

Structure et repliement des protéines

Cours 2

Dirk Stratmann (dirk.stratmann@upmc.fr)
<http://www.imPMC.upmc.fr/~stratmann/>

IMPMC, Sorbonne Université

octobre 2019

Plan

- 1 Rotamères des chaînes latérales
- 2 Stabilité de l'état natif, replié
- 3 Domaines
- 4 Structures tertiaires
- 5 Structures quartanaires
- 6 Outils
- 7 Docking protéine-protéine
- 8 Ma thématique de recherche actuelle
- 9 Bibliography

- 1 Rotamères des chaînes latérales
 - Rotamères des chaînes latérales

Rotamères des chaînes latérales

Conformations chaînes latérales

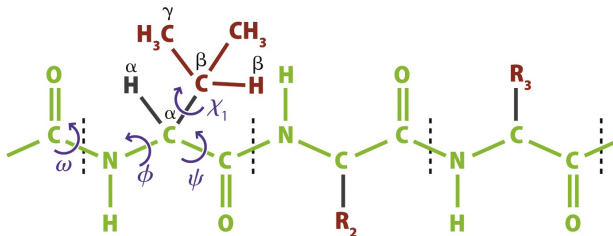


Figure 1.4 How Proteins Work (©2012 Garland Science)

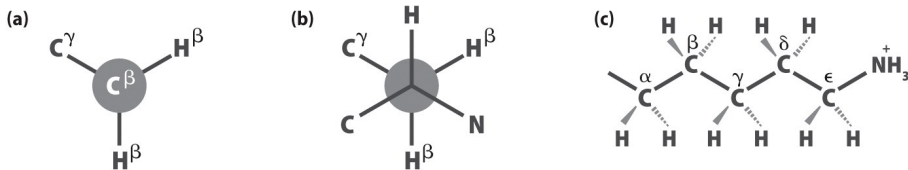
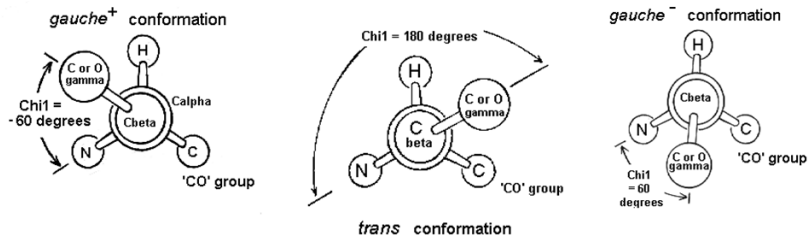


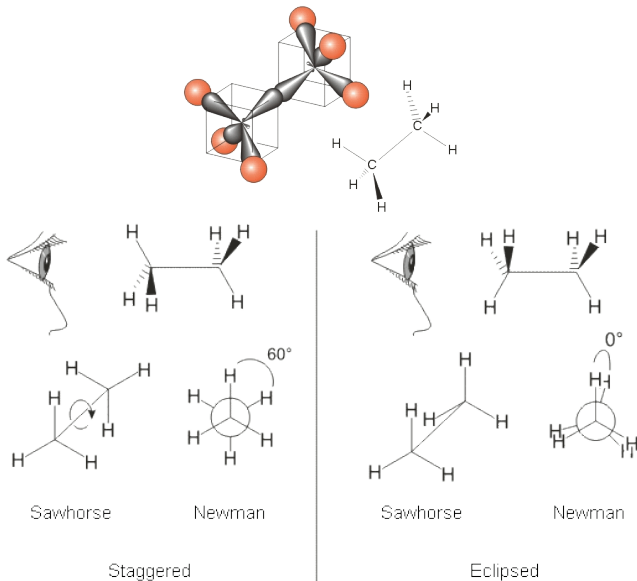
Figure 1.9 How Proteins Work (©2012 Garland Science)

Conformations chaînes latérales

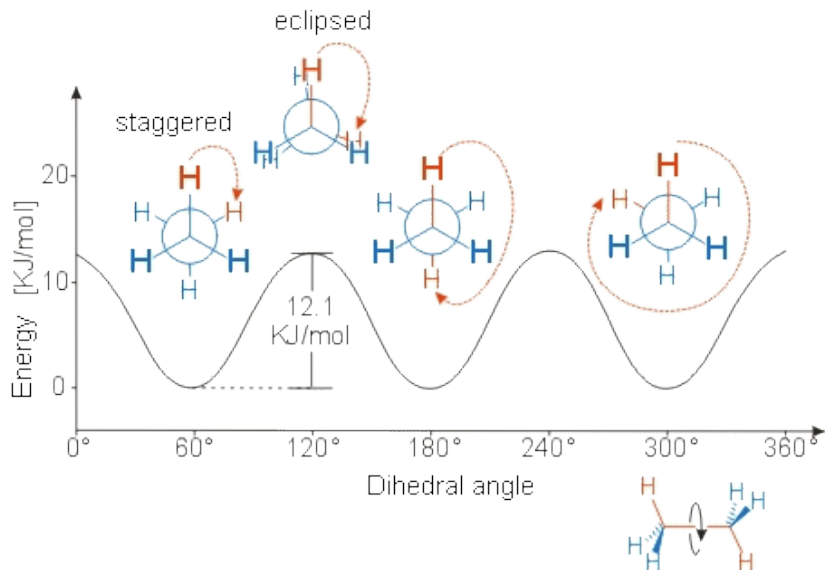


- Angles dièdres fréquence : *gauche*(+), *trans* puis *gauche*(-), instable
- Nomenclature $\chi_1, \chi_2 \dots$
- Banque de rotamères; souvent mêmes angles pour χ_1, χ_2 , formes décalées ($-60^\circ, 60^\circ, 180^\circ$); Large gamme de χ_2 pour Asn et Asp, qui forment liaisons hydrogènes

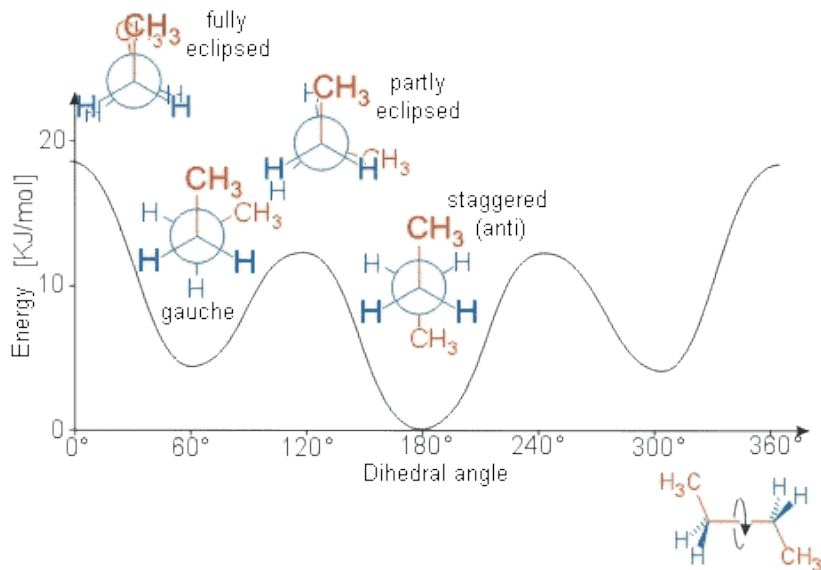
Rotamères de l'éthane



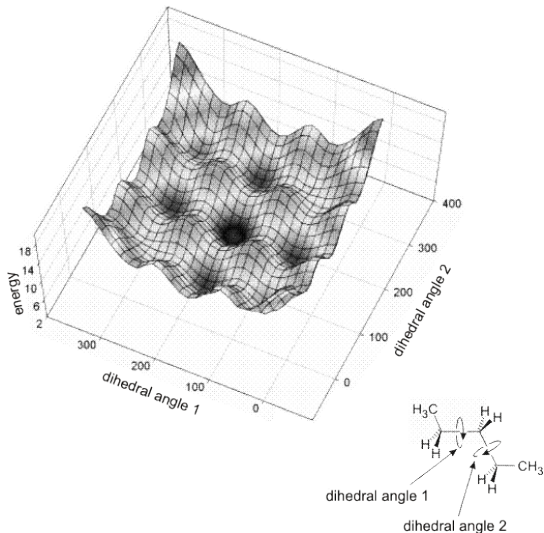
Rotamères de l'éthane



Rotamères du butane



Rotamères de pentane



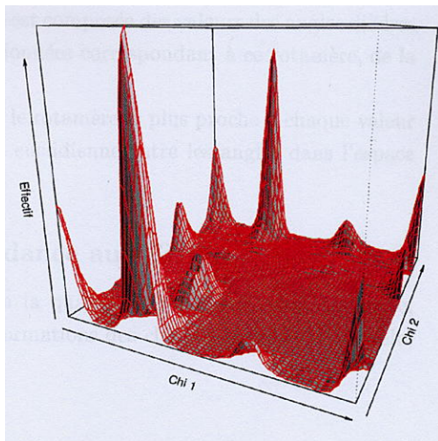
3D Models and properties of the energy minima of pentane

Tab.1 | Conformations of energy minima

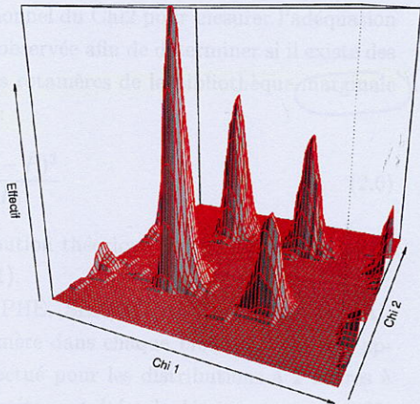
 Jmol	 Jmol	 Jmol
g^+g^+ $\theta_1 = 300^\circ$ $\theta_2 = 300^\circ$ C_2 symmetry	ag^+ $\theta_1 = 180^\circ$ $\theta_2 = 300^\circ$	g^+g^- $\theta_1 = 60^\circ$ $\theta_2 = 300^\circ$ C_S symmetry
 Jmol	 Jmol	 Jmol
g^+a $\theta_1 = 300^\circ$ $\theta_2 = 180^\circ$	aa $\theta_1 = 180^\circ$ $\theta_2 = 180^\circ$ C_{2v} symmetry	g^+a $\theta_1 = 60^\circ$ $\theta_2 = 180^\circ$
 Jmol	 Jmol	 Jmol
g^-g^- $\theta_1 = 300^\circ$ $\theta_2 = 60^\circ$ C_S symmetry	ag^- $\theta_1 = 180^\circ$ $\theta_2 = 60^\circ$	g^+g^- $\theta_1 = 60^\circ$ $\theta_2 = 60^\circ$ C_2 symmetry

Rotamères

ARG



PHE



Rotamères des acides aminés

Table I
Side-Chain Dihedral Angles Used as Starting Points for Energy Minimization^a

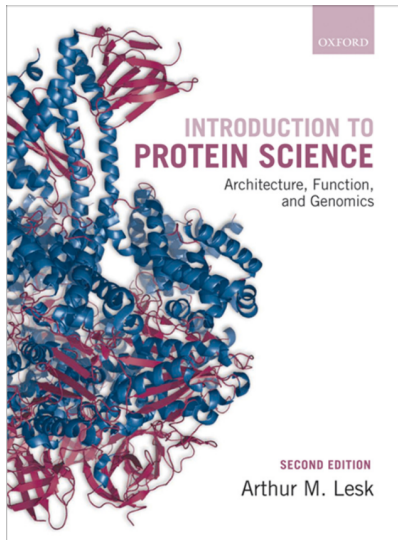
Residue	Side-chain dihedral angles, deg			
	χ^1	χ^2 ^b	χ^3 ^c	χ^4 ^d
Ala	60			
Arg	±60, 180	±60, 180	±60, 180	±60, 180 ^e
Asn	±60, 180	±30, ±90, ±150	180	
Asp	±60, 180	±30, ±90, ±150	0, 180	
Cys	±60, 180	±60, 180		
Gln	±60, 180	±60, 180	±30, ±90, ±150	180
Glu	±60, 180	±60, 180	±30, ±90, ±150	0, 180
Gly	<i>f</i>			
His ^g	±60, 180	±30, ±90, ±150		
Ile	±60, 180	±60, 180	60	60
Leu	±60, 180	±60, 180	60	60
Lys	±60, 180	±60, 180	±60, 180	±60, 180 ^h
Met	±60, 180	±60, 180	±60, 180	60
Phe	±60, 180	90, ±30		
Pro	<i>f</i>			
Ser	±60, 180	±60, 180		
Thr	±60, 180	±60, 180	60	
Trp	±60, 180	±30, ±90, ±150		
Tyr	±60, 180	±30, ±90, ±150	0	
Val	±60, 180	60	60	

2 Stabilité de l'état natif, replié

- Repliement des protéines - une bonne introduction
- Forces chimiques dans les protéines
- D'autres facteurs sur la stabilité des protéines

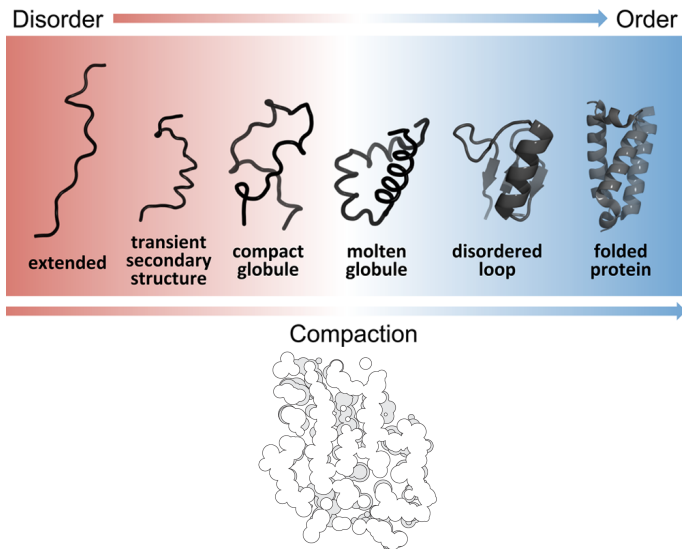
Repliement des protéines - une bonne introduction

Arthur M. Lesk - Introduction to protein science



Forces chimiques dans les protéines

Protéines globulaires



Forces chimiques dans les protéines

Chemical Interactions that Stabilize Polypeptides

Interaction	Example	Distance dependence	Typical distance	Free energy (bond dissociation enthalpies for the covalent bonds)
Covalent bond	$-\text{C}_{\alpha}-\text{C}-$	-	1.5 Å	356 kJ/mole (610 kJ/mole for a C=C bond)
Disulfide bond	$-\text{Cys}-\text{S}-\text{S}-\text{Cys}-$	-	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)

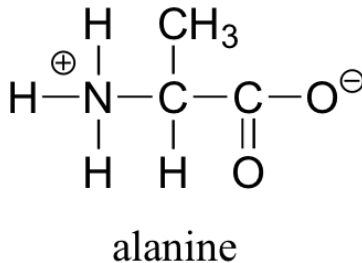
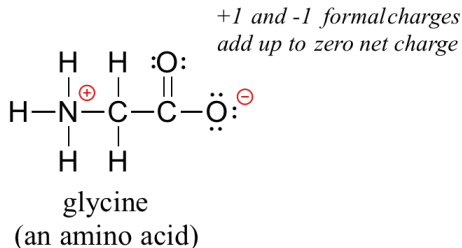
Forces chimiques dans les protéines

- Toutes les forces chimiques sont des forces électromagnétiques
- En théorie il suffirait d'écrire l'équation de Schrödinger puis résoudre le problème avec des méthodes de chimie quantique.
- Mais en pratique cela est impossible vu le grand nombre d'atomes d'une protéine
- On a besoin d'un modèle des interactions des centres des noyaux (appelé ici "atomes") sans traiter explicitement les électrons
- *Charges formelles*: montrent si un atome ou un ensemble d'atomes a perdu ou gagné un électron.

