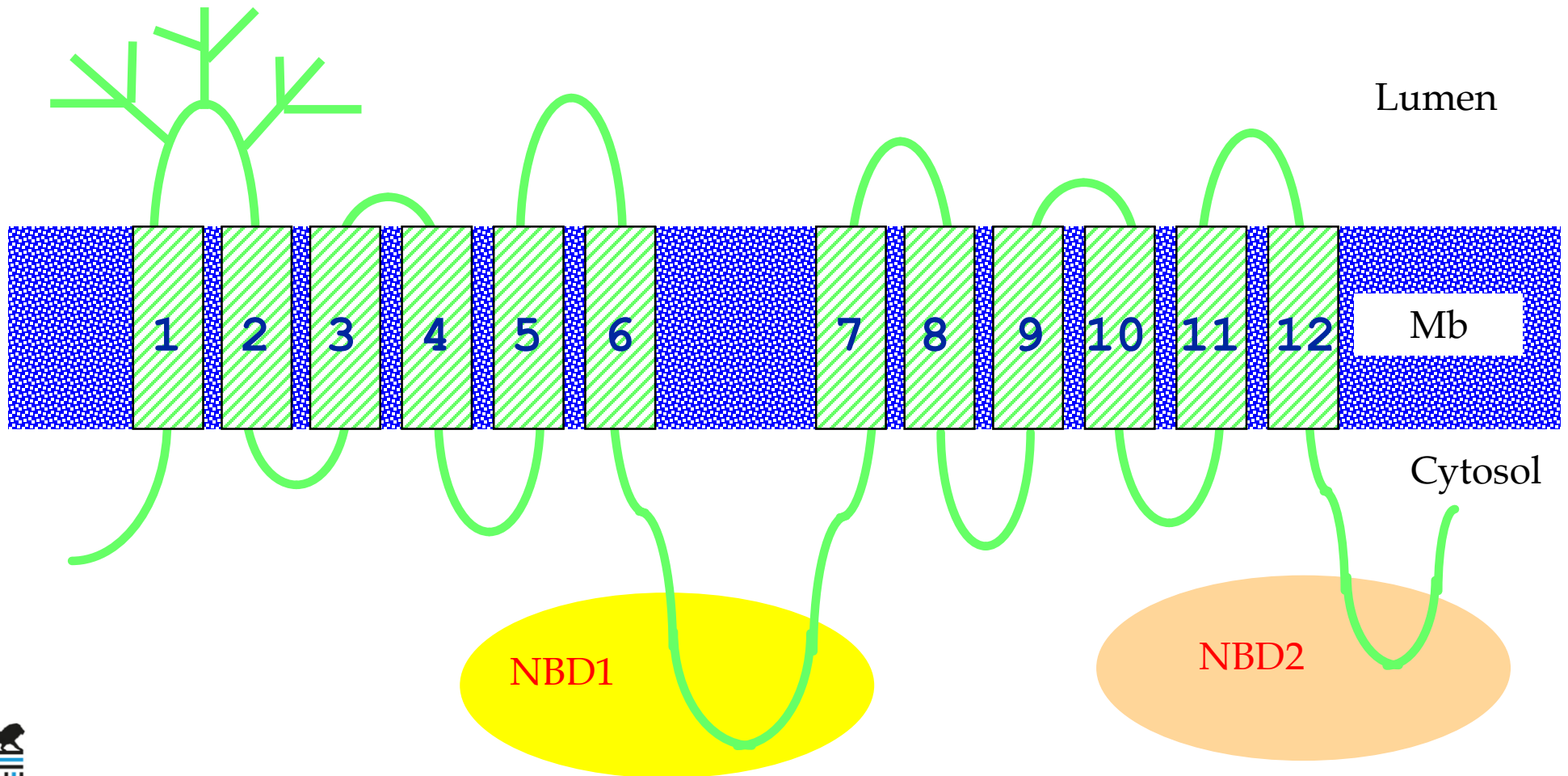
- F_SUPERPOSE, F_ALIGN, D_STAT, F_OUT_SUPER, F_OUT_RMSD, I_PAS1, I_PAS2, D_PDB, D_FICHER, CUT_OFF_DIST, CUT_OFF_HYDROPHOBE, CUT_OFF_HYDROGEN, CUT_OFF_SALIN, FREQ_OBS_INTERAC, N_STR_REF, OMEGA, CHI1, SSBRIDGE, WITHOUT MODEL, HELIX, SHEET, TURN, COIL, DISTANCE

- Paramètres internes à **X-PLOR**

- NRES_XPLOR, NASSIGN_XPLOR, NB_STR_XPLOR, NB_COOL_DGSA, NB_COOL_REFINE, NB_STEP_MINIM

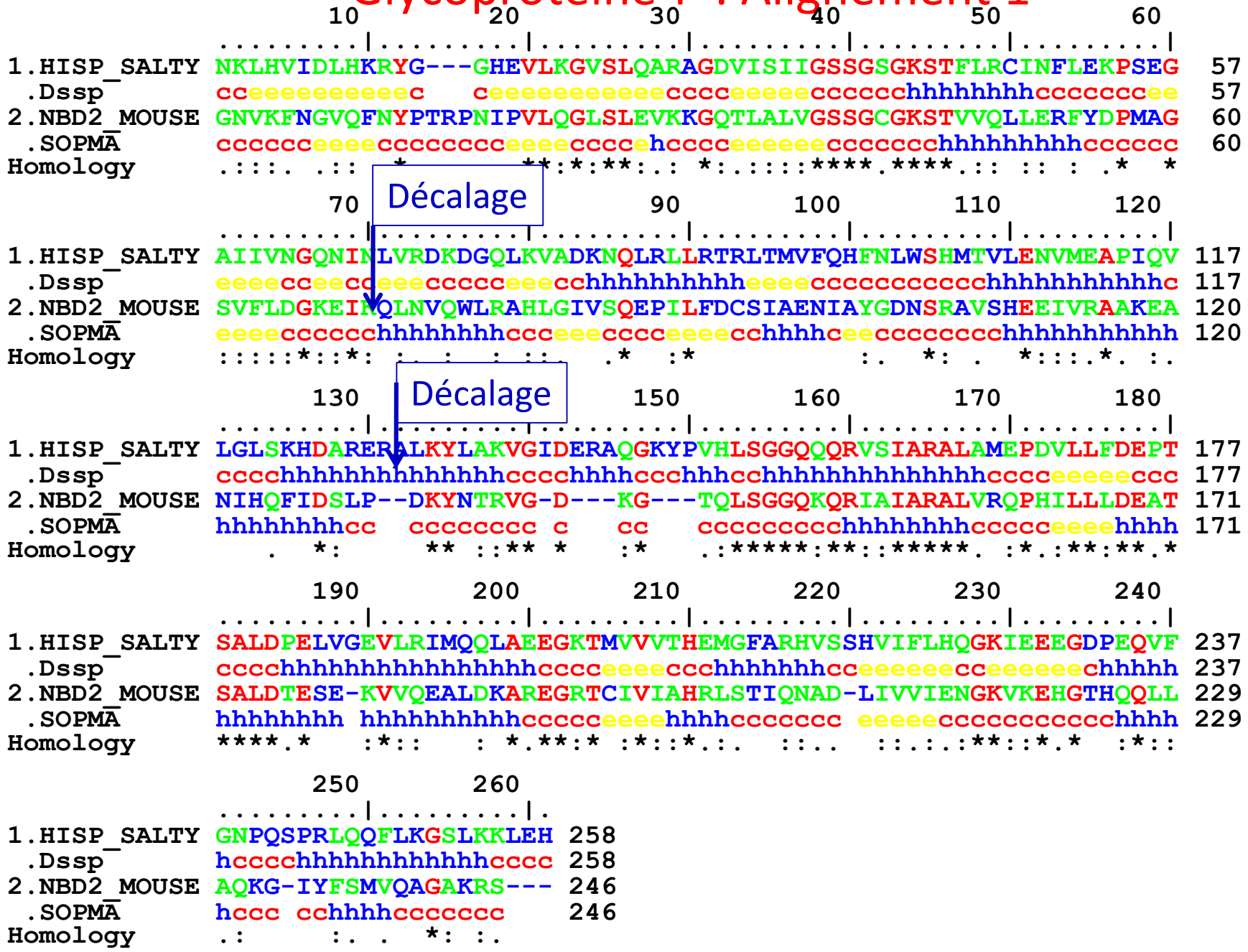
- Modélisation du 2^{ème} domaine de fixation de l'ATP de la Glycoprotéine-P (eucaryote)
- Protéine gp41 de VIH
- Protéine de capsid d'adénovirus AAV2
- Protéine NR-13 (protéine anti-apoptose)

- Protéine membranaire
 - Rôle de résistance aux agressions chimiques
 - Rejette les agents chimiques hors de la cellule
 - Transport actif peu spécifique dépendant de l'hydrolyse d 'ATP
- Cause de la résistance à la chimiothérapie au long cours



- Modélisation du 2^{ème} domaine de fixation de l'ATP de la Glycoprotéine-P (eucaryote)
 - Empreinte : HISP (Histidine transport ATP-binding protein from salmonella)
 - Optimisation de l'alignement en fonction
 - des prédictions de structures secondaires
 - de l'alignement du premier domaine avec la protéine HISP
 - Les régions avec gaps ou insertions dans une des deux séquences ne sont pas contraintes
 - Utilisation des contraintes sur les angles dièdres

Glycoprotéine-P : Aligement 1

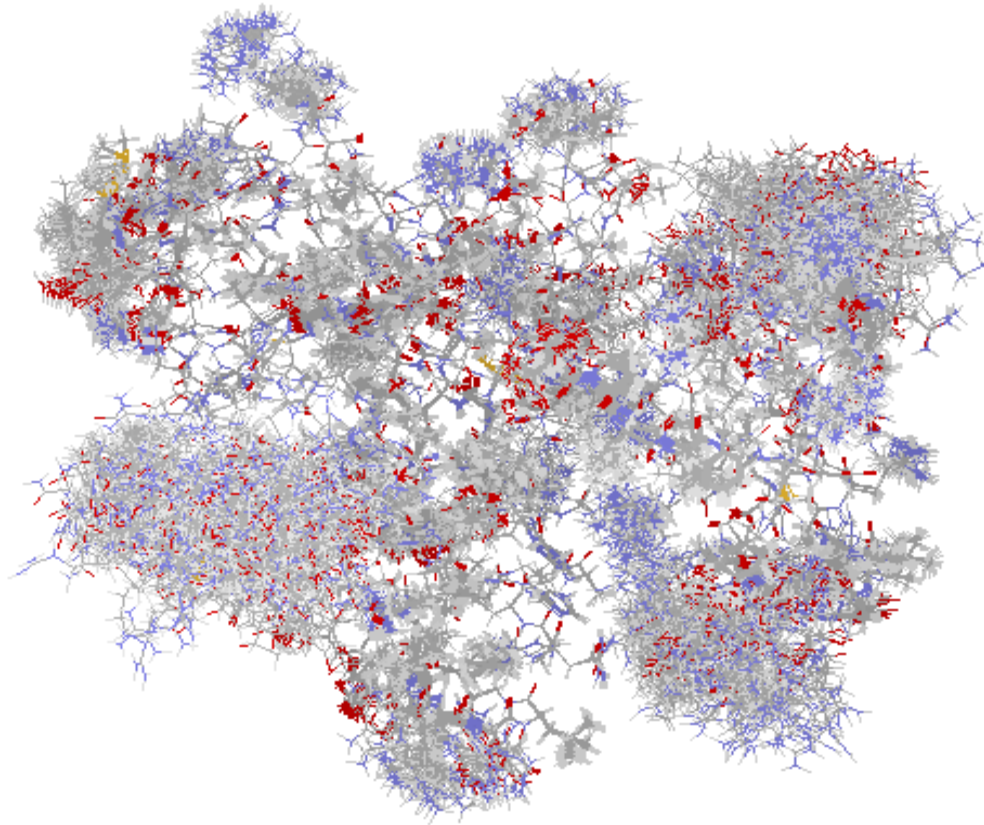


Glycoprotéine-P : Alignement 2

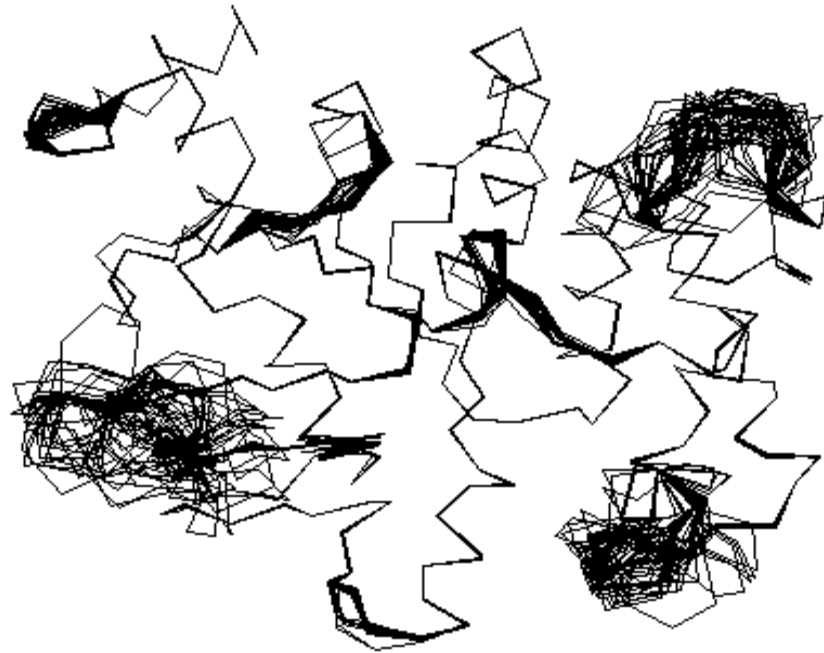
	10	20	30	40	50	60	
1.HISP_SALTY	NKLHV	IDLHKRYG	---	GHEVLKGVSLQARAGDVISIIGSSSGKSTFLRCINFLEKPSEG			57
.Dssp	cccccccccccc	cccccccccccc	cccccccccccc	cccccccccccc	chhhhhhhhh	cccccccc	57
2.NBD2_MOUSE	GNVKFNGVQFNYPTRPNI	PVLQGLSLEVKKGQTLALVGSSGCGKSTVVQLLERFYDPMAG					60
.SOPMA	cccccccccccc	cccccccccccc	cccccccccccc	cccccccccccc	chhhhhhhhh	cccccccc	60
Homology				**::: : * *		
<div style="border: 1px solid blue; padding: 5px; display: inline-block;">Déplacement d'insertion et regroupement</div>							
					110	120	
1.HISP_SALTY	AIIVNGQINLVRDKDGQLKVADKNQLRLLRTRLT	MVFQHFNLWSHMTVLENVMEAPIQ-					116
.Dssp	eeeeccccc	cccccccccccc	chhhhhhhhh	eeeecccccccc	chhhhhhhhh		116
2.NBD2_MOUSE	SVFLDGKEI	-----	KQLNVQWLR	HLGIVSQEP-ILFDCSIAENIAYGDNSR			106
.SOPMA	eeeeccccc		chhhhhhhhh	cccccccc	eeeecchhhh	cccccc	106
Homology	:::.*::*		:: : **::* * *	: * *	: : **:	.	
	130	140	150	160	170	180	
1.HISP_SALTY	-VLGLSKHDARERALKYLAKVG	---	IDERAQGKYPVHLSGGQQQRVSIARALAMEPDVLL				172
.Dssp	ccccchhhhhhhhhhhhhhhcc	cchhhh	ccchhh	ccchhhhhhhhhhhhhhhhh	ccccccc		172
2.NBD2_MOUSE	AVSHEEIVRAAKEANIHQFIDSLPDKYNTRVGDKGTQLSGGQKQRIARALVRQPHILL						166
.SOPMA	ccchhhhhhhhhhhhhhhhhhhhh	cccccccccccccccccccccccccccc	chhhhhhhhh	ccccccc			166
Homology	* . * :.* :	.	:	* . :*****:**:*****.	:*:* **		
	190	200	210	220	230	240	
1.HISP_SALTY	FDEPTSAIDPELVGEVLRIMQQLAEEGKTMVVVTHEMGFARHVSSHVIFLHQKIEEGD						232
.Dssp	eeeeccccchhhhhhhhhhhhhhhhh	cccccccc	ccchhhhhhh	cccccccc	cccccccc	cccccccc	232
2.NBD2_MOUSE	LDEATSAIDTESEKVVQEAIDKARE-GRTCIVIAHRLSTIQNAD-LIVVIENGKVKEHGT						224
.SOPMA	ehhhhhhhhhhhhhhhhhhhhhhh	ccccccc	hhhh	ccccccc	eeeecccccccc	cccccccc	224
Homology	::**.***.*	* . :::	* **	:::* * . .	::: .	::: . **:* *	
	250	260					
1.HISP_SALTY	PEQVFGNPQSPRLQQFLKGS	LKKLEH					258
.Dssp	hhhhhh	ccccchhhhhhhhhhhhhhh	cccc				258
2.NBD2_MOUSE	HQQLLAQKG-IYFSMVQAGAKRS	---					246
.SOPMA	chhhhh	cccc	cchhhh	ccccccc			246
Homology	::: . .	* : . .					



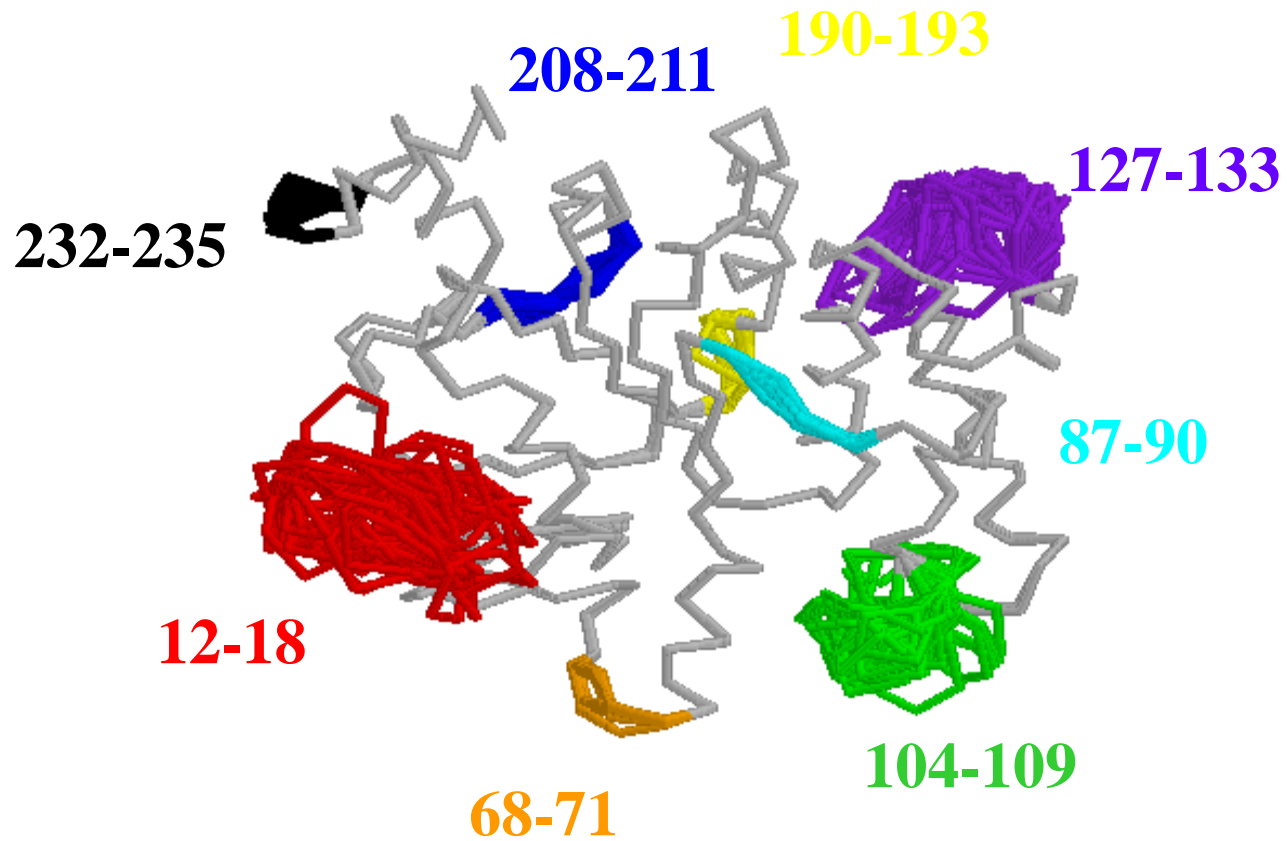
Glycoprotéine-P : Résultat



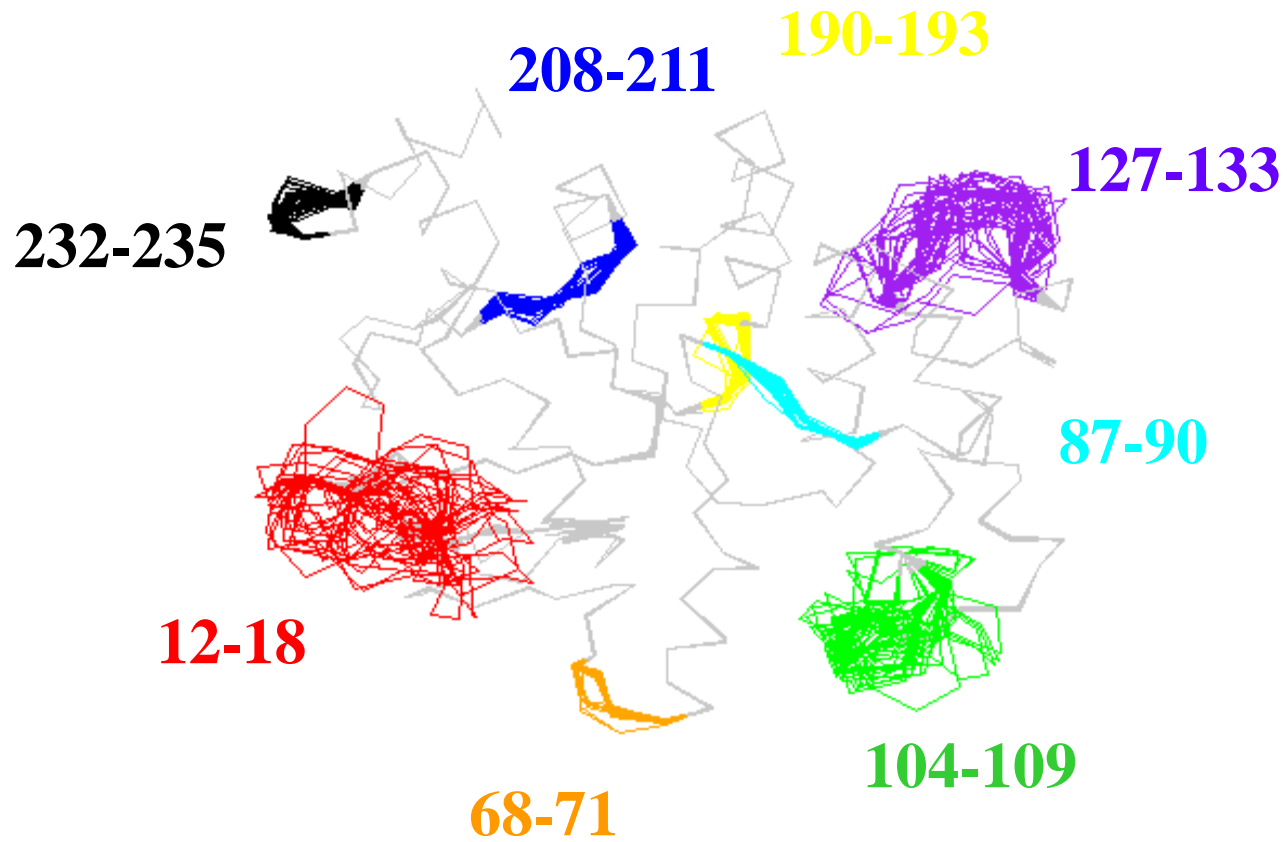
Glycoprotéine-P : Résultat



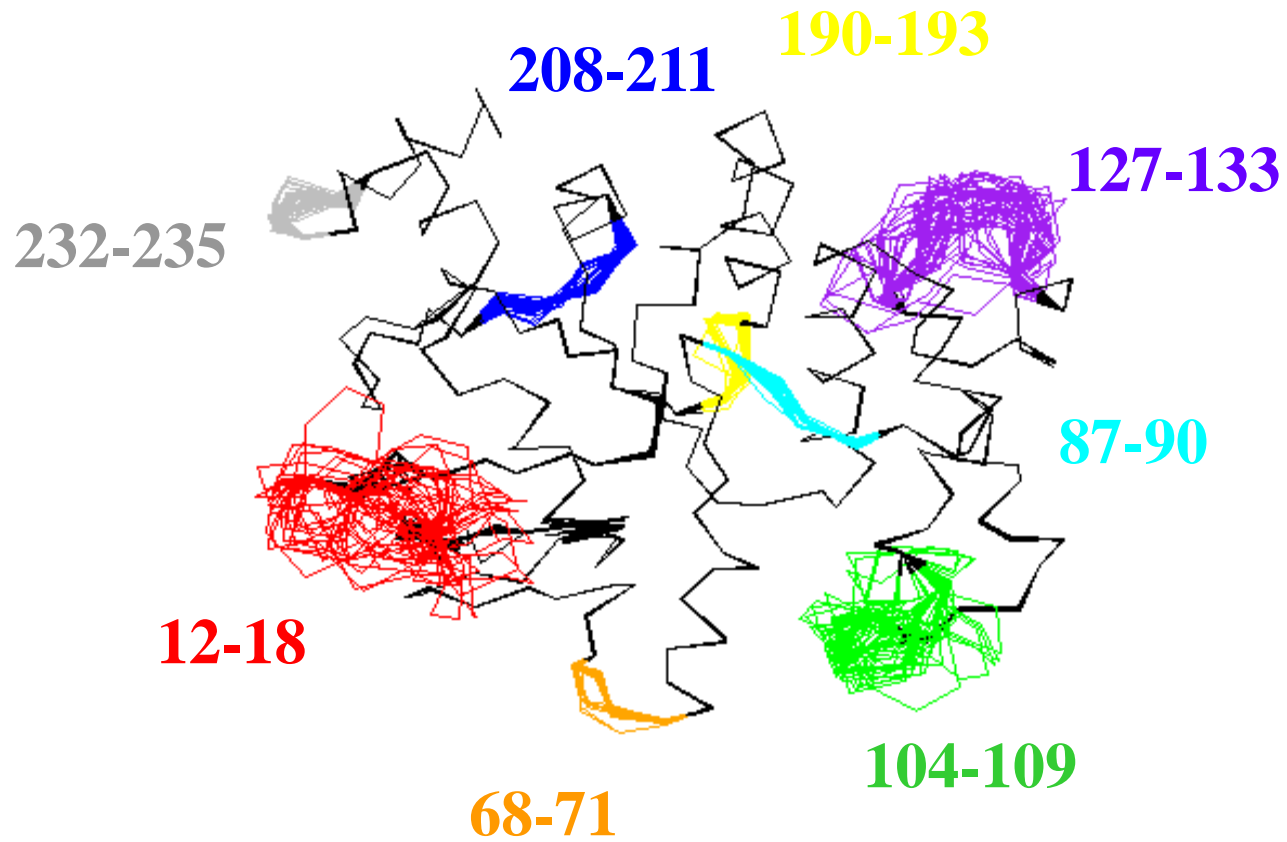
Glycoprotéine-P : Résultat



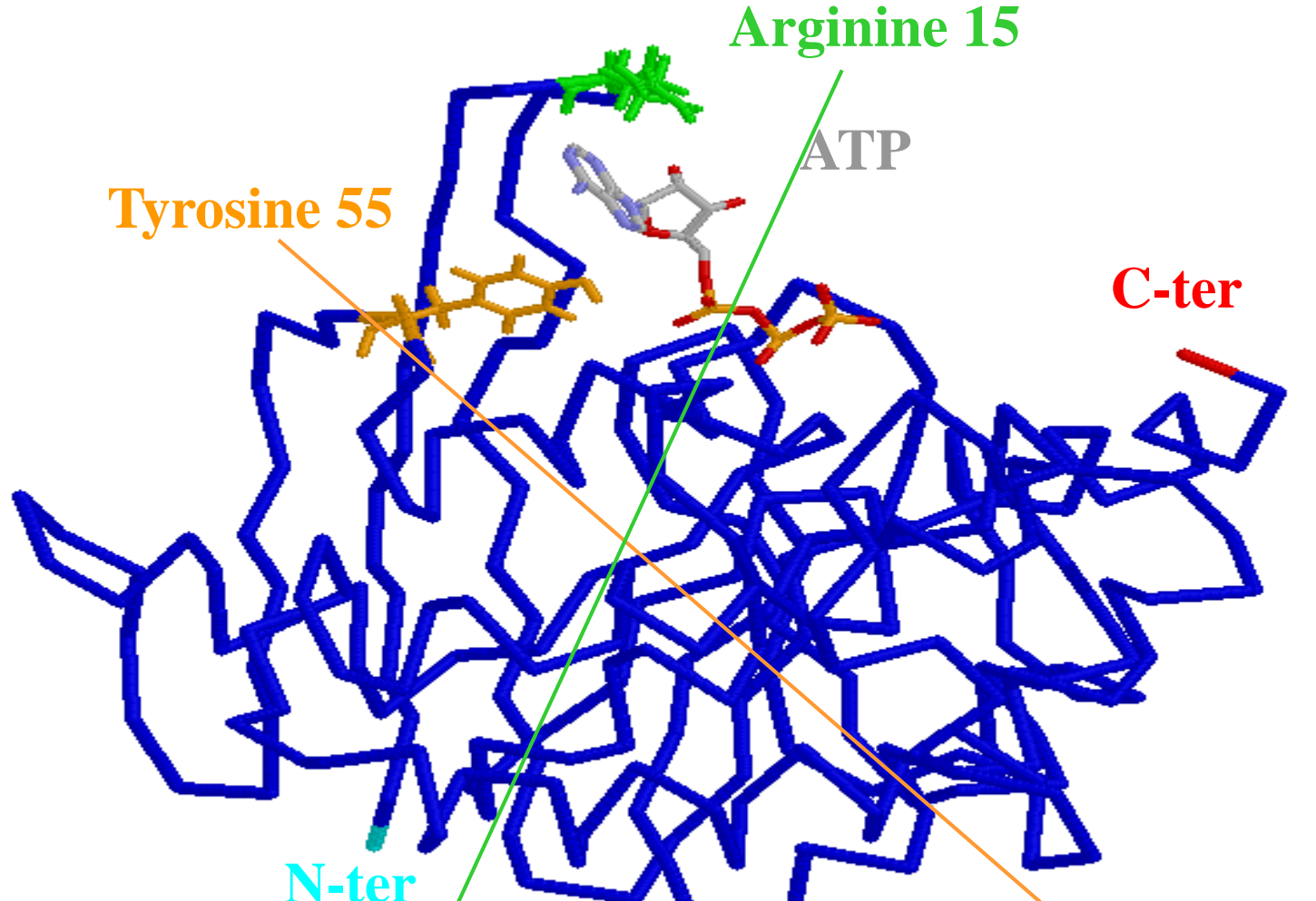
Glycoprotéine-P : Résultat



Glycoprotéine-P : Résultat

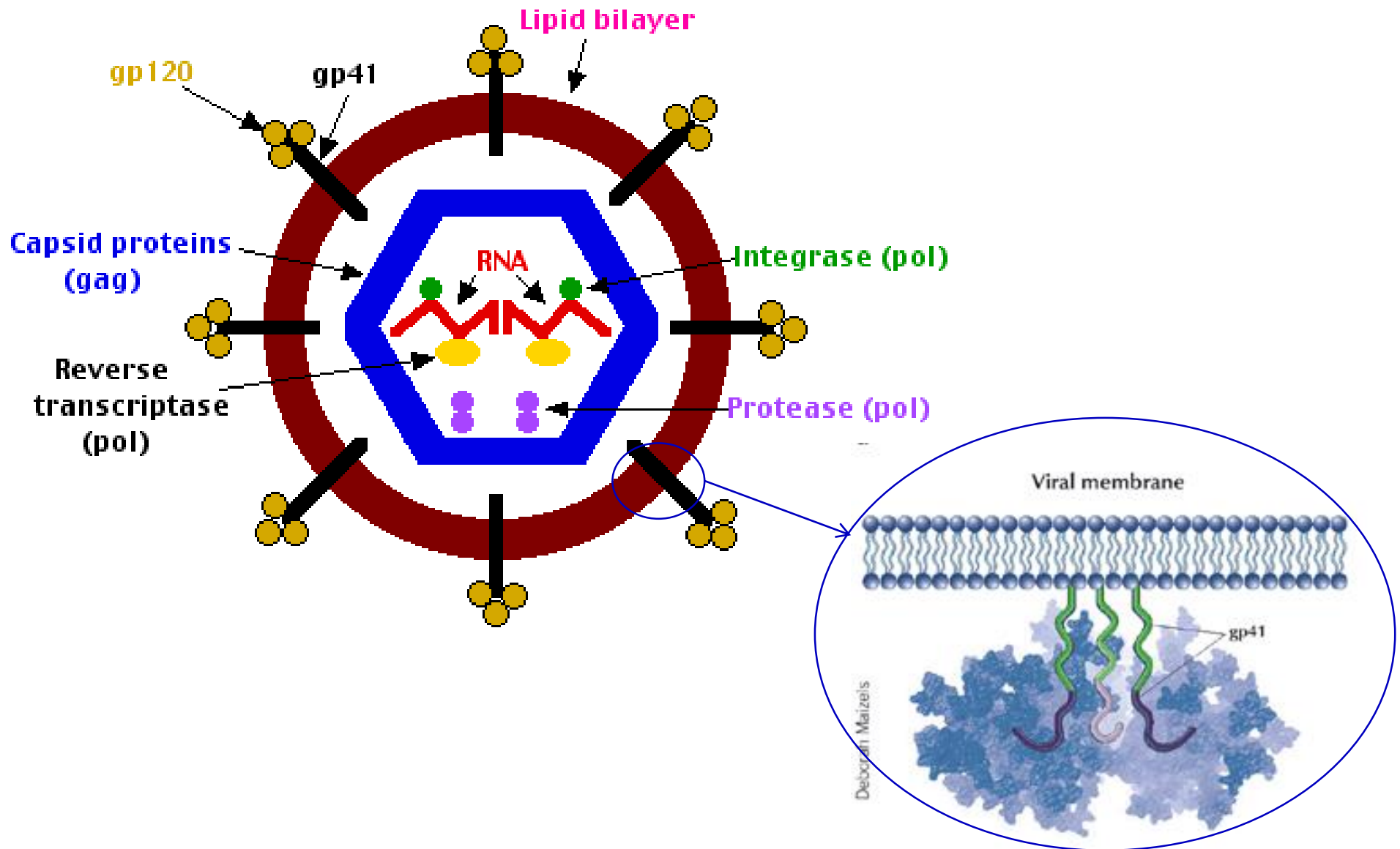


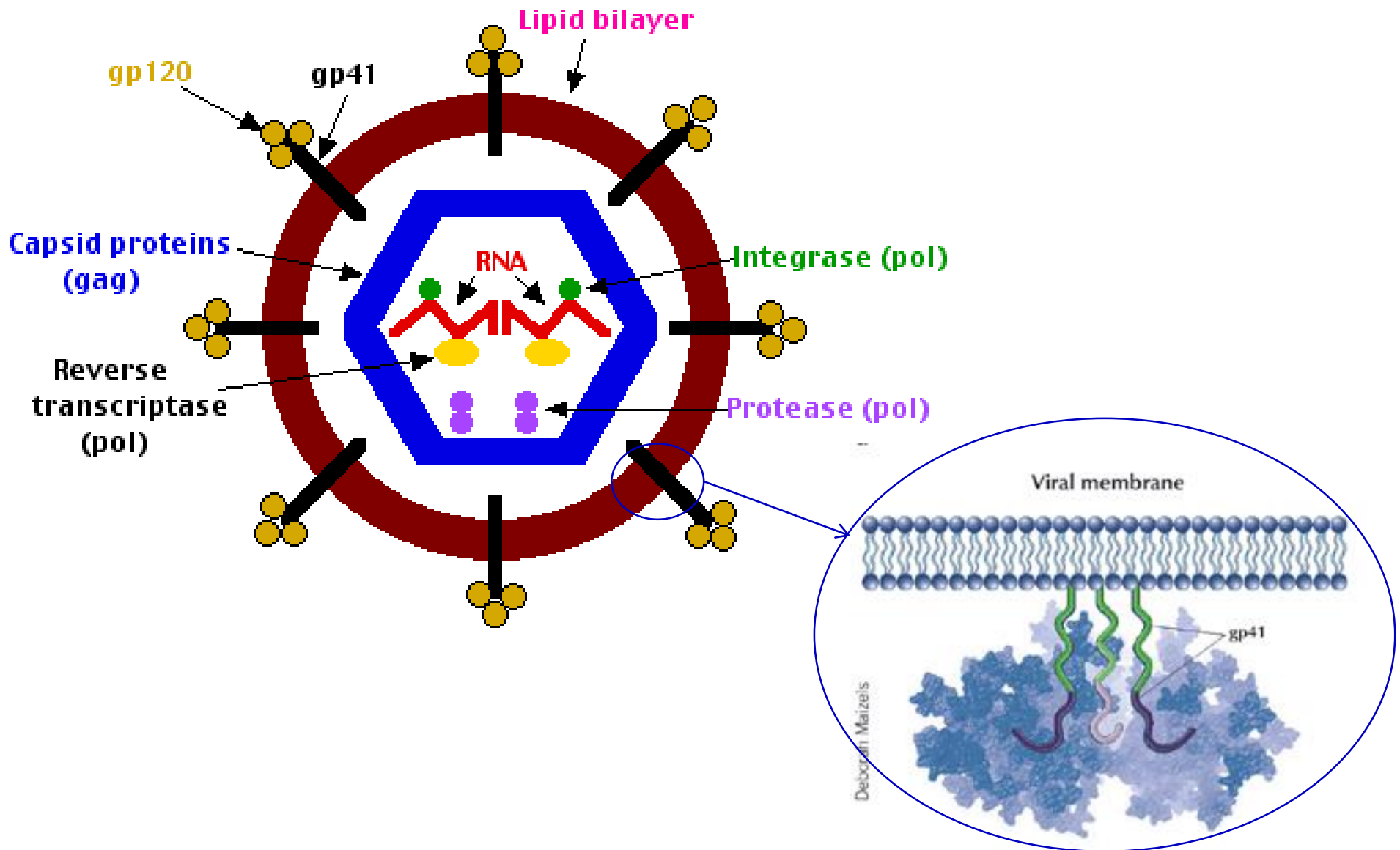
Glycoprotéine-P : Hypothèses



	10	20	30	40	50	60	
1.HISP_SALTY	NKLHVIDLHKRYG	--GHEVLKGVSLQARAGDVISIIGSSGSGKSTFLRCINFLEKPSEG					57
.Dssp	cccccccccccc	ccccccccccccccccccccchhhhhhccccccc					57
2.NBD2_MOUSE	GNVKFNGVQFNYPTRP	NIPLVQLGLSLEVKKGQTLALVGSSGCGKSTVVQLLRFYDPMAG					60
.SOPMA	ccccccccccccccccccccccccchccccccccccccccccchhhhhhccccccc						60
Homology	:: *	. **::**::	: *:::****.****.	:: : *	*	







- Pas de structure 3D disponible
- Organisation en domaines



Partie N terminale

Partie transmembranaire

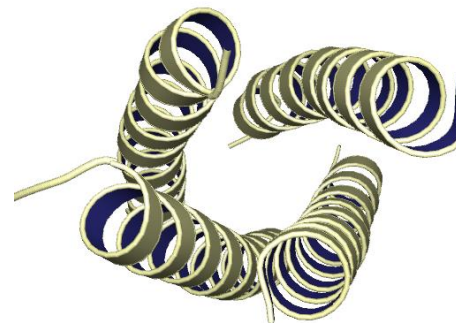
Partie C terminale

- Hypothèse d'une partie en « coiled-coil »
 - Base d'expérience
 - Analyse de séquences par bioinformatique prédictive (Lupas et Berger)

Question: gp41 est un trimère ou un tétramère?

