

Structure et repliement des protéines

Dirk Stratmann (dirk.stratmann@upmc.fr)

<http://www.imPMC.upmc.fr/~stratmann/>

IMPMC, Université Pierre et Marie Curie (Paris VI)

septembre 2016

Plan

- 1 Structures tertiaires
- 2 Structures quartanaires
- 3 Domaines
- 4 Outils
- 5 Docking protéine-protéine
- 6 Bibliography

1

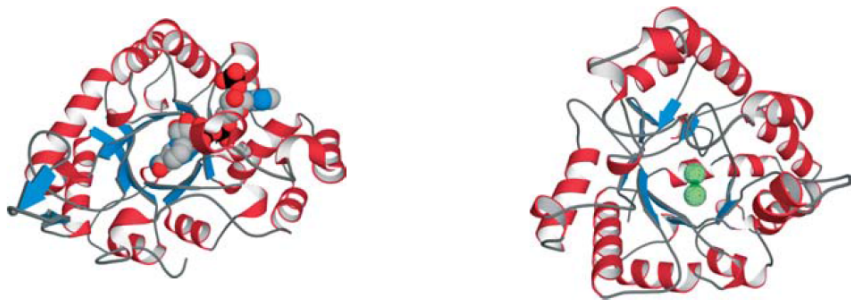
Structures tertiaires

- Structures tertiaires
- Structure $\leftrightarrow$ Fonction

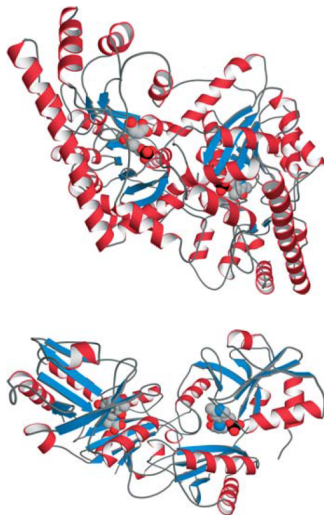
Structures tertiaires

Structure $\leftrightarrow$ Fonction

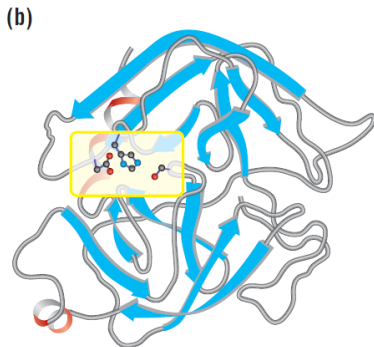
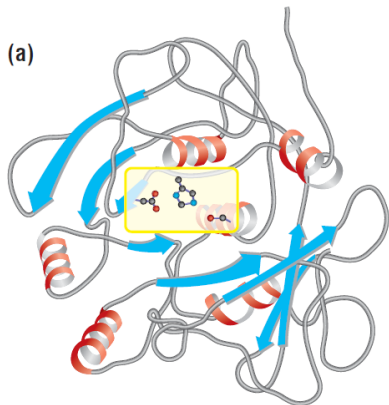
Structure 3D similaire, mais fonction différent



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

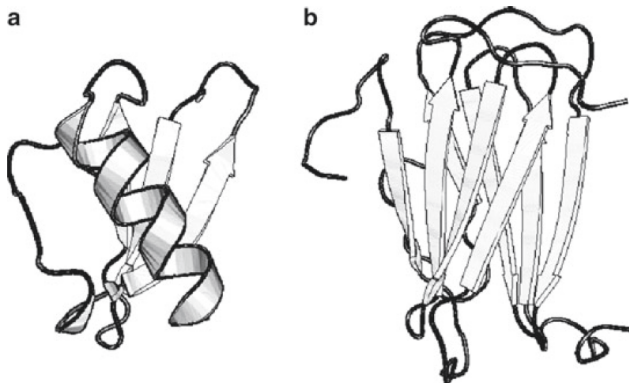


Fig. 5. The structures of two alternative folds of lymphotactin (Ltn10). **(a)** Monomeric Ltn10 (pdb 1j8i) and **(b)** dimeric Ltn10 (pdb 2jp1).