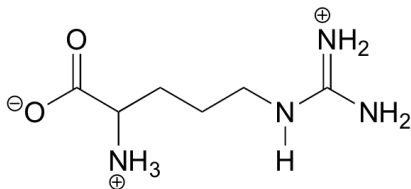
Charges formelles

● Charge formelle =

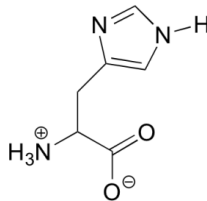
- nombre d'électrons de valence de l'atome isolé
- nombre d'électrons de valence de l'atome lié
- nombre d'électrons de valence de l'atome isolé
- nombre d'électrons non-liés de l'atome lié
- 1/2 nombre d'électrons liés de l'atome lié



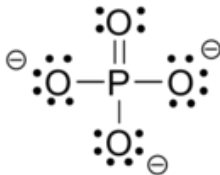
Charges formelles



arginine
 $\text{C}_6\text{H}_{15}\text{N}_4\text{O}_2^+$

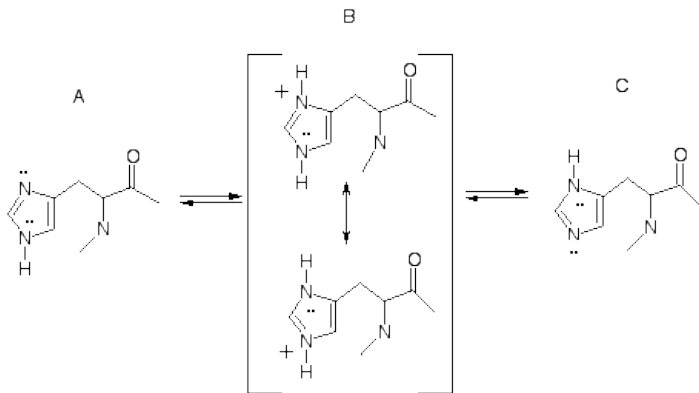


histidine



phosphate

Histidine à pH 7 chargé positivement



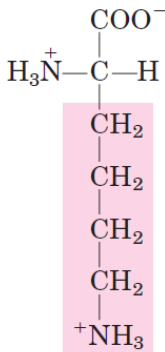
A or C: imidazole

B: imidazolium ion, resonance hybrid

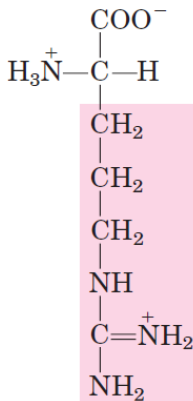
<https://spdbv.vital-it.ch/TheMolecularLevel/Goodies/Get2NoHistidine.html>

A.A. chargés positivement

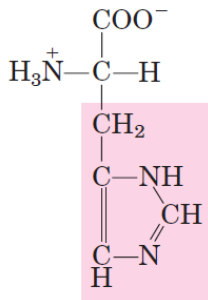
Positively charged R groups



Lysine

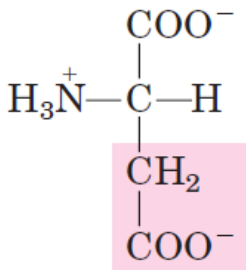


Arginine

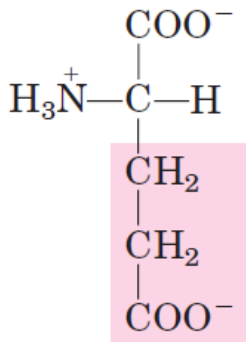


Histidine

A.A. chargés négativement

Negatively charged R groups

Aspartate

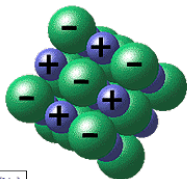


Glutamate

Pont salin (salt bridge, ion-pairing)

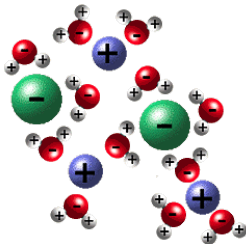
- interaction électrostatique entre deux (ou plus) résidus chargés
- peut inclure des liaisons hydrogènes, ce n'est pas un terme orthogonal à "liaison hydrogène"
- plus qu'une simple liaison hydrogène, forte interaction monopole
- ne dépend pas de la directionnalité
- atténué par l'eau qui fait écran
- pas souvent enfoui dans la protéine, malgré l'absence d'écran, car énergie libre favorable si en contact avec l'eau.

NaCl crystal structure

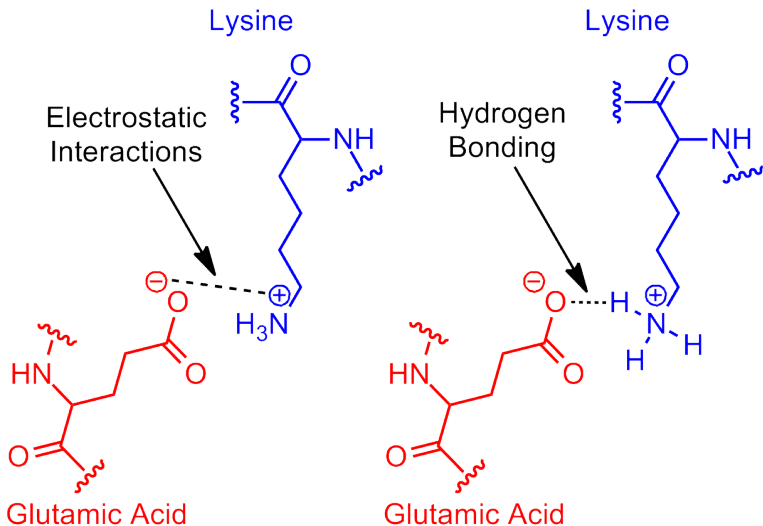


sodium (Na)
chlorine (Cl)

NaCl in water



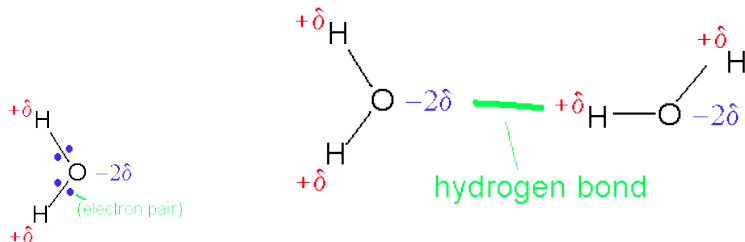
Pont salin (salt bridge or ion-pair)

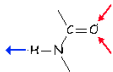


Liaisons hydrogènes (hydrogen bonds)

Liaisons hydrogènes:

- Interaction électrostatique entre des groupements sans charge formelle
- Polarisation: Électrons sont attirés par les atomes électronégatifs (O, N et un peu aussi S), mais restent partagés à travers la liaison covalente
- Ceci crée un moment dipolaire
- => Formation d'une liaison hydrogène
- Distance entre les deux O de H₂O: 2,8 Å





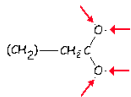
peptide



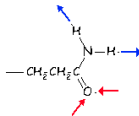
serine



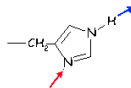
threonine



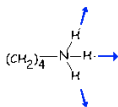
glutamic/aspartic acid



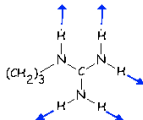
glutamine/asparagine



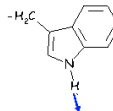
histidine



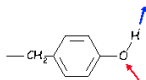
lysine



arginine



tryptophan



tyrosine

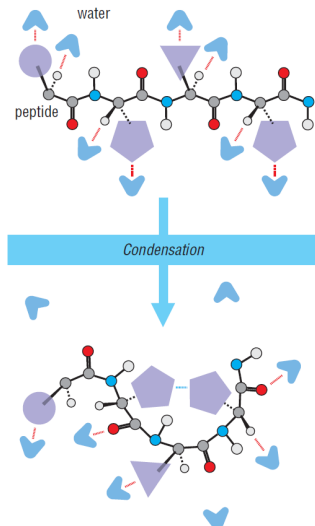
hydrogen bond donor →
hydrogen bond acceptor ←

Liaisons hydrogènes (hydrogen bonds)

Liaisons hydrogènes:

- Le squelette forme des liaisons hydrogènes, car il est polaire
- État dénaturée: liaisons hydrogènes avec l'eau
- État natif: liaisons hydrogènes entre les résidus => formation de structures secondaires
- Liaisons hydrogènes aussi avec des cofacteurs et des ligands

Liaisons hydrogènes (hydrogen bonds)



Liaisons hydrogènes (hydrogen bonds)

Liaisons hydrogènes:

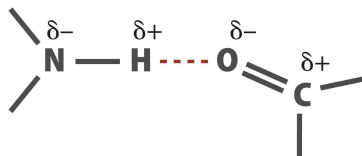


Figure 1.18 How Proteins Work (©2012 Garland Science)

- Charges partielles: N: -0.4, H: +0.2, O: -0.5, C: +0.5
- Très directionnel: Angle linéaire pour N-H...O et entre 120 et 180 pour C=O...H
- Liaison légèrement covalente, car on observe du couplage J en RMN à travers la liaison hydrogène.
- Énergie: 15-20 kJ / mol (valeur égale entre peptide-peptide et peptide-eau)

Liaisons hydrogènes (hydrogen bonds)

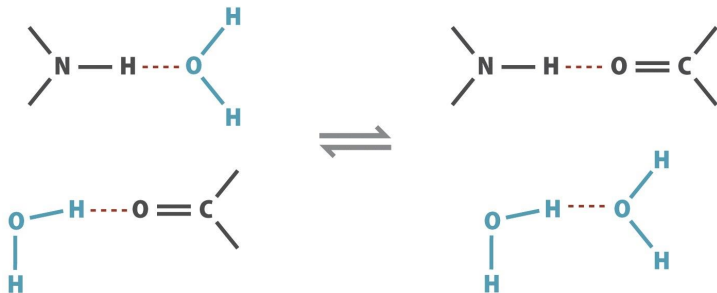
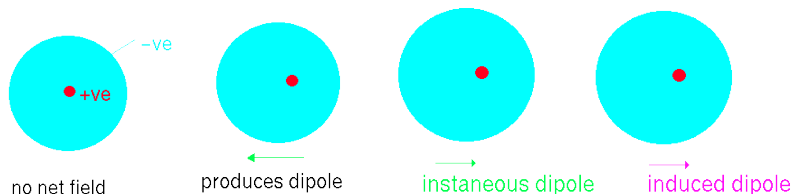


Figure 1.19 How Proteins Work (©2012 Garland Science)

Force de dispersion ou van der Waals



- Fluctuations de la densité des électrons autour du noyau d'un atome, créent un moment dipolaire temporaire.
- Ce moment dipolaire induit à un autre atome très proche aussi un moment dipolaire transitoire
- Résultat: force attractive $F_{disp} = -B_{ij}/r_{ij}^6$, faible mais additive

van der Waals - potentiel de Lennard-Jones

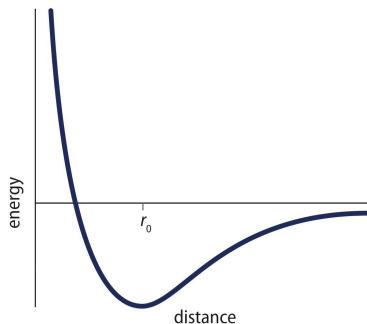


Figure 1.20 How Proteins Work (©2012 Garland Science)

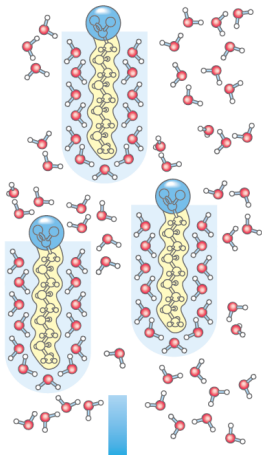
- Résultat: force attractive $F_{disp} = -B_{ij}/r_{ij}^6$, faible mais additive
- Répulsion (principe de Pauli): $F_{rep} = A_{ij}/r_{ij}^{12}$
- Total: $F_{total} = A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6$
- Rayons de van der Waals (minimum du potentiel r_0):
C (1,85Å), O (1,60Å), H (1,0Å), N (1,75Å)

D'autres facteurs sur la stabilité des protéines

Paramètres expérimentaux

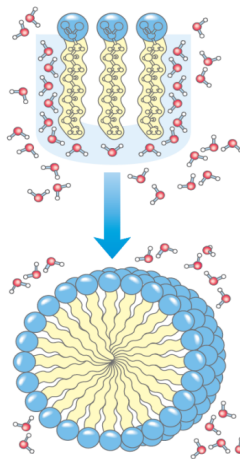
- Température : augmente énergies vibration & rotation
- Pression (modification du solvant, liaisons H rompues)
- pH (interactions électrostatiques) : dépliement ≤ 5 & ≥ 10
- Agent chimique :
 - Urée
 - TFE (2,2,2-trifluoro-ethanol, $\text{CF}_3\text{CH}_2\text{OH}$) favorise hélices
 - chlorure de guanidinium (GuHCl) très fréquent
- Mutations peuvent modifier la cinétique

Effet hydrophobe, ex: micelles



Dispersion of lipids in H₂O

Each lipid molecule forces surrounding H₂O molecules to become highly ordered.



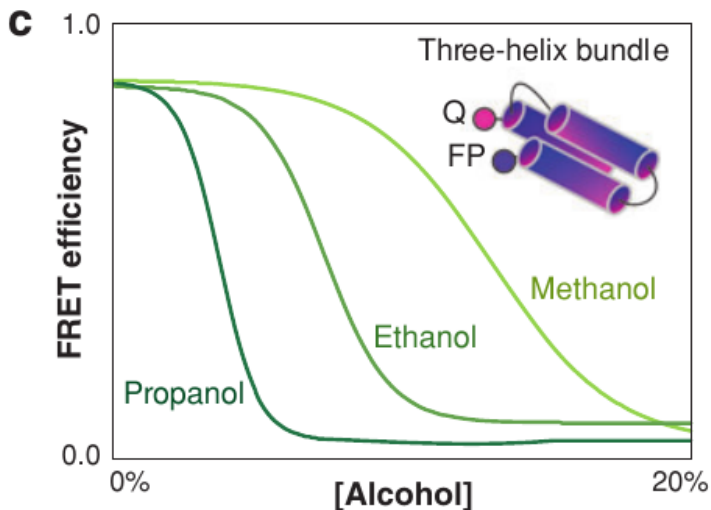
Clusters of lipid molecules

Only lipid portions at the edge of the cluster force the ordering of water. Fewer H₂O molecules are ordered, and entropy is increased.