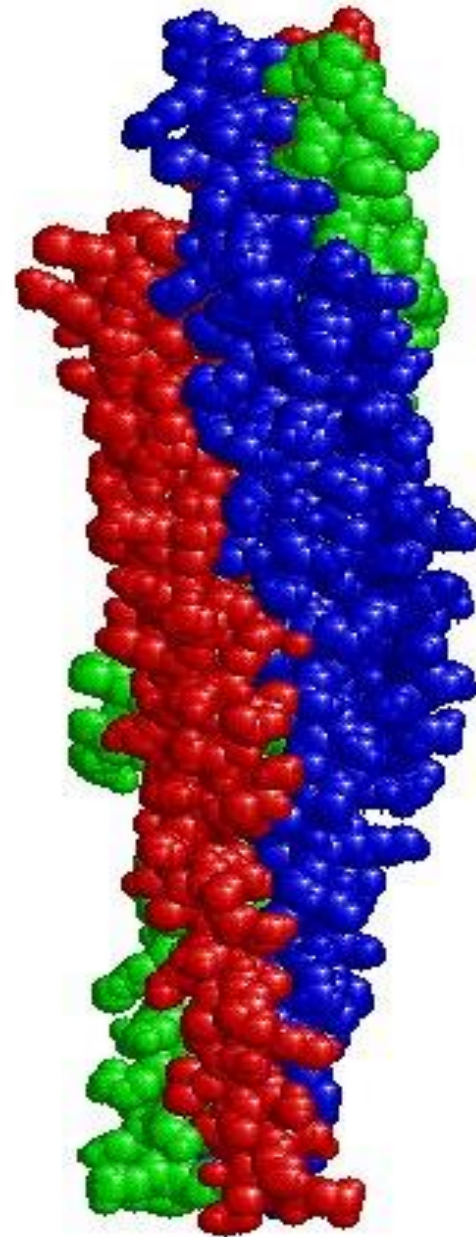
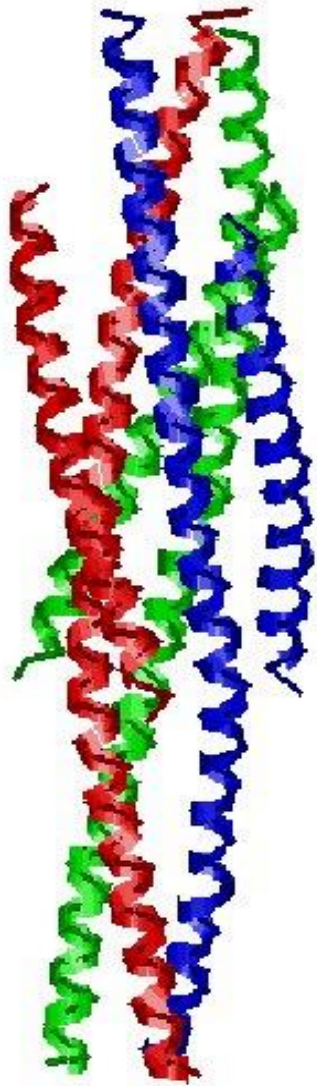


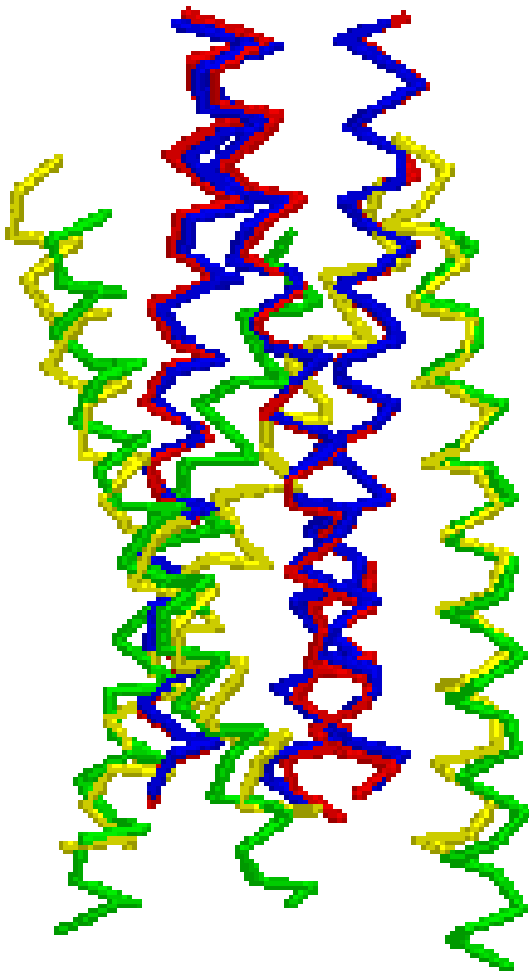
?



- **Choix de l'empreinte**
- **Mesures des contraintes sur l'empreinte**
 - 1164 contraintes de distances
 - 300 contraintes angles dièdres
- **Modélisation sous contraintes de type RMN**
 - Géométrie des distances
 - Régularisation
 - Affinement

Modèle 3D





Modèle 3D

trimère central

trimère externe

Structure 3D (Kim et al. 1997)

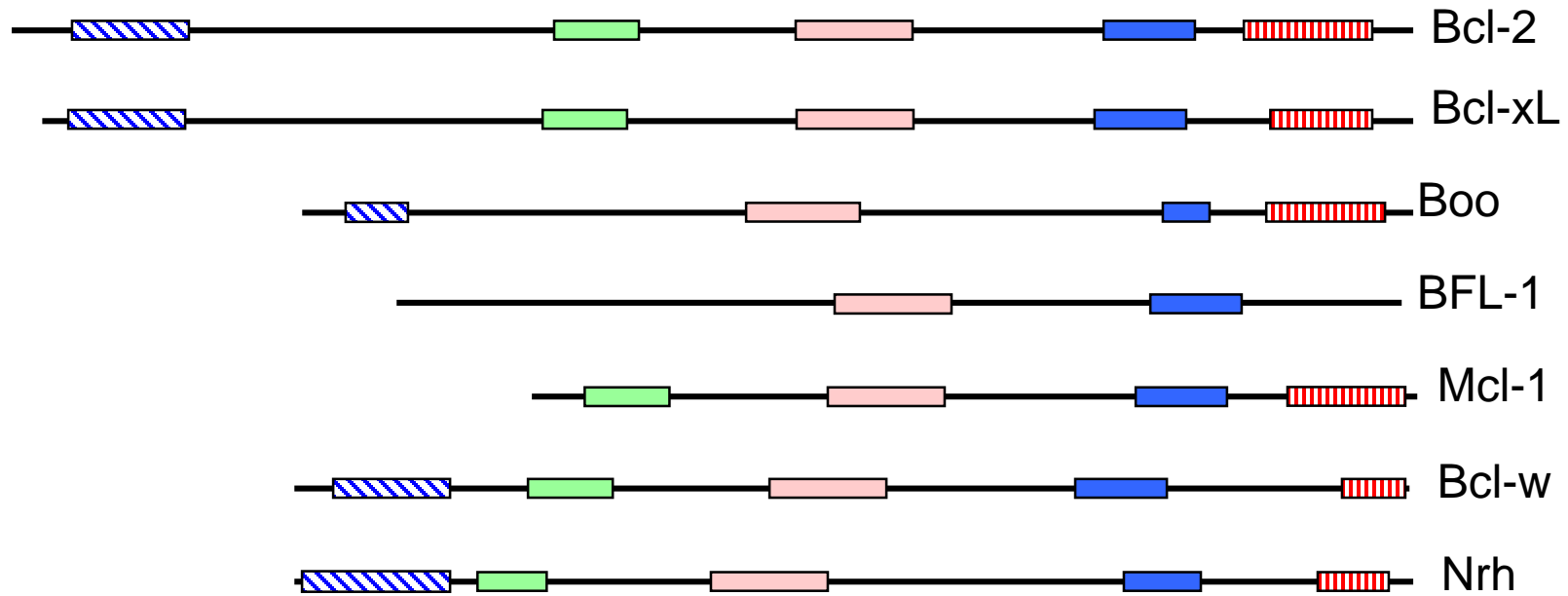
trimère central

trimère externe

Superposition

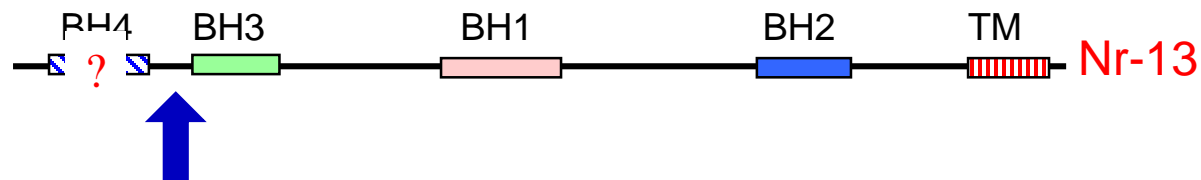
RMSD C α trimère central 1Å

Collaboration avec le groupe du Pr. Gillet



→ Fonction : Protéine impliquée dans la protection contre la mort cellulaire programmée (apoptose).

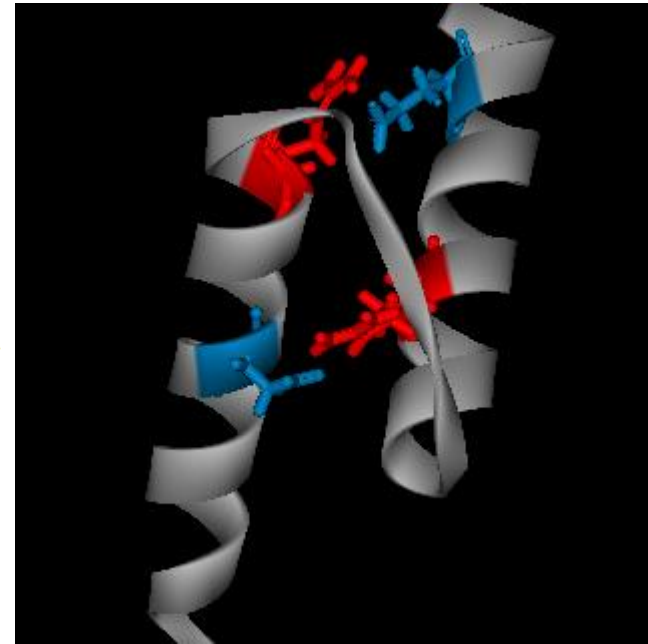
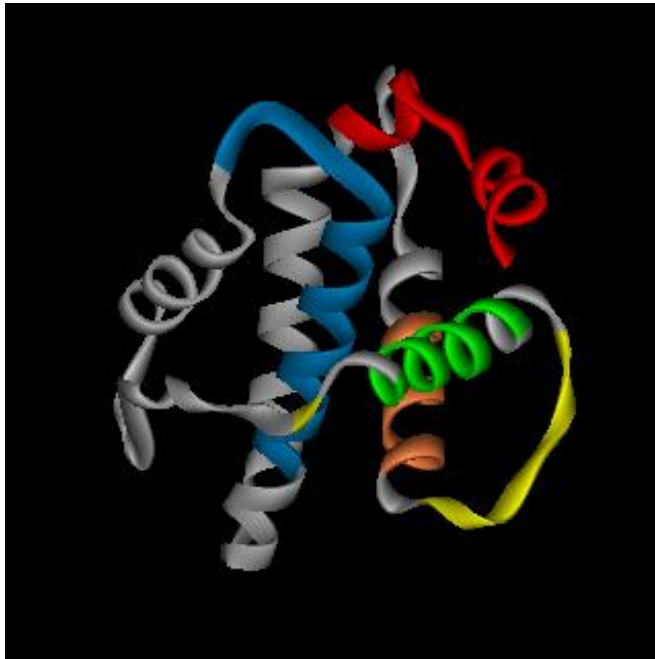
- Appartient à la même famille que les protéines Bcl2 et Bcl-xL.
- Cette famille comporte 3 à 4 régions conservées : BH4, BH3, BH2 et BH1.
- Une différence significative de NR-13 : Région N-terminale beaucoup plus courte.



→ Question : Cette protéine possède-t-elle une région BH4 malgré la connexion courte?

- Stratégie : Construction d'un modèle 3D de la protéine NR-13.
- A partir de la structure cristallographique de Bcl-xL (1maz).
 - Après avoir optimisé elle-même l'alignement entre les 2 protéines (sur la base de la structure secondaire).

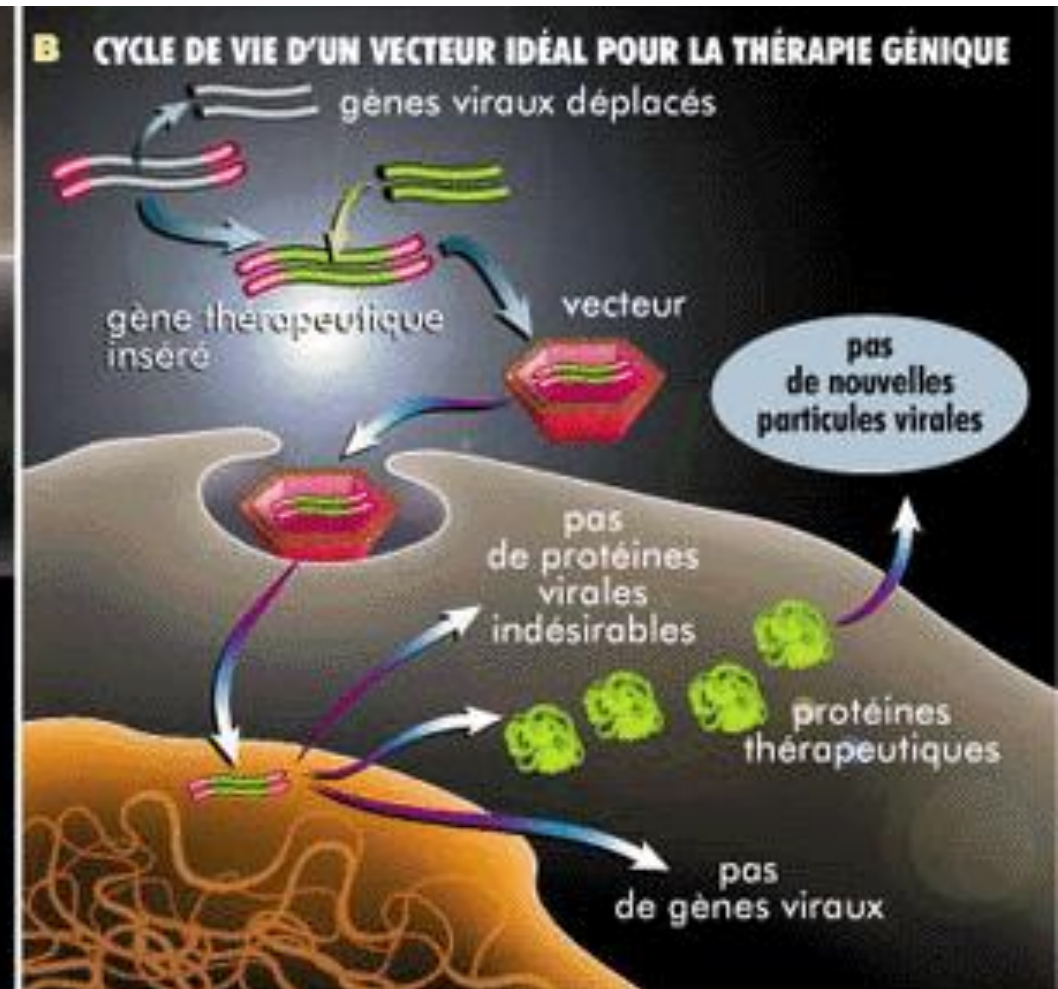
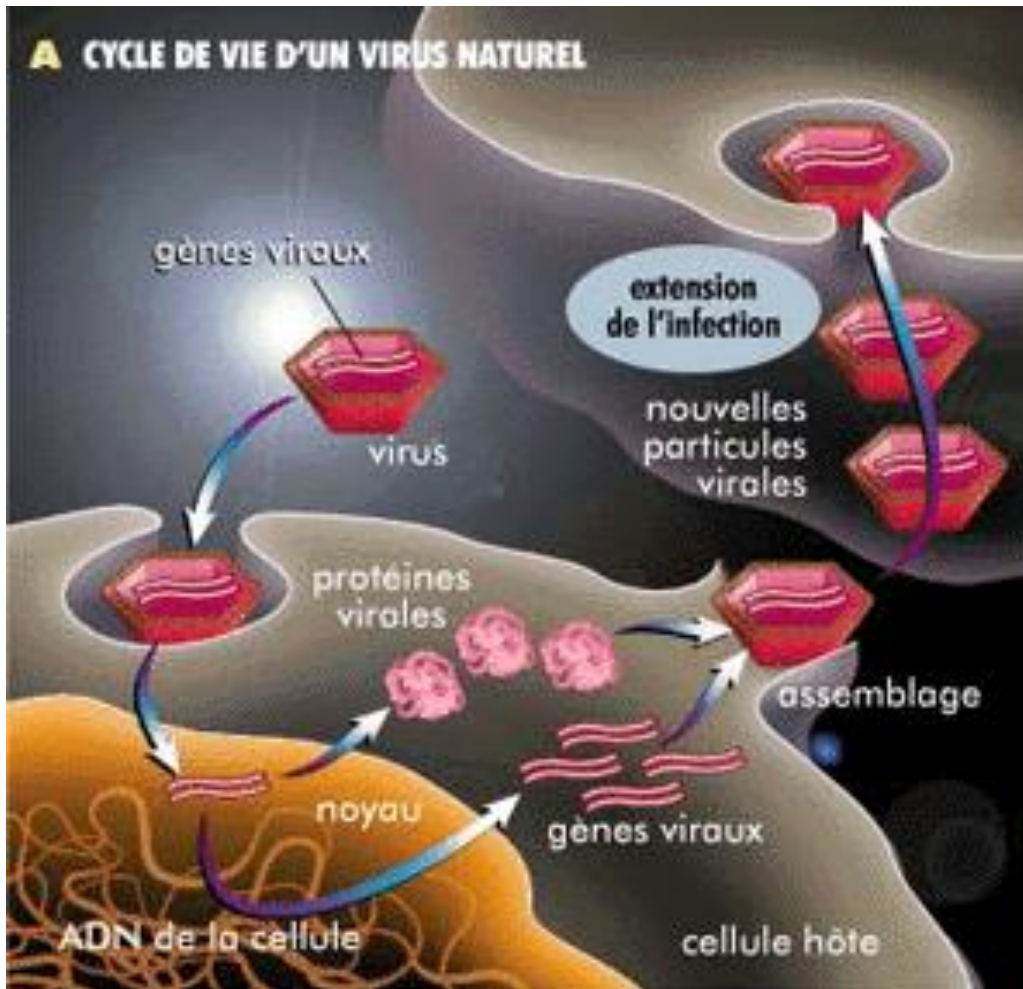
	10	20	30	40	50	60	
1. Bcl-XL						
.DSSP	----MSQSNRELVVDFLSYKLSQKGYSWQFSDVEENRTEAPEGTESEMETPSAINGNPS						56
2. Nr13	MPGSLKEETALLEDYFQHRAGGAALPP-----						28
.Sec. Cons.	cccchhhhhhhhhhhhhhhhhhhcccccccc						28
Homology	: : : : * : * : : : : . . .						
	70	80	90	100	110	120	
1. Bcl-XL						
.DSSP	WHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITP						116
2. Nr13	-----SATAAELRRAAAELERRERPFRRSCAPLARAEP						61
.Sec. Cons.	hhhhhhhhhhhhhhhhhhhhccccchccccccccch						61
Homology	: * . * * . * : * * * * * . : . : *						
	130	140	150	160	170	180	
1. Bcl-XL						
.DSSP	GTAYQSFEQVVNELFRDG-VNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAMATYLN						175
2. Nr13	REAAALLRKVAQLETDGGLNWGRLLALVVFAGTLAAALAESACEEGPSRLAAALTAYLA						121
.Sec. Cons.	hhhhhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh						121
Homology	* : : * . : * * * : * * * : * : . * . * : * . . : : : * * : * * : : * *						
	190	200					
1. Bcl-XL						
.DSSP	DHLEPWIQENGGWDTFVELYG		196				
2. Nr13	EEQGEWMEEHGGWDGFCRFFG		142				
.Sec. Cons.	hhhhhhhhhhccccchhhccc		142				



La modélisation a montré que :

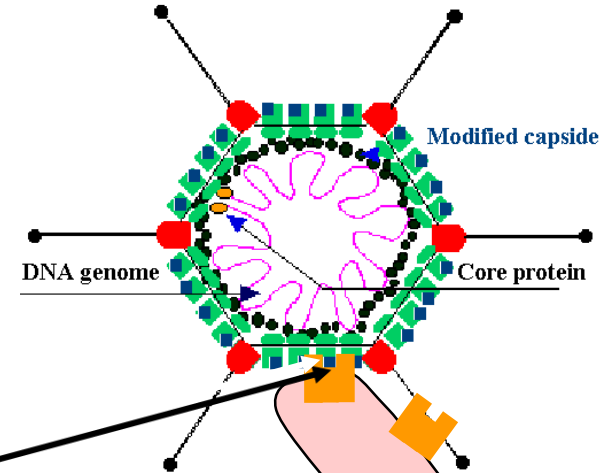
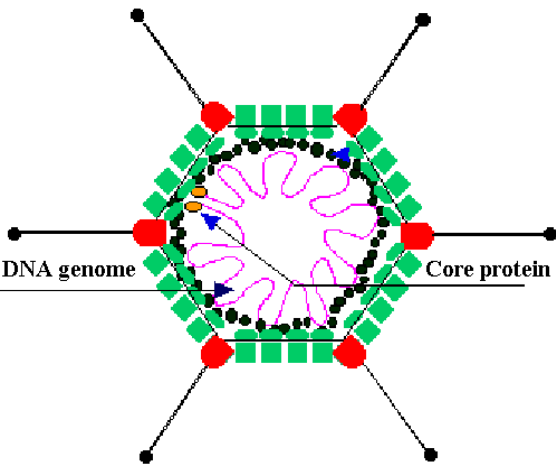
- l'existence de la région BH4 et d'une connexion très courte entre BH4 et BH3 est géométriquement possible chez NR-13.
- l'existence de deux interactions ioniques potentielles entre BH4 et BH3
 - Acide aspartique 15 - Arginine 36 et Arginine 20 - Acide glutamique 43
 - Proposition de mutations de ces acides aminés en alanine
- **La modélisation moléculaire a été validée par :**
 - Des expériences de mutagenèse réalisées dans le groupe de G. Gillet (interaction fonctionnelle D15-R36).
 - Des expériences de double mutagenèse D15-R36 et R20-E43

Lalle *et al.*, Biochem. J. (2002) **368**,213 -221



Collaboration avec Université de Munich Dr M. Hallek et A. Girod

Biologie moléculaire

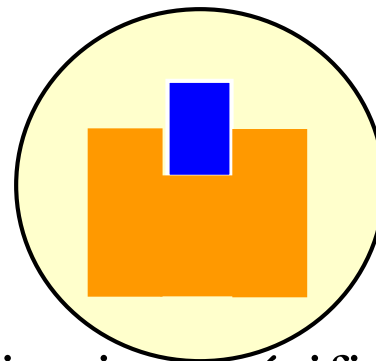


Virus natif n'infectant pas des cellules en culture

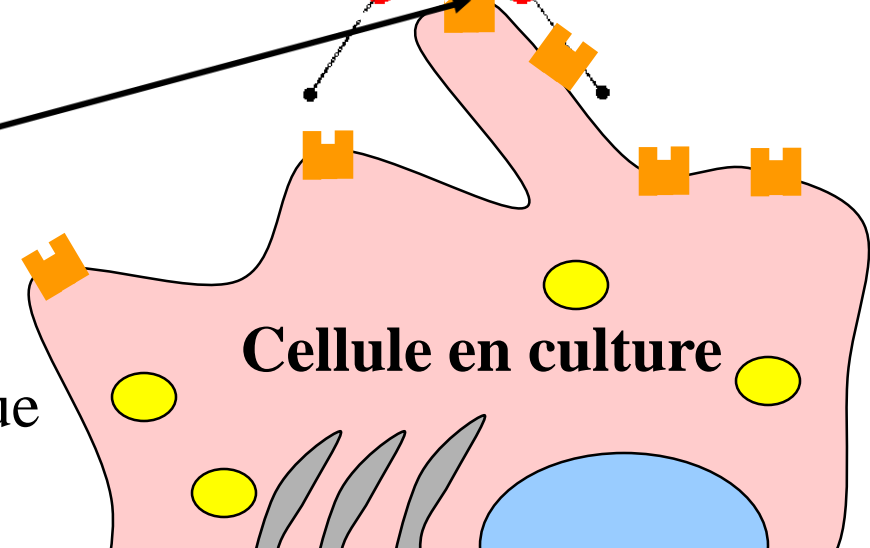
Récepteur d'intégrine 

Peptide d'adhésion cellulaire 

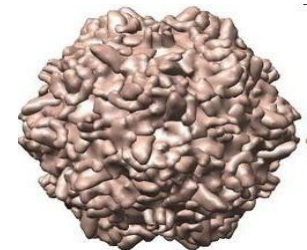
QAGTFALRGDNPQG



Fixation spécifique



Respecter l'intégrité de la structure de la protéine capsidiale
Ne pas empêcher l'assemblage du virus
Insérer le peptide sur la bonne face de la protéine



Comment faire sans structure 3D?

ANTHEPROT 2000 V5.2 by G. Deléage (g.deleage@ibcp.fr)

IBCP, 7 passage du Vercors, 69367 Lyon cedex, FRANCE

Date :04-05-2001 at :09:26:08

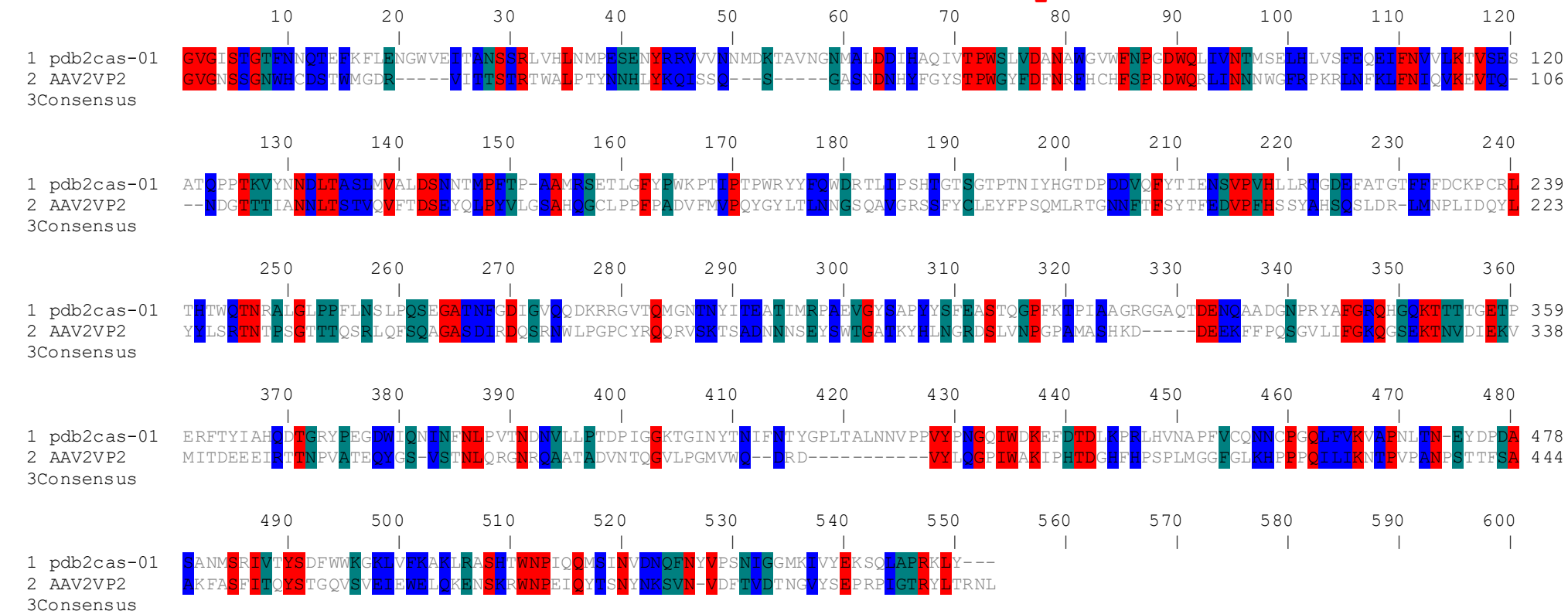
Matrix BLOSUM

Amino acids identity : : 100
 >= 75
 >= 50
 < 50

Secondary structure code :

Helix : |
 Sheet : =
 Turn : :
 Coil : :

Gap opening penalty 10 Gap extension penalty .05 Number of perfect matches 93 Identity 16.82%



100

200

300

400

500

DPM

DSC

GOR4

HNNC

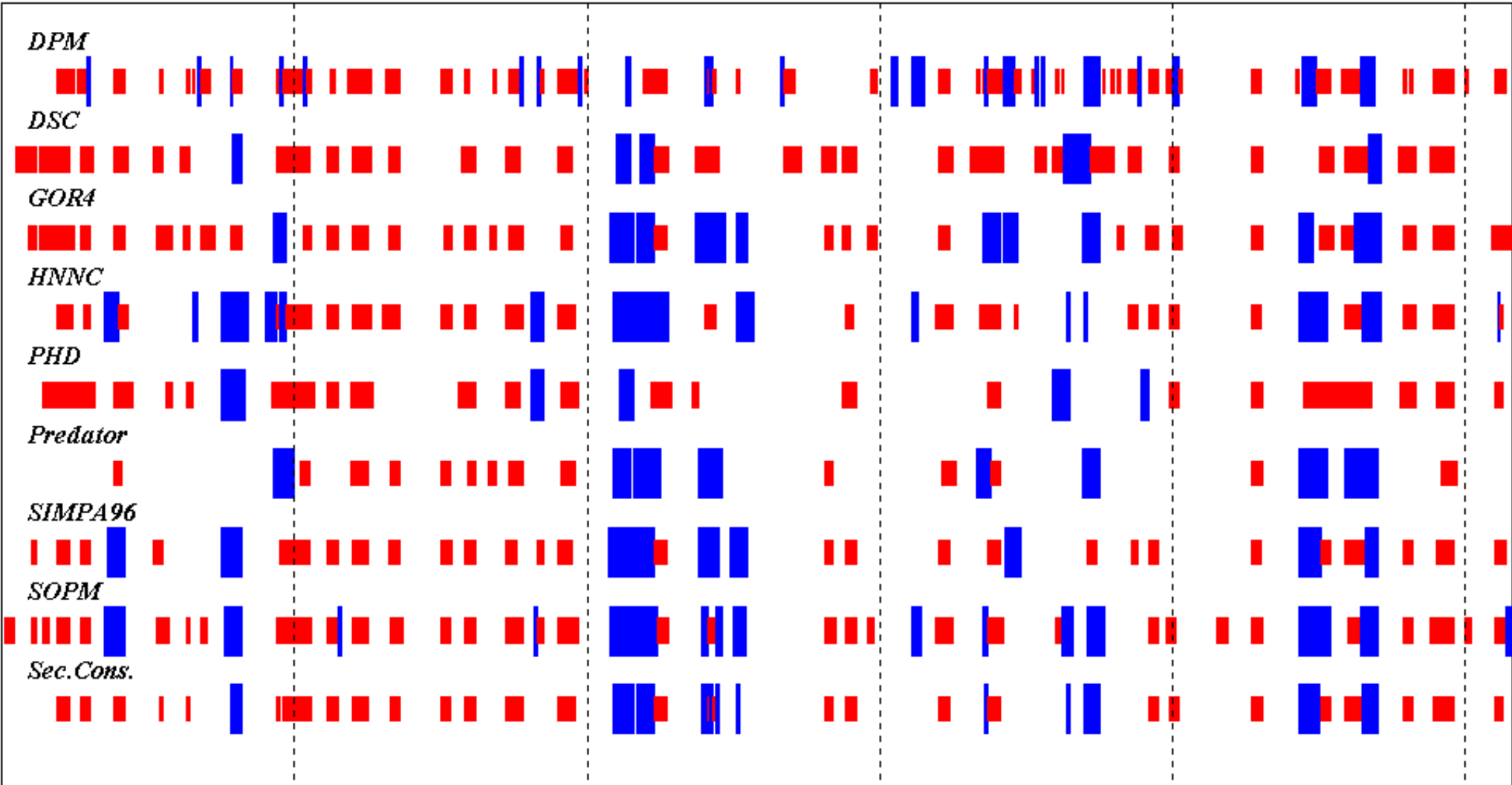
PHD

Predator

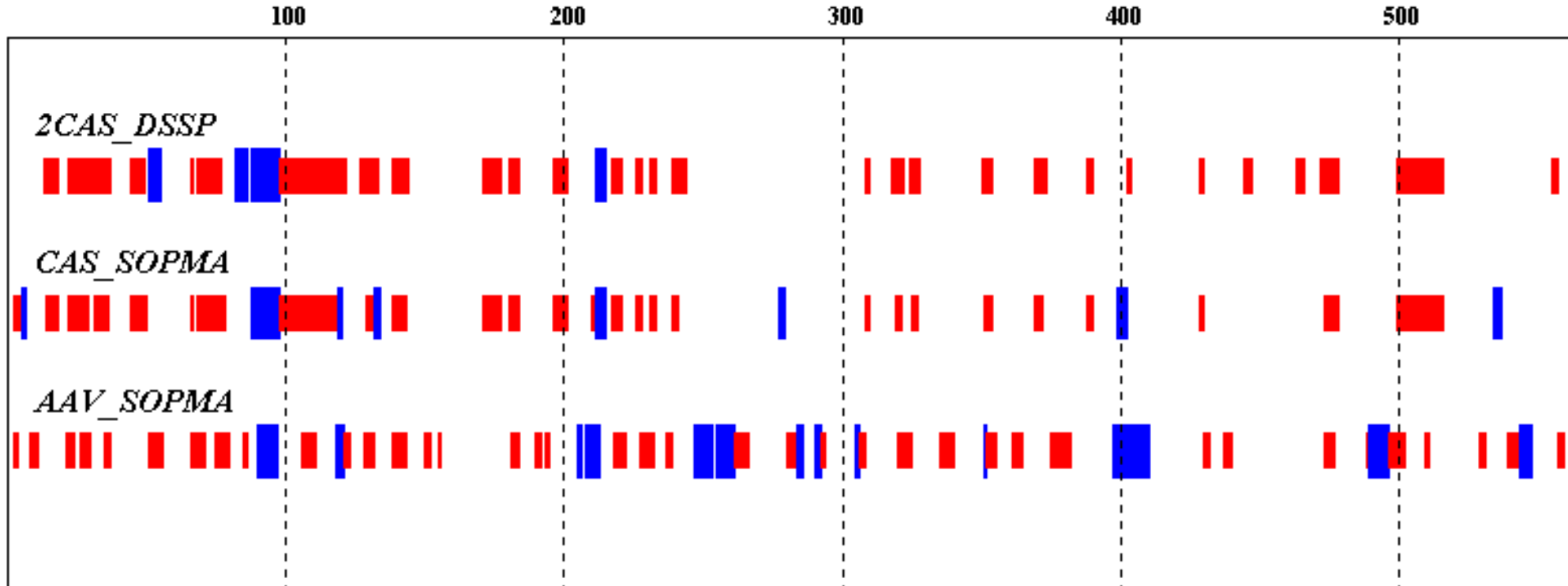
SIMPA96

SOPM

Sec.Cons.