Dogmes de la biologie structurale

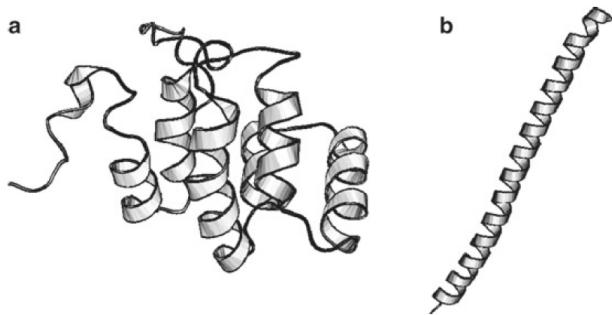


Fig. 6. The death domain of protein kinase Pelle (Pelle-DD) (a) solution structure, (b) crystal structure in MPD.

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

Dogmes de la biologie structurale

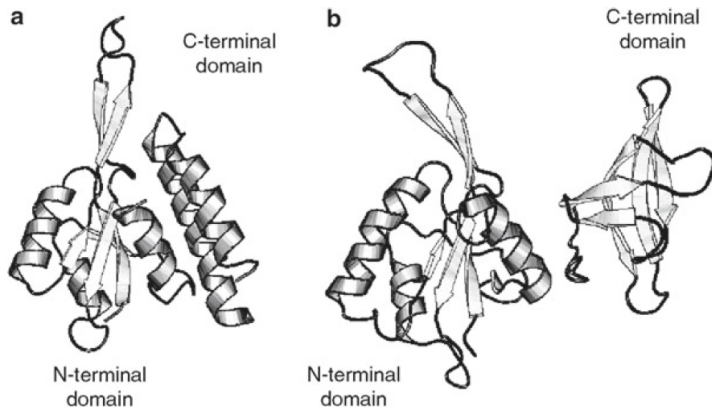


Fig. 10. Fold transition. Structures of (a) RfaH and (b) NusG.

- 2 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

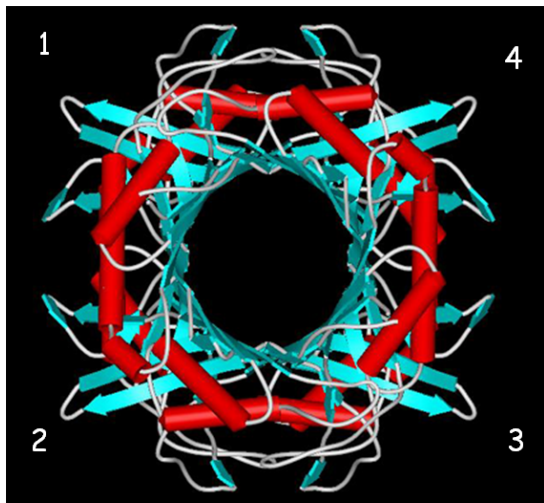
Très souvent, un domaine peut interagir avec plusieurs sortes de domaines, par l'intermédiaire de différentes surfaces d'interaction, tel que c'est le cas pour le domaine G Ras-like