Micelles

All hydrophobic groups are sequestered from water; ordered shell of H₂O molecules is minimized, and entropy is further increased.

Protéine dans l'alcool



Amino Acid	Group	Engleman et al.(3)	Kyte and Doolittle (2)	Hoop and Woods (4)
Phe	Nonpolar	3.7	2.8	-2.5
Met	Nonpolar	3.4	1.9	-1.3
Ile	Nonpolar	3.1	4.5	-1.8
Leu	Nonpolar	2.8	3.8	-1.8
Val	Nonpolar	2.6	4.2	-1.5
Cys	Unch/Polar	2.0	2.5	-1.0
Trp	Nonpolar	1.9	-0.9	-3.4
Ala	Nonpolar	1.6	1.8	-0.5
Thr	Unch/Polar	1.2	-0.7	-0.4
Gly	Nonpolar	1.0	-0.4	0.0
Ser	Unch/Polar	0.6	-0.8	0.3
Pro	Nonpolar	-0.2	-1.6	0.0
Tyr	Unch/Polar	-0.7	-1.3	-2.3
His	Charged (+)	-3.0	-3.2	-0.5
Gln	Unch/Polar	-4.1	-3.5	0.2
Asn	Unch/Polar	-4.8	-3.5	0.2
Glu	Charged (-)	-8.2	-3.5	3.0
Lys	Charged (+)	-8.8	-3.9	3.0
Asp	Charged (-)	-9.2	-3.5	3.0
Arg	Charged (+)	-12.3	-4.5	3.0

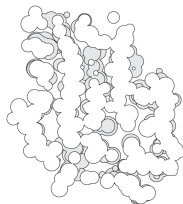
Accessible and buried surface area (ASA, BSA)

Effet hydrophobe => maximiser la surface enfouie

- Chaque $\text{\AA}^2$ de surface enfouie contribue à peu près à 6 J/mol pour ΔG .
- Formation d'oligomères: 1000 à 5000 $\text{\AA}^2$
- Très défavorable d'enfouir des chaînes latérales chargées, car elles forment déjà des liaisons très favorable avec l'eau.
- On peut avoir des surfaces hydrophobes => sites de liaisons protéine-protéine

Facteurs qui stabilisent l'état natif

Compactification:



- Améliore les forces attractives de van der Waals

Cooperativité

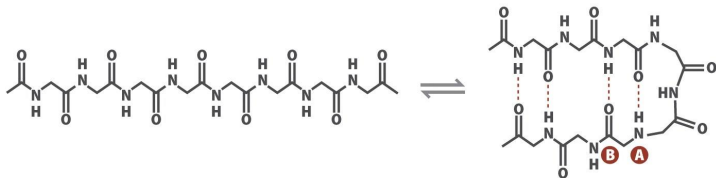


Figure 1.25 How Proteins Work (©2012 Garland Science)

Coopérativité

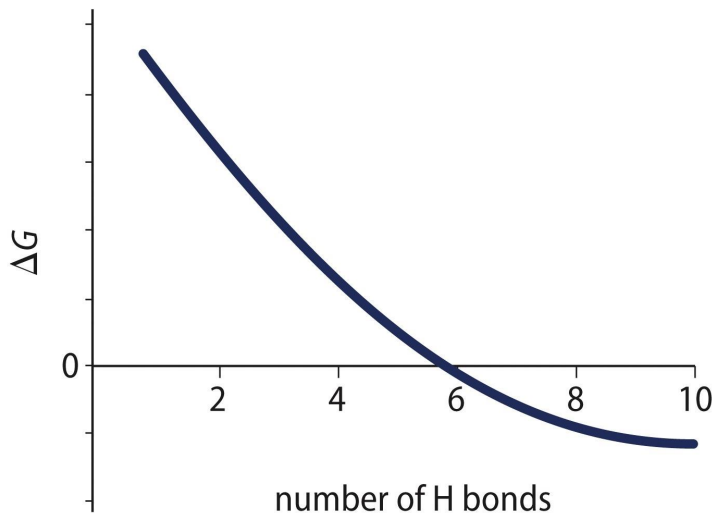
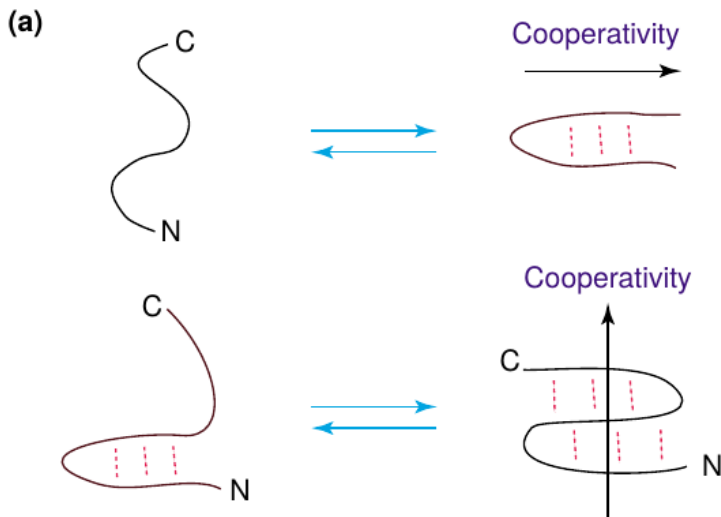
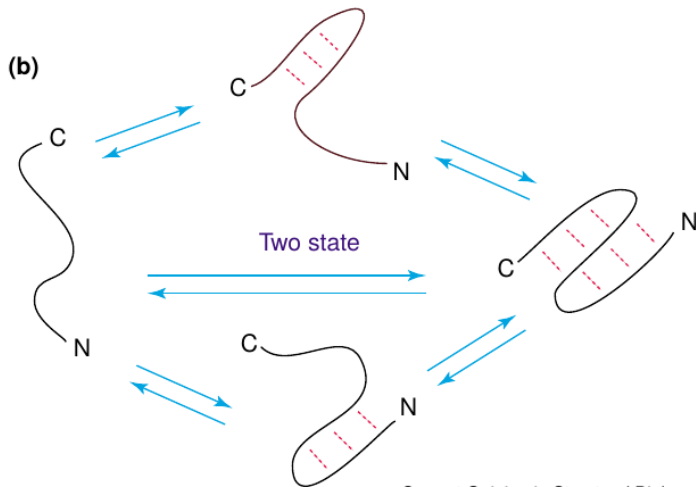


Figure 1.26 How Proteins Work (©2012 Garland Science)

Cooperativité



Two or four state model ?



Current Opinion in Structural Biology

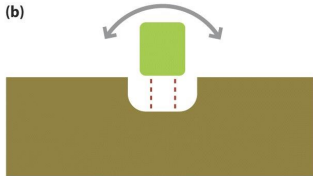
Compensation Enthalpie / Entropie

(a)



Figure 1.29 How Proteins Work (©2012 Garland Science)

(b)



Compensation Enthalpie / Entropie

Plusieurs facteurs possibles:

- solvant (eau)
- mobilité protéine et ligand
- protéine dénaturée = système non-corrélé avec beaucoup de degrés de liberté
- protéine native = système hautement corrélé avec peu de degrés de liberté

Difficile de savoir d'avance qui domine dans un cas concret.

Compensation Enthalpie / Entropie

Redistribution des forces de liaison (+uniforme, +mobile => +Entropie):

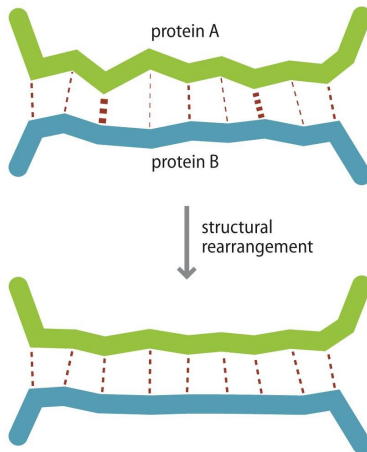
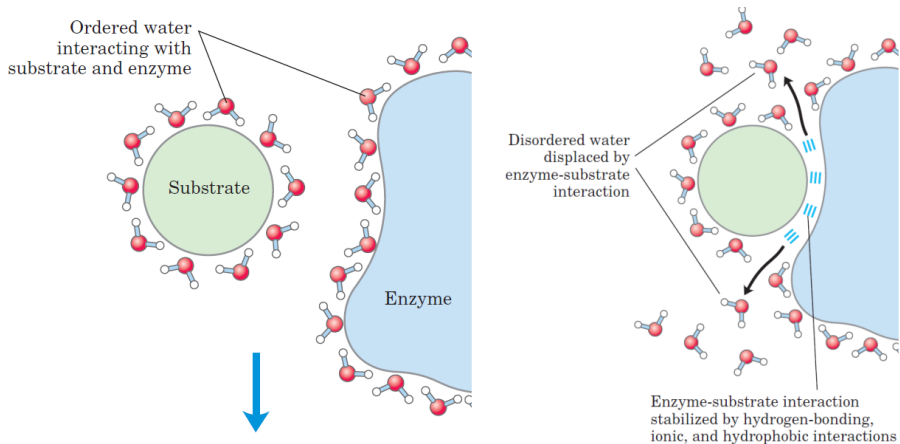


Figure 1.30 How Proteins Work (©2012 Garland Science)

Désolvatation

Désolvatation:



Facteurs qui stabilisent l'état natif

Liaisons covalentes: ponts di-sulfures, exemple Crambin (1CRN):

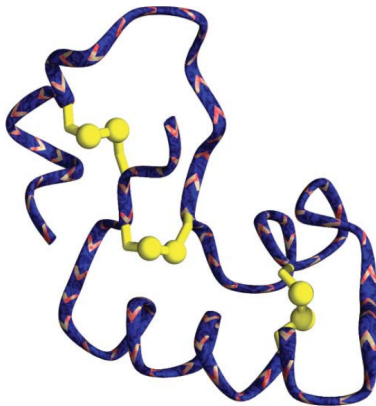
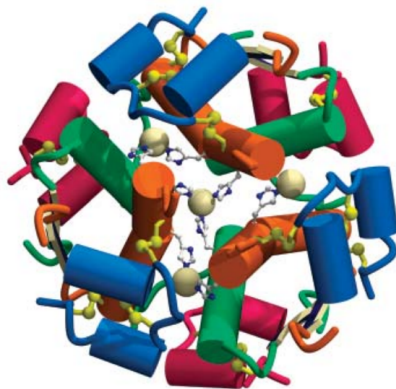


Figure 2.5 Crambin [1CRN]. The yellow double-lollypops are disulphide bridges.

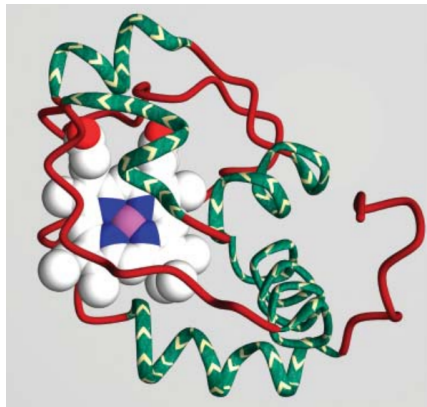
Facteurs qui stabilisent l'état natif

Ponts di-sulfures et ions métalliques, exemple insuline hexamer avec 4 zinc (4ZNI):



Facteurs qui stabilisent l'état natif

Ions métalliques dans un ligand, exemple cytochrome c (1CCR):



3

Domaines

- Domaines
- Annotations des séquences des protéines
- Classification structurale

Domaines

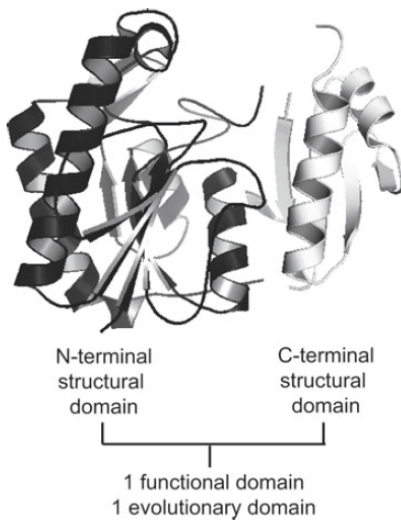
Domaines

Il existe trois types de domaines:

- 1 Domaine structural
- 2 Domaine fonctionnel
- 3 Domaine évolutif

Domaine structural:

- 1 compact, coeur hydrophobe
- 2 peut souvent se replier tout seul
- 3 chaque chaîne polypeptidique peut comprendre un ou plusieurs domaines



Domaines globulaires

- Chaîne polypeptidique peut comprendre plusieurs domaines, parfois appelés lobes
- Domaine défini par caractéristiques 3D: compacité
- Ne correspond pas nécessairement à une continuité dans la séquence
- Il peut y avoir une fonction par domaine

Fibronectine

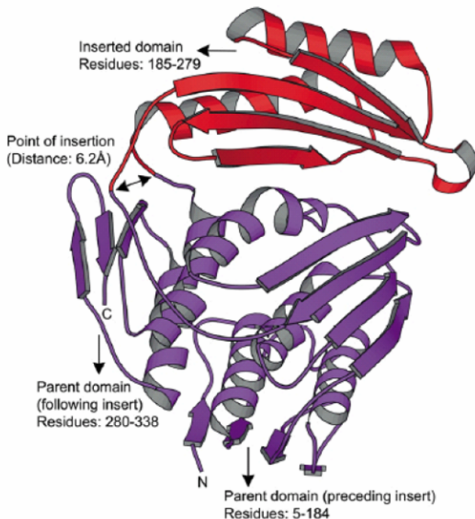


Code PDB : 1fnh. 3 domaines immunos

Liaison aux surfaces cellulaires et à la fibronectine,
l'héparine, l'actine...

80% domaines = segments contigu d'une seule chaîne
Parfois insertion d'un domaine dans un autre
(parent)

Domaine type
thioredoxine inséré
(rouge) dans un
domaine parent (bleu)
: 1qmhA



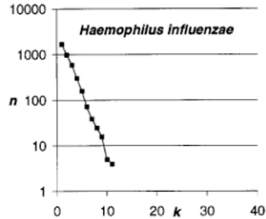
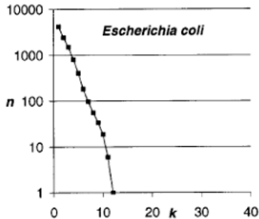
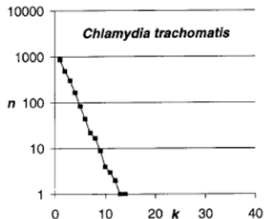
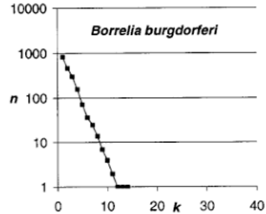
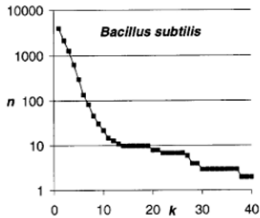
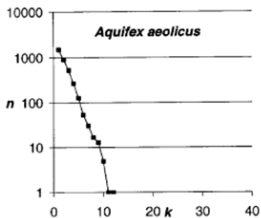
Protéines multidomaines

Grande proportion de gènes, jusqu'à 90% dans les eucaryotes, forment des chaînes de protéines multidomaines

Ceci augmente le nombre de gènes orphelins, si on tient compte des domaines

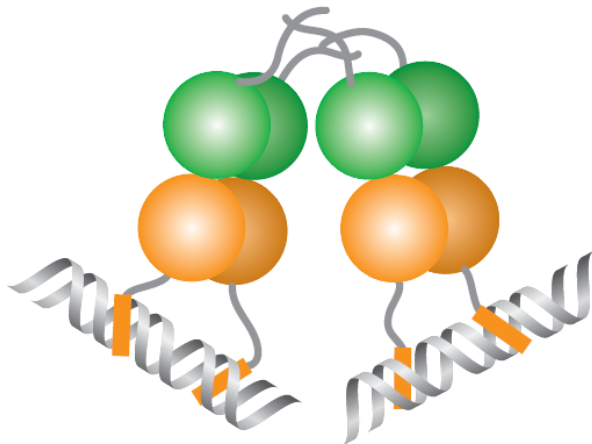
On ne peut que modéliser un domaine, pas toute une chaîne

Protéines multidomaines



Distribution des protéines multi domaines dans 17 génomes ; n = nombre de protéines de k domaines.