Sov 65.8%



GENO3D : AUTOMATIC MODELING OF PROTEINS THREE-DIMENSIONAL STR...

Fichier Edition

Adresse http://geno3d-pbil.ibcp.fr

Pôle Bio-Informatique Lyonnais
Geno3D
Geno3D is the [IBCP](#) contribution to [PBIL](#) in Lyon, France

[\[HOME\]](#) [\[GENO3D\]](#) [\[HELP\]](#) [\[REFERENCES\]](#) [\[NEWS\]](#)

Monday, March 5th 2001 : **GENO3D** is now available ([see news](#))

GENO3D : AUTOMATIC MODELING OF PROTEINS

[\[Abstract\]](#) [\[GENO3D help\]](#) [\[Original server\]](#)

Database :

Sequence name (optional) :

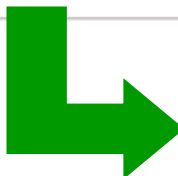
Paste a protein sequence below : [help](#)

```

DPDDVQFYTIENSVPVHLLRTGDEFATGTFDFDCKPCR
LPPFLNSLPQSEGATNFGDIGVQQDKRRGVTQMGNTNY
GYSAPYYSF EASTQGPFKTP IAAGRGGQAQTDENQAADG
KTTTTGETPERFTYIAHQDTGRYPEGDWIQININFNLPV
GGKTGINYTMIFNTYGPLTALMNVPPVYPNGQIWDKEF
PFVCCQNNCPGQLFVKVAPNL TNEYDPDASANMSRIVTY
AKLRASHTWNP IQQMS INVDNQFNYPVPSNIGGMKIVYE
    
```

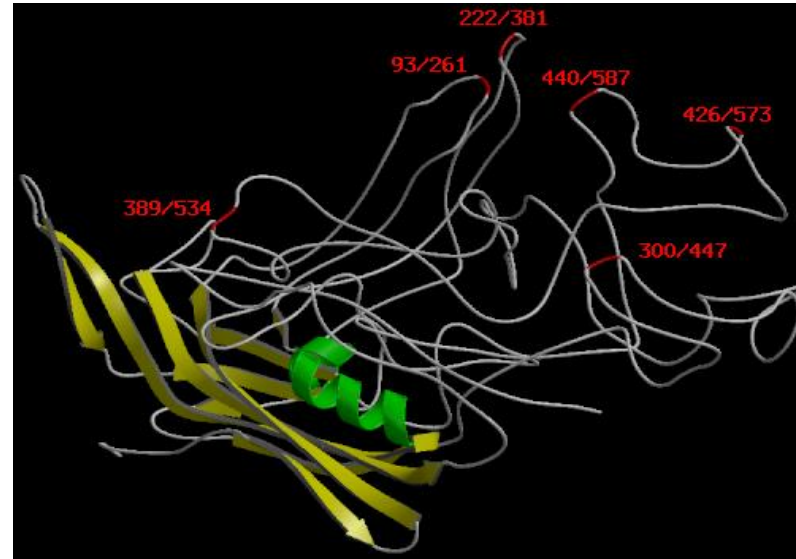
Number of PSI-BLAST run

Modèles de capside AAV2

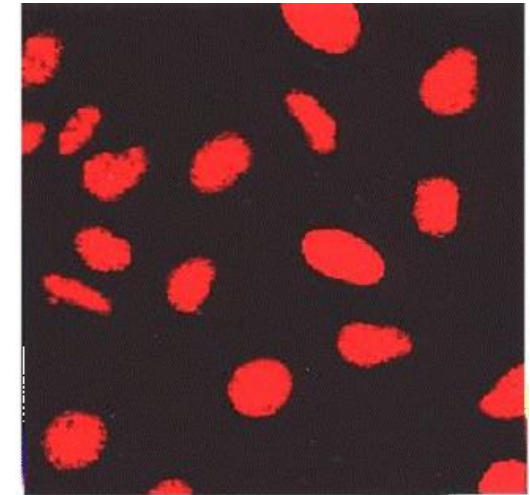


6 sites d'insertions retenus

Témoin pas infection
virus sauvage



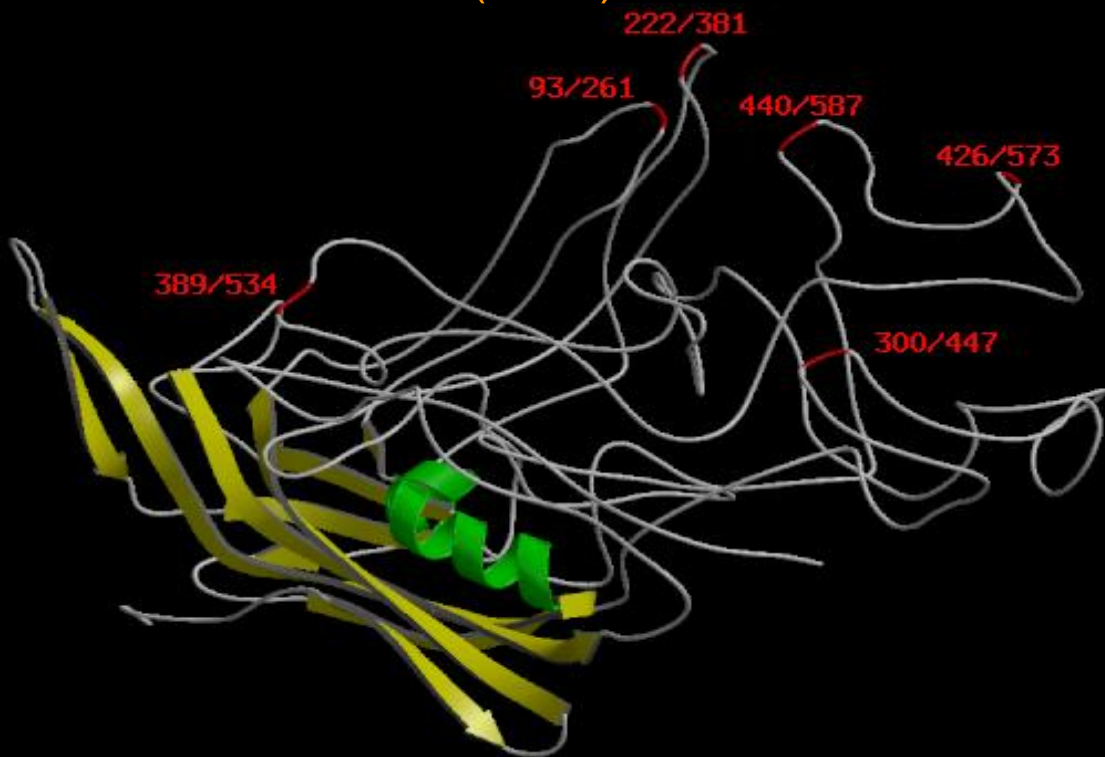
Infection
I587 L14



- 6 mutants /6 virus mutés sont encapsidés correctement
- 3/6 expriment le peptide L14 à leur surface
- 1/6 (I587/L14) se fixe spécifiquement sur le récepteur des intégrines

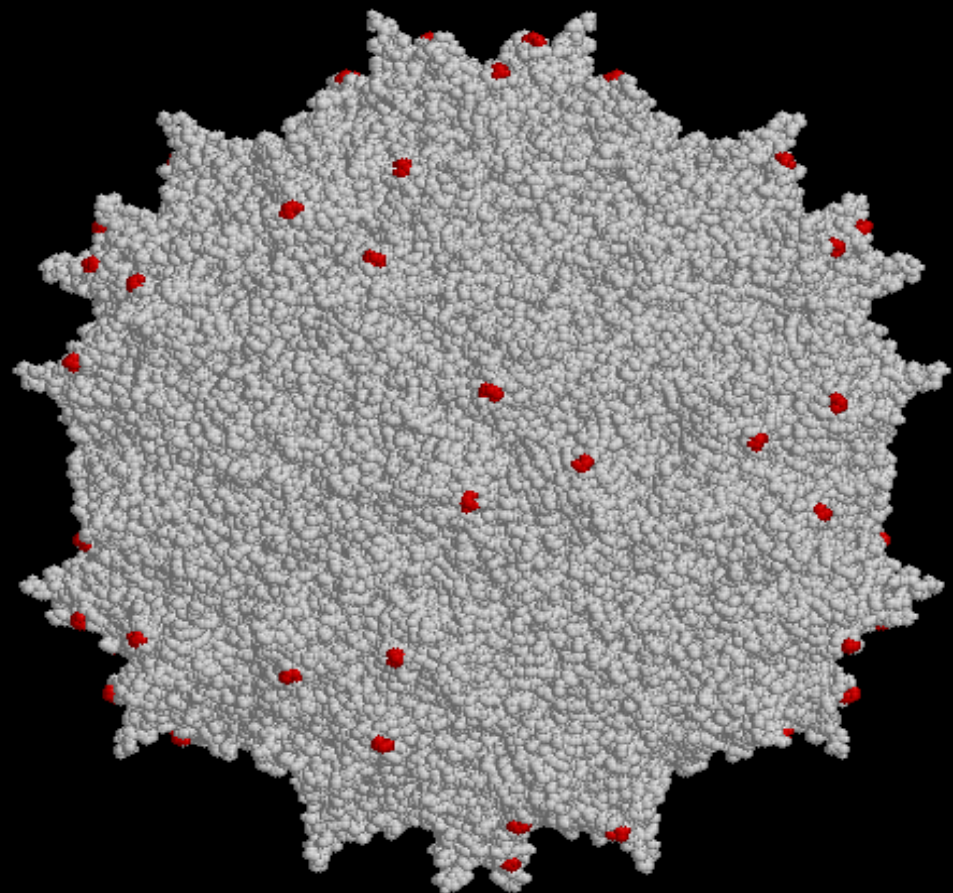
Girod, A. Ried, M. Wobus, C. Lahm, H. Leike, K. Kleinschmidt, J. Deléage, G. and Hallek M. *Nature Medicine* (1999) 5, 1052-1056
Brevet N°DE19827457 sur la possibilité d'insérer un peptide dans la capsid AAV2 en région 587

Prédictions modèle (2cas)

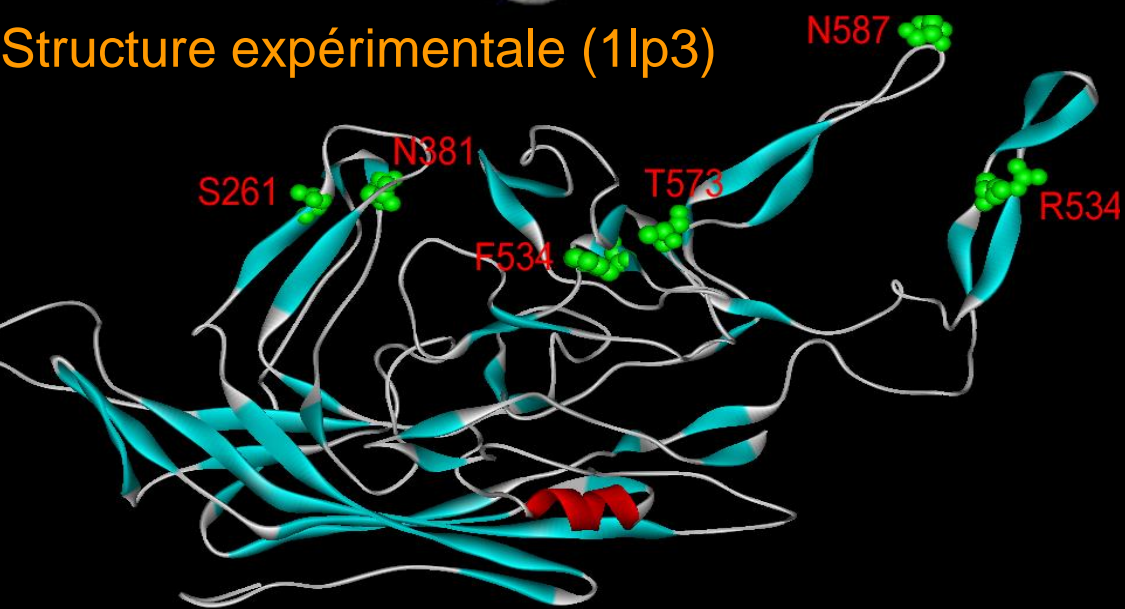


Capside expérimentale

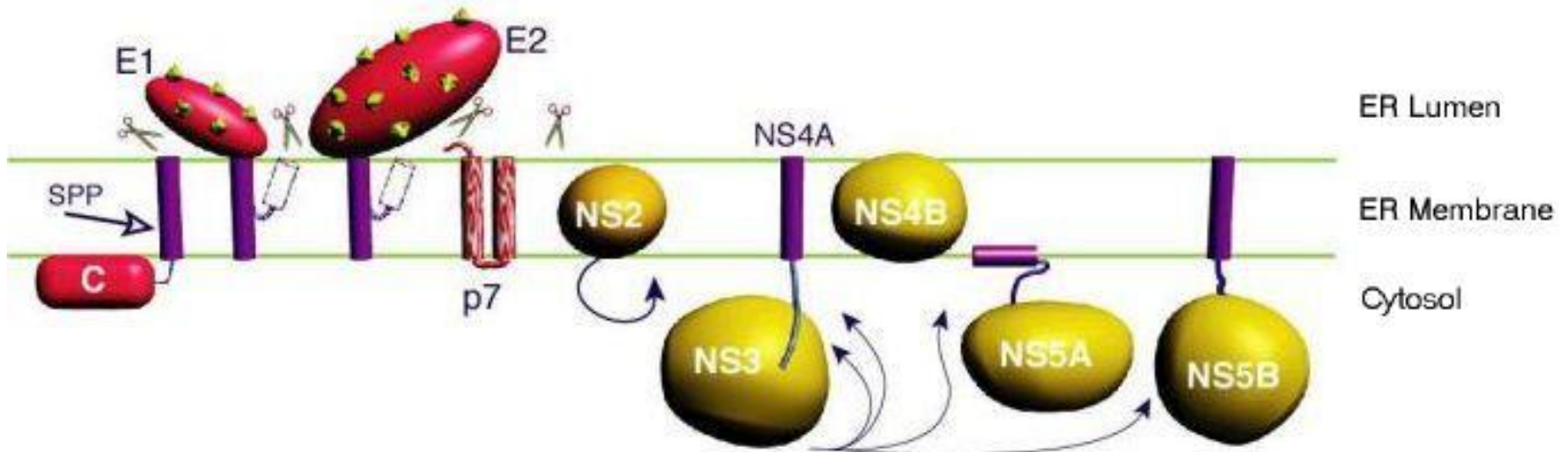
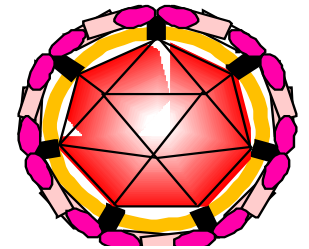
N587





Structure expérimentale (1lp3)



Positive strand RNA, 9,6 kb (3010-3033 aa)



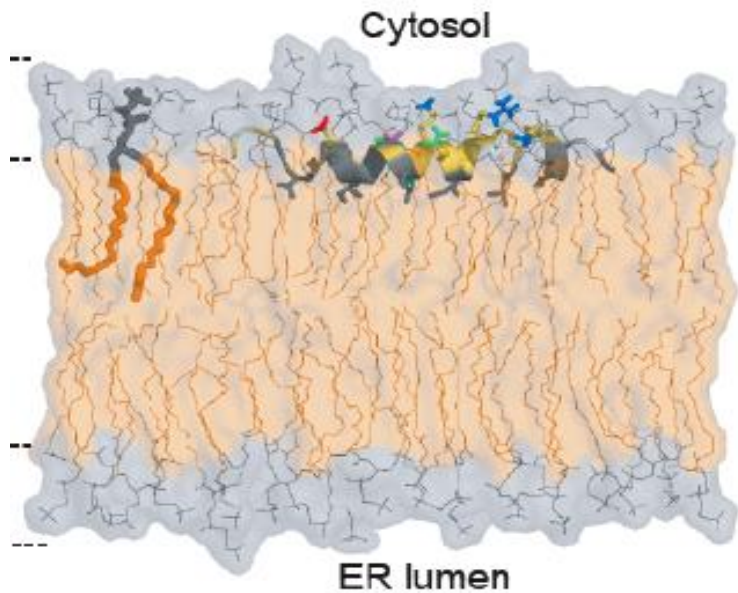
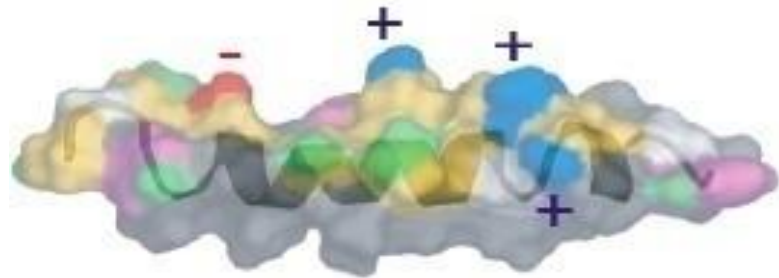
-  Interaction of hepatitis C virus proteins with host cell membranes and lipids (*Review*)
 J. Dubuisson, F. Penin, and D. Moradpour.
Trends in Cell Biol. (2002) 2 : 517-523.
-  Structural biology of hepatitis C virus. (*Review*)
 Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D. and Pawlotsky, J. M.
Hepatology (2004) 39 : 5-19

Objectifs

Analyser les relations structure-activité des protéines du VHC impliquées dans les processus d'entrée virale, de traduction, de formation du complexe de réplication et d'assemblage du virus pour identifier de nouvelles cibles thérapeutiques, aider à la conception de vaccin, et/ou concevoir de nouvelles stratégies de lutte contre l'infection virale.

In-plane membrane anchor

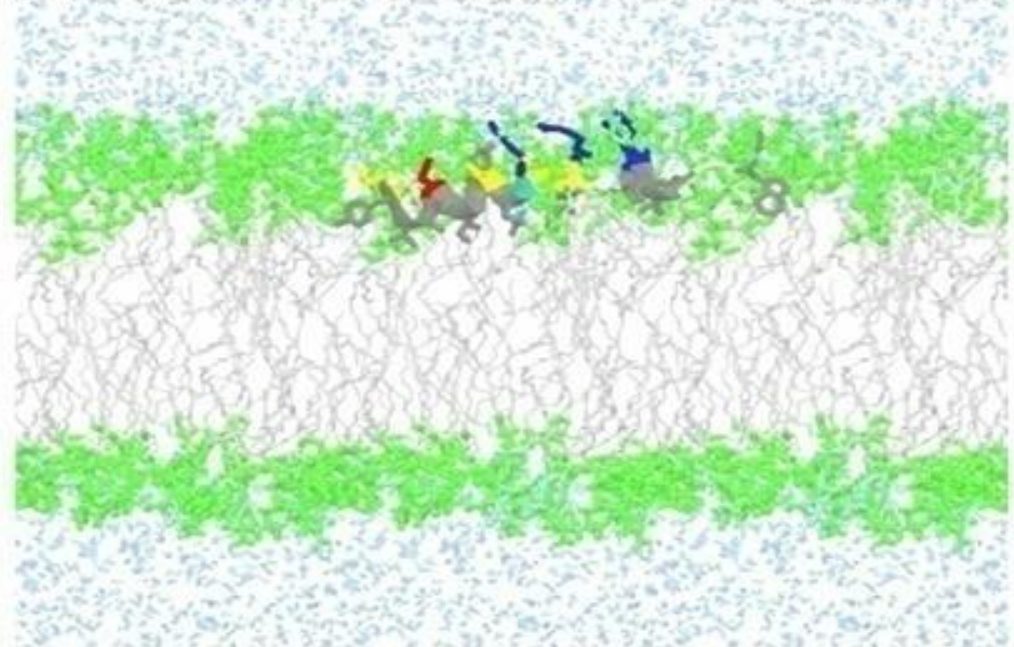
VHC NS5A[1-31] **S**G**S**W**L**R**D**I**W**D**W**I**C**E**V**L**S**D**F**K**T**W**L**K**A**K**L**M**P**Q**L**
 ||.: : | | .: .::: || ::
 BVDV NS5A[1-28] **S**G**N**Y**--**V**L**D**L**I**Y**S**L**H**K**Q**I**N**R**G**L**K**K**I**V**L**G**W**A**



RMN, CD, ATR-TFIR, DL, DM

NMR Structure and Molecular Dynamics of the In-plane Membrane Anchor of Nonstructural Protein 5A from Bovine Viral Diarrhea Virus
 Sapay, N. et al. (2006) **Biochemistry** 45 : 2221-2233

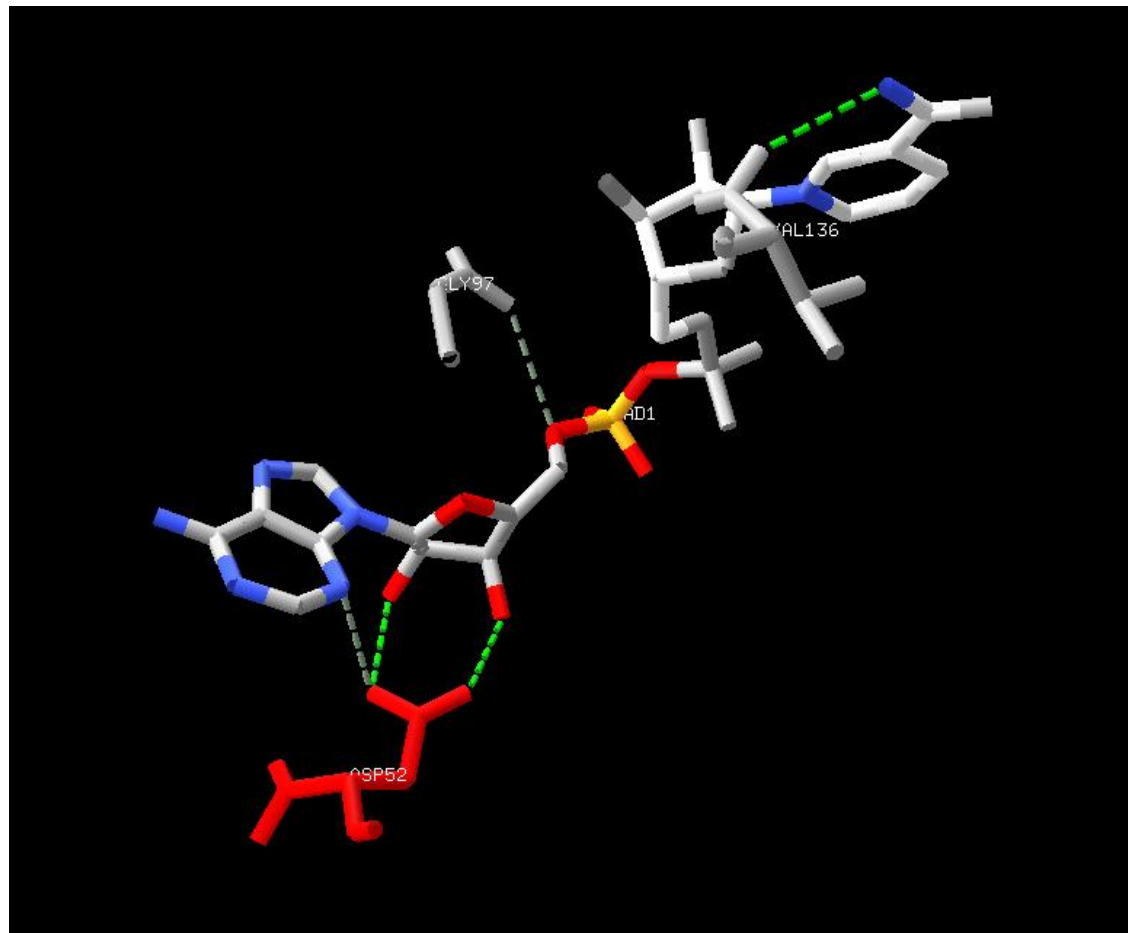
NS5A in-plane membrane domain from BVDV
 simulation of BVDV NS5A[1-28] in a POPC membrane bilayer



Interactions moléculaires 3D

Analyse de la fixation d'un ligand NAD sur une protéine 1LDM

- **Montrer les liaisons H (Tools=> Compute Hydrogen bonds)**
- **Sélectionner le ligand dans la fenêtre de contrôle**
- **Limiter les liaisons H au ligand (la sélection)**
- **Limiter la visualisation aux groupes possédant des liaisons H**
- **Centrer, réorienter, afficher les « labels ». Calculer des distances**



Wallace A C, Laskowski R A, Thornton J M (1996). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8, 127-134. [PubMed id: [7630882](#)]

LIGPLOT home page - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

LIGPLOT home page

www.ebi.ac.uk/thornton-srv/software/LIGPLOT/

Les plus visités Débuter avec Firefox À la une

EMBL-EBI Find Help Feedback

Databases Tools Research Training Industry About Us Help Site Index

- Home
- Manual
- Reference
- FAQ

Click to enlarge

EMBL > Groups > Thornton > Software > LIGPLOT

LIGPLOT v.4.5.3 - Program for automatically plotting protein-ligand interactions

Written by **Andrew Wallace** and **Roman Laskowski**

Automatically generates schematic diagrams of protein-ligand interactions for a given PDB file. (Click on the example on the left).

The interactions shown are those mediated by **hydrogen bonds** and by **hydrophobic contacts**. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

Availability

The official LIGPLOT download site is UCLB E-LUCID:

[LigPlot for Unix/linux](#) [LigPlot for Windows](#)

Alternatively, commercial users may wish to contact our distributor [Ebisu](#).

Contact

If you have any questions about LIGPLOT or problems using it, please check out the [FAQ page](#), or contact Roman Laskowski at roman@ebi.ac.uk.

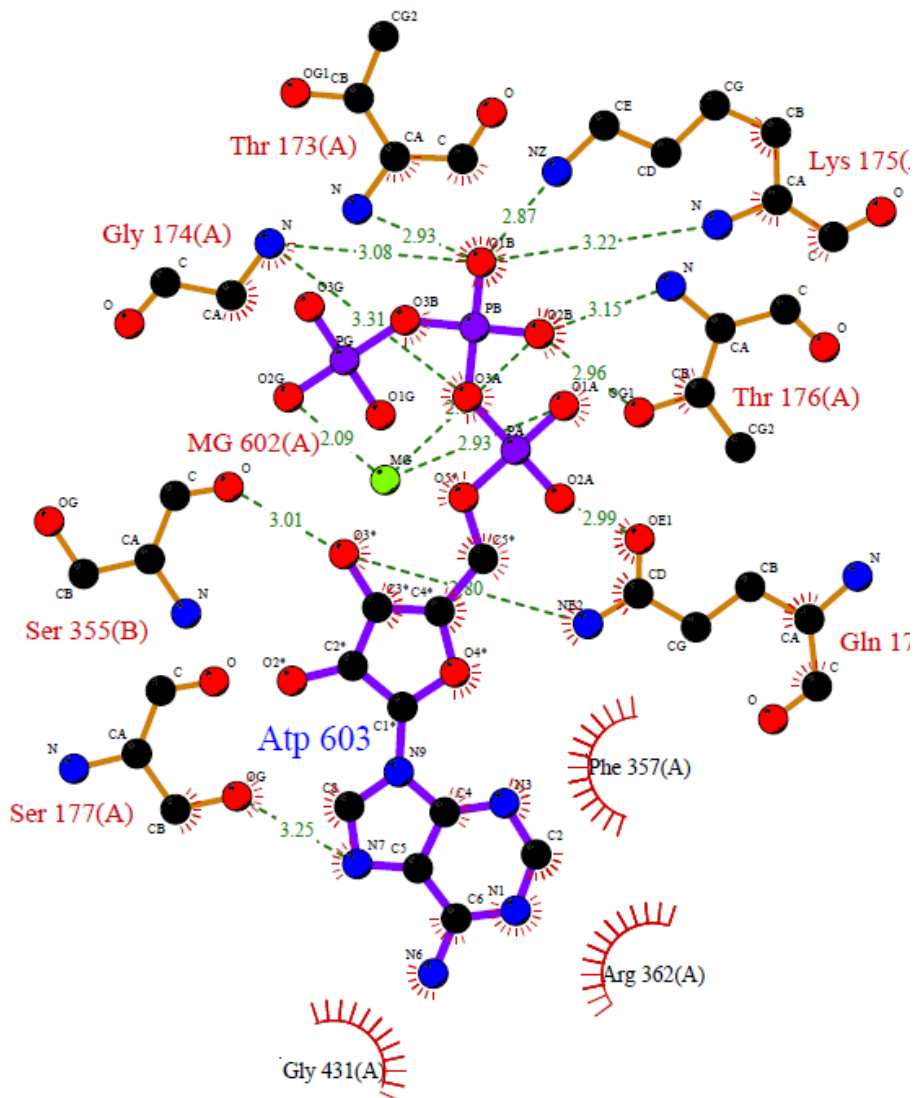
Last modified: 22 March 2010

Notes

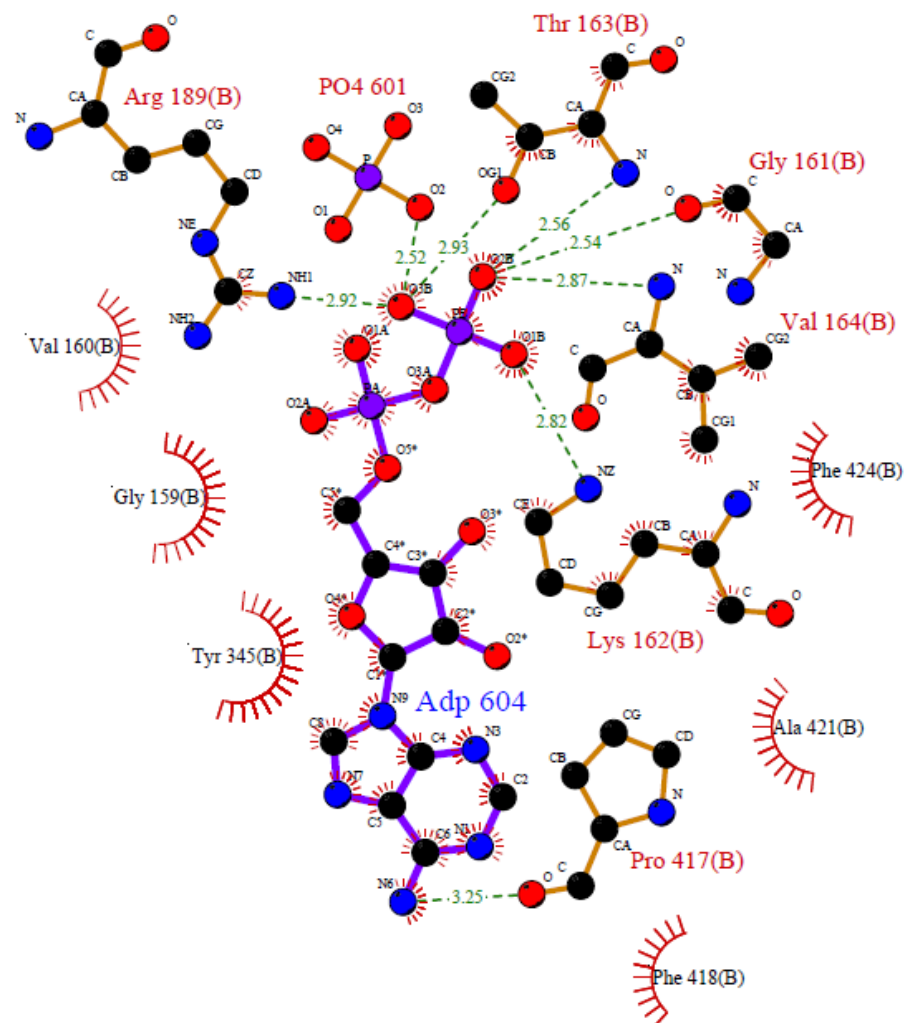
➡ A new, GUI-based version of **LIGPLOT**, called **LigPlot⁺**, is now available.