Exemple: lac repressor



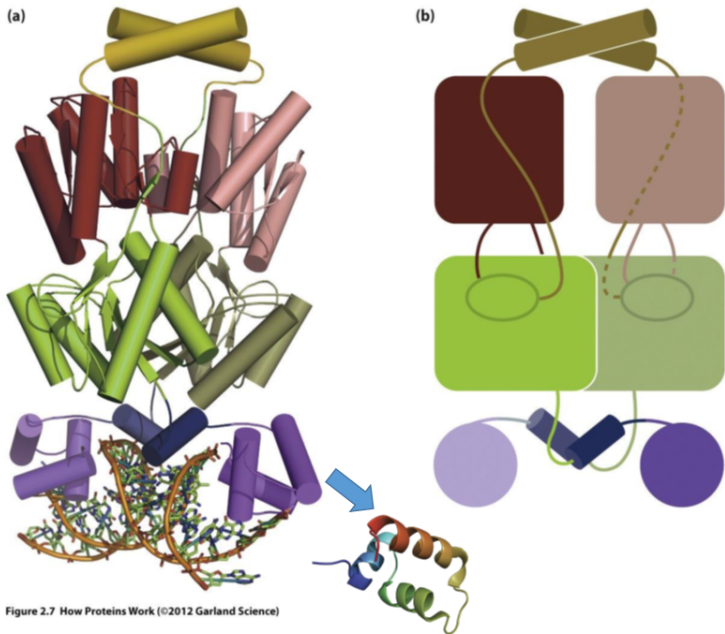
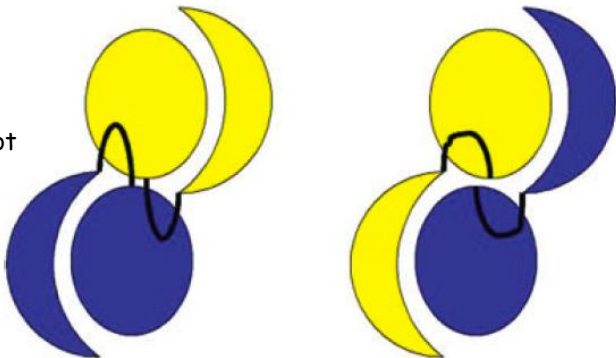


Figure 2.7 How Proteins Work (©2012 Garland Science)

Domain Swapping

- Echange de domaines (SSR) entre monomères pour polymériser
- En général, fragments échangés aux extrémités

Liu & Eisenberg, Prot
Sci 11 (2002) 1285



Annotations des séquences des protéines

Outils d'annotation

Annotation (séquence, structure, fonction):

- InterProScan: Identification domaines, motifs, familles
- PsiPred: Prédiction structure secondaire
- DisoPred: Prédiction désordre (=> IUP)
- MEMSAT: Prédiction segments transmembranaires

Classification structurale

CATH - Class, Architecture, Topology, Homology

- classement automatique avec vérification manuelle
- organisation hiérarchique avec 7 niveau:
 - 1 *Homologous superfamily (H-level)* = similitude au niveau séquence, structure ou fonction
 - 2 *Topology (T-level)* = topologie et arrangement similaire des structures secondaires
 - 3 *Architecture (A-level)* = arrangement similaire des structures secondaires, indépendamment de la connectivité
 - 4 *Class (C-level)* = composition en structure secondaire
- *A-level*: spécifique à CATH, les trois autres niveaux ont un niveau équivalent en SCOP.

CATH

Viewed CATH
Protein Structure Classification

CATH DHS Gene3D Impala FTP Internal

Home > Top > Class 1 > 10 > 210 > 10 > 1 > 1 > 1 > 1utg00 View this page as XML

Search

Go!

PDB Code
CATH Code
General Text

Goto

SSAP Server
GRATH Server
DHS
Gene3D

Navigation

Home
Top of hierarchy

CATH Domain 1utg00

Classification

- Class 1
- Mainly Alpha
- Architecture 1.10
 - Orthogonal Bundle
- Topology 1.10.210
 - Uteroglobin
- Homologous Superfamily 1.10.210.10
 - Uteroglobin
- Sequence Family (S35) 1.10.210.10.1
 - Uteroglobin
- Non-identical (S95) 1.10.210.10.1.1
 - Uteroglobin
- Identical (S100) 1.10.210.10.1.1.1
 - Uteroglobin

PDB Information

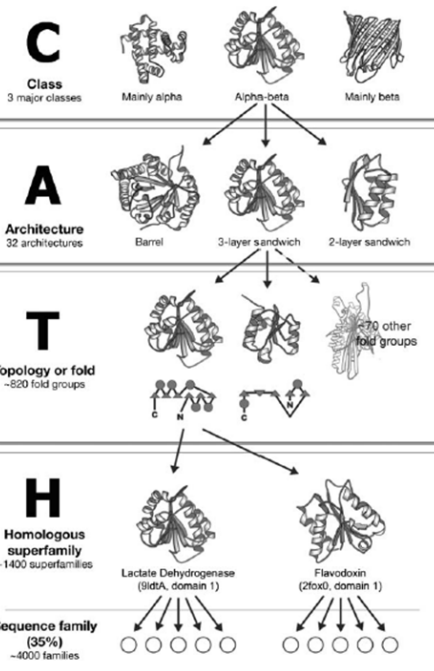
PDB Code	1utg
PDB Header	Uteroglobin (oxidized)
PDB Source	Rabbit (oryctolagus cuniculus) female genital tract

Domain Information

Domain Sequence	GICPRFAHVIEIENLLGTPSSYETLSKEFEPDDTMKDAGM QMKKVLDLSPQTTREINIMKLTEKIVKSPCLM
-----------------	--

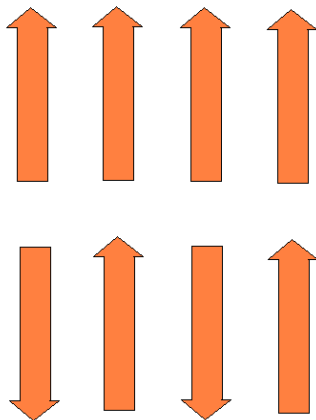
CATH considers structural domains as semi-independent folding units. It is quite common that a structural domain is made up of more than one sequence segment (i.e. non-sequential stretch of peptide). The table below provides information for each of the segments that make up the structural domain.

Segment Number	Segment Length	PDB Start	PDB Stop
1	70	1	70



Examples of protein architecture

Architecture refers to the arrangement and orientation of SSEs, but not to the connectivity.

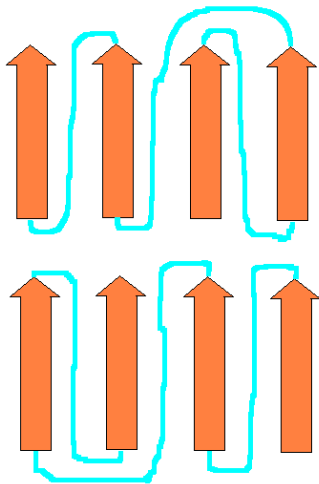


β -sheet with all pairs of strands parallel

β -sheet with all pairs of strands anti-parallel

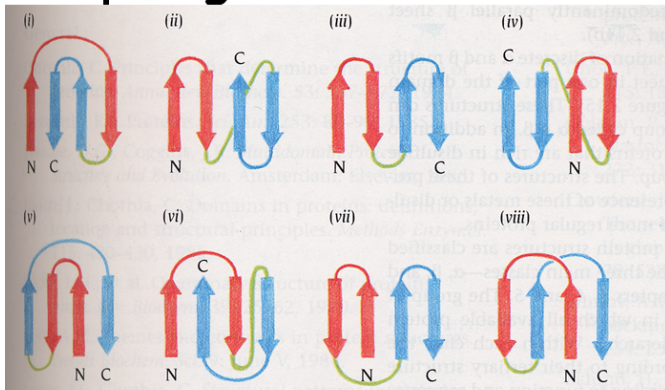
Examples of protein topology

Topology refers to the manner in which the SSEs are connected.



Two β -sheets (all parallel) with different topologies.

Topologies des feuilletés à 4 brins



- Feuillet 4 brins = 2 épingles à cheveux beta
 - 24 topologies possibles
 - 8 dans la PDB
 - Deux très fréquentes : clef grecque (i et v)
- Validation modèles!

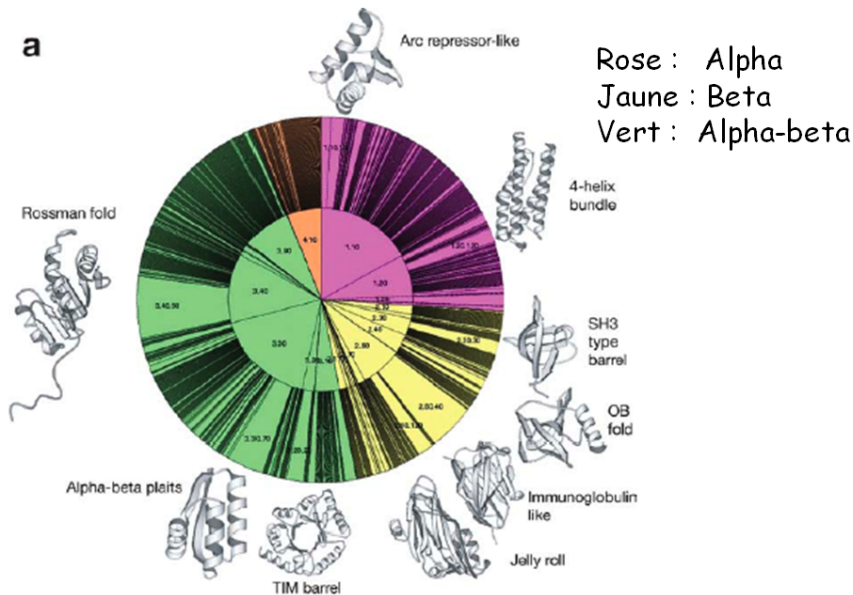
CATH: taxinomie structurale

CLASSES

- **alpha** (principalement hélices); pas nécessairement tout alpha
- **beta** (principalement brins)
- **alpha+beta**, dissociation des domaines α et β
- **alpha/beta**, alternance, unités $\beta\alpha\beta$
- **few**, séquence courte, stabilisation non effectuée par les SSR mais par les ponts disulfure

Roue CATherine

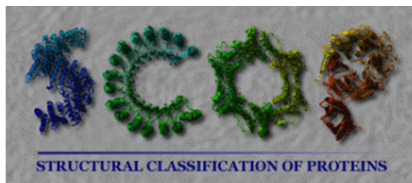
a



SCOP - Structural Classification of Proteins

- classer les protéines dans un cadre évolutif cohérent, basé sur la conservation de la séquence et de la structure
- organisation hiérarchique avec 6 niveau:
 - 1 *Species* = au niveau de l'espèce
 - 2 *Protein* = groupe de protéines avec la même fonction dans différentes espèces ou des protéines isoformes dans le même organisme.
 - 3 *Family* = séquences homologues, mais fonction différente
 - 4 *Superfamily* = groupe de familles avec une propriété fonctionnelle ou structurale commune
 - 5 *Folds* = superfamilles qui sont similaires au niveau de la structure
 - 6 *Classes* = composition en structure secondaire
- classification fait manuellement => bonne référence, mais en retard par rapport à la PDB.

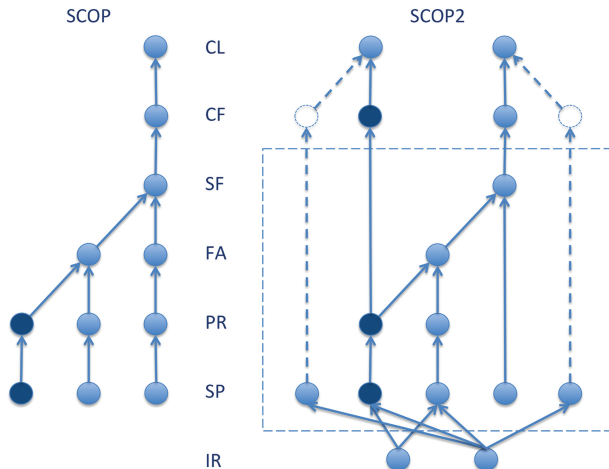
SCOP



	2002	2009 Fev
Repliements tout alpha	151	284
Repliements tout beta	110	174
Repliements alpha/beta	113	147
Repliements alpha+beta	208	376
Multi domaines	34	66
Protéines membranaires	12	58
Petites protéines	58	90
Total de repliements	686	1195

SCOP 2

<http://scop2.mrc-lmb.cam.ac.uk/>



SCOP 2

<http://scop2.mrc-lmb.cam.ac.uk/>

The screenshot displays the SCOP 2 website interface, which is used for the Structural Classification of Proteins 2. The interface includes a main navigation menu at the top with links for About, Browser, Graph, Download, and Support. A color legend is located on the left side, and a search bar is on the right. The central part of the page features a hierarchical tree diagram showing the classification of proteins, with nodes representing different levels of structural organization. A sidebar on the left lists various relations and nodes, including Category, Attribute, Type, Class, Hyperfamily, Fold, Superfamily, Family, Protein, Species, Event, and Interrelationship. A pop-up window on the right shows details for a specific protein structure (1R-8X0370-SCQA), including its domain, comment, segments, and a representative sequence. The interface is annotated with red boxes and arrows pointing to various features:

- Alternative ways for database access:** Points to the About, Browser, Graph, Download, and Support links in the main navigation menu.
- Main navigation menu:** Points to the same set of links in the main navigation menu.
- Action buttons:** Points to the Configure, Share, Reset, and Help buttons in the top navigation bar.
- Navigation tabs:** Points to the Structure, Parents, Children, and Comments tabs in the pop-up window.
- Action buttons:** Points to the Focus, Remove, and Browse buttons in the pop-up window.
- Links to external sequence and structural resources:** Points to the View in PDBe, View in PDB, and View in UniProt links in the pop-up window.
- Click to bring up the node pop up window:** Points to a node in the hierarchical tree diagram.