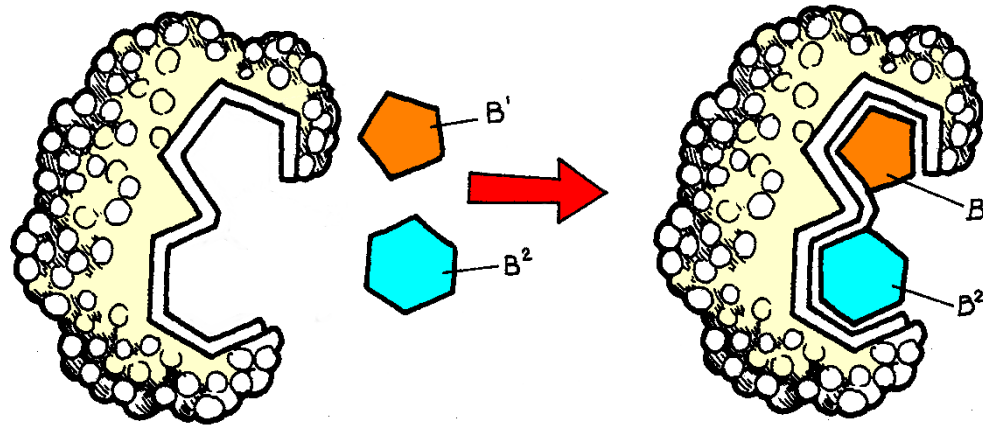
Fixation d'ATP +Mg



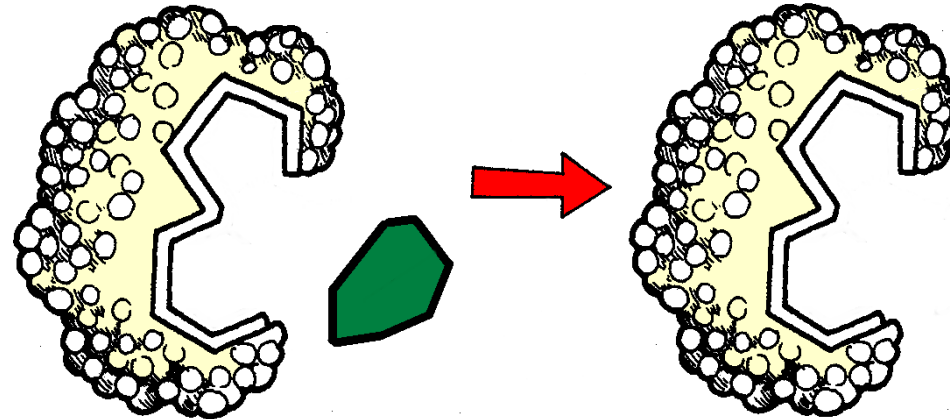
Fixation d'ADP + PO4



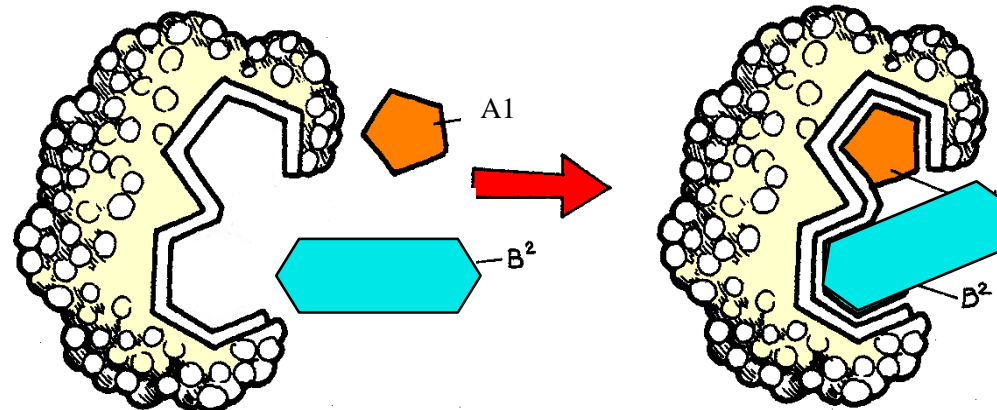
Complémentarité
de forme



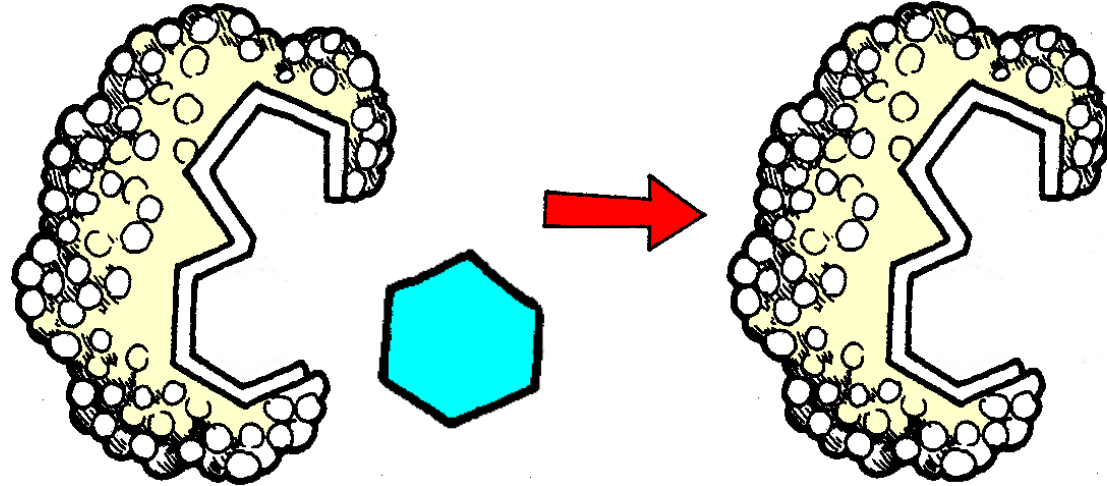
Pas de fixation



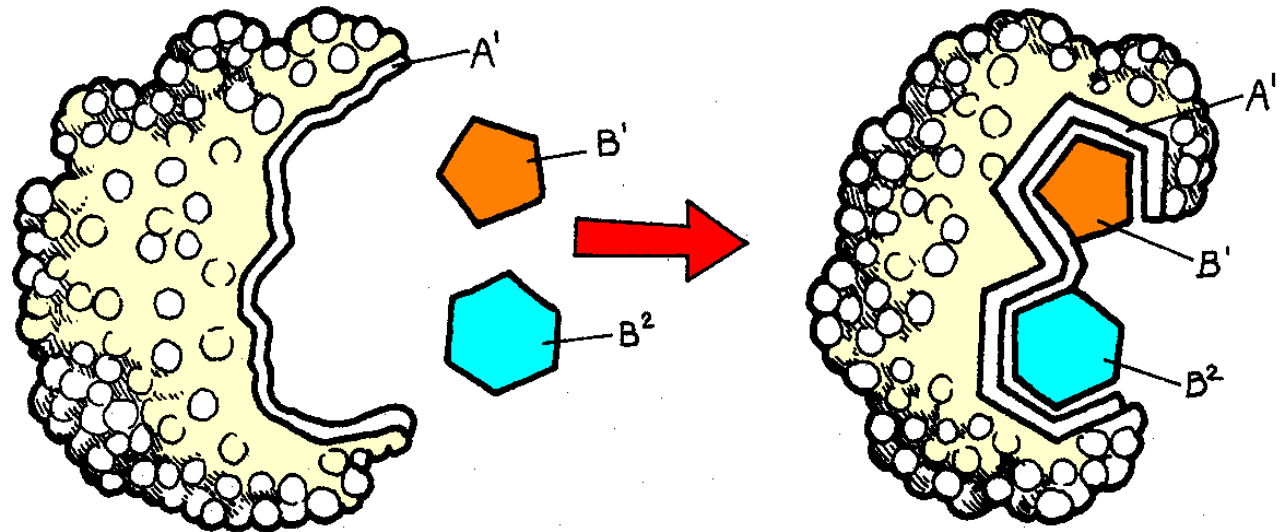
Systèmes ordonnés

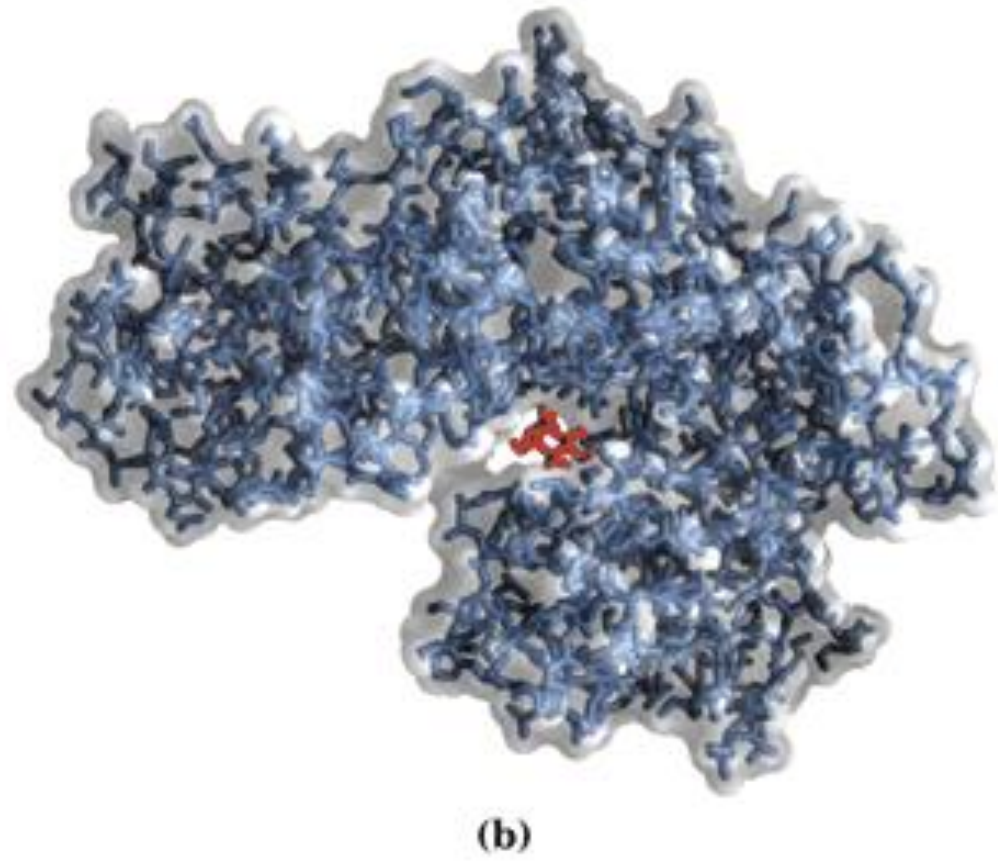
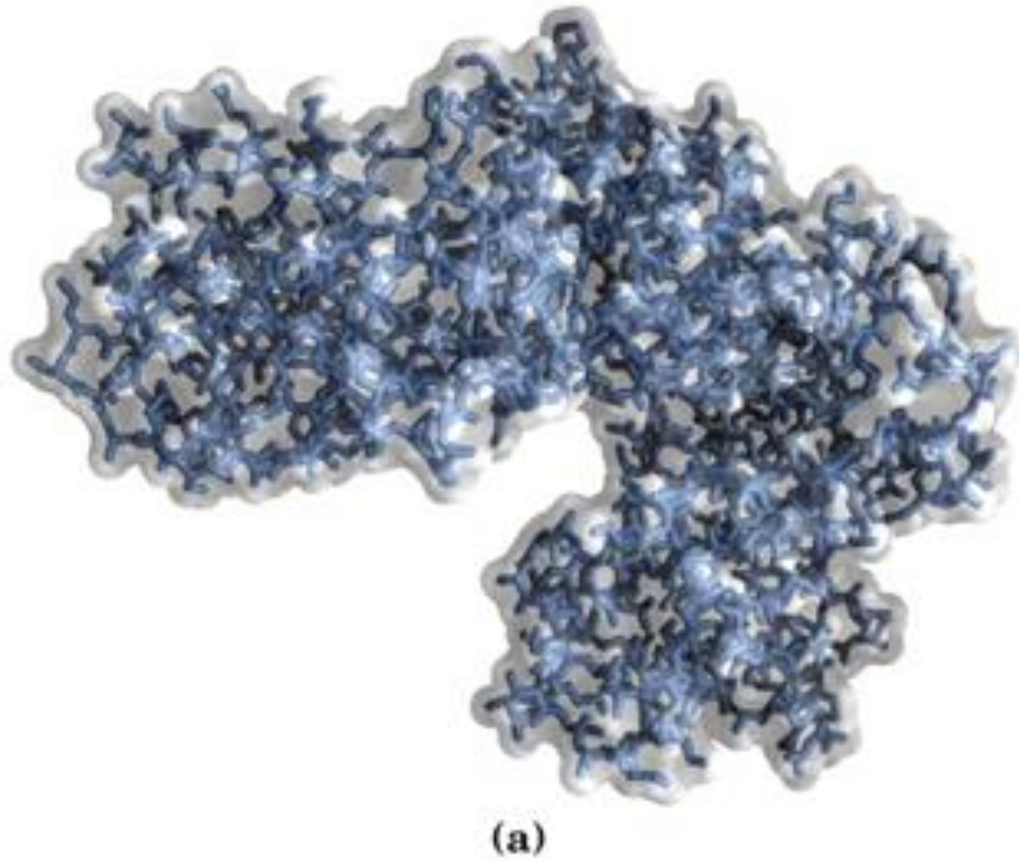


Peu d'adaptation

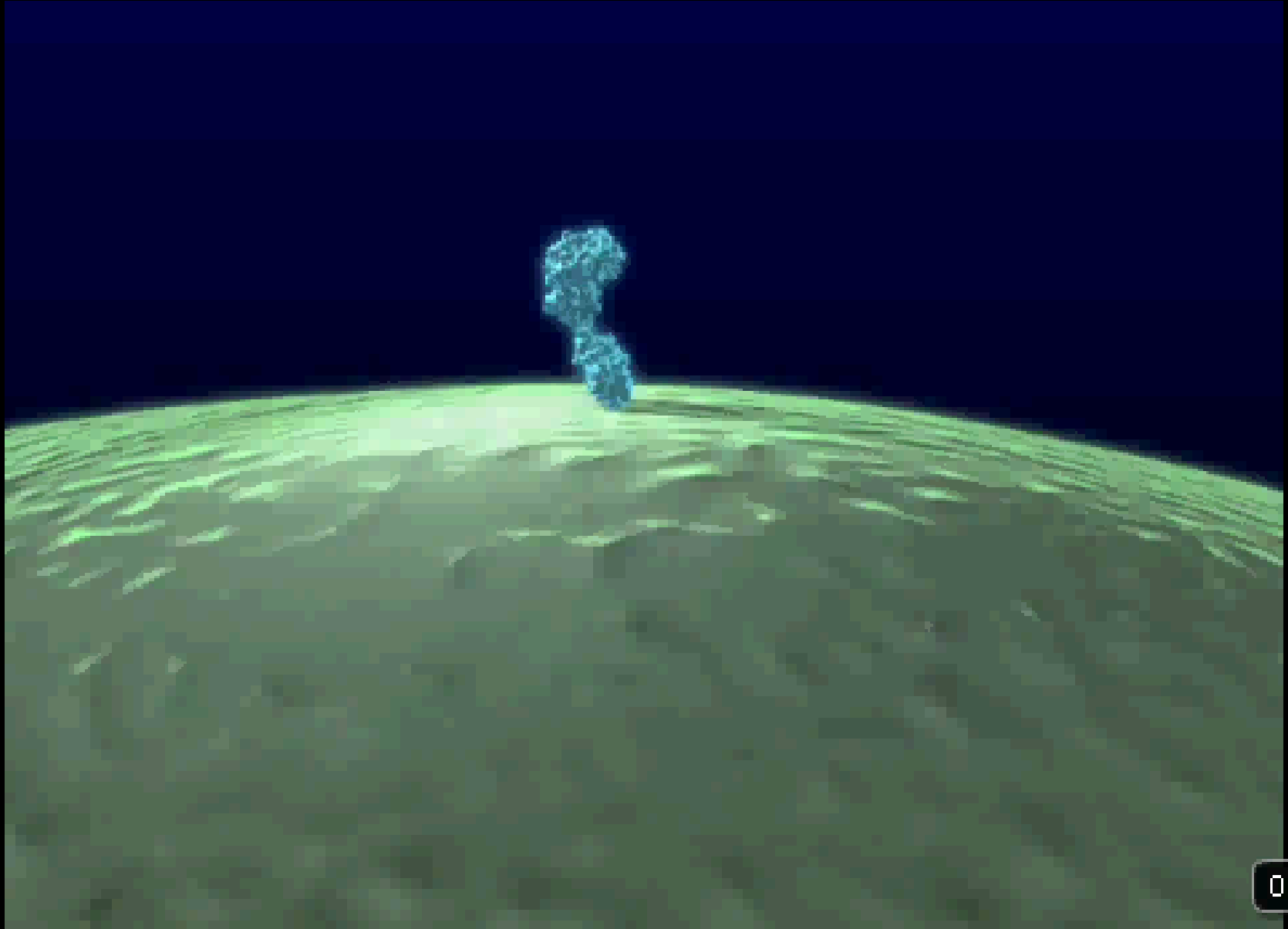


Adaptation induite
Induced fit





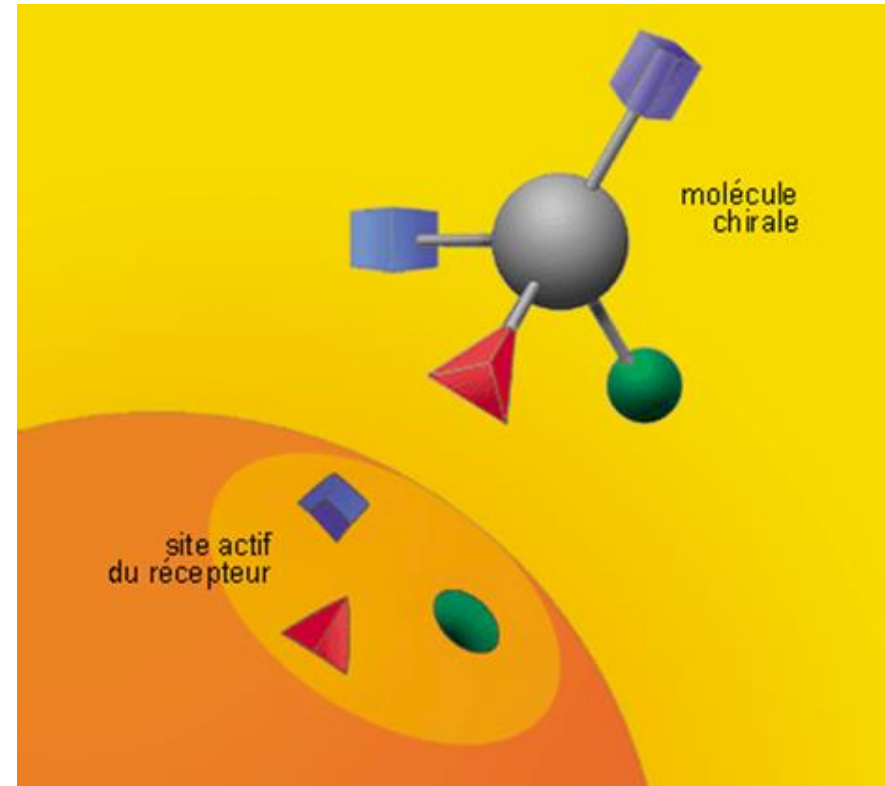
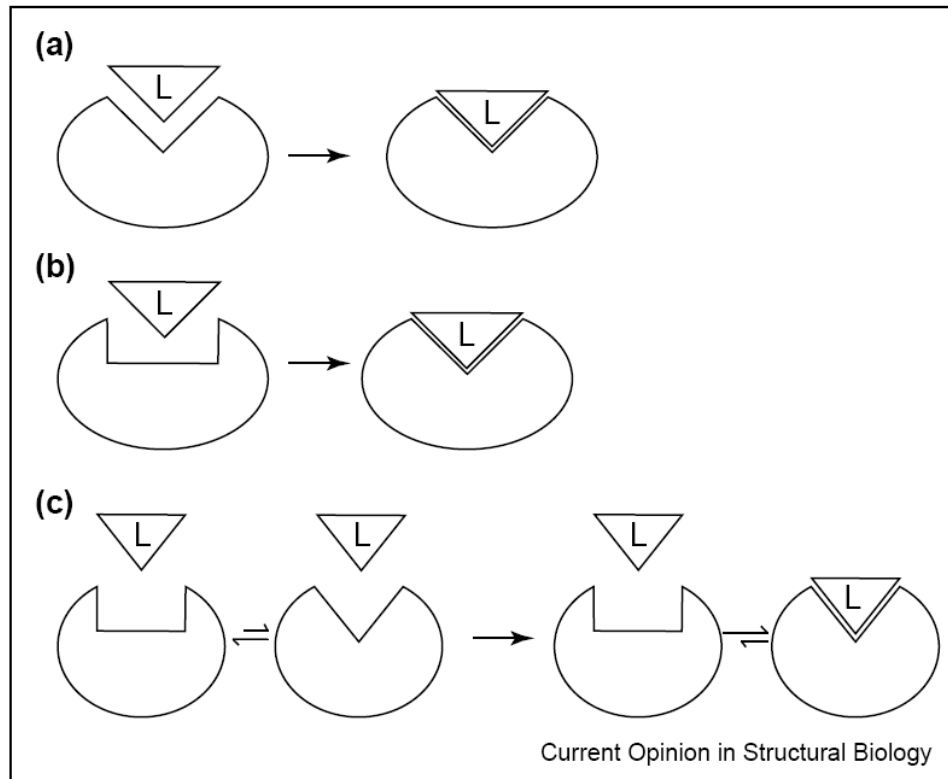
Exemple d'induced fit interleukin-recepteur



Enzymes : inhibiteurs compétitifs, non compétitifs, allostériques
Récepteurs : agonistes, antagonistes, agonistes inverses

Les modèles moléculaires doivent rendre compte des mécanismes d'interaction !

Figure 1



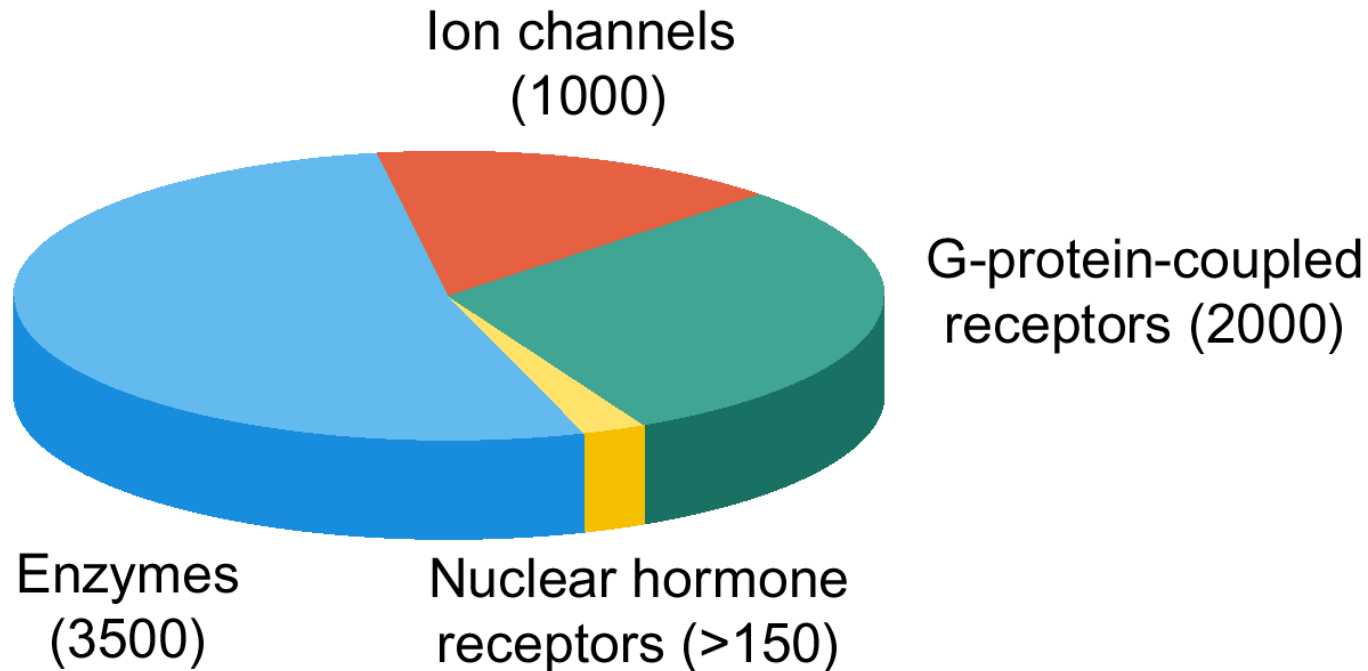
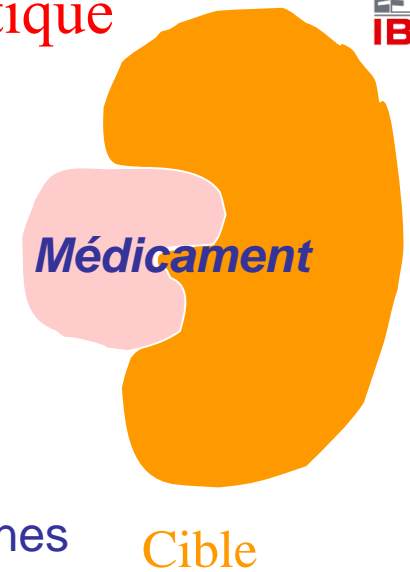
Models of protein binding mechanisms. **(a)** Lock and key model. **(b)** Induced-fit model. **(c)** Pre-existing equilibrium model. L, ligand.

Acides nucléiques

- triples hélices , ARN antisens

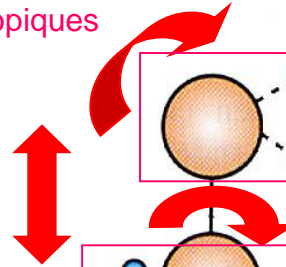
Protéines

- Canaux ioniques
- transporteurs
- Enzymes : protéases, kinases, caspases ...
- Récepteurs : nucléaires, trans-membranaires GPCR
- Récepteurs de petites molécules ou interactions protéines-protéines



$$K_{bind} = \frac{k_1}{k_{-1}} = \frac{[C]}{[P][L]}$$

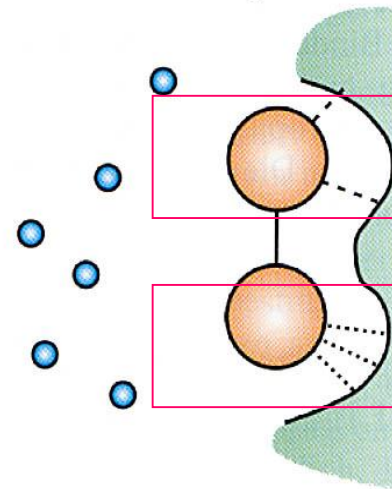
effets
entropiques



Interactions polaires
avec le solvant

Interactions non-polaires
avec le solvant

$$\Delta G_{bind} = -RT \ln K_{bind}$$



Interactions polaires
avec le ligand

interactions
protein-ligand non-
polaires

$$\Delta G_{binding} = -RT \ln(K_{affinity})$$

$$= \Delta G_{complex/solv} - \Delta G_{ligand/solv} - \Delta G_{protein/solv} + \Delta G_{interaction} - T\Delta S + \Delta \lambda$$

$$\Delta G = \Delta H - T(\Delta S)$$

Pour un processus spontané, ΔG doit être négatif

4 scénarios possibles

ΔH	ΔS	ΔG négatif?	Pronostic
(-)	(+)	toujours	toujours spontané
(+)	(-)	jamais	jamais spontané
(+)	(+)	si $T(\Delta S) > \Delta H$	favorable si T élevée
(-)	(-)	si $T(\Delta S) < \Delta H$	favorable si T basse



Etapes

 ΔH
 ΔS

Désolvatation du ligand

 >0
 <0

Désolvatation du récepteur

 >0
 <0

 Ligand \rightarrow conformation active en général défavorable

 >0

 Récepteur \rightarrow conformation de liaison

 >0
 >0

Liaison ligand-récepteur

doit être favorable <0
 >0

Maximiser les étapes favorables et minimiser les étapes défavorables

Minimiser l'enthalpie de désolvatation

- Le Ligand ne doit pas être trop hydrophile
- Pas trop de groupes pour faire des liaisons Hydrogènes (Règles de Lipinski)

Maximiser l'entropie de désolvatation du récepteur

- Le ligand doit remplir toute la cavité → déplacer toutes les molécules d'eau

Minimiser le coût enthalpique pour adopter une conformation active

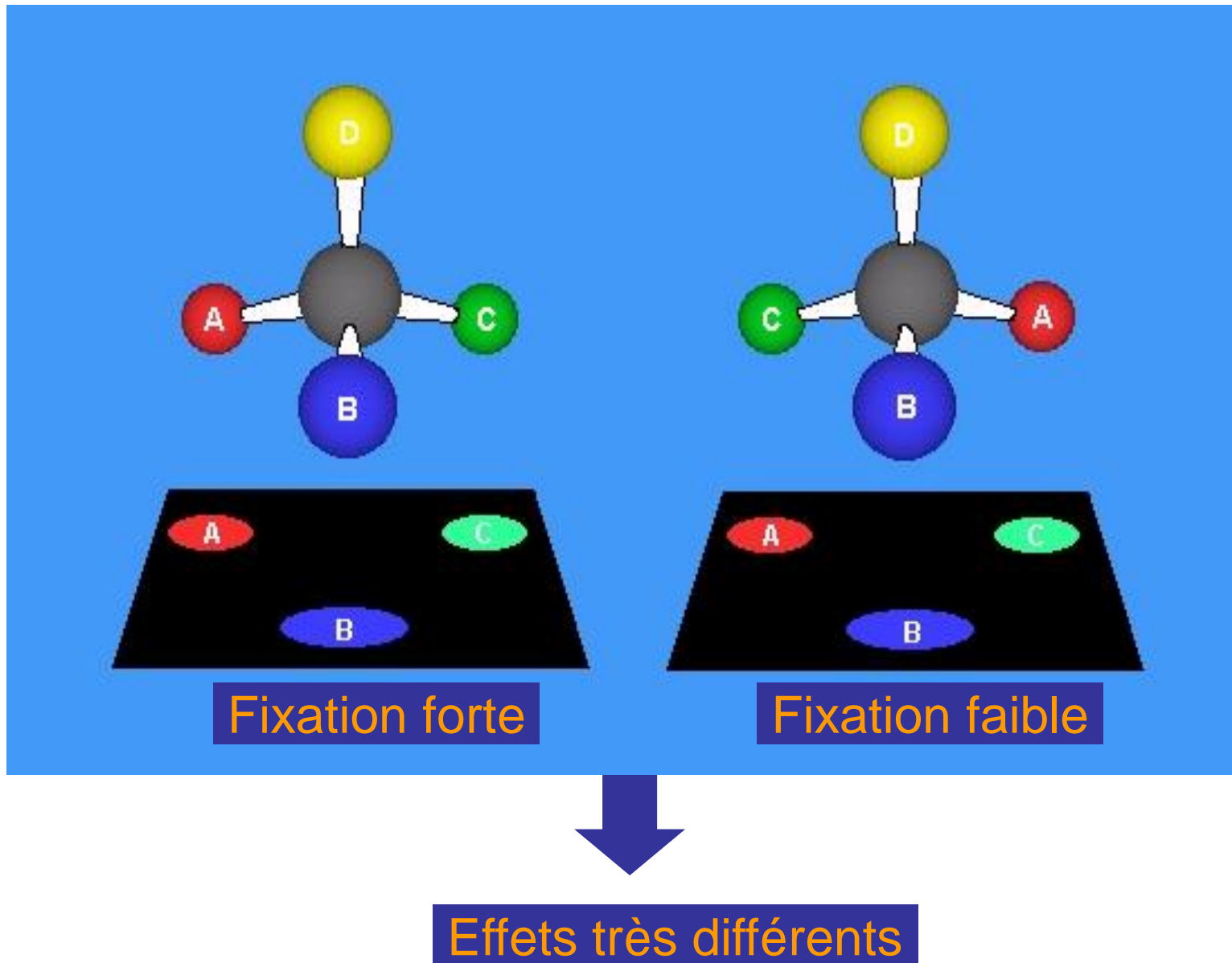
- La conformation en interaction doit être de basse énergie

Minimiser le coût entropique pour adopter une conformation active

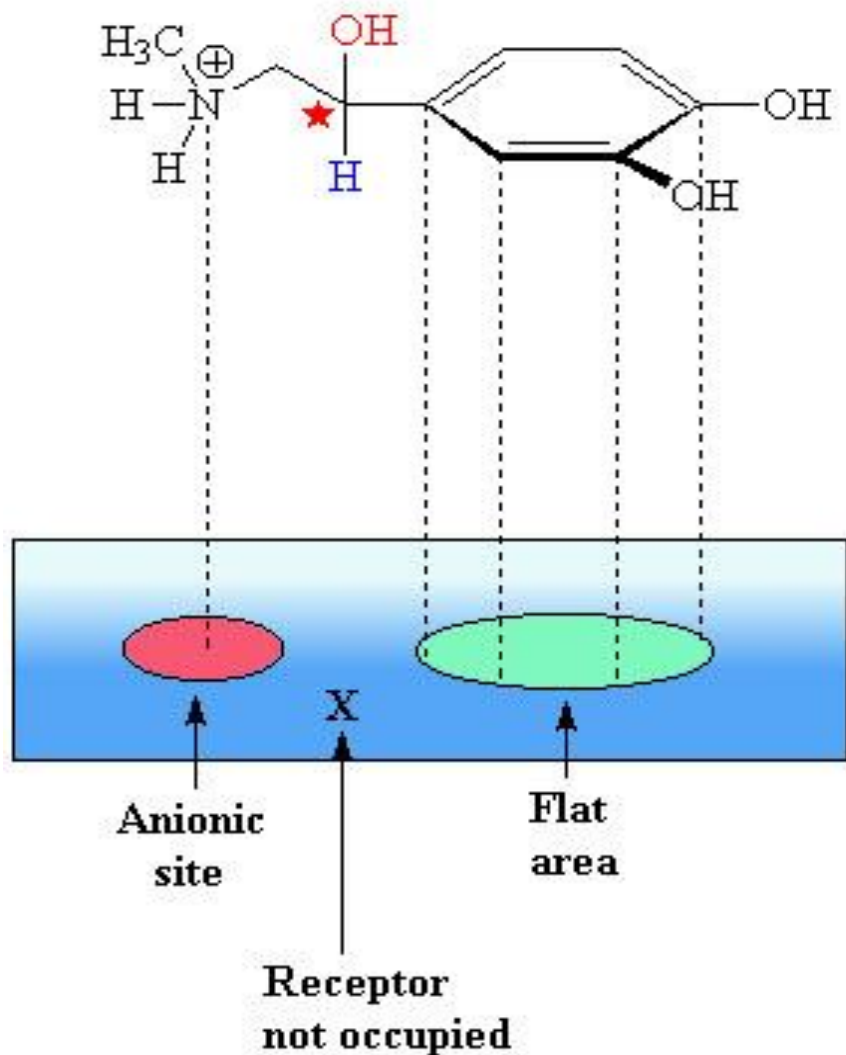
- Le ligand doit être assez rigide (pas trop, la plupart des drogues sont semi-rigides)

Maximiser l'enthalpie de liaison ligand-récepteur

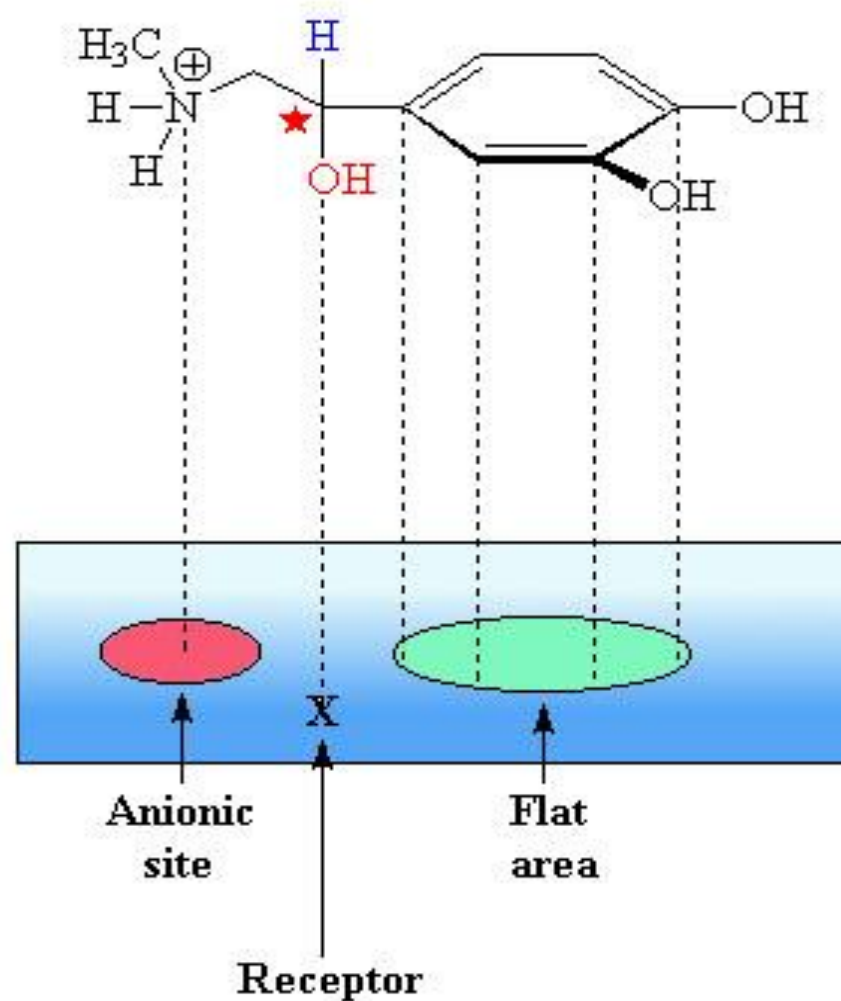
- Surfaces hydrophobes du ligand en regard des surfaces hydrophobes du récepteur
- surfaces hydrophiles du ligand en regard des surfaces hydrophiles du récepteur
- Complémentarité des liaisons H entre récepteur et ligand



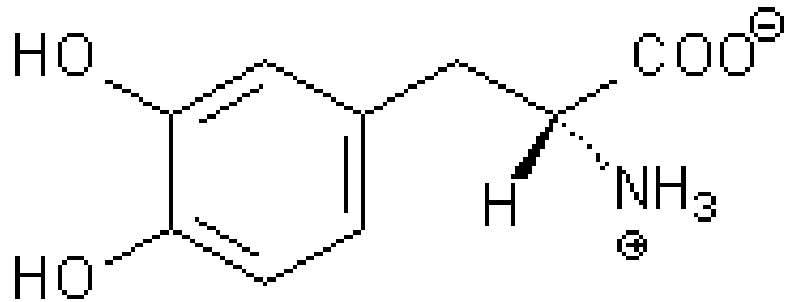
Exemple de préférence chirale



(+) Epinephrine - less active



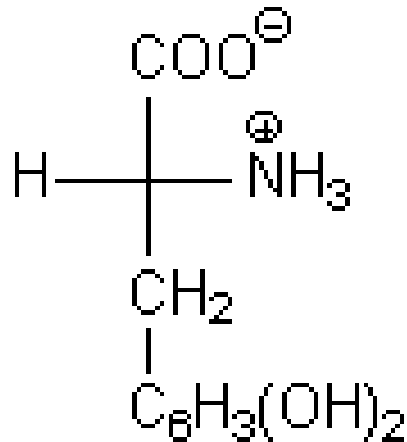
(-) Epinephrine - more active



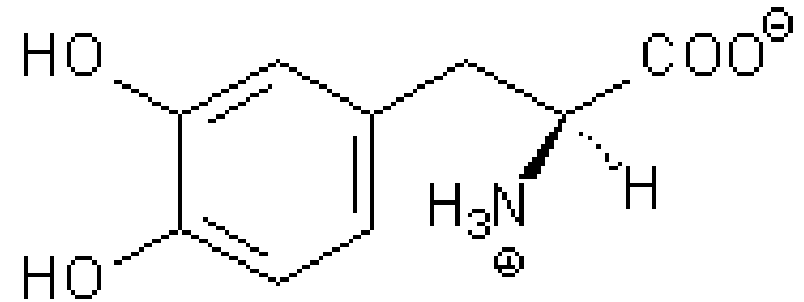
D-dopa



Pas d'effet biologique



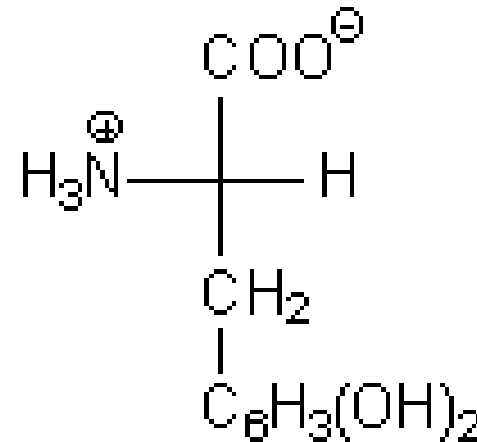
D-dopa



L-dopa

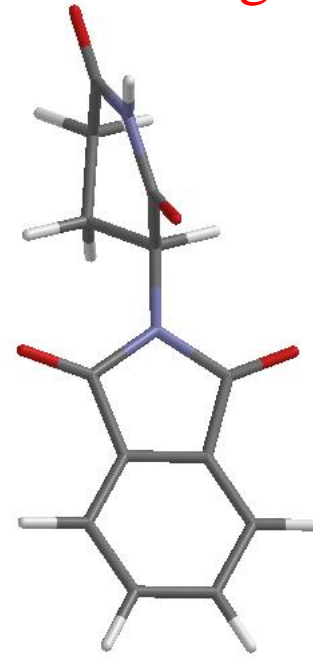
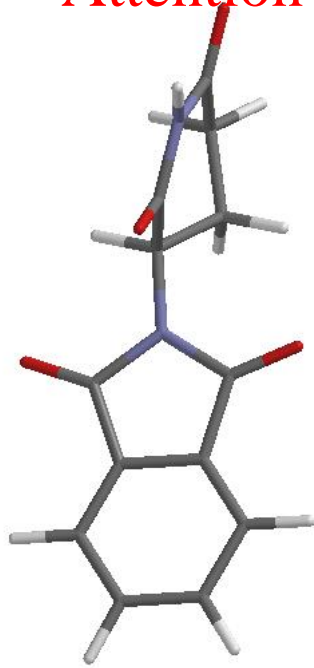


Agent anti-parkinsonien

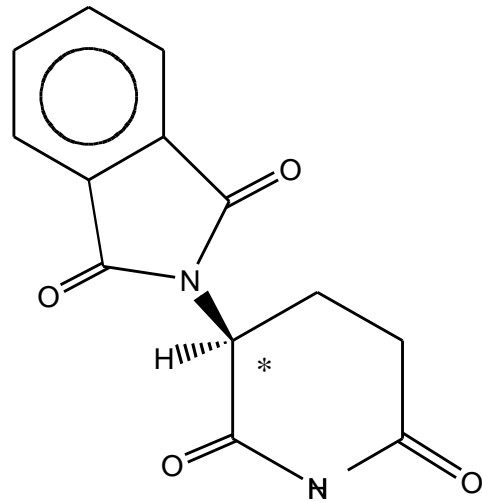


L-dopa

Attention à la stéréochimie des ligands !

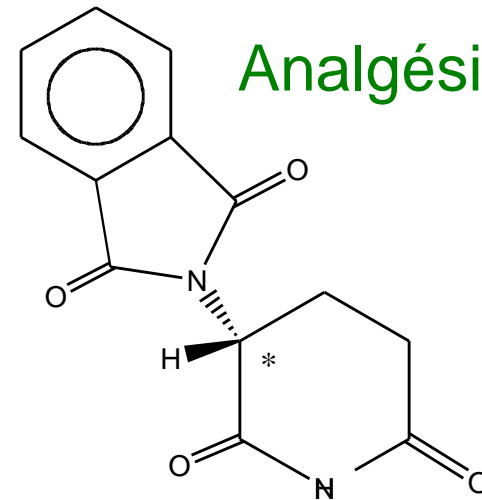


Tératogène

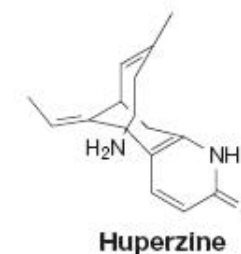
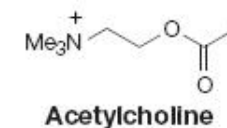
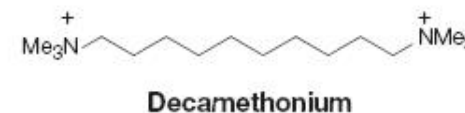
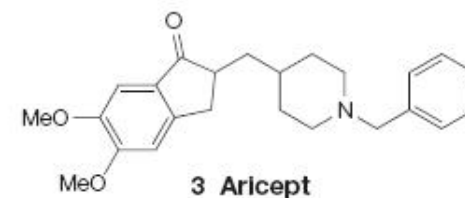
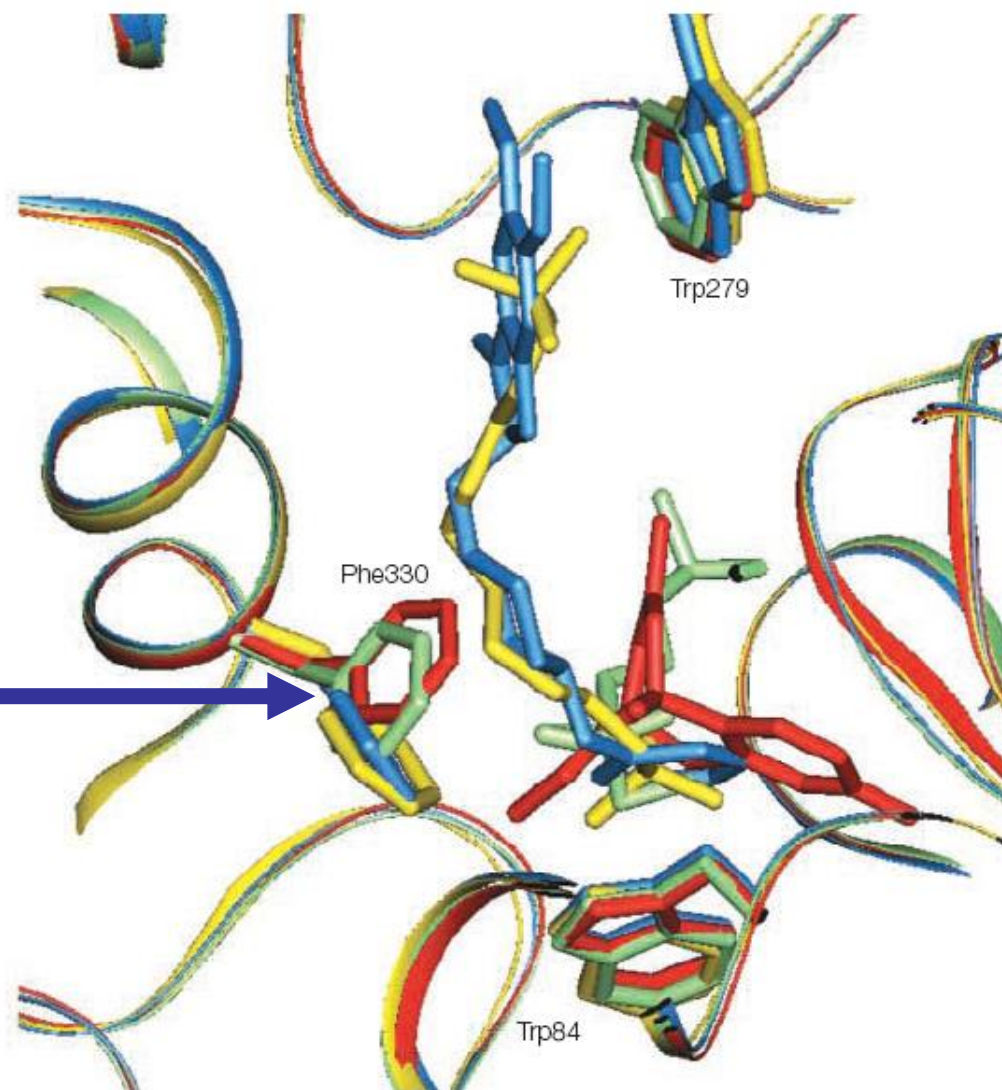


(S)-Thalidomine

Analgésique



(R)-Thalidomine



Phe 330
porte mobile

Figure 5 | **Multiple conformations of a single residue.** Overlay of native and three *Torpedo californica* acetylcholinesterase–ligand complexes using the protein C- α atoms (Protein Data Bank codes 2ACE, 1EVE, 1VOT and 1ACL). Key protein side chains are indicated by thick lines, as are the inhibitors. The colour codes are: donepezil (Aricept, blue), decamethonium (yellow), native (green), huperzine (red). The flexibility of Phe330, in comparison with the rigidity of the rest of the gorge, is highlighted.

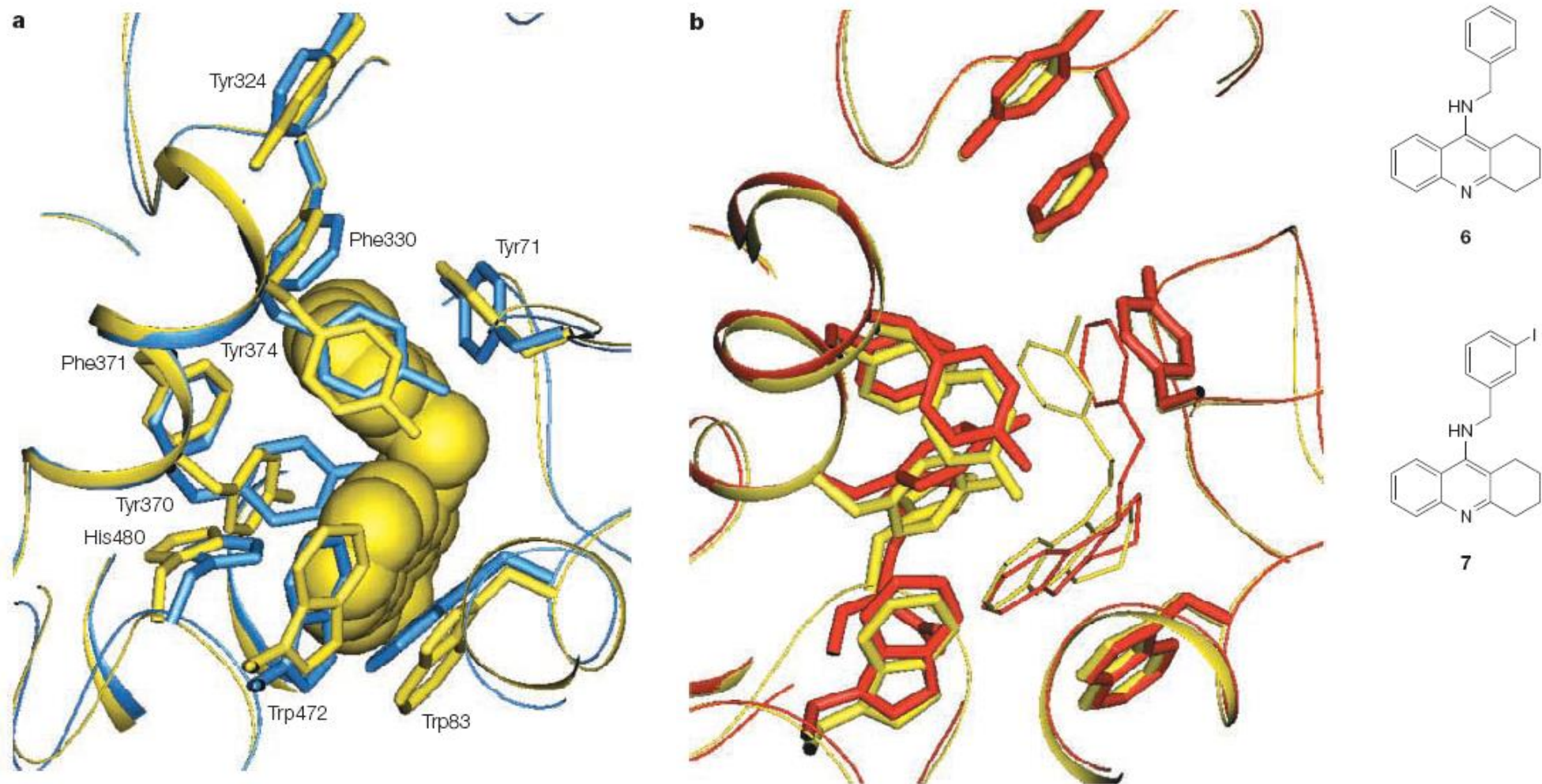
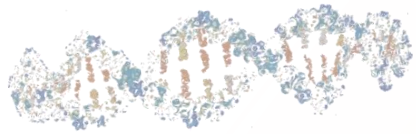


Figure 7 | **Movement of a large number of residues: example 1.** **a** | The complex between *Drosophila melanogaster* acetylcholinesterase and compound (6) in yellow showing displacement of nine aromatic residues when compared with the native structure in blue (Protein Data Bank codes 1QO9 and 1DX4). **b** | Overlay of the complexes between *Drosophila melanogaster* acetylcholinesterase complexed with compounds (6) in red and (7) in yellow. Both the side chains of the protein and the positions of the inhibitors are altered (Protein Data Bank codes 1QON and 1DX4).



Gènes

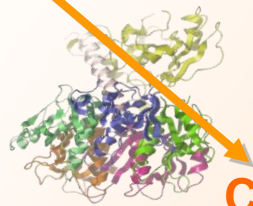
```

-----TNNLMASGGACC
AGWNAYNNLMA--DGTC
AGWNAYIDNMA--DGTC
    
```

Génomique et protéomique

Outils bioinformatiques

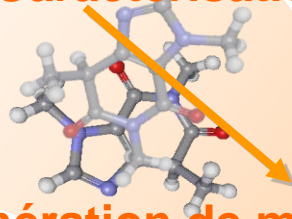
Identification de cibles



*Détermination et relation
Structure-fonction*

**RMN
Diffraction rayons X
Modélisation par homologie
Comparaison structurale
Dynamique Moléculaire**

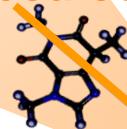
Caractérisation de cibles



Identification de ligands

**Screening virtuel
Conception de ligands
Pharmacophore 3D
QSAR
Sélection et diversité**

Génération de molécules actives



Optimisation

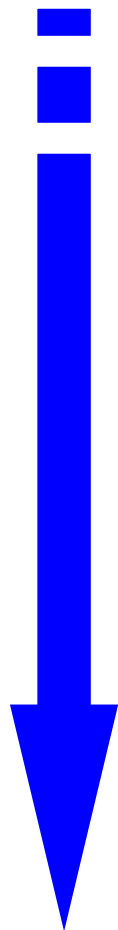
**Profil de sélectivité
Profil ADME*/Tox**

Lead

* Absorption
Distribution
Métabolisme
Élimination



Recherche



Sélection de cibles

Structure 3D de la cible connue, faisabilité ?
ligands connus, pharmacophore ?

Recherche de touche «hit»

Virtual screening, database filtering

De la touche au « chef de file »

docking, sélectivité

Optimisation de la molécule

virtual ADMET, chemogenomics

Développement



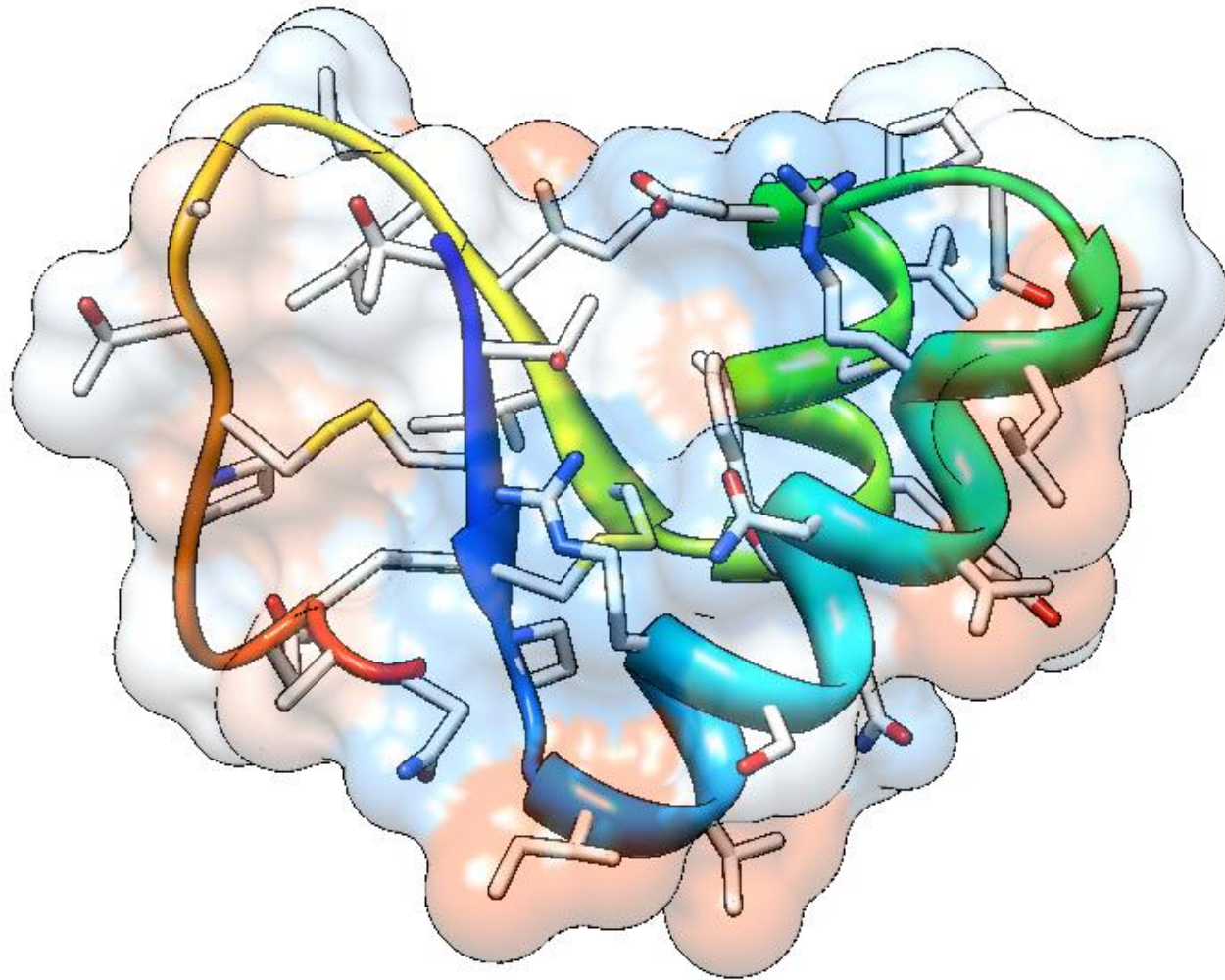
- → 1970s
 - Leads = produits naturels, hasard
 - Optimisation tributaire de la chimie, fondée sur hypothèses
 - **Limitations : expérimentation animale**
- → 1990s
 - Modélisation moléculaire
 - Tests in-vitro (inhibition d'enzymes, liaison au récepteur)
 - **Limitations : synthèse chimique**
- → 2000
 - Génie génétique (production de protéines)
 - Structure-based drug design
 - Chimie combinatoire (mélanges)
 - High-throughput screening (HTS)
 - **Limitations : propriété ADME* des leads**
- 2000 →
 - Génomique, protéomique, pharmacogénomique, bioinformatique
 - Animaux transgéniques proof of concept (POC)
 - Chimie combinatoire (1 composé par puit, bibliothèques focalisées ou diverses)
 - Structure-based drug design
 - Data mining (screening virtuel de très grande bibliothèque)
 - ADMET profiling (HTS et virtual ADMET, chemo-génomique)
 - data pipelining
 - **Limitations : validation de cibles, déluge de données**

Quelques causes d'échec en drug design

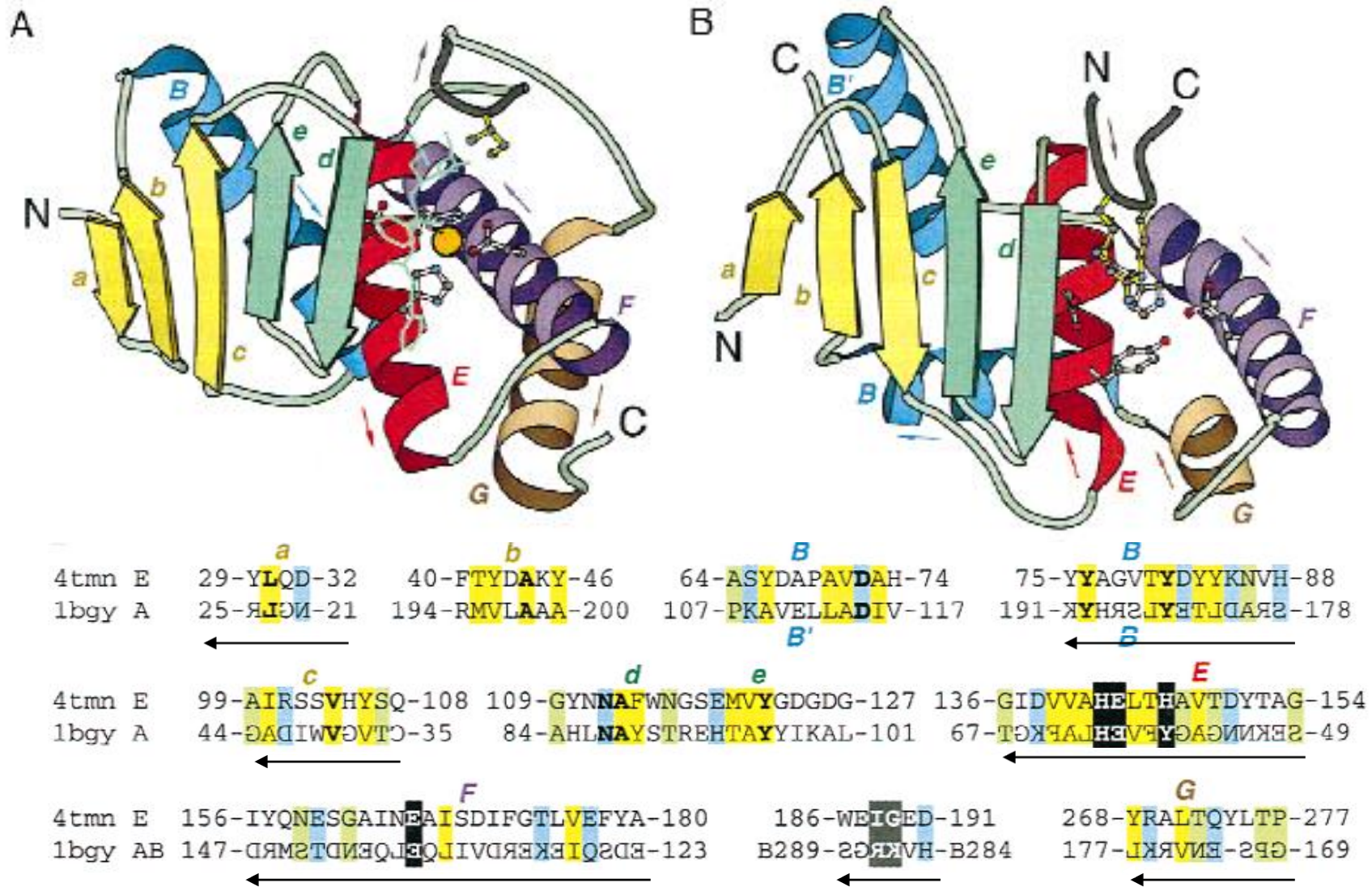
- ✓ mauvaise pharmacocinétique (mauvais profil ADMET* chez l'homme, métabolites problématiques)
- ✓ Faibles activités cliniques
- ✓ Effets secondaires, toxicité (métabolites, sélectivité)
- ✓ Stratégie marketing
- ✓ Plus l'échec est tardif plus il coûteux !

* Absorption
Distribution
Métabolisme
Élimination
Toxicité

- Conception d'inhibiteurs à partir de la structure du substrat
- Identification de Pharmacophores
- Peptidomimétiques
- Structure-based drug design
Meilleure prise en compte de l'Affinité, de la sélectivité
- Computer-aided drug design
- Combinatorial drug design
- caractère drug-like, biodisponibilité (e.g., Lipinski's *Rule of Five*)



A partir d'une structure 3D
Structures proches fonctions différentes
Structures différentes = fonctions identiques

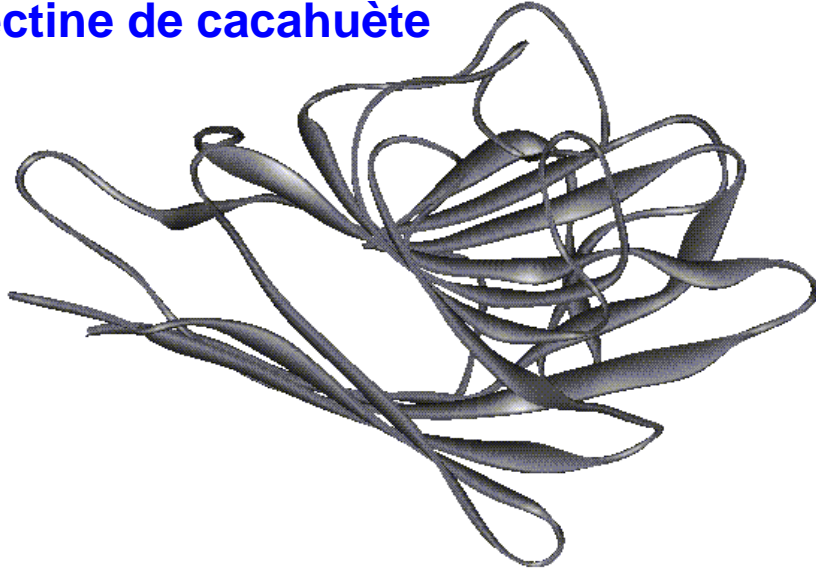


Les connexions sont différentes (topologies différentes) mais l'architecture est la même
Fonction identique

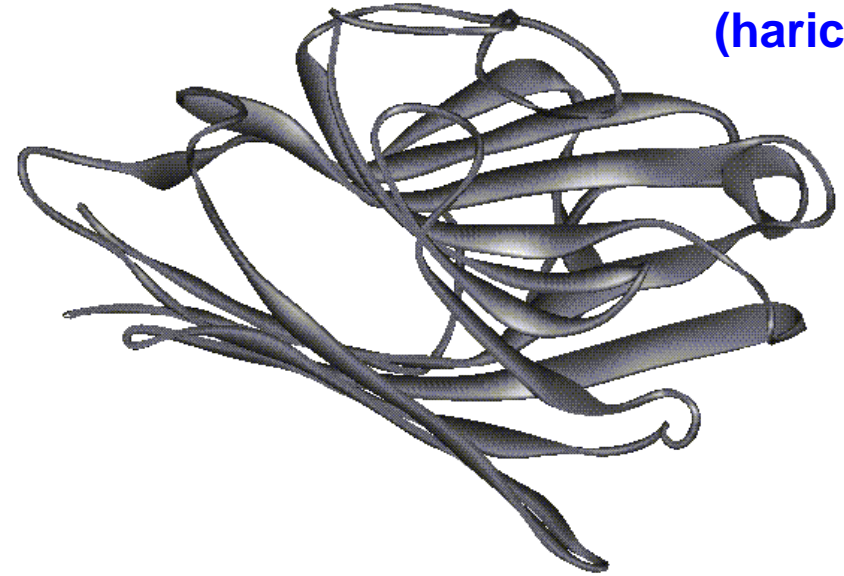
(d'après Grishin et Makarova, 1999 Prot Science)

Evolution divergente: Séquences similaires,
mêmes repliements, fonctions un peu différentes

Lectine de cacahuète



Arceline
(haricot)

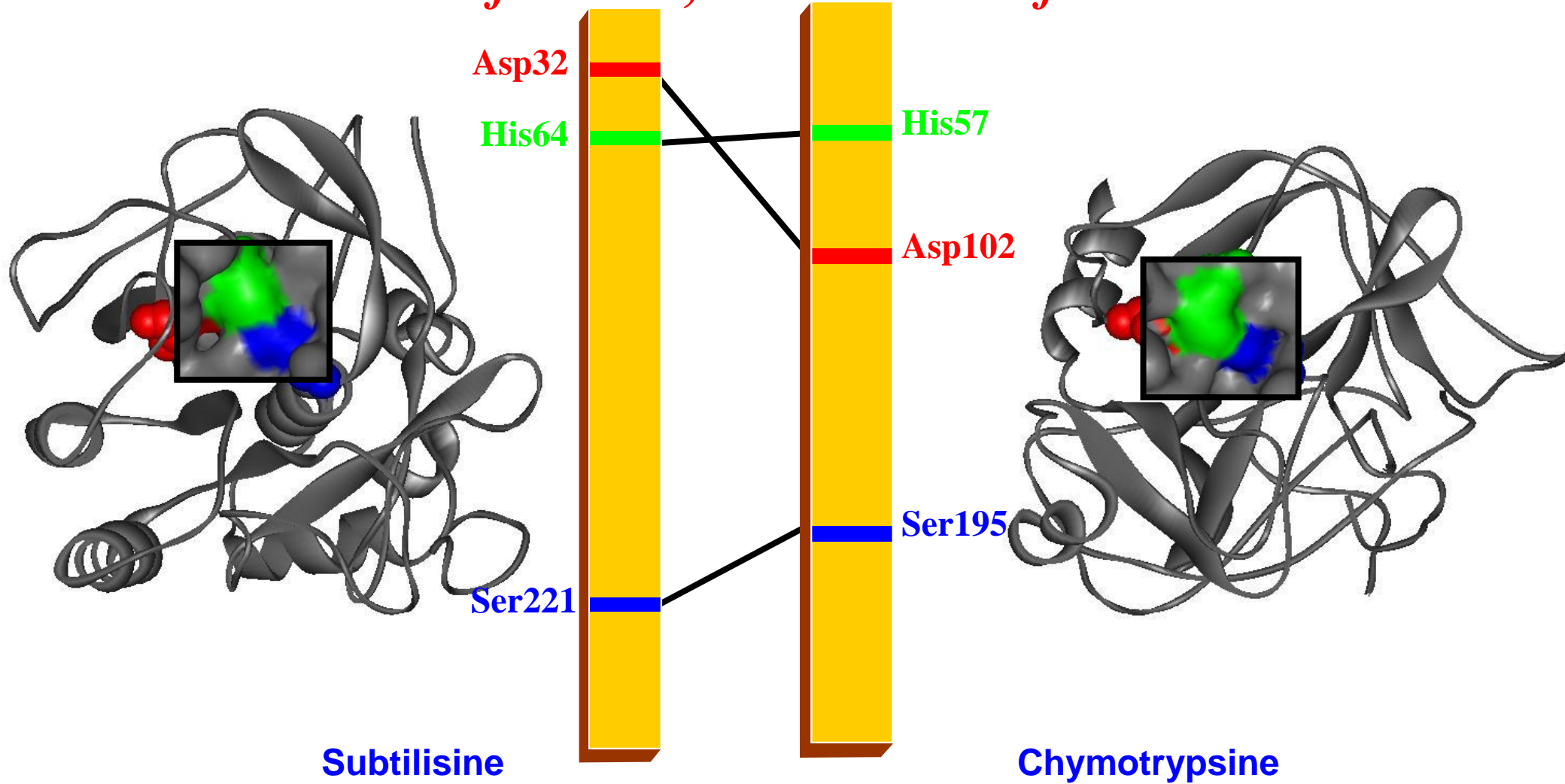


• *RMSD sur positions conservées = 4,1 Å*

	10	20	30	40	50	60	70	80	90	100	110	120	
2PELx0	AETVSFNFNSFSEGNPAINFQGDVTVLSNGNIQLTNLN	-----KNSVGRVLYAMPVRIWSSATGNVASFLLTSFSFEMKDIKDYPADGIIFFIAPEDTQIPAGSIGGGTLGVSDTKGAGHFVGV											
1IOAx1	ATETSFNFPNFHTDDKLI-LQGNATISSKGQLQLTGVGSNELPRVDSLGRAFYSDPIQIKD	--SNNVASFNTNFTFIIRAKNQSI SAYGLAFALVPVNSPPQKKQEFGLGIFNTNNPEPNARTVAV											
Prim.cons.	* .**** .*	: * :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	* :*:* :	
	135	145	155	165	175	185	195	205	215	225	235	245	
2PELx0	EFDYNSSEYNDPPTDHVGIDVNSVDSVKTVPWNSVSGAVVKVTVIYDSSTKTLSVAVTNDNGDIT	-IAQVVDLKAKEPERVKFGFSASG--SLGGRQIHILRSWSFTSTLITTRRS-----											
1IOAx1	VFNTFKNR--IDFDKNFIKPYVN	----ENCDFHKYNGEKTVDVQITYDSSNNDLRVFLHFTVSQVKCSVSATVHLEKEVDEWVSVGFSPTSGLTEDTTETHDVLWSFS											
Prim.cons.	* :*:* *	:*:* **	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	:*:* *	
	2F2T22N2EY2D22222222VNSVDSV22222222G2222V222YDSS222L2V2222222222S2222V2L22222E2V22GFS222GL222222H222SWSF2S22222222L22222222												

• *26 % d'identité entre les séquences*

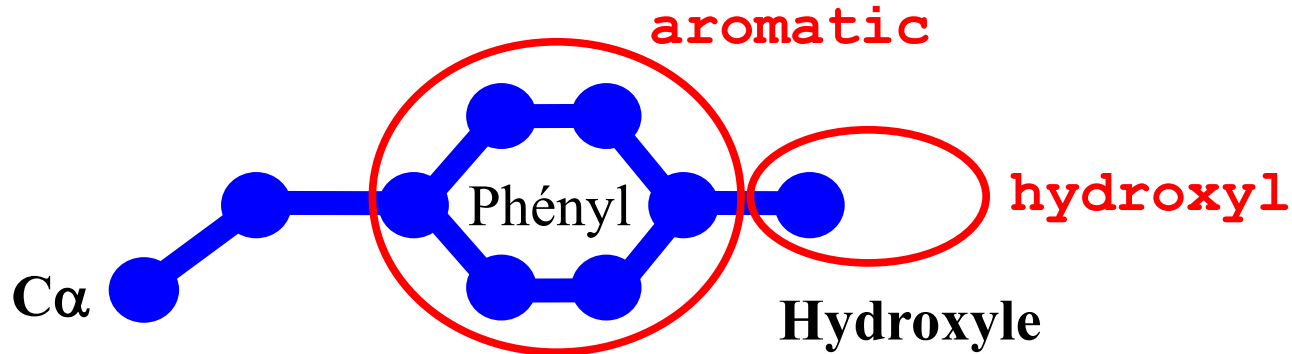
Evolution convergente : repliement différent, même fonction, même site actif



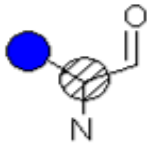
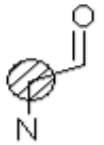
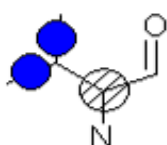
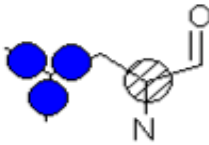

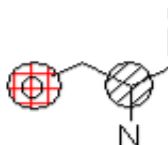
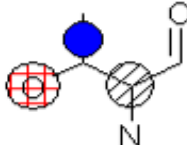
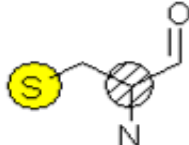
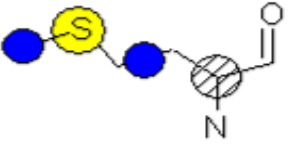
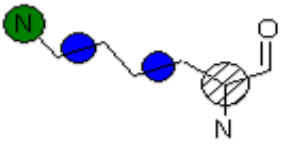
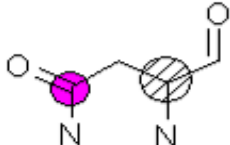
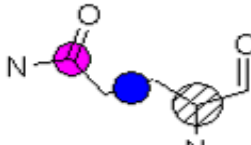
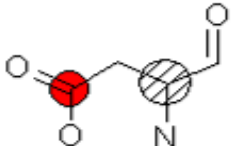
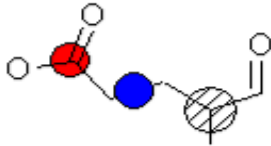
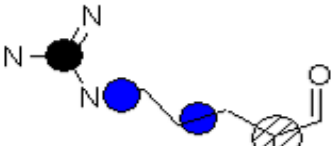
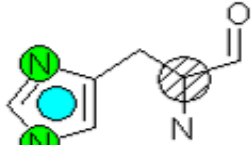
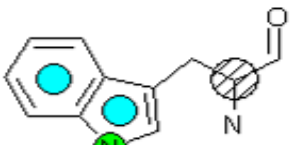
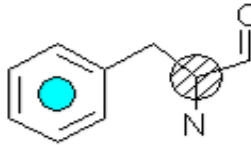
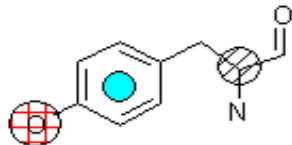
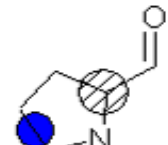