SCOP 2 - July 2013 © MRC Laboratory of Molecular Biology

Contact | Sitemap | Top of page

`http://ekhidna.biocenter.helsinki.fi/dali`

Dali Database

Institute of
Biotechnology

[SERVICES & TOOLS](#)[GROUP MEMBERS](#)[NEWS & VACANCIES](#)[RESEARCH](#)[PUBLICATIONS](#)

Dali structural neighbours

The Dali Database is based on all-against-all 3D structure comparison of protein structures in the Protein Data Bank (PDB). The structural neighbourhoods and alignments are automatically maintained and regularly updated using the [Dali](#) search engine.

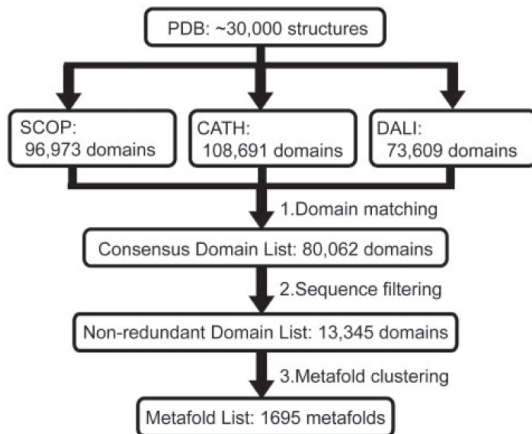
- Please note that PDB structures released after the last update will not be in the database! If you wish to find structural neighbours of these proteins, you are advised to submit the structure to the [Dali Server](#) instead.
- If you want to superimpose two particular structures, you can do it in the [pairwise DaliLite](#) server.

* Last Update: 7 March 2011
Update frequency: twice a year

Enter PDB identifier: chain: (optional)

(Keyword search for PDB identifiers)

Consensus Domain Dictionary (CDD)



Consensus Domain Dictionary (CDD)

Table 1. Summary statistics of the SCOP, CATH and Dali domain dictionaries used in the v2003 and v2009 CDD

	Chains (C)	Domains (D)	Folds ^a	D/C ^b
v2003				
SCOP	27 308	35 095	783	1.29
CATH	25 622	36 480	1453	1.42
Dali	21 493	35 492	1088	1.65
v2009				
SCOP	74 608	96 973	1280	1.29
CATH	74 240	108 691	1110	1.46
Dali	52 740	73 609	2783	1.39

^aNumber of unique folds at the chosen level within each domain dictionary.

^bNumber of distinct domains (D) per distinct chain (C).

Nombre de folds

Table 1. Estimates for the number of protein folds and superfamilies by year

Year	Folds	Superfamilies	Reference
1992	<1000	1500	Chothia (1992)
1994	<7700	23 100	Orengo <i>et al.</i> (1994)
1994	6727	—	Alexandrov and Go (1994)
1996	455	—	Zhi-Xin (1996)
1997	<920	920	Brenner <i>et al.</i> (1997)
1997	≤5200	17 175	Zhang (1997)
1998	650	1150	Zhi-Xin (1998)
1998	836	—	Zhang and DeLisi (1998)
1999	3756	—	Govindarajan <i>et al.</i> (1999)
2000	~1000	4000–7000	Wolf <i>et al.</i> (2000)
2002	10 000	50 000	Coulson and Moulton (2002)
2007	1613	—	Levitt (2007)
2009	~1700 ± 400	~4000	Sadreyev <i>et al.</i> (2009)

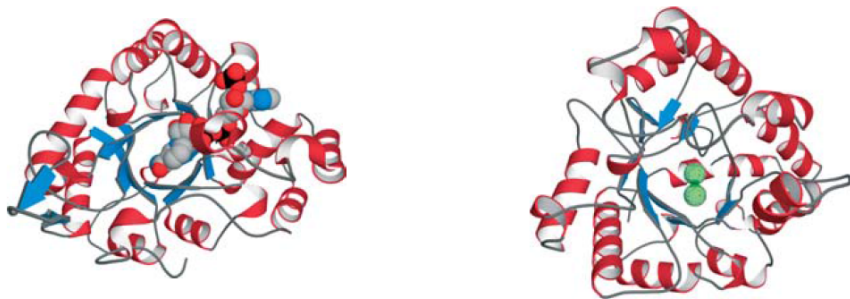
4

Structures tertiaires

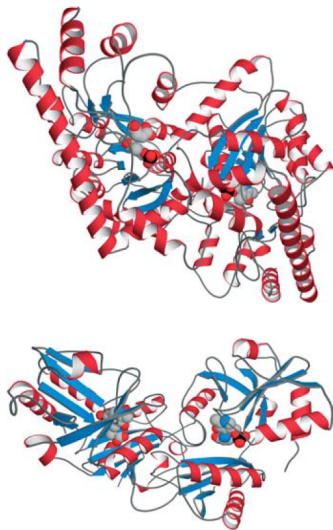
- Structure $\leftrightarrow$ Fonction

Structure $\leftrightarrow$ Fonction

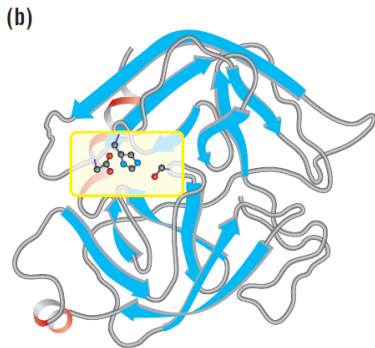
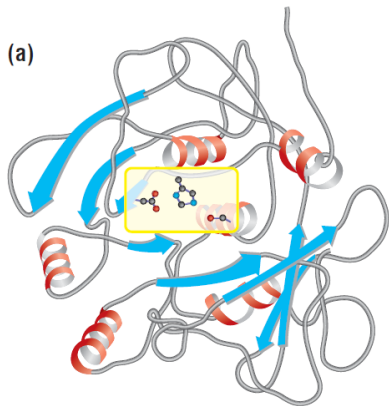
Structure 3D similaire, mais fonction différent



Même fonction, mais pas Structure 3D



Même fonction, mais pas Structure 3D



Dogmes de la biologie structurale

- Séquence => Structure => Fonction
- Une séquence - une structure

Dogmes de la biologie structurale

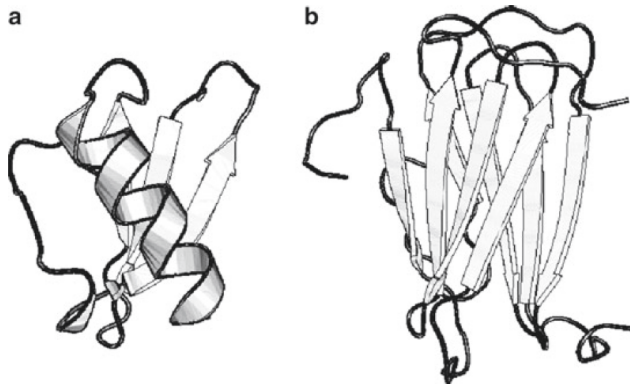


Fig. 5. The structures of two alternative folds of lymphotactin (Ltn10). (a) Monomeric Ltn10 (pdb 1j8i) and (b) dimeric Ltn10 (pdb 2jp1).

Dogmes de la biologie structurale

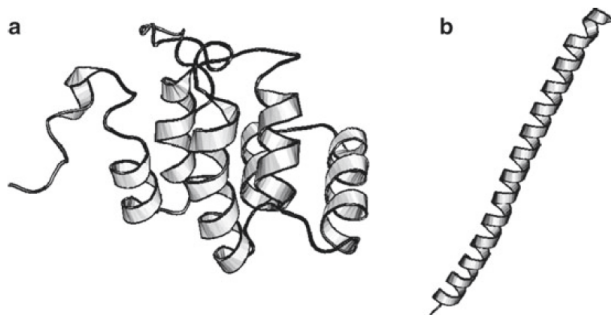


Fig. 6. The death domain of protein kinase Pelle (Pelle-DD) (a) solution structure, (b) crystal structure in MPD.

Andreeva, "Homology Modeling", ch. 1, Methods in Mol. Biol.(2012)

Dogmes de la biologie structurale

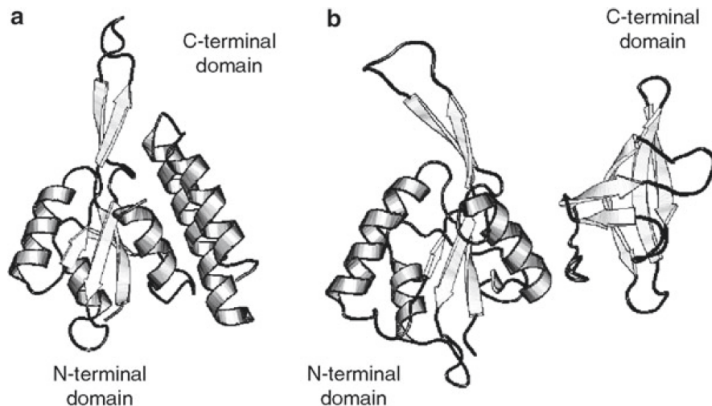


Fig. 10. Fold transition. Structures of (a) RfaH and (b) NusG.

- 5 Structures quartanaires
 - Structures quaternaires

Structures quaternaires

Structure quaternaire

- Résultat de l'agrégation de plusieurs chaînes polypeptidiques pour former une protéine biologiquement fonctionnelle
- Homopolymères ou hétéropolymères
- Cavités peuvent exister entre les monomères

Structure quaternaire

Intérêt d'une structure à plusieurs chaînes:

SAV

facilité de réparation plus grande qu'avec une seule chaîne gigantesque

Lego

Principe de la construction en préfabriqué et assemblage sur place : les sites de fabrication des sous unités ne sont pas nécessairement les mêmes que celui de l'assemblage complet

Structure quaternaire

Très souvent, un domaine peut interagir avec plusieurs sortes de domaines, par l'intermédiaire de différentes surfaces d'interaction, tel que c'est le cas pour le domaine G Ras-like

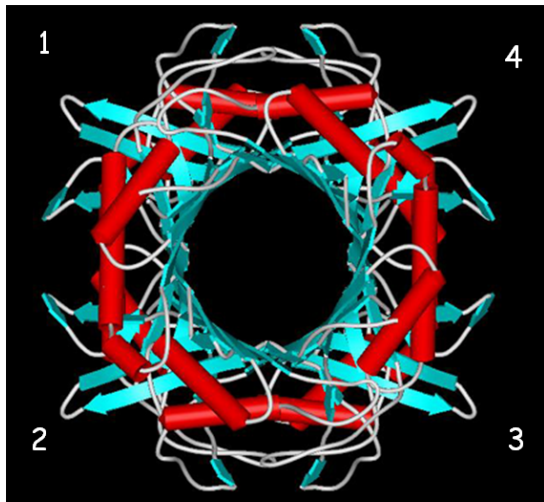
Pour des protéines qui appartiennent à une même famille, l'interaction se fait habituellement de manière similaire

Kiel Annu Rev Biochem (2008) 77:1-5

Urate Oxydase



monomère



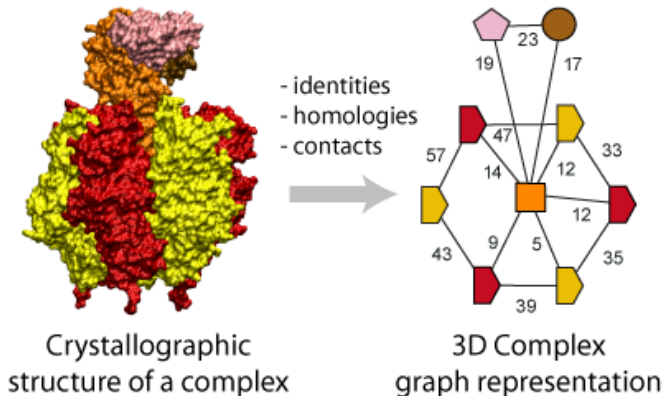
tétramère

30

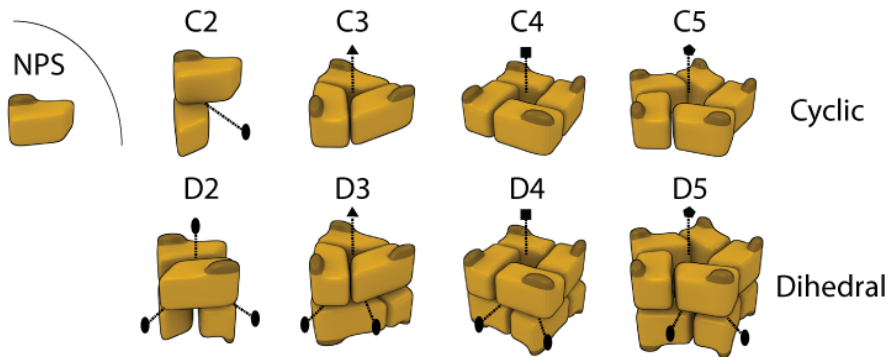
3D complex

- Classement des structures de complexes de protéine
- Complexes homo- et hétéromer
- Relations évolutives
- Organisation hiérarchique avec 12 niveau selon 5 critères:
 - 1 Topologie du complex, représentée par le nombre de chaînes et leurs contacts
 - 2 Architecture en domaines de chaque chaîne comme donnée dans SCOP
 - 3 Nombre de chaînes non-identiques par architecture en domaines
 - 4 Similarité de séquence entre les chaînes d'un complexe
 - 5 Symétrie du complexe

Graph des assemblages



Symétries des homomères



6 Outils

- Banques de structures 3D expérimentales
- Modèles par homologie

Banques de structures 3D expérimentales

PDB, PDBsum, PDBe

- PDB <http://www.rcsb.org>
- PDBsum <http://www.ebi.ac.uk/pdbsum/>
- PDBe <http://www.ebi.ac.uk/pdbe/>



Search

Advanced

Browse

Everything

Author

Macromolecule

Sequence

Ligand



2PE5

[Search History](#) , [Previous Results](#)

Summary

3D View

Sequence

Annotations

Seq. Similarity

3D Similarity

Literature

Biol. & Chem.

Methods

Links

Crystal Structure of the Lac Repressor bound to ONPG in repressed state

DOI: [10.2210/pdb2pe5/pdb](#) NDB ID: PD1093

2PE5

 Display Files ▾
 Download Files ▾

Primary Citation

Structural analysis of lac repressor bound to allosteric effectors.

Daber, R. , Stayrook, S. , Rosenberg, A. , Lewis, M.