Protéases à sérine

Détection de sites dans les structures 3D

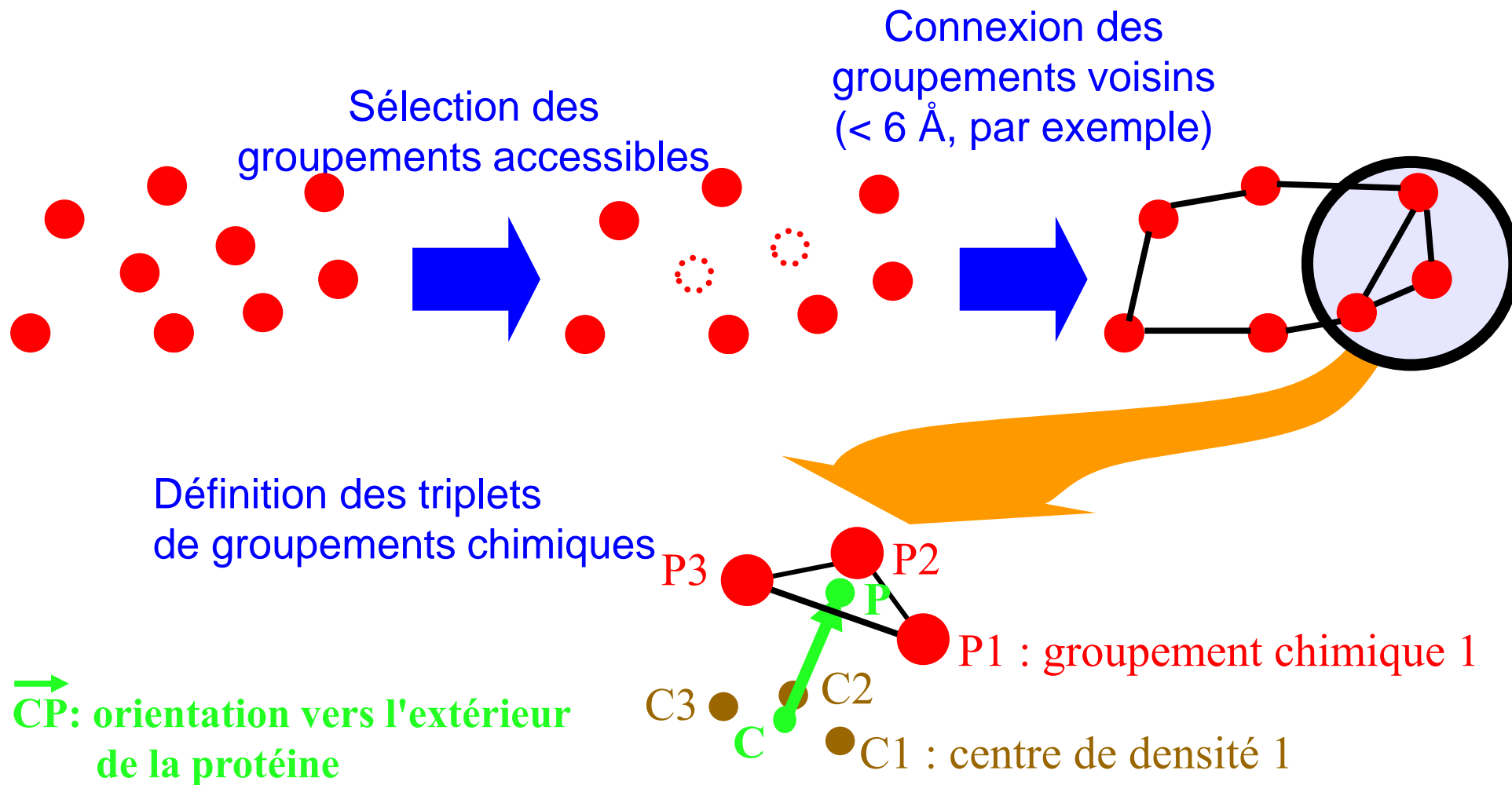
Etape 1 : découpage en groupements chimiques



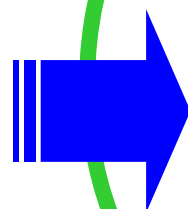
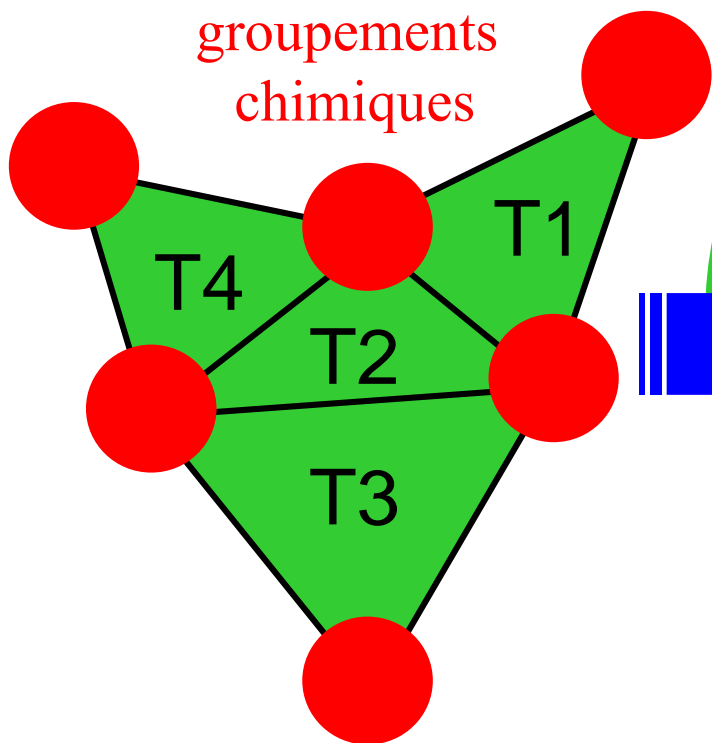
Exemple : tyrosine = **aromatic,**
hydroxyl

			
<u>Alanine (A)</u>	<u>Glycine (G)</u>	<u>Valine (V)</u>	<u>Leucine (L)</u>
			
<u>Isoleucine (I)</u>	<u>Serine (S)</u>	<u>Threonine (T)</u>	<u>Cysteine (C)</u>
			
<u>Methionine (M)</u>	<u>Lysine (K)</u>	<u>Asparagine (N)</u>	<u>Glutamine (Q)</u>
			
<u>Aspartic acid (D)</u>	<u>Glutamic acid (E)</u>	<u>Arginine (R)</u>	<u>Histidine (H)</u>
			
<u>Tryptophan (W)</u>	<u>Phenylalanine (F)</u>	<u>Tyrosine (Y)</u>	<u>Proline (P)</u>
 hydrophobic aromatic	 hydroxyl	 Amide	
 hydrophobic	 carbon alpha	 guanidinium	
 amine	 thiol		
 amine aromatic	 carboxylic acid		

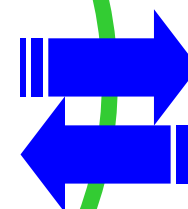
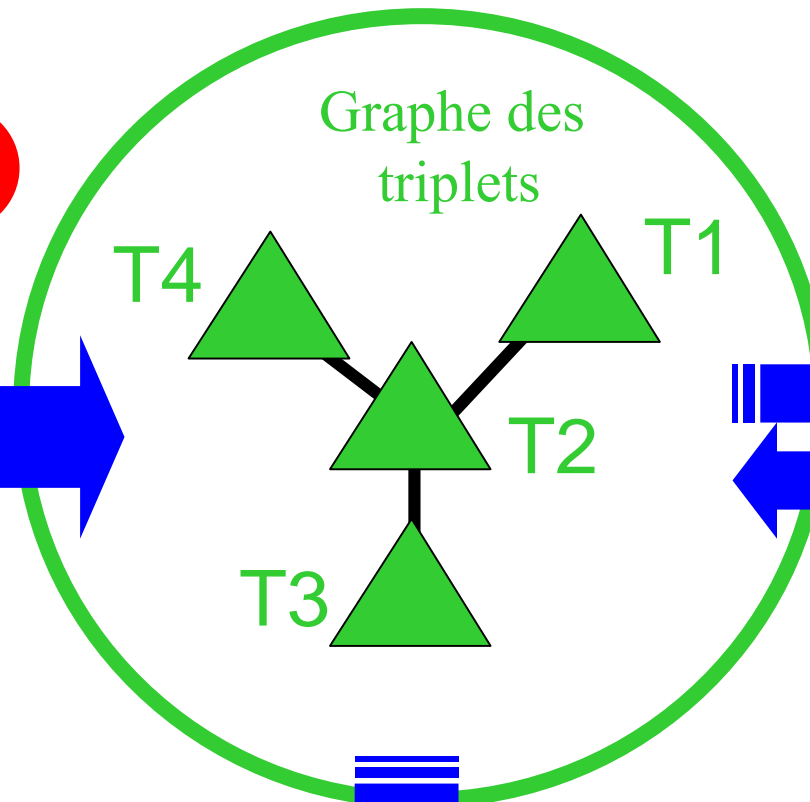
Etape 2 : génération d'un graphe de triplets de groupements chimiques



Graphe des
groupements
chimiques



Graphe des
triplets

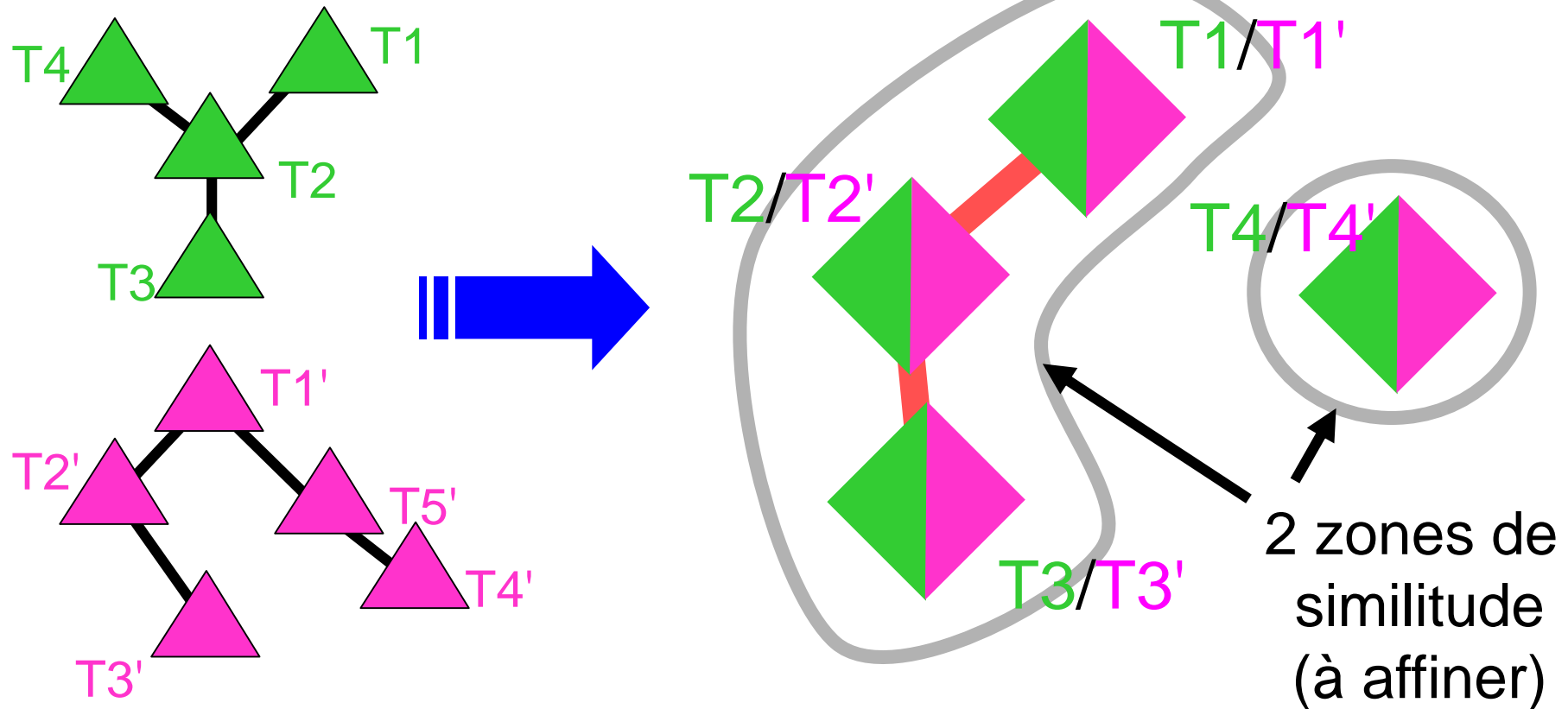


Base de
données

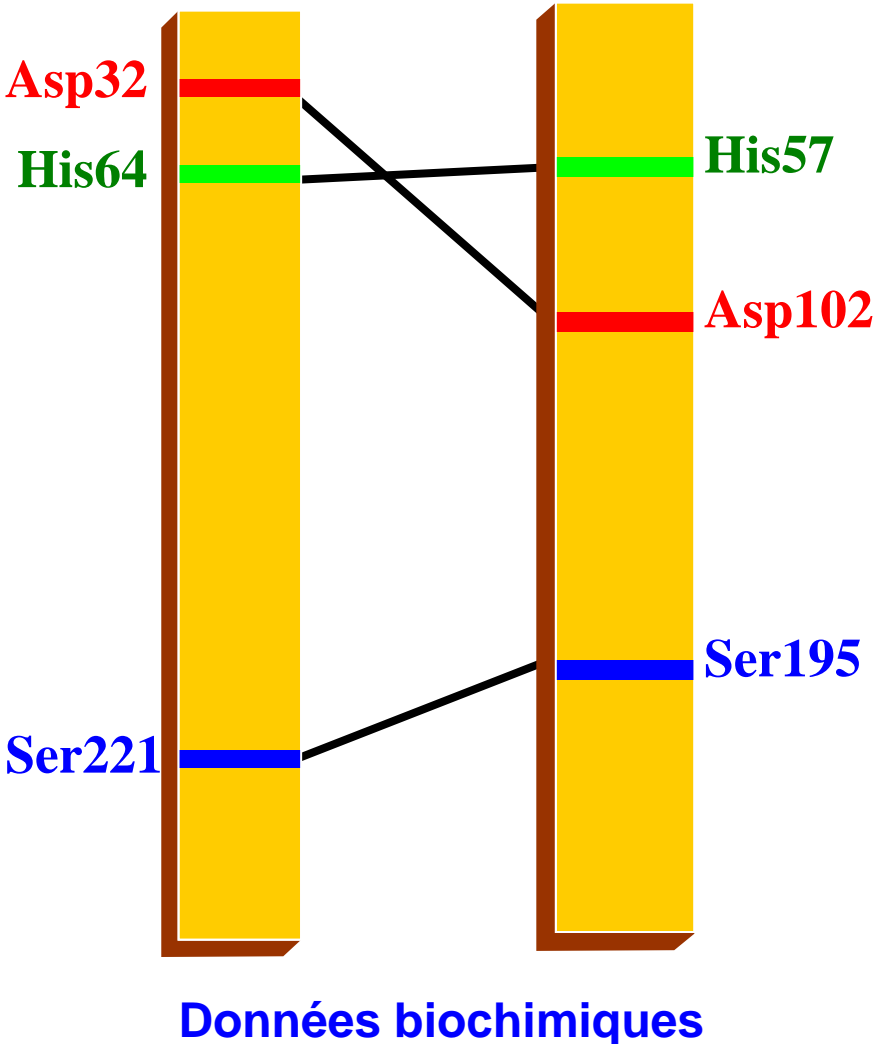
Utilisation
(comparaisons)

Principe de la comparaison entre 2 molécules

- Identification de toutes les paires de triplets similaires
- Connexion des paires



Exemple application : protéases à sérine



Comparaison subtilisine 1SBC / chymotrypsine 1AFQ



Résultat :

[Comparison result of (1AFQ,1SBC) :

```
----- Patch number 1 -----
Number of groups: 4
Score = 1.116 [RMSD = 0.708] [penalty = 0.408]
```

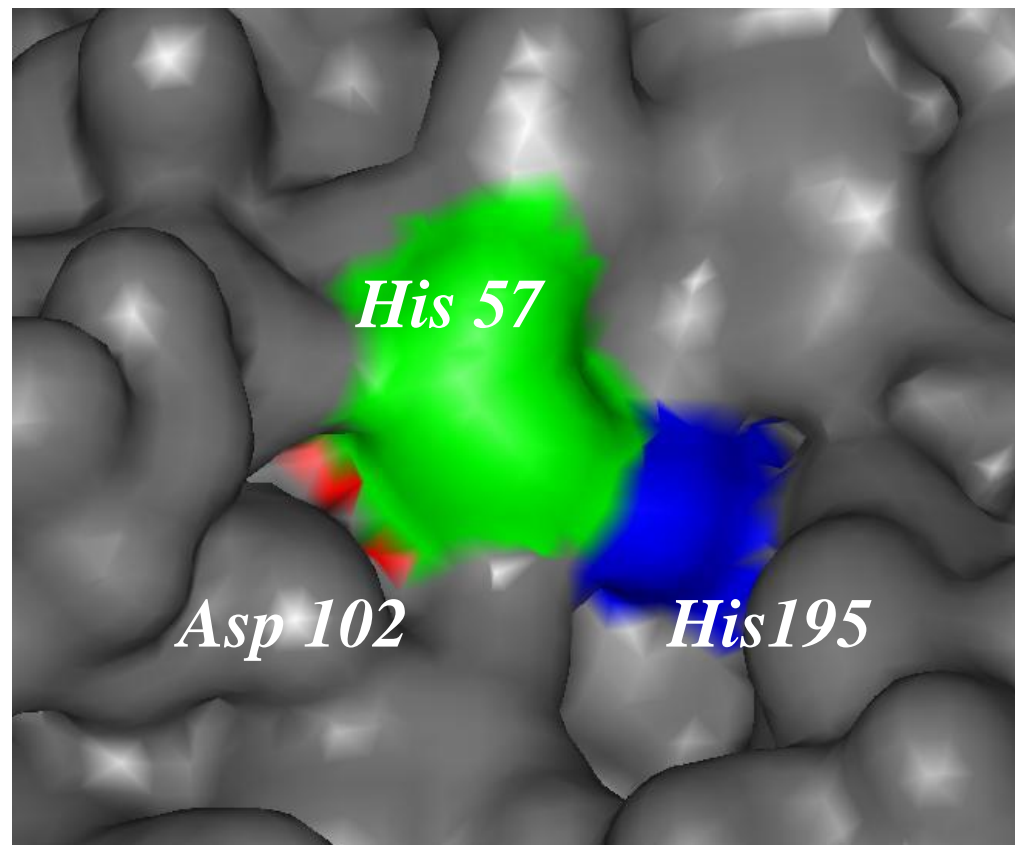
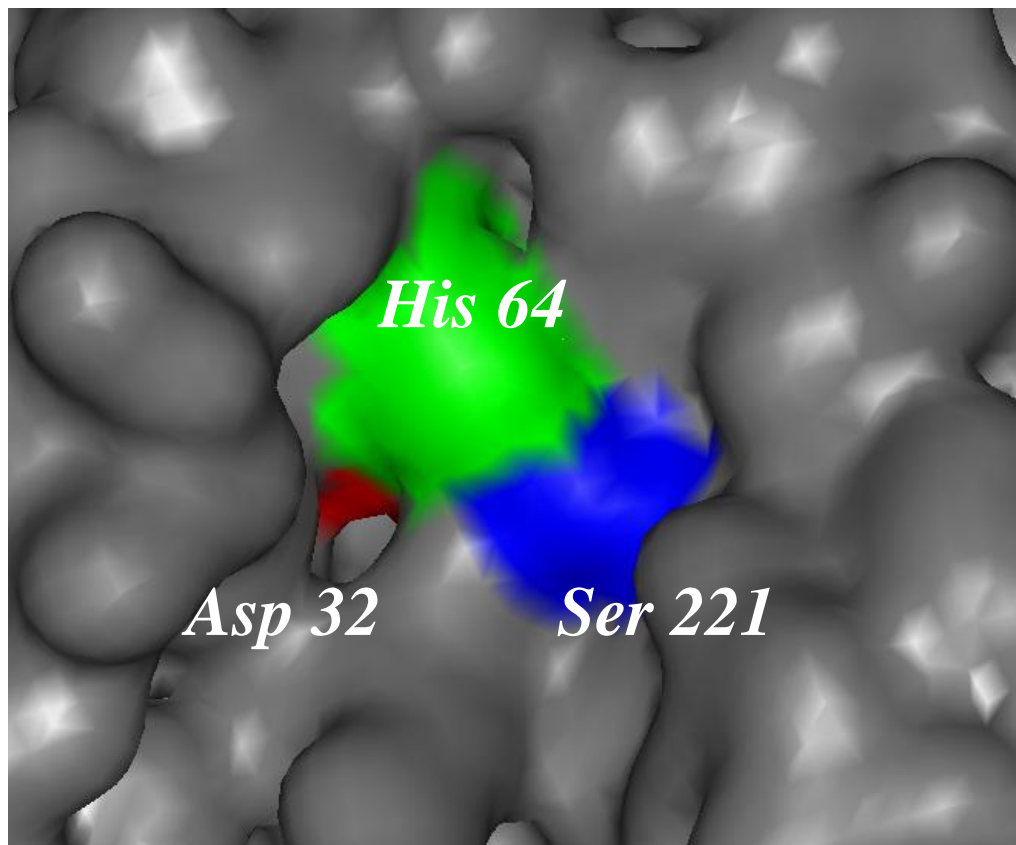
Selected pairs of groups:

ammonium	B	HIS	57		0.55		30.245	2.96...
ammonium		HIS	64		0.54		30.407	2.70...
aromatic	B	HIS	57		0.56		31.077	3.61...
aromatic		HIS	64		0.56		31.309	3.27...
acyl	B	ASP	102		0.67		32.459	7.49...
acyl		ASP	32		0.66		31.730	7.17...
hydroxyl	C	SER	195		0.59		27.628	2.51...
hydroxyl		SER	221		0.57		27.964	3.42...

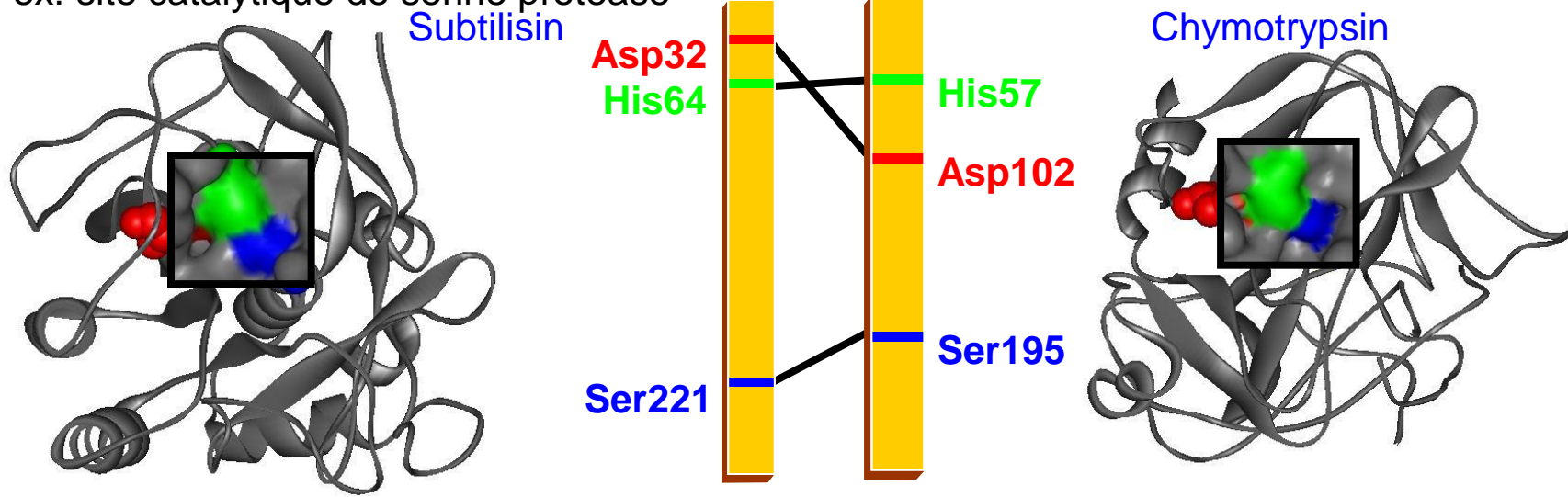


Subtilisine (1SBC)

Chymotrypsine (1AFQ)

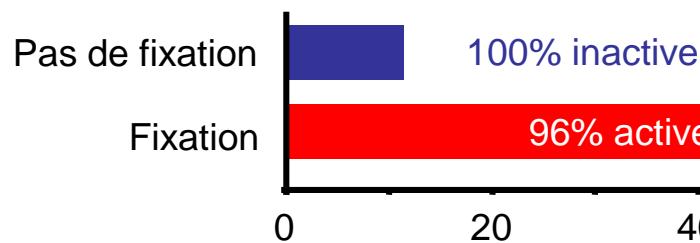
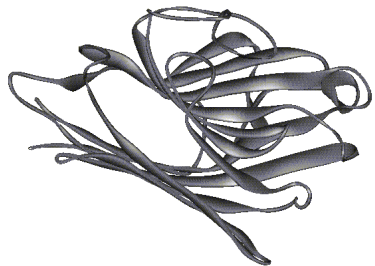


- ✓ Comparaison entre structure 3D de protéines ayant des folds et des séquences différents
ex: site catalytique de serine protease



- ✓ Chercher un site 3D (décrits en terme de groupement chimiques) dans une librairie de structures 3D
- ✓ Discrimination entre des structures 3D fonctionnelles et non fonctionnelles.

1IOA : Ne fixe pas les sucres

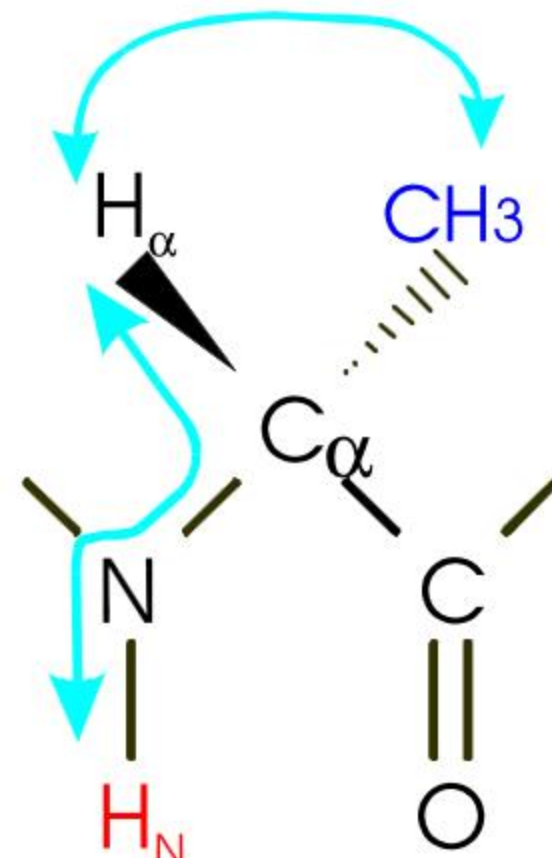
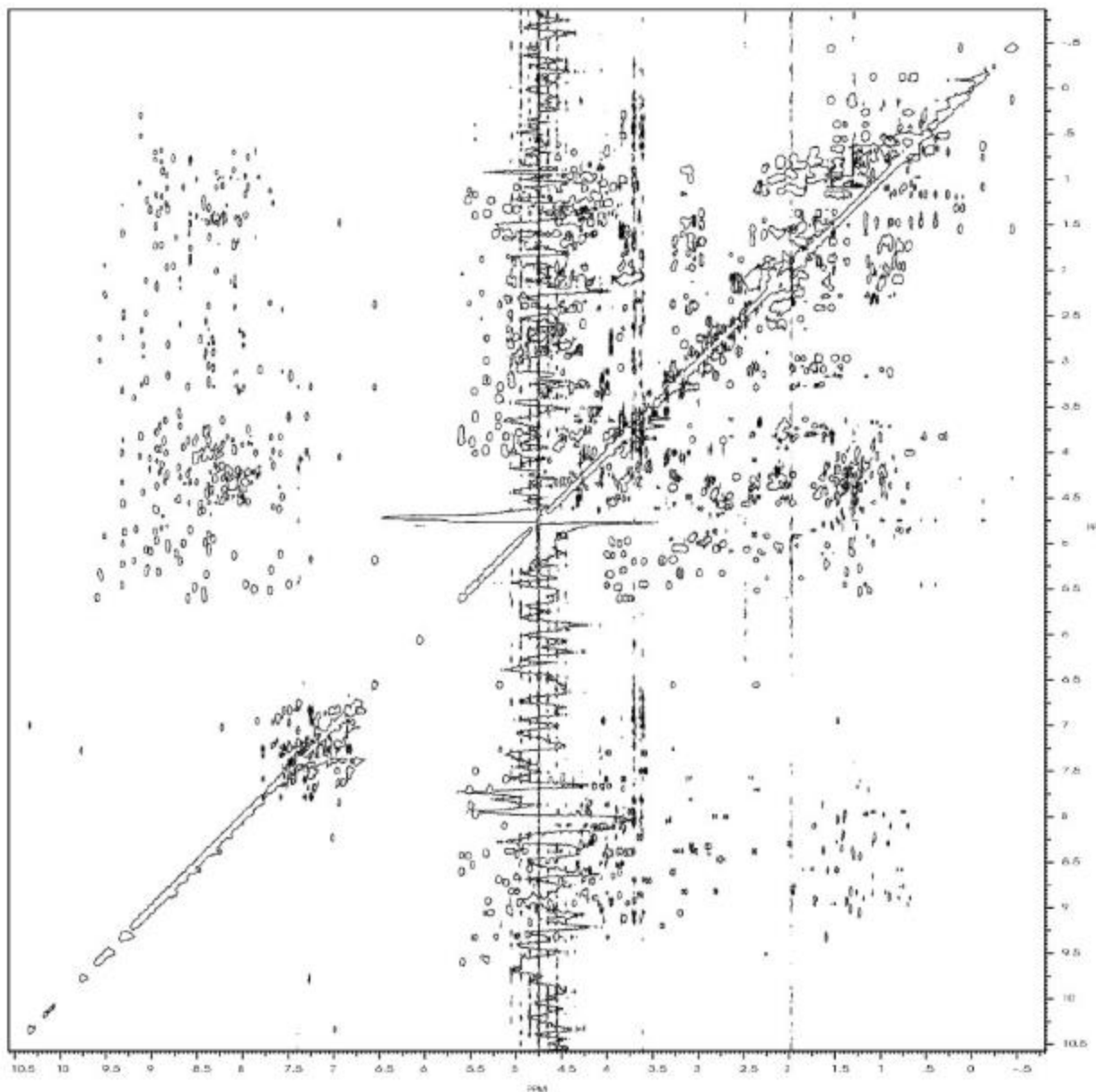


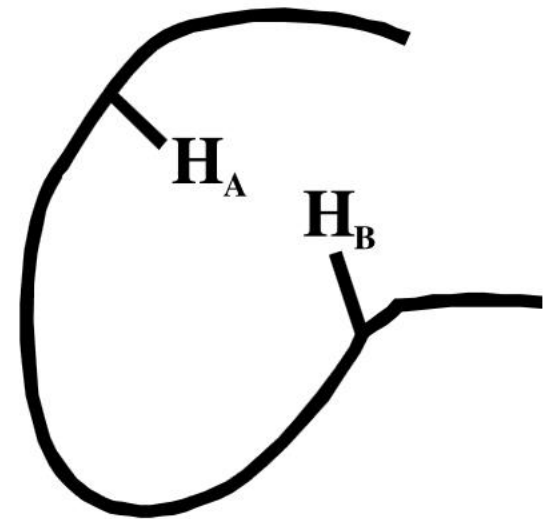
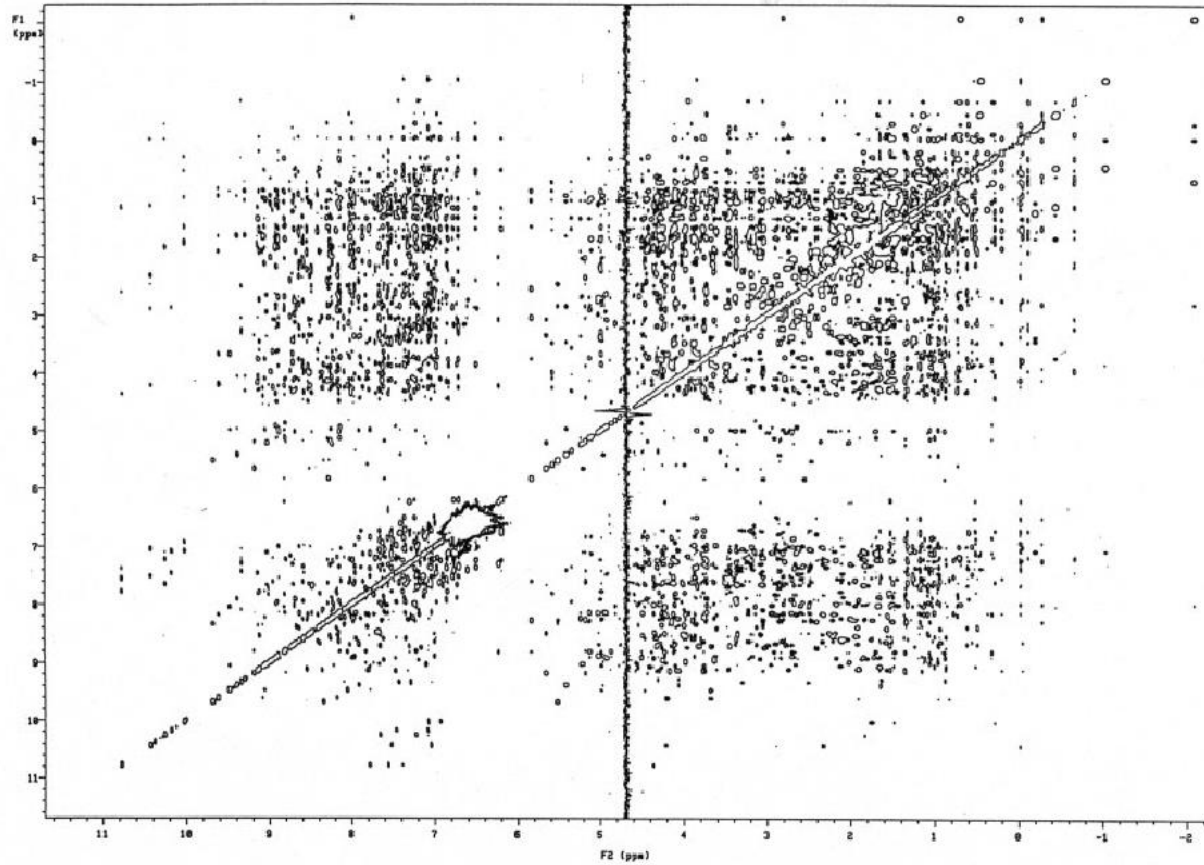
2PEL : fixe les sucres



- ✓ Affranchissement de la séquence et du repliement
- ✓ Comparaison locale entre structure 3D de protéines ayant des repliements et séquences quelconques
- ✓ Comparaisons de structures 3D proches avec des fonctions subtilement différentes
- ✓ Comparaison d'1 protéine de fonction inconnue avec un ensemble de protéines (PDB)
- ✓ Description des ligands avec une approche similaire de description cartographique du site de liaison
- ✓ Balayage des banques avec un ligand
- ✓ Identification de nouvelles fonctions pour des protéines de structures 3D connues

Modélisation sous contraintes RMN





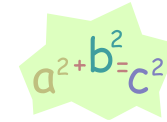
- **Paramètre mesuré nuclear Overhauser effect ou nOe:**

- η_{ij} intensité nOe entre 2 protons i et j

- **Approximation η_{ij} est proportionnelle à la vitesse de relaxation**

$$\sigma_{ij}: \sigma_{ij} = f(\tau_c) 1/r_{ij}^6$$

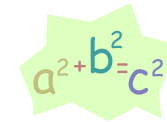
où τ_c = temps de corrélation du vecteur r_{ij}



$$a^2 + b^2 = c^2$$

- **Paramètre structural obtenu: r_{ij} tel que $r_{ij} \leq 5 \text{ \AA}$**

$$r_{ij} = r_{\text{ref}} (\eta_{\text{ref}} / \eta_{ij})^{(1/6)}$$



$$a^2 + b^2 = c^2$$

où η_{ref} = intensité nOe du pic de référence

r_{ref} = distance de référence

Introduction d'intervalles de valeurs possibles [d_{min} - d_{max}]

1. Définition de classes de contraintes

Exemple :

Fort	1.8 -> 2.6 Å
Moyen	2.6 -> 3.8 Å
Faible	3.8 -> 5.0 Å

2. Contraintes de distances introduites avec une marge d'erreur (20-25%) par rapport aux valeurs déterminées.

$$d \pm \Delta d$$

- **Quantitative**

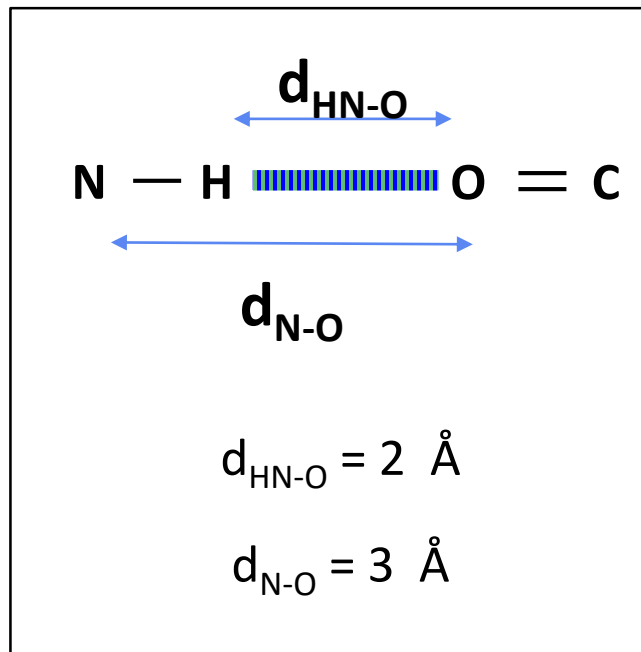
- Nombre insuffisant en théorie $n \times (n-1)/2$ pour n atomes
- Une protéine de 50 aa soit 809 atomes soit 327240 contraintes nécessaires
- Environ 3 contraintes longues portées /résidu au mieux (115)

- **Qualitative**

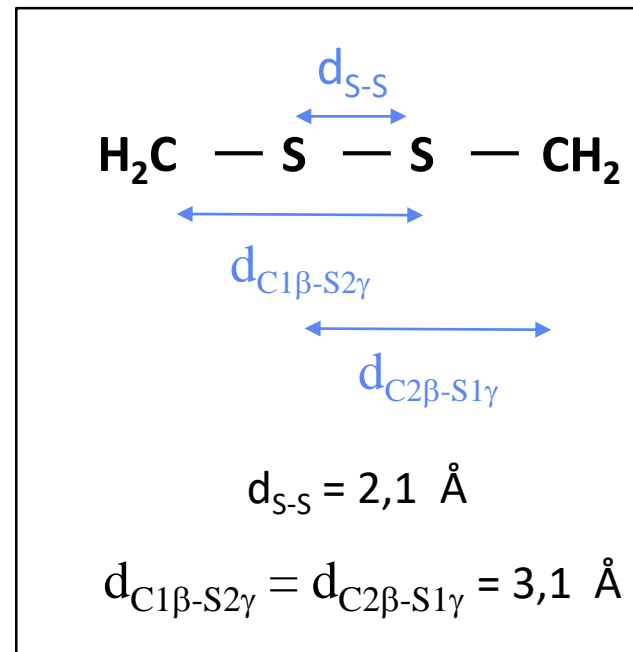
- 1 seul type atomique (hydrogènes)
- Génération de toutes les coordonnées
- Répartition inégale des contraintes le long de la structure
- Imprécision sur les distances

- **Contraintes de distances NOE**
- **Contraintes angulaires déduites des constantes de couplage**
- **Autres contraintes de distances (H, SS)**
 - Vitesse d'échange 1H/2D des protons amides
 - + détermination des éléments de structures secondaire (déplacements chimiques)

Position des liaisons H



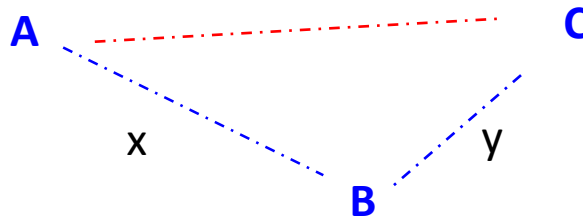
Ponts disulfures



- Inégalité triangulaire

- Augmentation du nombre de contraintes

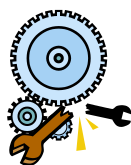
si A est distant de x de B et B distant de y de C alors A et C sont distants de moins de $x + y$



- Tirage au sort de distances dans l'intervalle de distances

- Génération d'un ensemble de structures dont certaines ne seront pas de bonne qualité

```
assign (resid 3 and name HA ) (resid 4 and name HN ) 2.5 .7 .2
assign (resid 3 and name HB ) (resid 4 and name HN ) 3.2 1.4 .3
assign (resid 3 and name HG* ) (resid 4 and name HN ) 3.6 1.8 2.8
assign (resid 4 and name HN ) (resid 4 and name HA ) 2.7 .9 .3
assign (resid 4 and name HN ) (resid 4 and name HB# ) 2.8 1.0 1.3
assign (resid 4 and name HN ) (resid 4 and name HG ) 3.4 1.6 .3
assign (resid 4 and name HN ) (resid 4 and name HD1# ) 3.8 2.0 1.4
assign (resid 4 and name HN ) (resid 4 and name HD2# ) 4.0 2.2 1.4
assign (resid 4 and name HA ) (resid 5 and name HN ) 3.4 1.6 .3
assign (resid 2 and name HN ) (resid 3 and name HN ) 3.2 1.4 .3
assign (resid 59 and name HA ) (resid 62 and name HB# ) 3.6 1.8 1.4
assign (resid 55 and name HD* ) (resid 19 and name HE# ) 3.4 1.6 3.3
assign (resid 55 and name HE* ) (resid 19 and name HE# ) 3.0 1.2 3.3
assign (resid 55 and name HE* ) (resid 16 and name HG2# ) 3.6 1.8 3.4
assign (resid 55 and name HN ) (resid 52 and name HA ) 3.8 2.0 .4
```

Spectres RMN multidimensionnels

Intensités des NOE
Constantes de couplage

Conversion des intensités NOE en
distances, et des constantes en angles

Contraintes de distances
Contraintes angles dièdres

Géométrie des distances

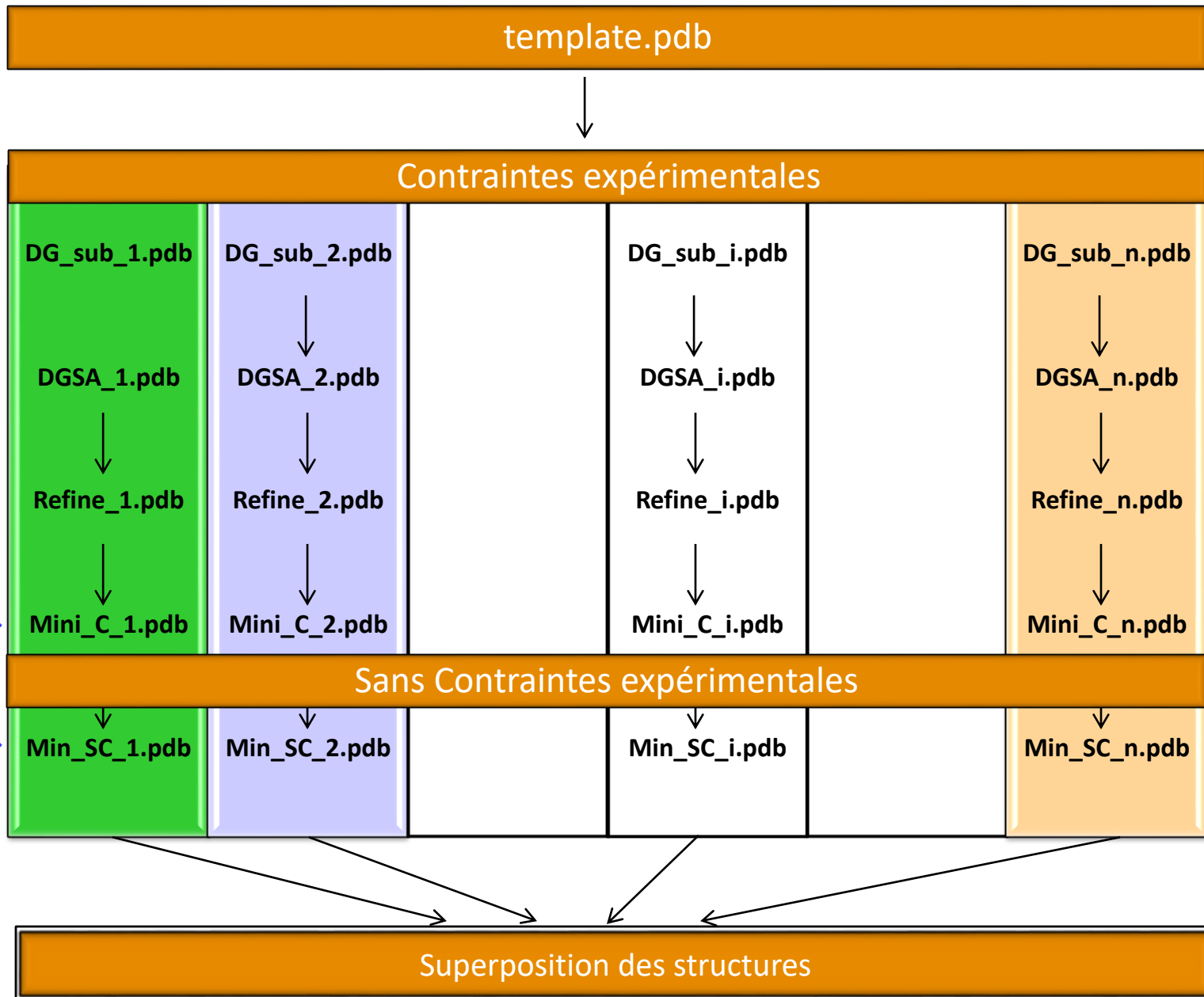
Génération d'un grand nombre
de structures compatibles avec les données
RMN

Dynamique moléculaire (recuit simulé)

Optimisation des structures issues
de la géométrie des distances

Faisceaux de structures

Etape d'affinement des structures



Séquence protéine et champ de force

Génération de fichiers de paramètres structuraux et de topologie décrivant la molécule dans le champ de force utilisé

INFORMATIONS MOLECULAIRES (psf)

Fichier : generate_psf.inp

Génération d'une molécule chimiquement correcte
STRUCTURE INITIALE avec la nomenclature atomique du champ de force

Fichiers : generate_tmp.inp et template.pdb

Contraintes expérimentales (NOE, dièdres)

Géométrie des distances

Génération de x structures avec des coordonnées partielles (« embedding »)

Fichiers : dg_sub_embed.inp et DG_SUB_x.PDB

Régularisation de structures par dynamique moléculaire (« recuit simulé »)

Génération de x structures avec toutes les coordonnées

Fichiers : dgsa.inp et DGSA_x.PDB

Affinement des structures par dynamique moléculaire

Fichiers : refine.inp et REFINE_x.PDB

Minimisation de l'énergie des structures sous contraintes RMN avec parmallh3x.pro

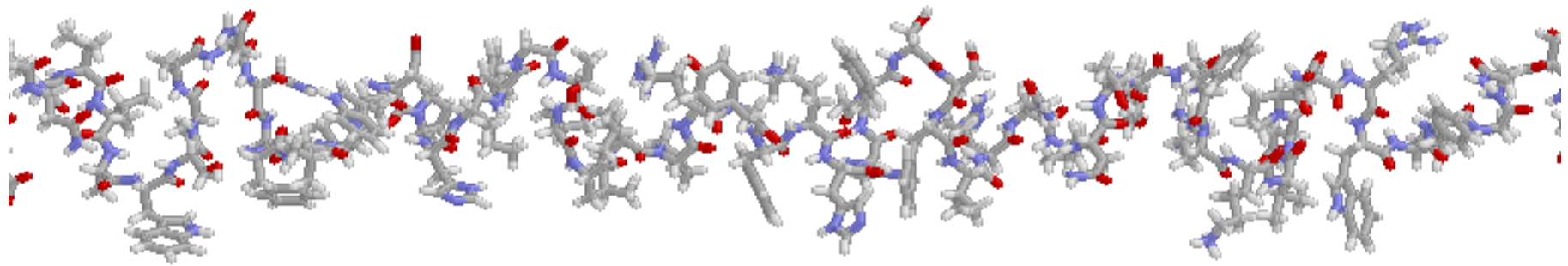
Fichiers : minimize_ctr.inp et MINIM_x.PDB

Minimisation finale sans contrainte avec parmallh3x.pro

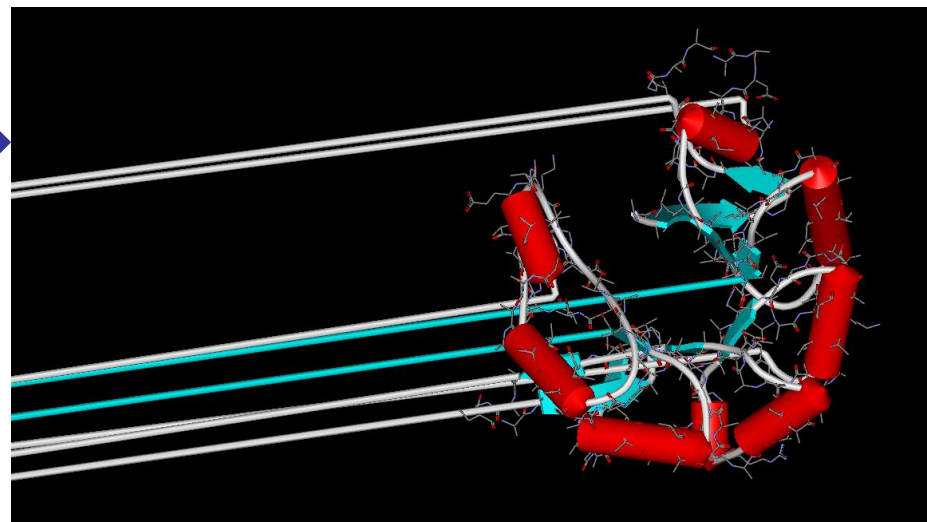
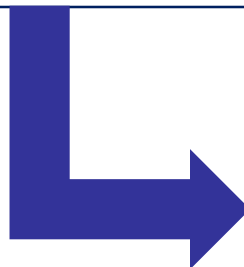
Fichiers : minimize.inp et MINIM_x.PDB

Analyse des structures finales (sélection, superposition, classement en famille)

Molécule de départ sans autre contrainte que la chimie



ATOM	1669	CA	ARG	25	63.030	16.219	43.069	1.00	34.41	C
ATOM	1670	C	ARG	25	62.431	17.565	42.648	1.00	34.73	C
ATOM	1671	O	ARG	25	62.626	18.556	43.339	1.00	34.03	O
ATOM	1672	CB	ARG	25	61.969	15.373	43.832	1.00	35.43	C
ATOM	1675	N	PRO	26	61.717	17.617	41.526	1.00	35.19	N
ATOM	1676	CA	PRO	26	61.136	18.885	41.075	1.00	35.37	C
ATOM	1677	C	PRO	26	59.920	19.291	41.904	1.00	34.90	C
ATOM	1678	O	PRO	26	58.895	18.618	41.854	1.00	35.25	O
ATOM	1679	CB	PRO	26	60.706	18.806	39.596	1.00	36.26	C
ATOM	1675	N	LYS	27	9999.9999999	9999999.999	9999999.999	1.00	35.19	N
ATOM	1676	CA	LYS	27	9999.9999999	9999999.999	9999999.999	1.00	35.37	C
ATOM	1677	C	LYS	27	9999.9999999	9999999.999	9999999.999	1.00	34.90	C
ATOM	1678	O	LYS	27	9999.9999999	9999999.999	9999999.999	1.00	35.25	O
ATOM	1675	N	TYR	28	9999.9999999	9999999.999	9999999.999	1.00	35.19	N
ATOM	1676	CA	TYR	28	9999.9999999	9999999.999	9999999.999	1.00	35.37	C
ATOM	1677	C	TYR	28	9999.9999999	9999999.999	9999999.999	1.00	34.90	C
ATOM	1678	O	TYR	28	9999.9999999	9999999.999	9999999.999	1.00	35.25	O
ATOM	1675	N	ILE	29	9999.9999999	9999999.999	9999999.999	1.00	35.19	N
ATOM	1676	CA	ILE	29	9999.9999999	9999999.999	9999999.999	1.00	35.37	C



- | **809 atomes pour 50 résidus**

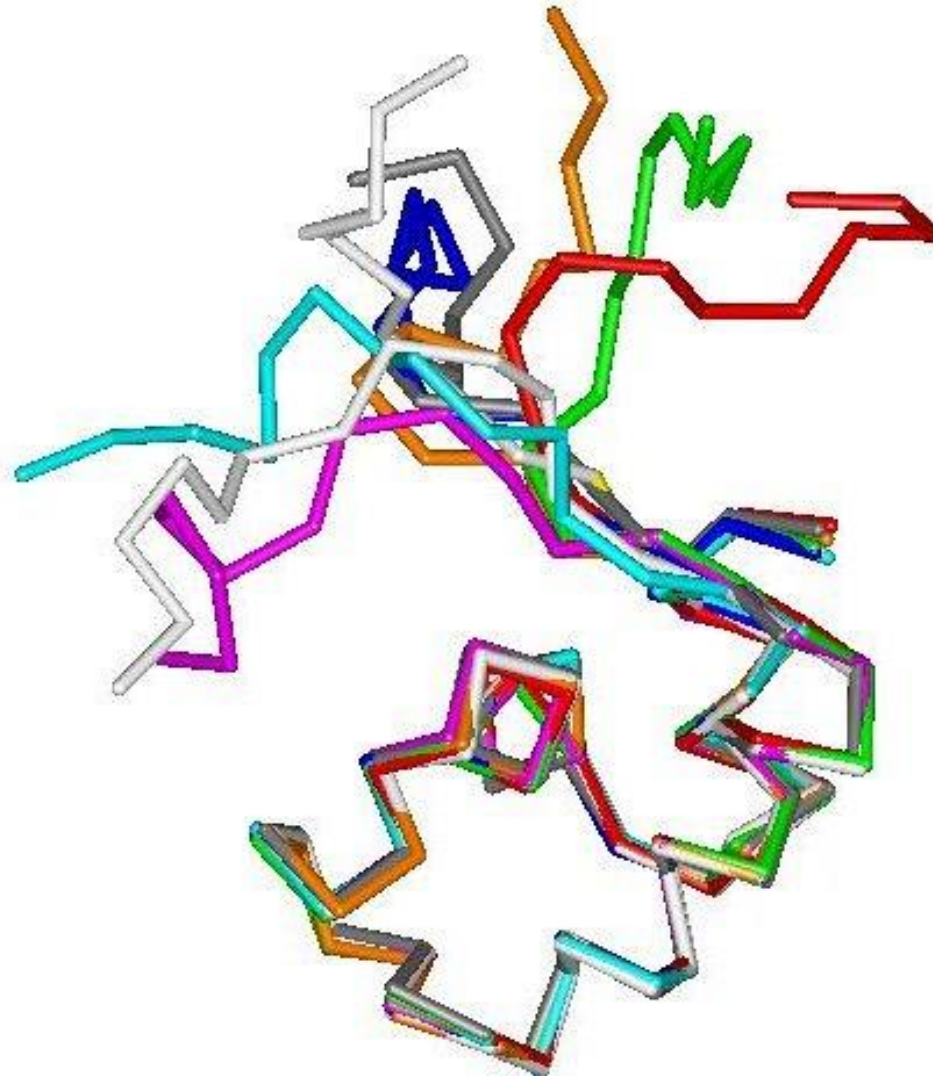
- | **Contraintes utilisées sur FruR**
 - | **Distance restraints**

Intra-residue	232
Sequential ($ i-j = 1$)	242
Medium rang ($ i-j \leq 4$)	187
Long range	148
Total distance restraints	809

 - | **Dihedral angle constraints**

ϕ	11
--------	----

- Géométrie Ramachandran, angles χ
- Compacité de la structure - Critères visuels
- RMSD entre les structures superposées
- Minimisation d'énergie
- Peu de violations
- Pas de mauvais contacts



Superposition des résidus 1- 47 et sur les C α

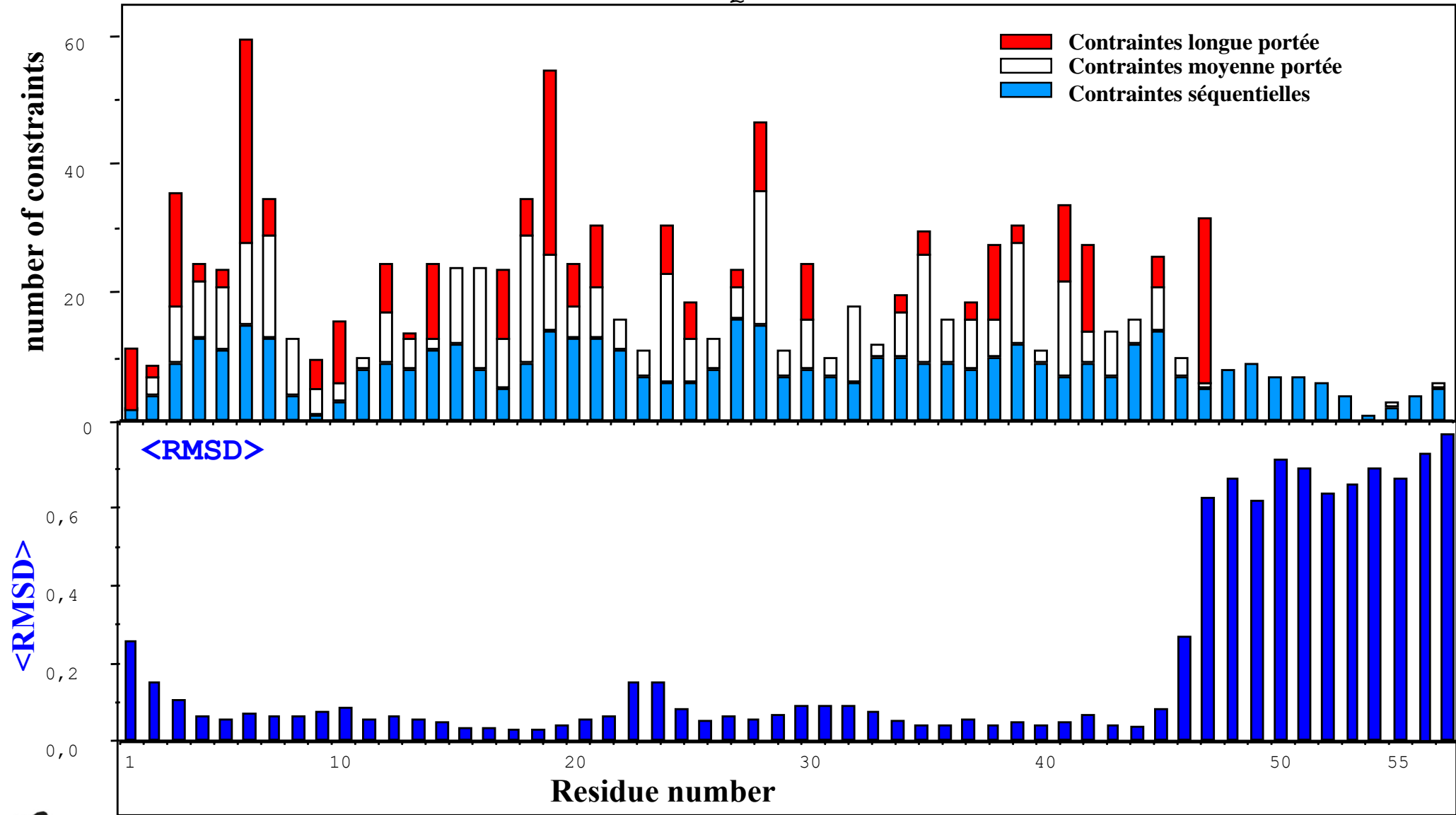


Superposition des résidus 1- 47 et sur les C α 0,50 Å

	X-PLOR energy (kcal mol⁻¹)		-269±12
	NOE violations		
	Number > 0.5 Å		none
	Rms deviation		0.046±0.002
	Dihedral violations		
	Number > 5°		none
	Rms deviation		0.96±0.48
	Deviation from idealized covalent geometry		
	Angles (°)		1.2±0.03
	Impropers (°)		0.2±0.01
	Bonds (Å)		0.005±0.0001
	Rms deviation en Å		
	Backbone (C', C ^α , N)	residues 1-47	0.50
		helix	0.33
	All heavy atoms	residues 1-47	1.31
		helix	1.22
	Ramachandran data		
	Residues in most-favored regions		81.6%±3.2
	Residus in allowed regions		16.4%±2.6
	Residus in generously regions		0.9%±1.5
	Residues in disallowed regions		1.1%±1.3



1 10 20 30 40 50 55
MKLDE IAR LAGV SRTTA SYV INGKAKQYRV SDKTVEKVM AVVREHNYHPNAVAAG LR



| **1713 atomes pour 104 résidus**

| **Contraintes utilisées sur HSF**

| **Distance restraints**

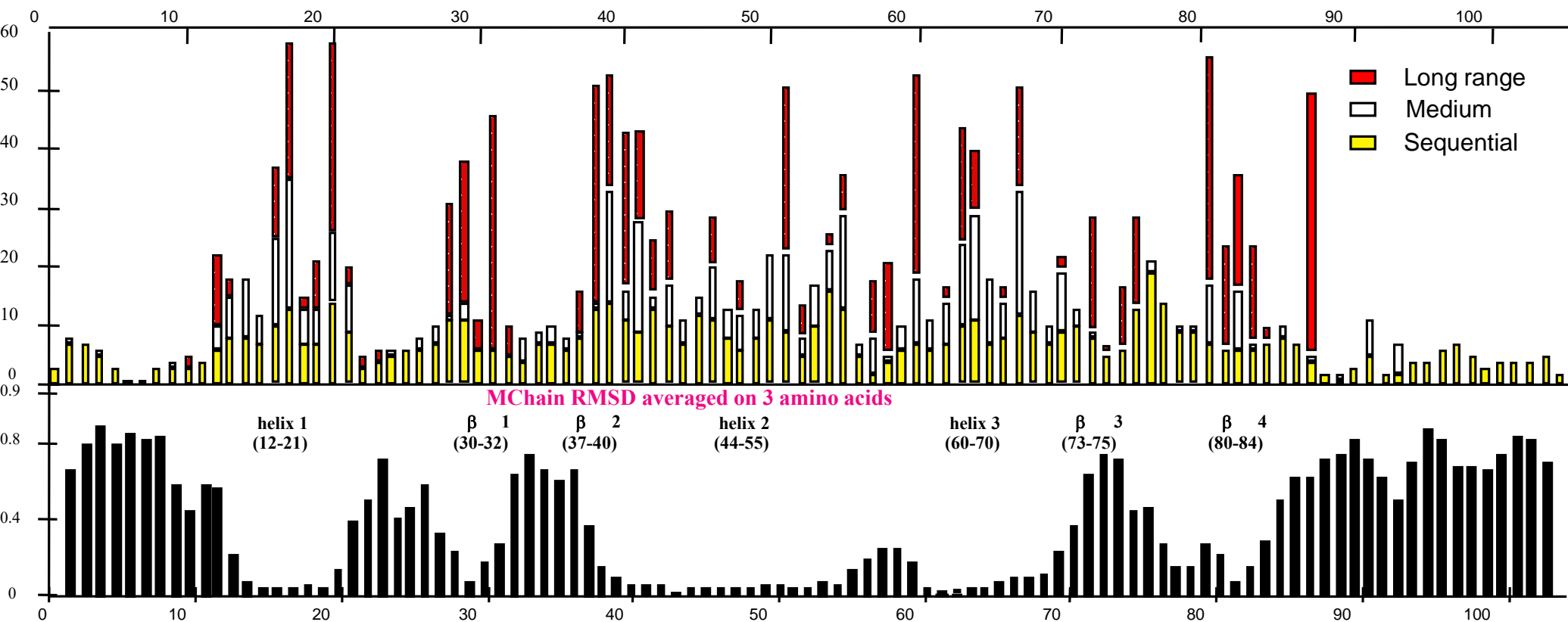
Intra-residue	362
Sequential ($ i-j = 1$)	367
Medium rang ($ i-j \leq 4$)	216
Long range	332
Total distance restraints	1277

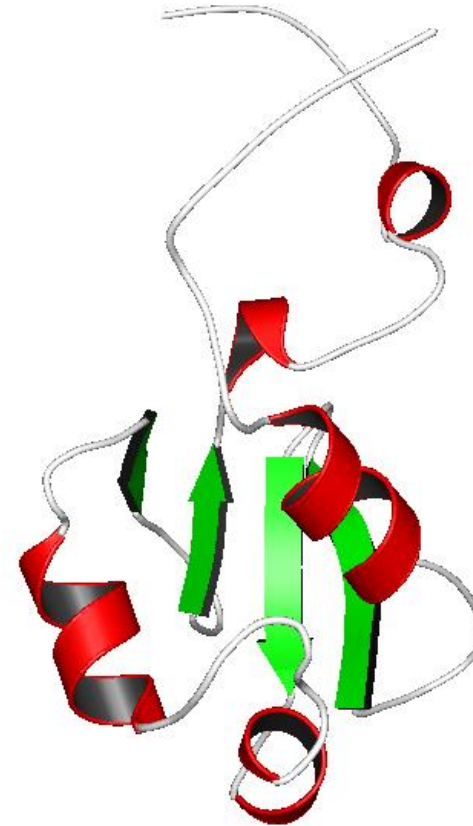
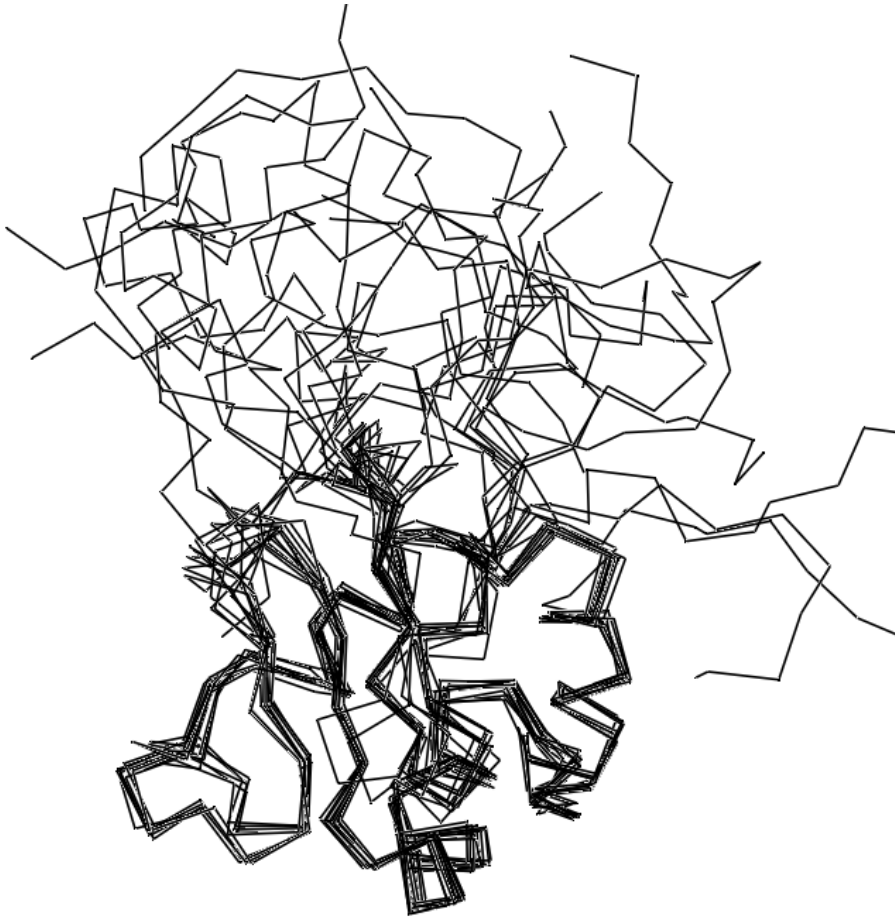
| **Dihedral angle constraints**

ϕ	24
--------	----

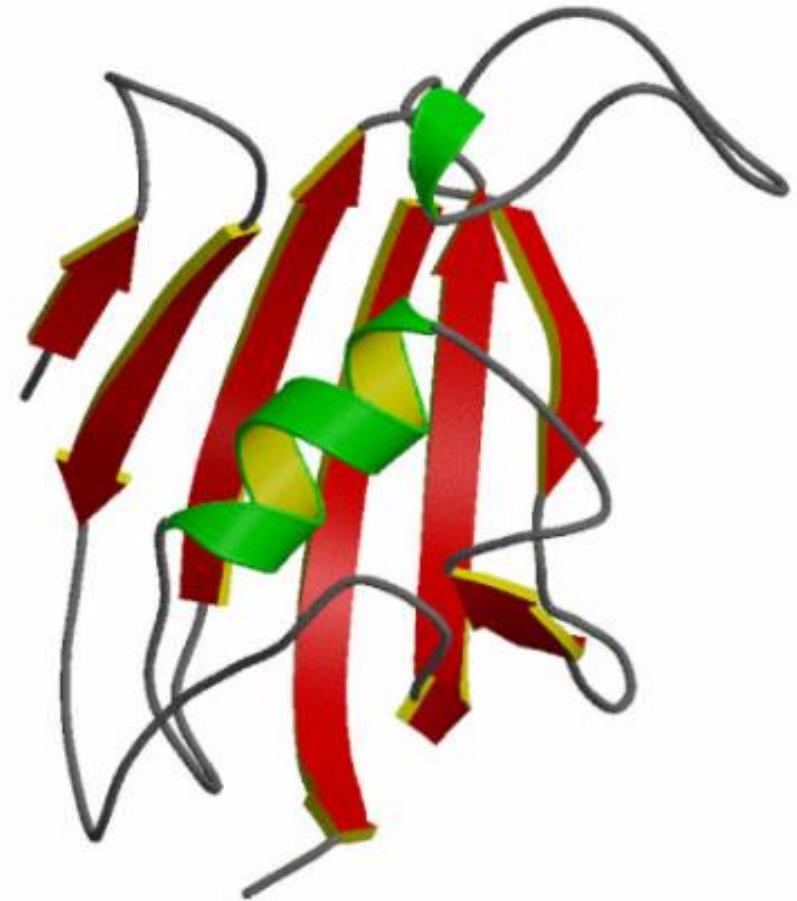
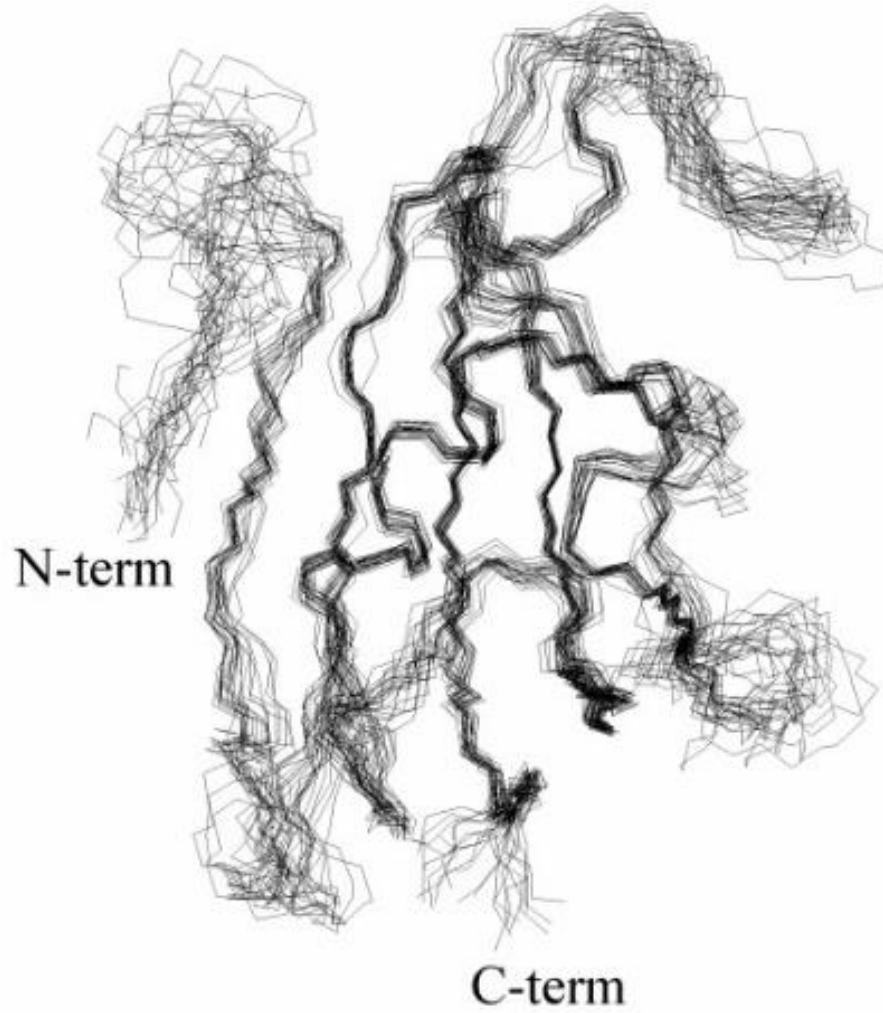
	X-PLOR energy (kcal mol⁻¹)	-406±46
	NOE violations	
	Number > 0.5 Å	0.2±0.4
	Rms deviation	0.068±0.0037
	Dihedral violations	
	Number > 5°	4.2 ± 3.7
	Rms deviation	0.96±0.48
	Deviation from idealized covalent geometry	
	Angles (°)	0.71±0.02
	Improper (°)	0.49±0.03
	Bonds (Å)	0.0045±0.0002
	Rms deviation en Å	
	Backbone (C', C ^α , N) residues 12-84	1.43
	Structure secondaires	0.73
	All heavy atoms residues 12-84	2.48
	Structure secondaires	2.07
	Ramachandran data	
	Residues in allowed regions	97,8±1,7
	Residus in disallowed regions	2,2±1,7

MDVLHDGSSP PFLTKTYDMVDDPTTNAVVSWSAANNFVWDPHI FGTVLL PRYFKHNNFS SFVRQLNTYGF RKVDPDRWE FANEE FLRGQRHL LKNIRRRKLN





Superposition des résidus en structure secondaire sur les C α



- 1200 repliements connus sur un total de 4000-6000



EGF 53 aa



Kringle domain 85aa (3 ponts SS)