Journal: (2007) J.Mol.Biol. **370**: 609-619PubMed: [17543986](#) PubMedCentral: [PMC2715899](#) DOI: [10.1016/j.jmb.2007.04.028](#) [Search Related Articles in PubMed](#)

PubMed Abstract:

The lac operon is a model system for understanding how effector molecules regulate transcription and are necessary for allosteric transitions. The crystal structures of the lac repressor bound to inducer and anti-inducer molecules provide a model for how these small... [[Read More & Search PubMed Abstracts](#)]

† Molecular Description

Hide

Classification: Transcription/dna

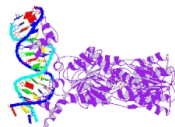
Structure Weight: 125535.28

Molecule: DNA
(5'-D(*DAP*DAP*DTP*DTP*DGP*DGP*DAP*DGP*DCP*DGP*DCP*DTP*DCP*DAP*DCP*DAP*DAP*DTP*DT)-3')

Polymer: 1 Type: dna Length: 20

Chains: D, E, F

Biological Assembly 1



3D View

[More Images...](#)

Symmetry: C2 view

Stoichiometry: Homo 2-mer - A2

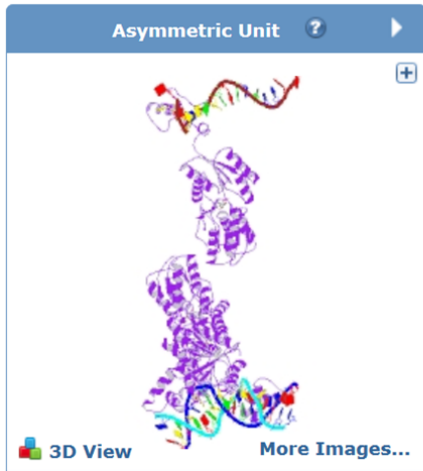
Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:

[Simple Viewer](#)[Protein Workshop](#)

2PE5

 Display Files ▾
 Download Files ▾



2PE5

 Display Files ▾
 Download Files ▾

Download

Asy

- FASTA Sequence
- PDB File (Text)
- PDB File (gz)
- mmCIF File
- mmCIF File (gz)
- PDBML/XML File
- PDBML/XML File (gz)
- Structure Factor (Text)
- Structure Factor (gz)
- Biological Assembly 1 (gz) (A+S)
- Biological Assembly 2 (gz) (A+S)

Fichier PDB

Coordonnées atomiques disponibles dans une banque de structures, la Protein Data Bank <http://www.rcsb.org/pdb/>

```
HEADER DNA-BINDINGPROTEIN 20-MAY-94 1ENH 1ENH 2
COMPND ENGRAILEDHOMEODOMAIN 1ENH 3
SOURCE (DROSOPHILA MELANOGASTER) RECOMBINANT FORM EXPRESSED IN 1ENH 4
SOURCE 2 (ESCHERICHIA COLI) 1ENH 5
AUTHOR N.D.CLARKE,C.R.KISSINGER,J.DESJARLAIS,G.L.GILLILAND,C.O.PABO 1ENH 6
REVDAT 1 31-AUG-94 1ENH 0 1ENH 7
SCALE3 0.000000 0.000000 0.008466 0.000000 1ENH 61
ATOM 1 N ARG 3 2.937 44.573 53.291 1.00 62.68 1ENH 62
ATOM 2 CA ARG 3 3.220 44.968 51.871 1.00 61.88 1ENH 63
ATOM 3 C ARG 3 1.922 45.475 51.229 1.00 62.67 1ENH 64
ATOM 4 O ARG 3 0.984 44.702 51.017 1.00 65.49 1ENH 65
ATOM 5 CB ARG 3 3.758 43.763 51.101 1.00 58.73 1ENH 66
ATOM 6 CG ARG 3 3.642 43.884 49.610 1.00 57.06 1ENH 67
ATOM 7 CD ARG 3 3.776 42.528 48.965 1.00 54.58 1ENH 68
ATOM 8 NE ARG 3 5.083 42.365 48.340 1.00 56.07 1ENH 69
ATOM 9 CZ ARG 3 6.183 41.961 48.980 1.00 57.06 1ENH 70
ATOM 10 NH1 ARG 3 6.141 41.670 50.274 1.00 57.63 1ENH 71
ATOM 11 NH2 ARG 3 7.335 41.841 48.325 1.00 57.77 1ENH 72
```

Natom Nrésidu X Y Z Q B

Fichier PDB

- Pas d'information explicite sur les liaisons entre atomes : c'est déterminé par le calcul des distances interatomiques et le nom des atomes (nomenclature internationale)
- Pas d'information sur la charge des atomes: doit être calculée

Crystal Structure of the Lac Repressor bound to ONPG in repressed state

Structural Similarities for the Entities in PDB 2PE5

Comparison Tool

Calculate pairwise sequence or structure alignments.

PDB 1:

Chain 1: ...

PDB 2:

Chain 2: ...

- Select Method -



Compare

[Align custom files](#) (Launches a Java Web Start application)

Documentation

Banques de structures 3D de protéine et leurs outils :

En général : chercher par code à 4 lettres/chiffres: ex: 2HBH

RCSB PDB

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

En haut de la page on trouve plusieurs onglets (Summary, Sequence, Annotations, ...):

1. Summary: Publication, Liste des chaines, ...
2. Sequence:
 - Séquence, structure 2D => bien pour voir les trous dans la structure 3D
 - Avec *Add Annotations* on peut ajouter des annotations de domaine de type CATH, SCOP, InterPro, ...
3. 3D Similarity => Comparison Tool (en bas à droite):
 - Permet de comparer deux structures de la PDB au niveau de la séquence et/ou la structure 3D
 - Sous *Align custom files* on peut aussi comparer ses propres fichiers PDB (ex: modèles par homologie) ou encore faire une recherche 3D dans toute la PDB.

A côté du code PDB en grand on trouve à droite deux menus pour afficher « Display Files » ou télécharger « Download Files » le fichier PDB (structure 3D, choisir « PDB File (Text) ») ou le fichier FASTA (séquence).

PiQSi, PISA: Bio ou cristallisé ?

Assemblage biologiques des oligomères:

Automatique:

PISA:

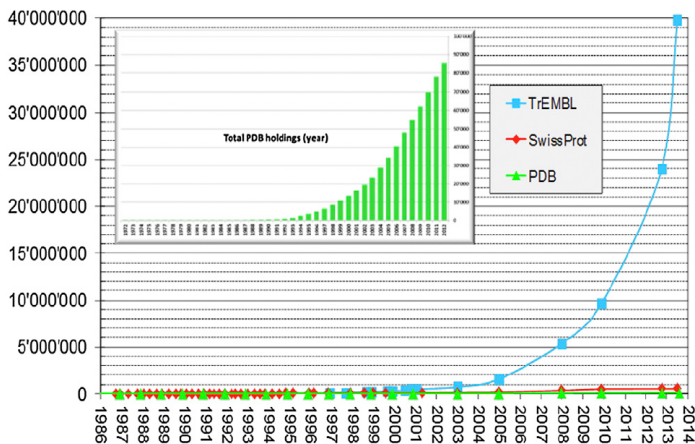
http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

Corrigé manuellement (que homomères):

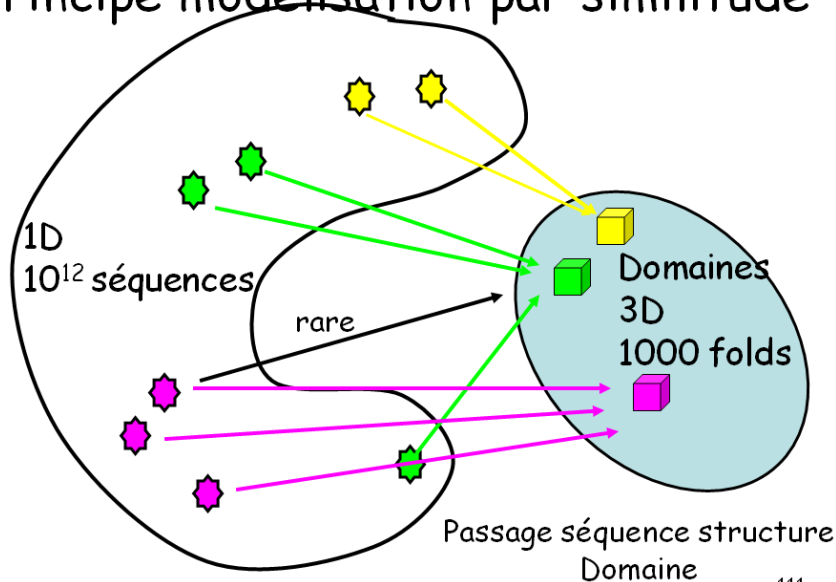
PiQSi: http://supfam.mrc-lmb.cam.ac.uk/elevy/piqsi/piqsi_home.cgi

Modèles par homologie

Protein structure gap



Principe modélisation par similitude



CMBI courses

<http://swift.cmbi.ru.nl/teach/B1SEM/>

The screenshot shows a web browser window with the title 'Bioinformatics Seminars'. The left sidebar contains a navigation menu with the following items: LAY-OUT, Introduction (with sublinks: The course, Dependencies, Goals, Logistics, Intro practicals, Molecular graphics), Homology Modelling (with sublinks: Intro, Video, Practical, Article), Validation (with sublinks: Intro, Video, Practical, Article), Force Field (with sublinks: Intro, Seminar, Practical 1, Practical 2), Drug Design (with sublinks: Intro, Seminar, Practical), Energy calculations, Miscellaneous (with sublinks: HELP, Exercise files, Seminars, Wiki). The main content area is titled 'Homology Modelling: Intro' and contains the following text:

After the Homology Modelling section you will:

- Be able to perform homology modelling using web-based servers;
- Understand the theoretical problems associated with Homology Modelling;

[The homology modelling seminar;](#)
[The homology modelling article;](#)

Homology Modelling is a technique to predict the structure of a protein from its sequence using the coordinates of a homologous protein with known structure.

We will explain homology modelling as an 8-step process. That is just a choice. Other people use three steps, very many steps, or even no steps at all.

It is nowadays sometimes also possible to predict the structure of a protein without the use of a homolog with known structure. This field is not yet developed far enough yet to teach about it because what we would teach today might be called old-stuff tomorrow.

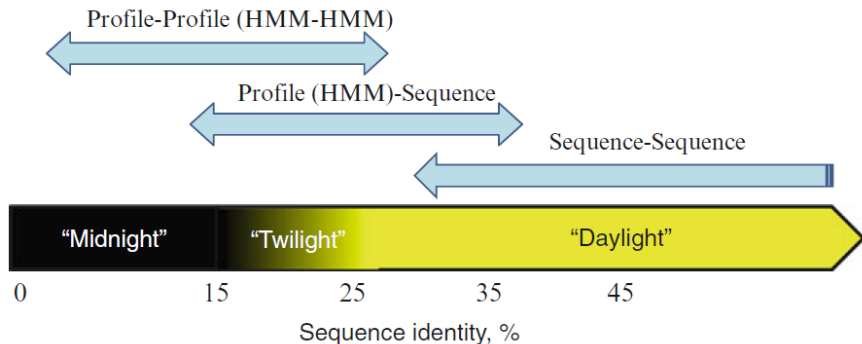
Please be aware that about every step in Homology Modelling includes Force Field computations. We will mention a few of them, but later during the course -after the Force Field seminar- the Homology Modelling seminar will be quickly repeated with the inclusion of many of these Force fields.

Navigation icons at the bottom right include a left arrow, a house icon, and a right arrow.

Article:

http://swift.cmbi.ru.nl/teach/B1SEM/HTML/hanka_modelling.pdf

Identité de séquence

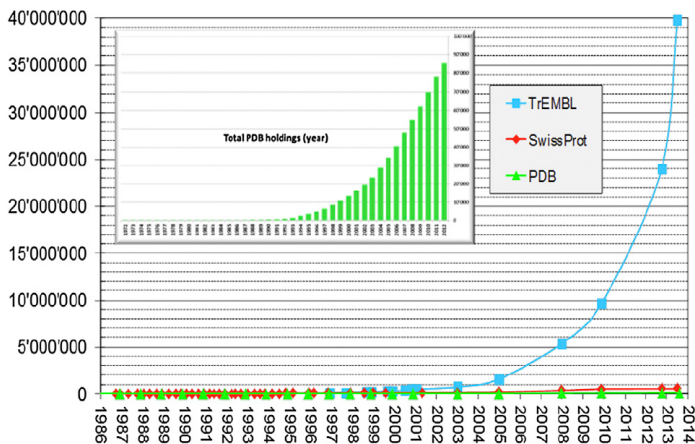


Venclovas, "Homology Modeling", ch. 3, Methods in Mol. Biol.(2012)

Application de la modélisation par homologie

- BLAST database: "Non-redundant protein sequences" ("nr"):
> 20 millions de séquences (29/09/2012)
<http://blast.ncbi.nlm.nih.gov>
- PDB database:
60091 structures 3D (29/09/2012)
- Combler l'écart grandissant : Modélisation par homologie

Protein structure gap



Génomique structurale

Problème: trop d'erreurs en modélisation pour $< 30\%$ identité de séquence

Solution: résoudre expérimentalement les structures 3D d'un sous-ensemble des séquences bien sélectionnées pour que chaque séquence a au moins un template structural avec une identité de séquence $> 30\%$

=> Besoin de > 16000 structures

Ref: Brenner, 2000, Nat. Struct. Bio. Vitkup, 2001, Nat. Struct. Bio.

Structural coverage

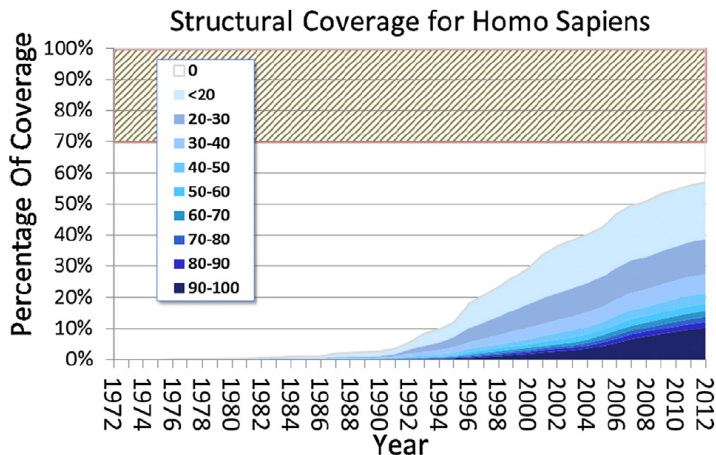


Table 1. Commonly Used Tools and Services for Protein Structure Modeling and Prediction

Tool or Service	Web Site
Protein Model Portal	http://www.proteinmodelportal.org (Arnold et al., 2009; Haas et al., 2013)
Model Archive	http://modelarchive.org
HHpred	http://toolkit.tuebingen.mpg.de/hhpred (Hildebrand et al., 2009)
IMP	http://www.salilab.org/imp (Russel et al., 2012; Yang et al., 2012)
IntFOLD	http://www.reading.ac.uk/bioinf/IntFOLD/ (Roche et al., 2011)
I-Tasser	http://zhanglab.cmb.med.umich.edu/I-TASSER/ (Zhang, 2013)
ModBase	http://salilab.org/modbase/ (Pieper et al., 2011)
Modeler/ModWeb	http://salilab.org/modeller/ (Pieper et al., 2011; Yang et al., 2012)
Pcons.net	http://pcons.net/ (Larsson et al., 2011)
PHYRE2	http://www.sbg.bio.ic.ac.uk/phyre2/ (Kelley and Sternberg, 2009)
Robetta	http://rosetta.bakerlab.org/ (Raman et al., 2009)
Rosetta	https://www.rosettacommons.org (Das and Baker, 2008)
SWISS-MODEL Repository	http://swissmodel.expasy.org/repository (Kiefer et al., 2009)
SWISS-MODEL Workspace	http://swissmodel.expasy.org/workspace/ (Arnold et al., 2006; Bordoli and Schwede, 2012)