Ca⁺⁺ binding domain

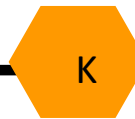
Modules domaines de base



Chymotrypsin 245 aa en 2 domaines

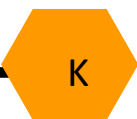
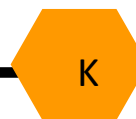
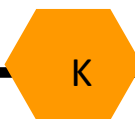
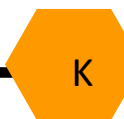
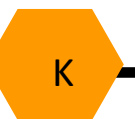


Urokinase



Facteur de coagulation IX

Réutilisation - légo



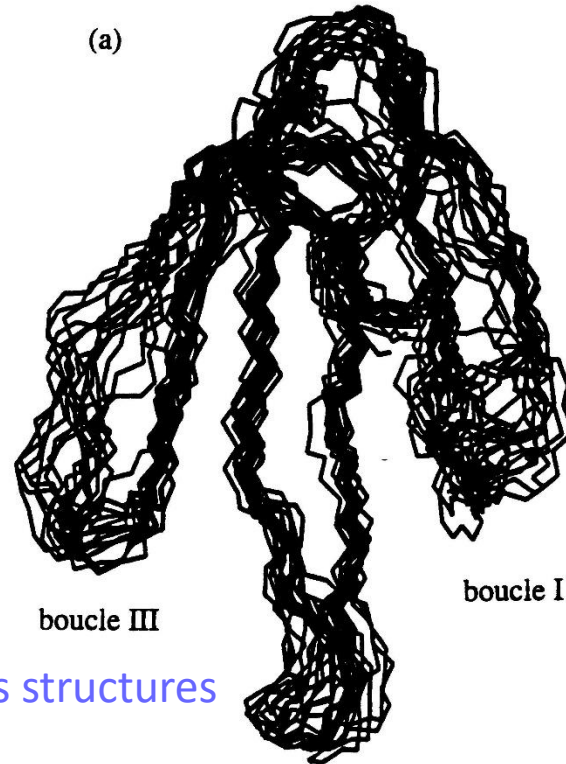
Plasminogen

✓ Construction à partir des données RMN

- ✓ DISGEO (Havel, 1984)
- ✓ DSPACE (Nerdal, 1986)
- ✓ EMBED, VEMBED (Kuntz, 1989)
- ✓ DGEOM (Blaney, 1990)
- ✓ DGII (Havel, 1991)
- ✓ XPLOR/DG (Brünger, 1992)

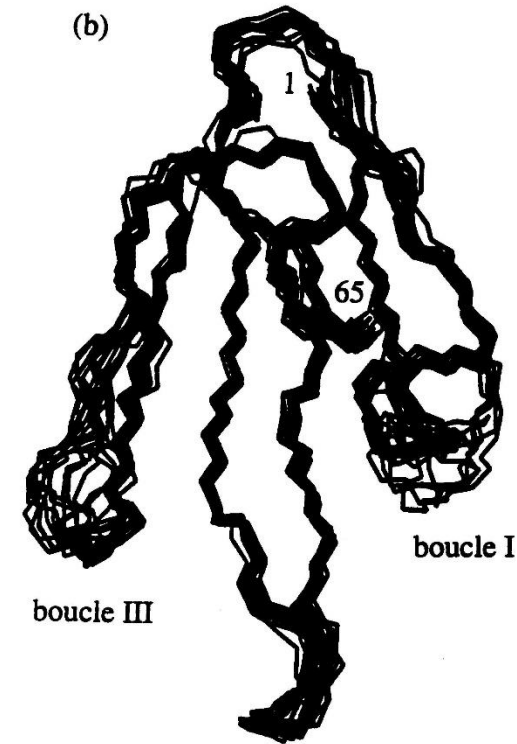
✓ Simulation de spectres NOESY à partir des structures

- ✓ CORMA (Keepers, 1984)
- ✓ IRMA (Boelens, 1988)
- ✓ MARDIGRAS (Borgias, 1990)
- ✓ X-PLOR (Brünger, 1990)



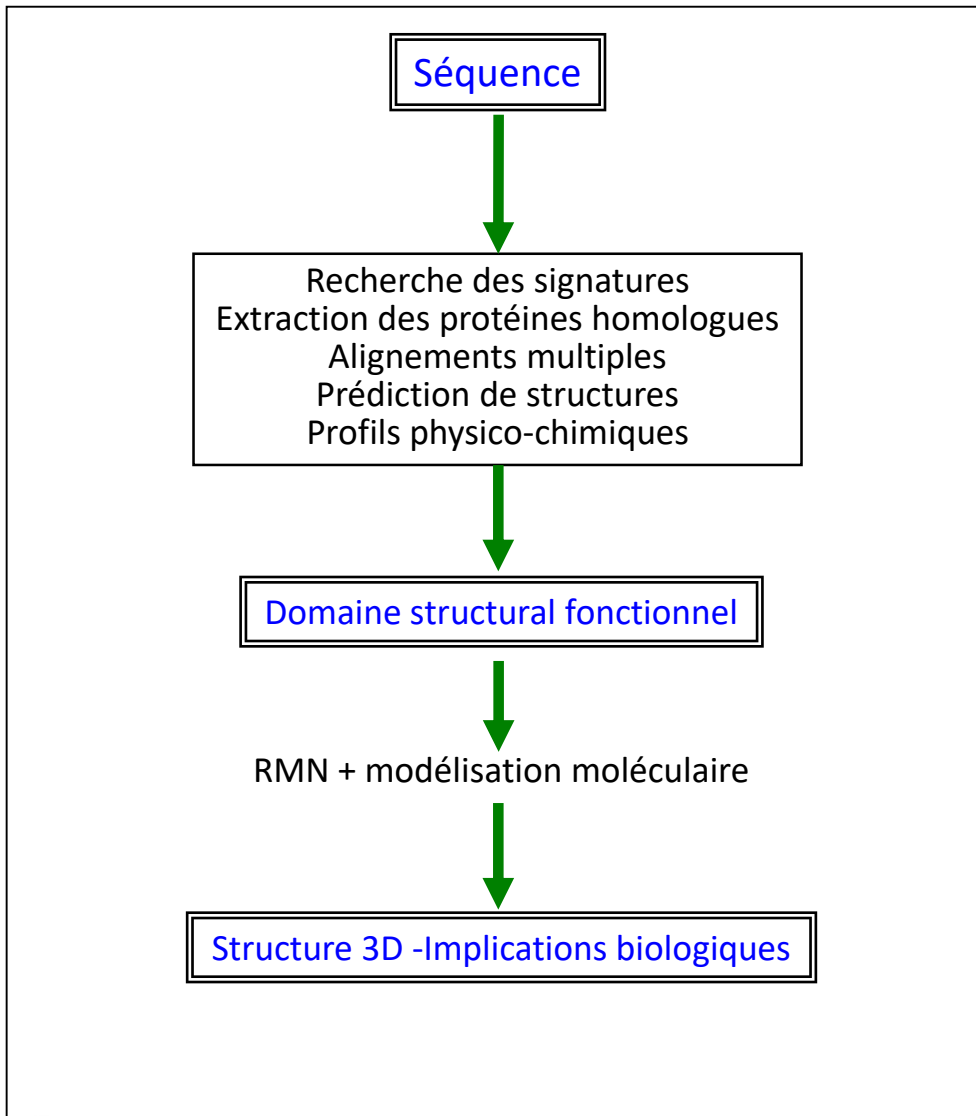
boucle II

Diana

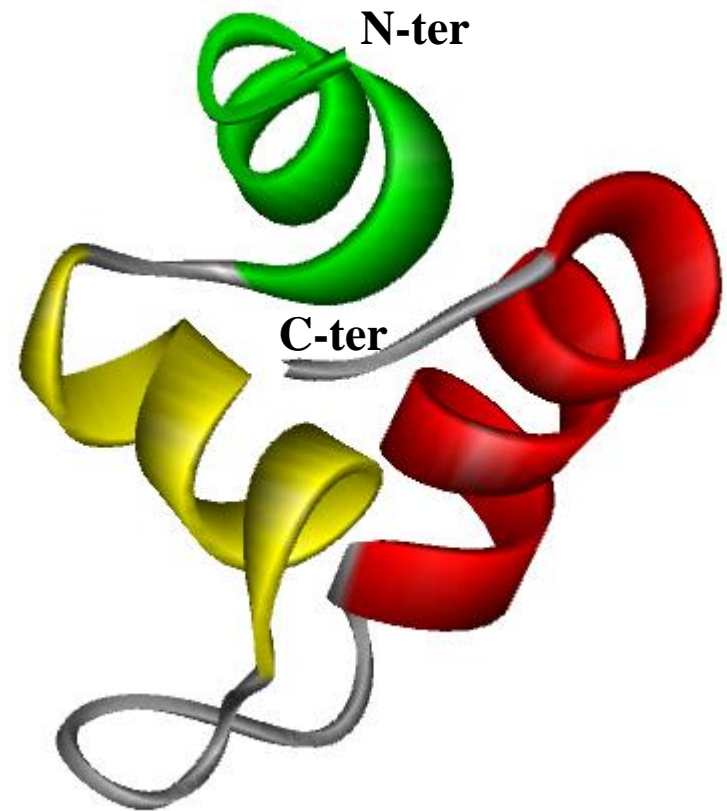


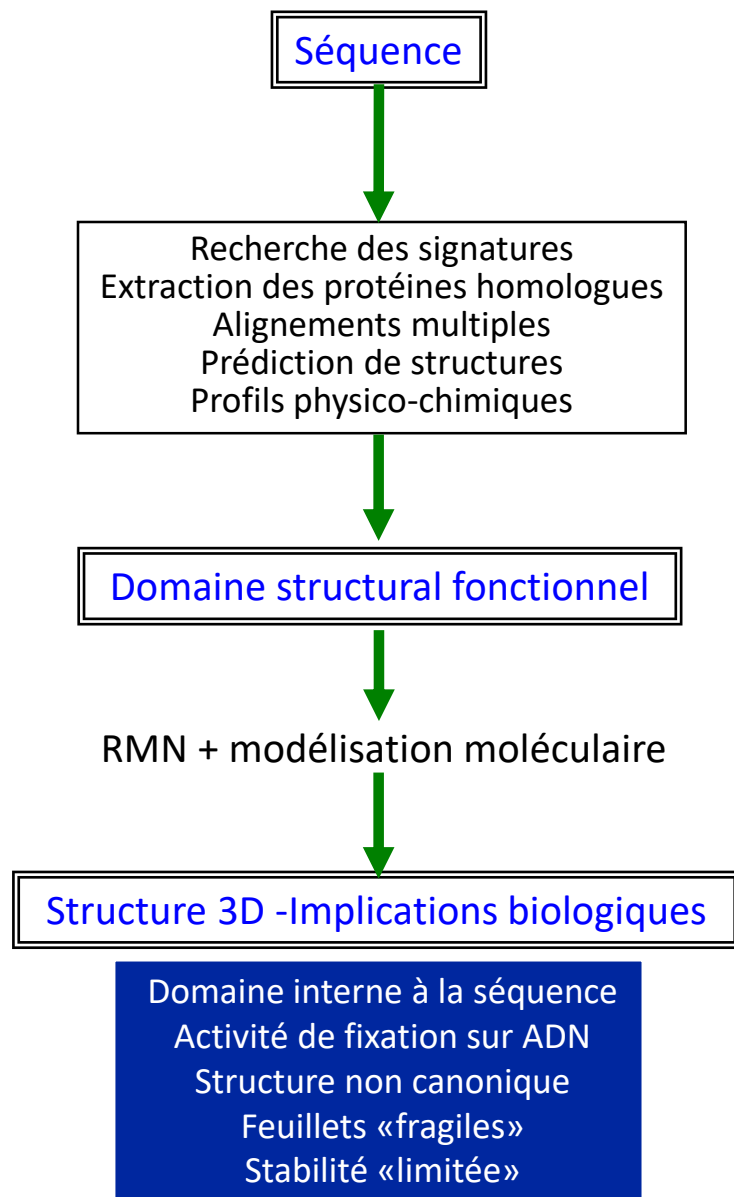
boucle II

X-PLOR

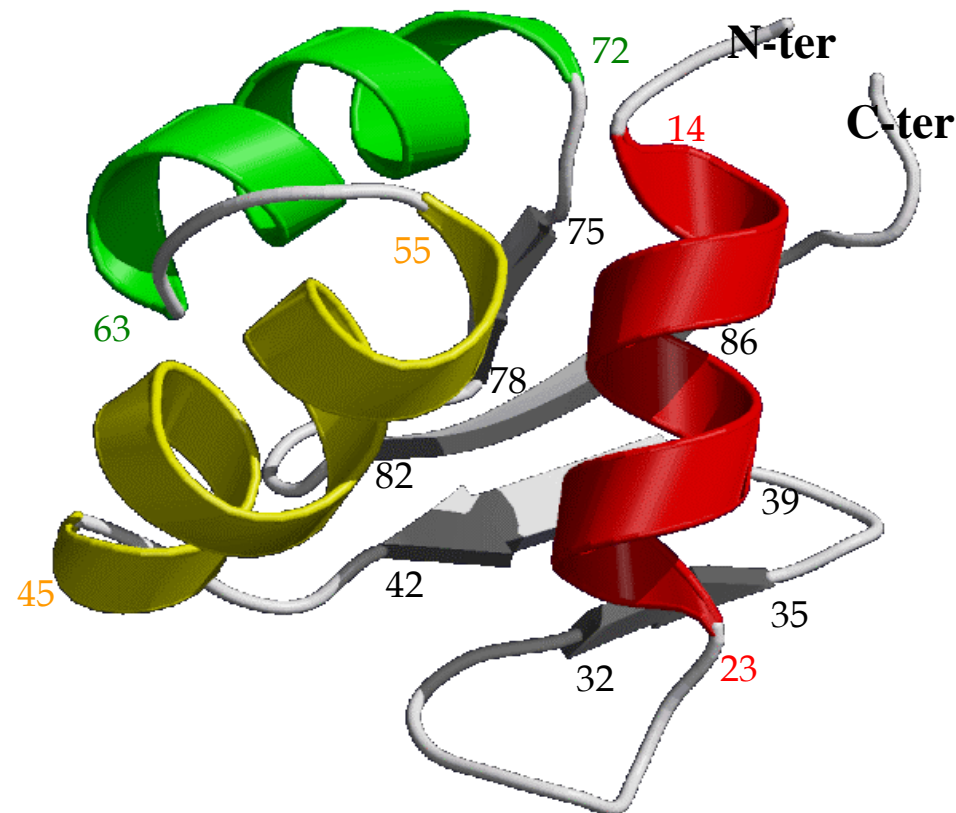


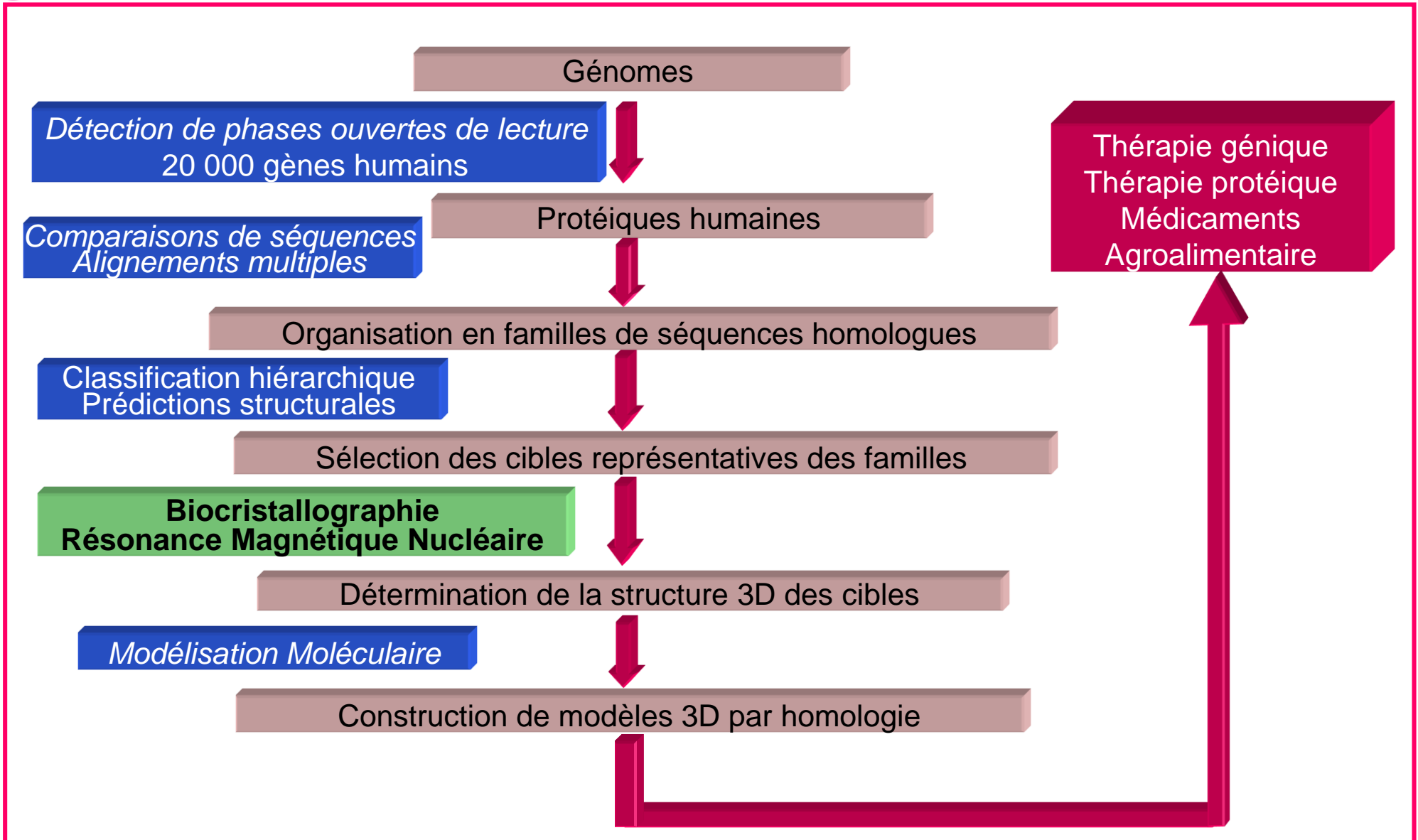
Structure RMN du domaine d'interaction à l'ADN de FruR (1-57)
Collaboration avec le Pr. AJ. Cozzone



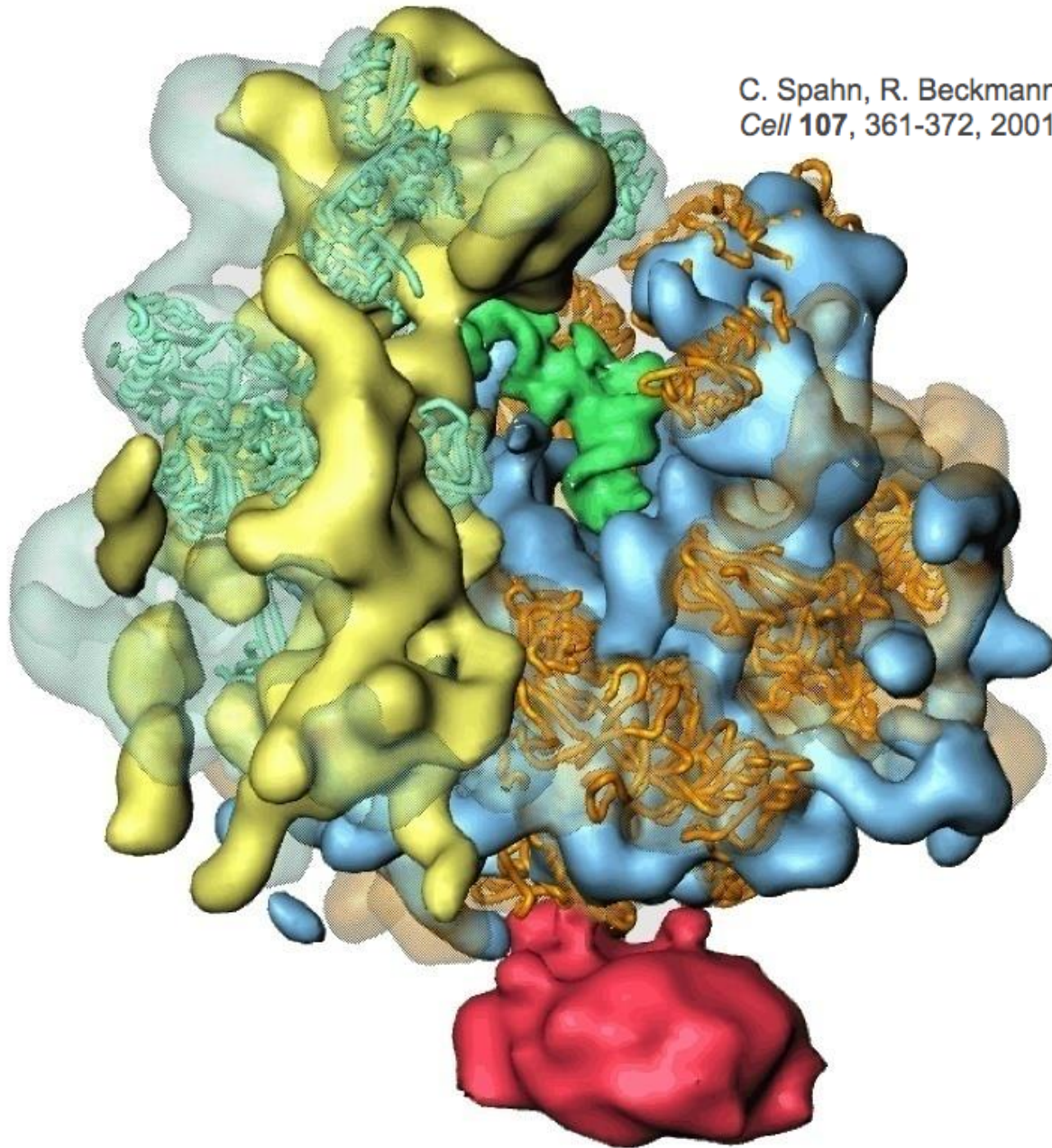


Structure RMN du domaine d'interaction à l'ADN d'HSF de maïs
Collaboration avec le Pr. C.DUMAS

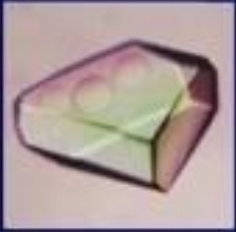
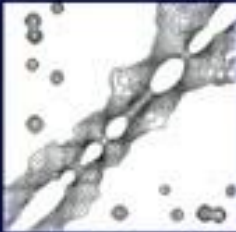
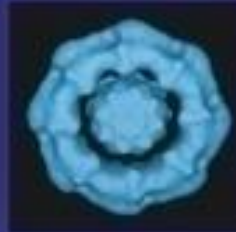

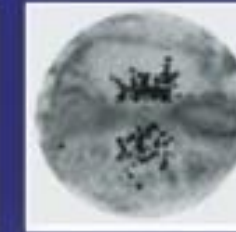
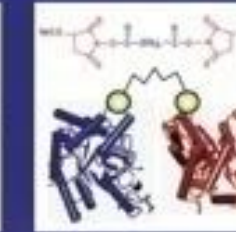



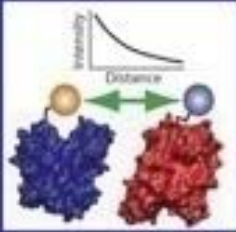
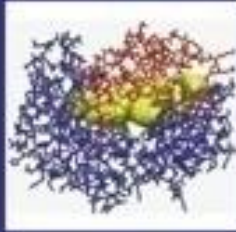
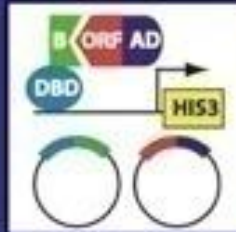



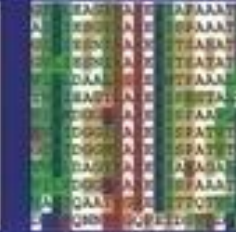


C. Spahn, R. Beckmann, N. Eswar, P. Penczek, A. Sali, G. Blobel, J. Frank.
Cell **107**, 361-372, 2001.



(d'après A. Sali)

						
X-ray crystallography	NMR spectroscopy	2D & single particle electron microscopy	electron tomography	immuno-electron microscopy	chemical cross-linking	affinity purification mass spectroscopy
subunit structure	subunit structure				subunit structure	
subunit shape	subunit shape	subunit shape	subunit shape			
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact		subunit-subunit contact	subunit-subunit contact
subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity	subunit proximity
subunit stoichiometry	subunit stoichiometry					
assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry	assembly symmetry		
assembly shape	assembly shape	assembly shape	assembly shape			
assembly structure	assembly structure					

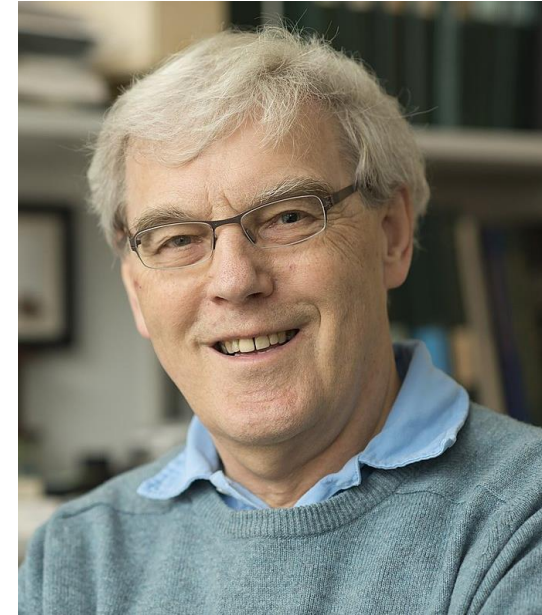
						
FRET	site-directed mutagenesis	yeast two-hybrid system	gene/protein arrays	protein structure prediction	computational docking	bioinformatics
subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit-subunit contact	subunit structure	subunit-subunit contact	subunit-subunit contact
subunit proximity		subunit proximity	subunit proximity	subunit shape		



Jacques Dubochet



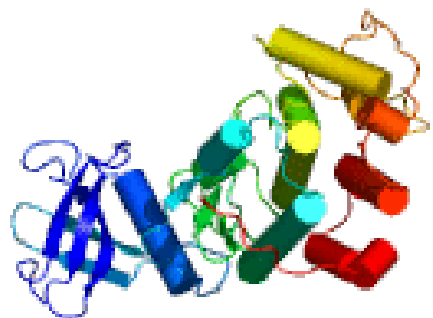
Joachim Frank



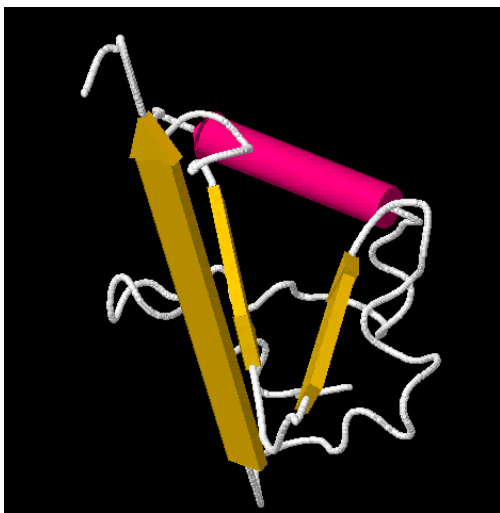
Richard Henderson

En 1990, Henderson, a le premier produit une image par microscopie de protéine en 3D avec une résolution atomique. Joachim Frank a perfectionné cette technique et l'a rendue plus facile à utiliser. Jacques Dubochet est parvenu à vitrifier l'eau, ce qui permet aux biomolécules de conserver leur forme naturelle.

Fit induit



CDK2/Cyclin A
[[1jst](#) --> [3qhr](#)]



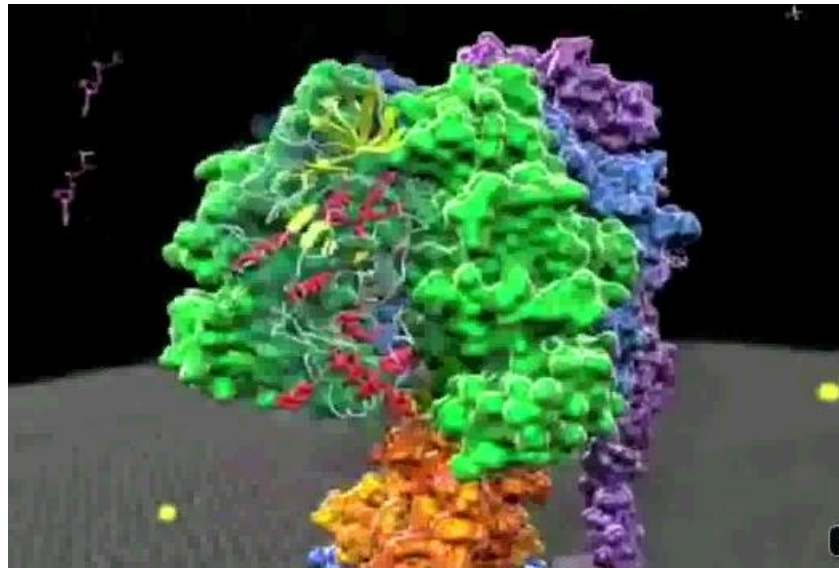
Déplacement de domaine



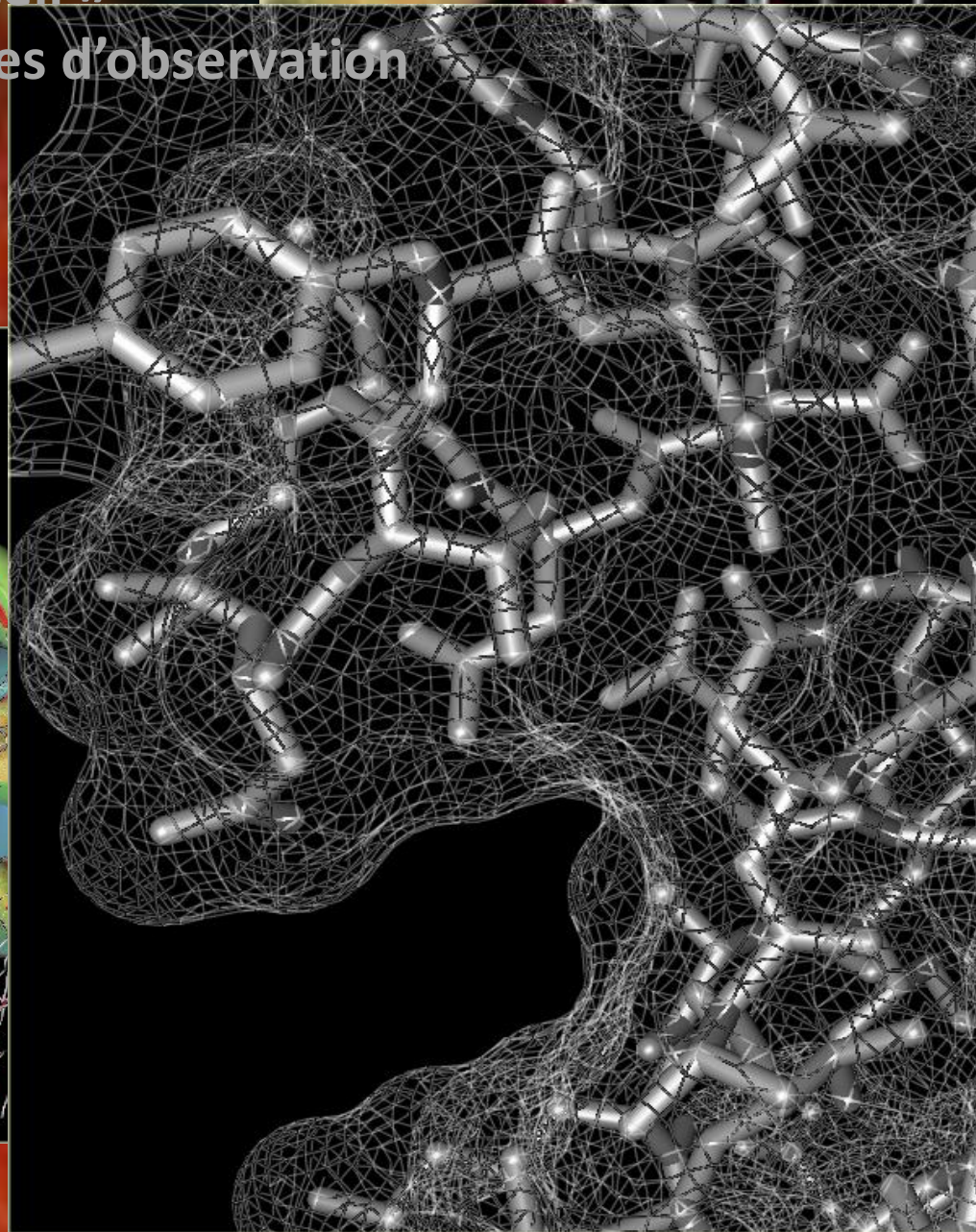
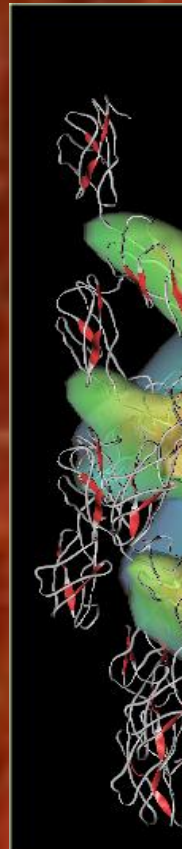
Mouvements globaux (respiration)



Canal sodique

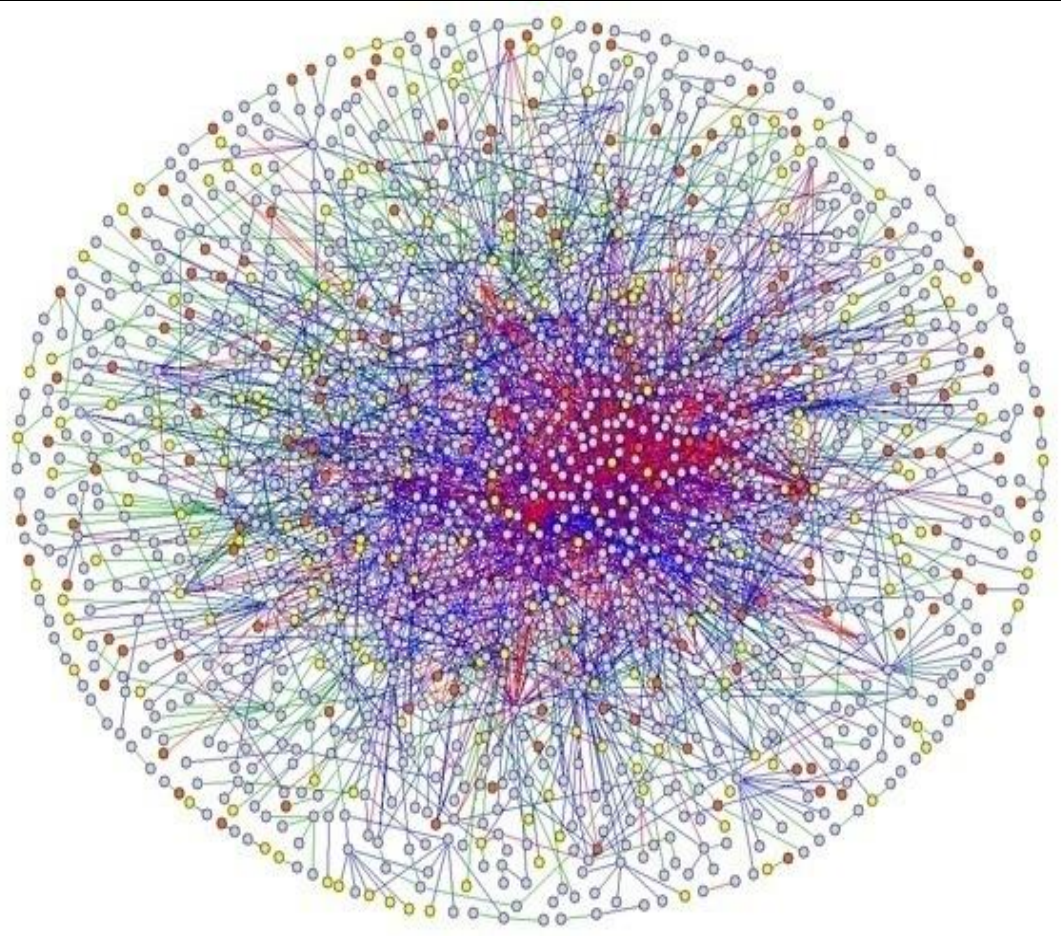


Après Google maps « Google cell »
Vers une continuité des échelles d'observation



Vous avez dit interactions?

3 200 interactions entre 1 700 protéines humaines.



- Ne croyez pas toujours les programmes (ils donneront toujours un résultat)
- Ne croyez pas toujours les bases de données (elles sont truffées d'erreur)
- Ayez un sens critique vis-à-vis des auteurs (certains sont optimistes, d'autres incompetents et très peu sont malhonnêtes).
- Différence entre signification mathématique et signification biologique
- Combiner les approches (utiliser la notion de faisceau de présomptions)
- Connaissez votre sujet (lire la bibliographie, les longs dimanches pluvieux...)
- Le meilleur modélisateur est le biologiste qui connaît bien sa protéine

Les 2 derniers sujets : 1 Corrigé
2 Corrigé