SWISS-MODEL

<http://swissmodel.expasy.org/>

- Web-serveur automatisé pour la modélisation par homologie
- Existe depuis 20 ans
- Dernière version: SWISS-MODEL Workspace (2014)
- Accessible pour les non-experts en bioinfo
- Automatisation permet une meilleure reproductibilité
- Sur mon site il y a un énoncé de TP sur SWISS-MODEL (qu'on ne fera pas cette année)

Meilleures méthodes de validation d'après CASP 10

- IntFOLD2: <http://www.reading.ac.uk/bioinf/IntFOLD/>
- ProQ2clust2 (pas encore disponible)
- ProQ2: <http://www.bioinfo.ifm.liu.se/ProQ2/>

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

- But: Promouvoir l'utilisation efficace de modèles moléculaires en recherche biomédicale
- Combine la PDB + modèles 3D théoriques de plusieurs sources
- Interface pour générer des modèles 3D à partir de plusieurs serveurs en parallèle:
ModWeb, M4T, SWISS-MODEL, I-TASSER, HHpred
- Validation en parallèle: ModEval, ModFOLD, QMEAN

Integrative Modeling Platform (IMP)

Nuclear Pore Complex (NPC) - 50 MDa, 456 protéines

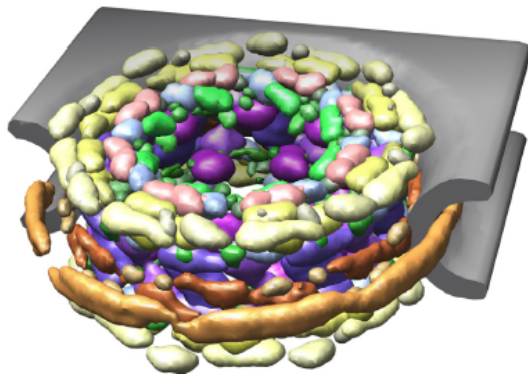


Figure 4. Integrative Structure Model of the NPC

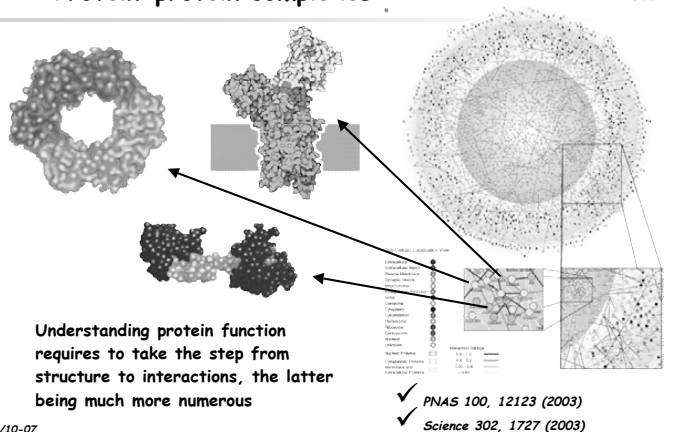
The molecular architecture of the approximately 50 MDa transmembrane NPC

- 7 Docking protéine-protéine
- Interactions protéine-protéine
 - Docking with experimental data
 - Docking with NMR data: CS-HADDOCK

Interactions protéine-protéine

Protein function

Protein-protein complexes



AB/10-07

Free proteins - Structural genomics

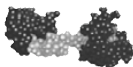
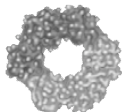
- 3D structure of a large number of unbound/free proteins solved => PDB
- Only about 1000 types of folds, almost all known.
- => Comparative modeling / Homology modeling

Protein-protein complexes

- Number of types of protein-protein interactions at least 10x times greater (> 10.000) than number of folds (1000).
- Experimental difficulties to solve protein-protein 3D structures.

Models of Protein Complexes

What can we learn from 3D structures
(models) of complexes?

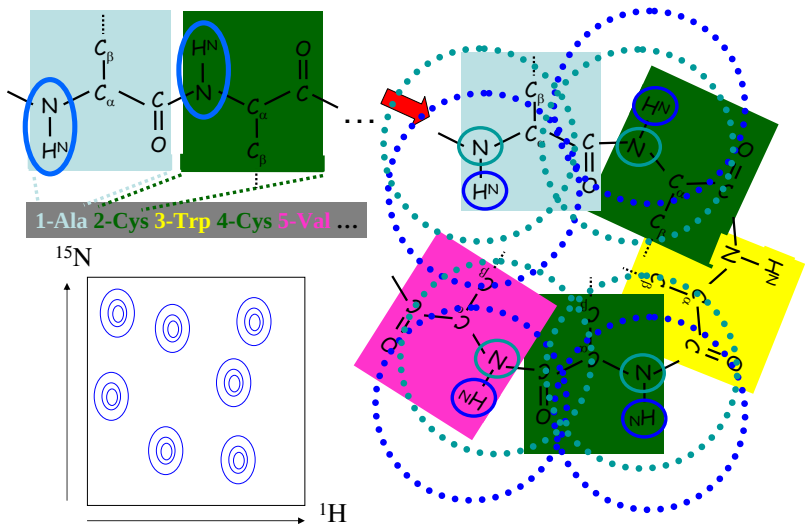


- Models provide structural insight into function and mechanism of action
- Models can drive and guide experimental studies
- Models can help understand and rationalize the effect of disease-related mutations
- Models provide a starting point for drug design

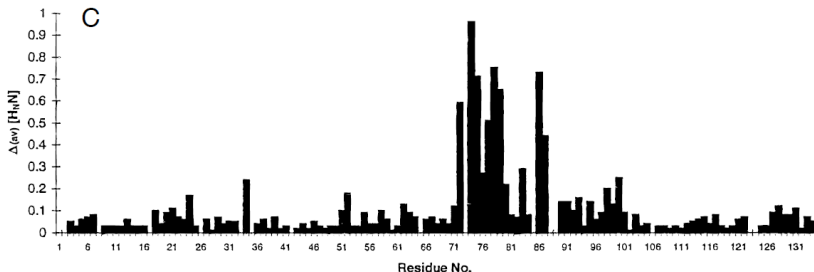
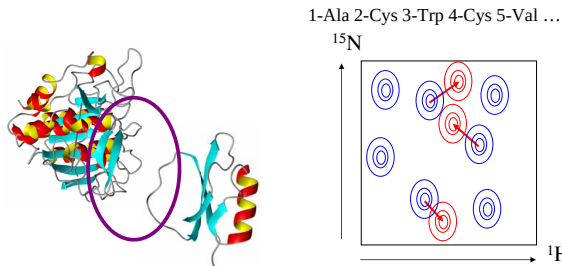
AB/10-07

Docking with experimental data

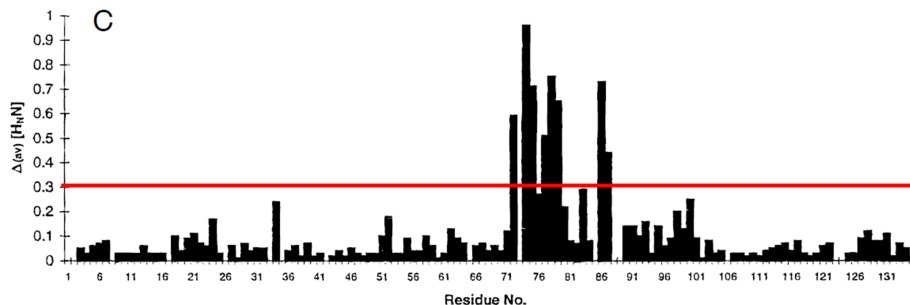
Chemical shift



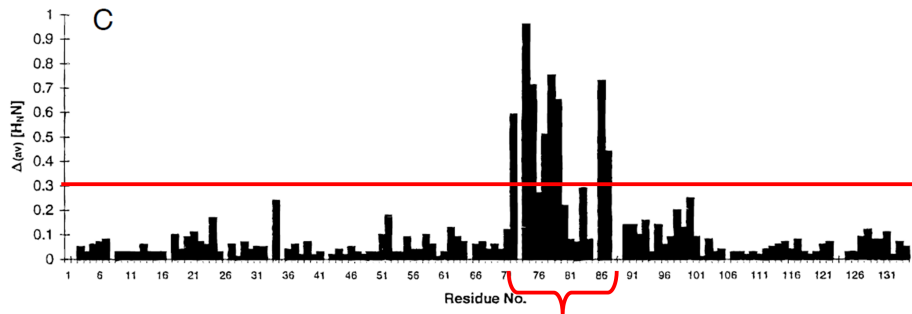
Chemical Shift Perturbation (CSP)



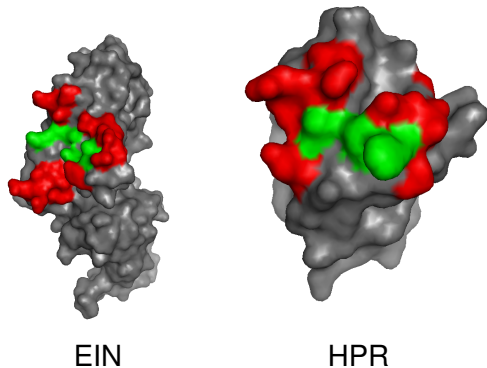
Chemical Shift Perturbation (CSP)



Chemical Shift Perturbation (CSP)

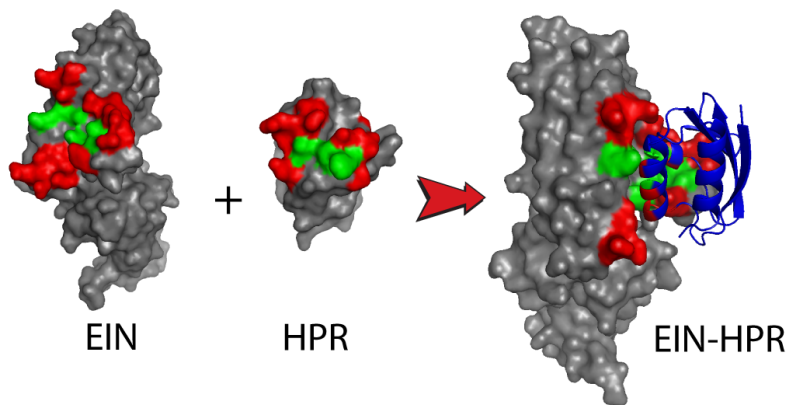


Interface localization on 3D structures



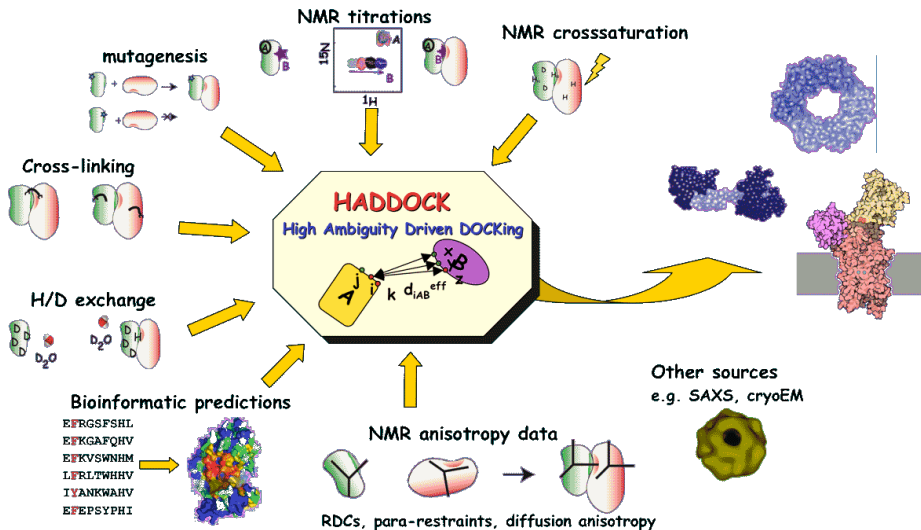
red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

Docking

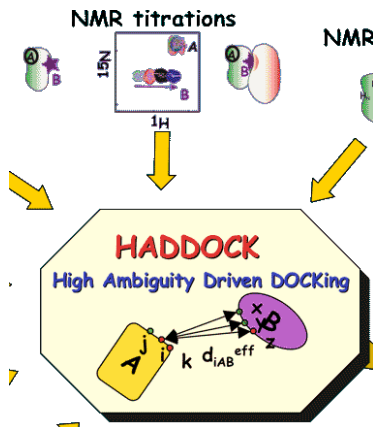


red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

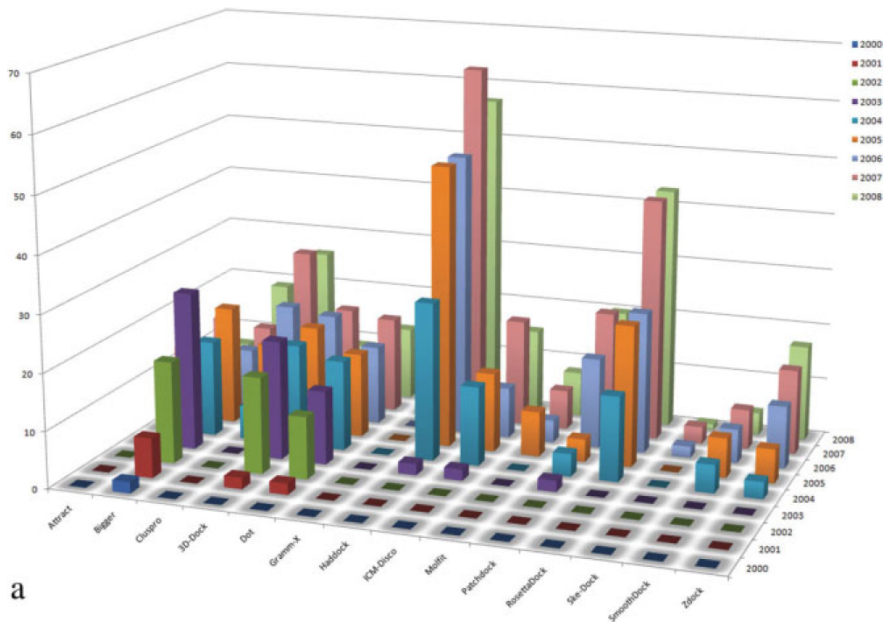
Haddock - <http://haddock.chim.uu.nl>



Haddock - <http://haddock.chem.uu.nl>



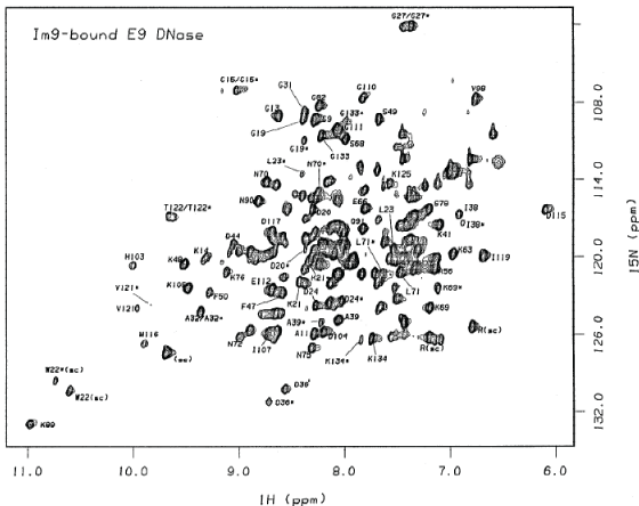
$$E_{Haddock} = E_{vdW} + E_{elec} + E_{AIR} + E_{desolv}$$

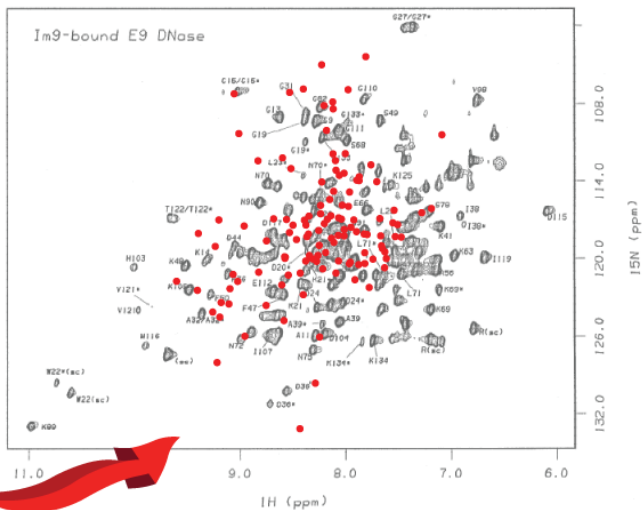
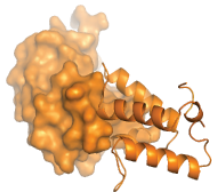


a

Docking with NMR data: CS-HADDOCK

3D to CS





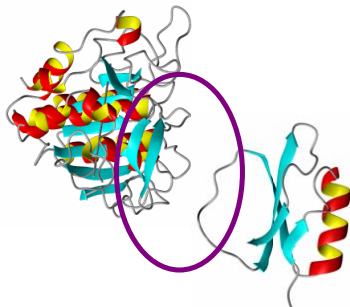
CS-predictor

3D to CS with ShiftX

Contributions to calculated CS δ_{calc} :

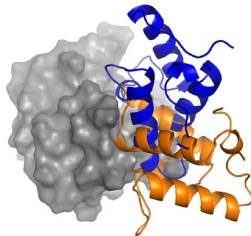
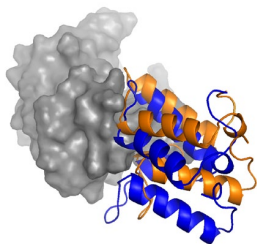
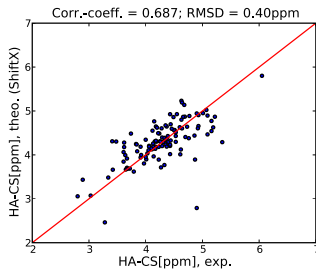
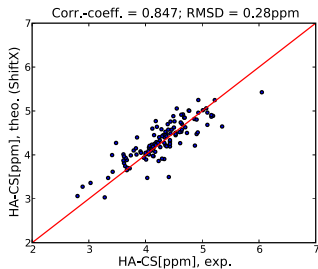
$$\delta_{calc} = \delta_{coil} + \delta_{RC} + \delta_{EF} + \delta_{HB} + \delta_{HS}$$

- δ_{coil} - random coil (amino acid type)
- δ_{RC} - ring current
- δ_{EF} - electric field
- δ_{HB} - hydrogen bonding
- δ_{HS} - empirical hypersurfaces (backbone dihedral angles)

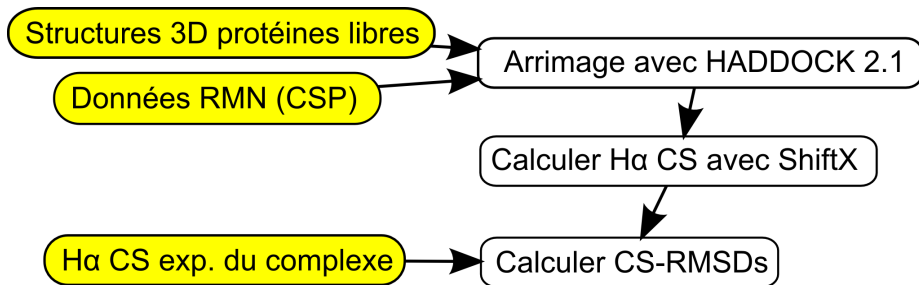


Neal et al., *J. Biomol. NMR* 26: 215-240, 2003

RMSD between δ_{calc} and δ_{exp} for $^1H^\alpha$ -CS



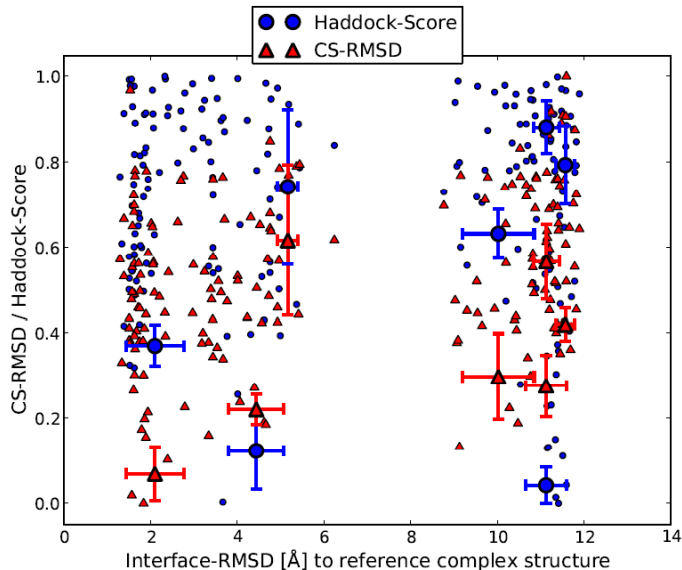
Protocole d'arrimage CS-HADDOCK



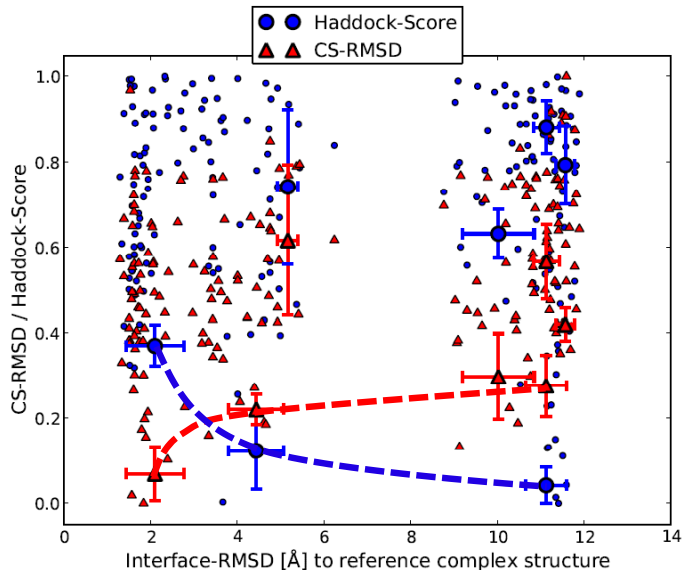
CS-RMSD =

$$\frac{\sqrt{\frac{\sum_{i=1}^{n_A} (\delta_i^{exp} - \delta_i^{theo})^2}{n_A}} + \sqrt{\frac{\sum_{i=1}^{n_B} (\delta_i^{exp} - \delta_i^{theo})^2}{n_B}}}{2}$$

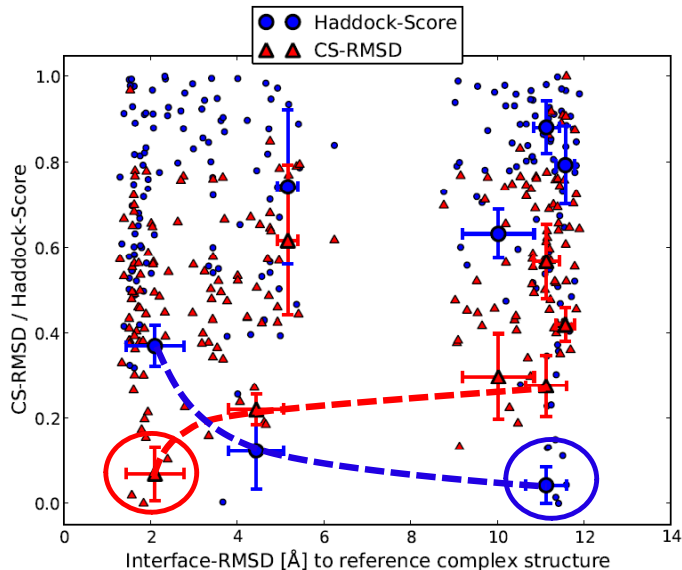
CS-HADDOCK vs HADDOCK



CS-HADDOCK vs HADDOCK

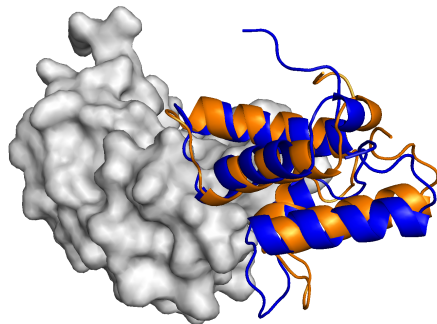


CS-HADDOCK vs HADDOCK

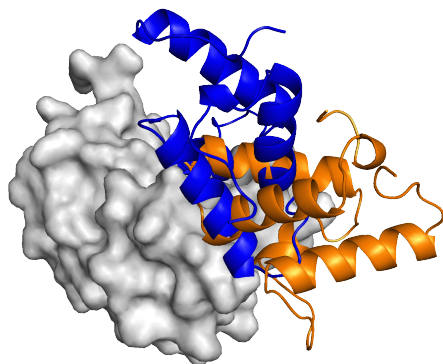


CS-HADDOCK vs HADDOCK

Meilleure structure (en bleu) par rapport à la référence (en orange):



(a) CS-RMSD score



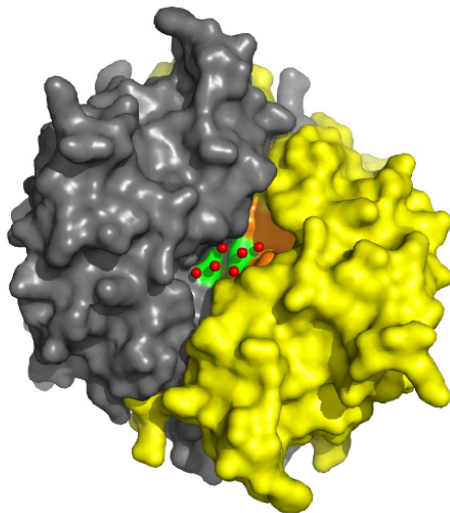
(b) HADDOCK score

8 Ma thématique de recherche actuelle

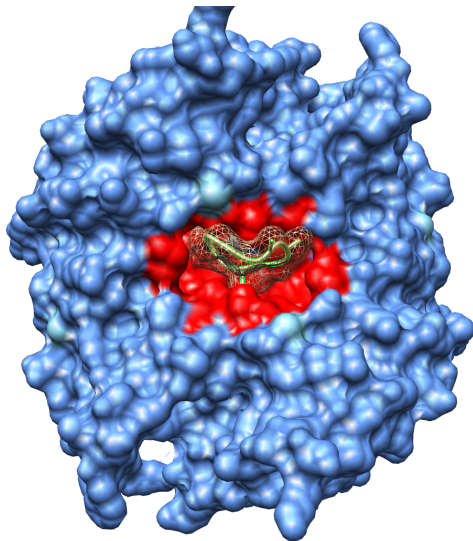
- Peptides cycliques ciblant des interactions protéine-protéine

Peptides cycliques ciblant des interactions protéine-protéine

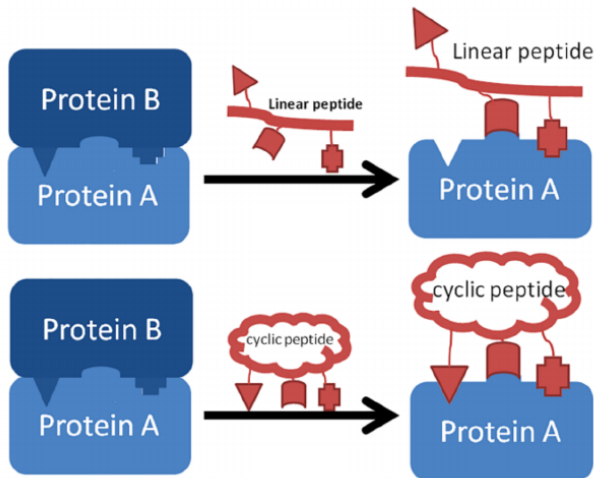
Drug design: cibler les interactions protéine-protéine



Drug = peptide cyclique

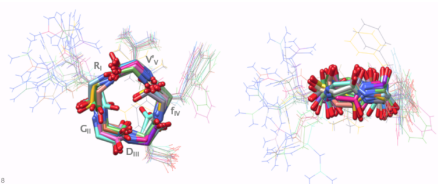
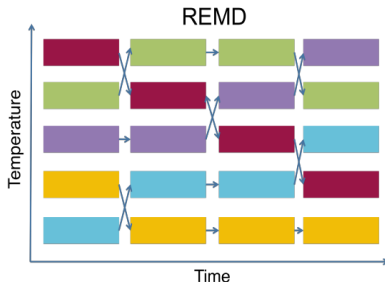
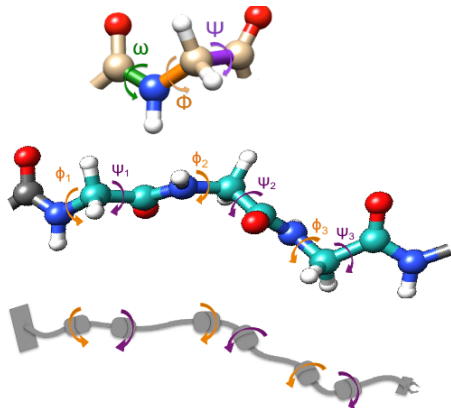


Drug = peptide cyclique



Peptide backbone as a robotic manipulator

Replica-exchange molecular dynamics (REMD)



Jusot, M., Stratmann, D., Vaisset, M., Chomilier, J., and Cortés, J., Exhaustive exploration of the conformational landscape of small cyclic peptides using a robotics approach. *Journal of Chemical Information and Modeling* 2018 58 (11), 2355-2368

<https://doi.org/10.1021/acs.jcim.8b00375>

The end

- MERCI pour votre attention!

9 Bibliography



Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008). The Protein Folding Problem. *Annual Review of Biophysics*, 37(1):289–316.



Schaeffer, R. D. and Daggett, V. (2011). Protein folds and protein folding. *Protein Engineering Design and Selection*, 24(1-2):11–19.



Schaeffer, R. D., Jonsson, A. L., Simms, A. M., and Daggett, V. (2011). Generation of a consensus protein domain dictionary. *Bioinformatics*, 27(1):46–54.



Schwede, T. (2013). Protein Modeling: What Happened to the “Protein Structure Gap”? *Structure*, 21(9):1531–1540.



Searle, M. S. and Ciani, B. (2004). Design of beta-sheet systems for understanding the thermodynamics and kinetics of protein folding. *Current Opinion in Structural Biology*, 14(4):458–464.



Stratmann, D., Boelens, R., and Bonvin, A. M. J. J. (2011). Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins*, 79(9):2662–2670.
PMID: 21744392.