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

Urate Oxydase



monomère

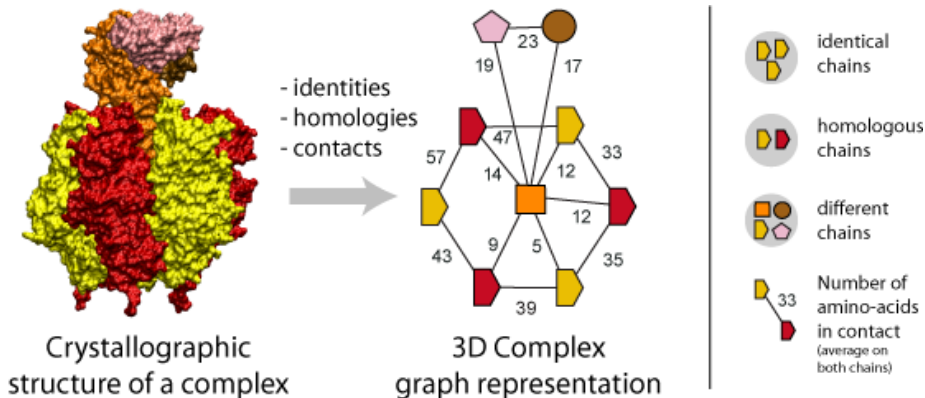


tétramère

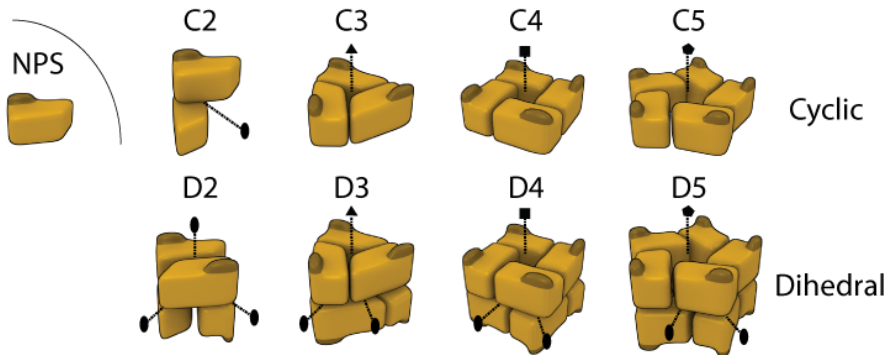
3D complex

- 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



3

Domaines

- Domaines
- Annotations des séquences des protéines
- Classification structurale

Domaines

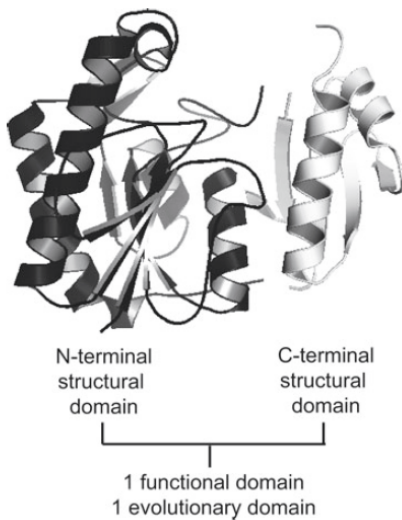
Domaines

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

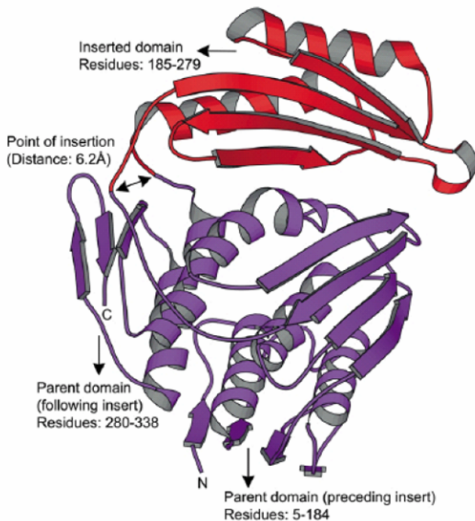


Code PDB : 1fnh. 3 domaines immunos

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



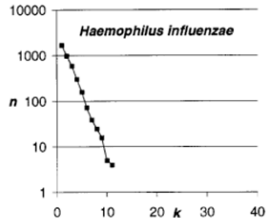
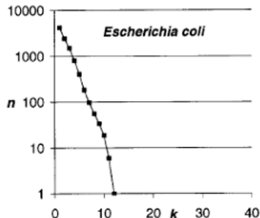
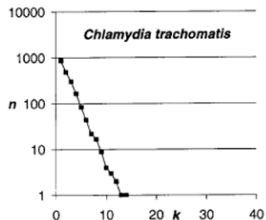
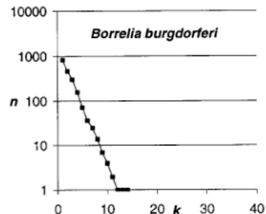
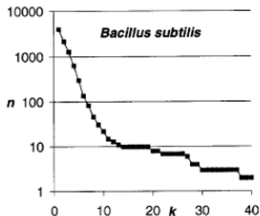
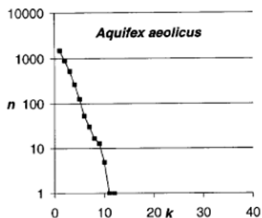
Protéines multidomaines

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

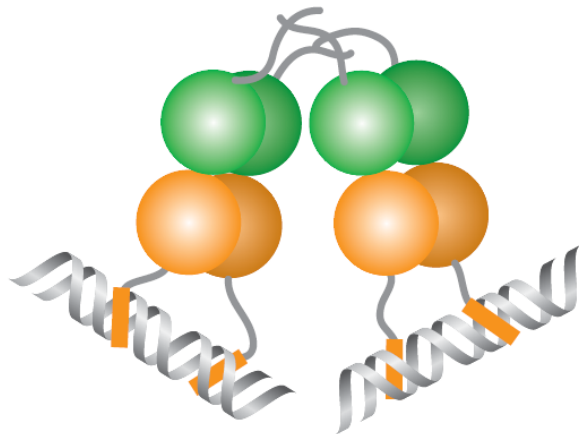
On ne peut que modéliser un domaine, pas toute une chaîne

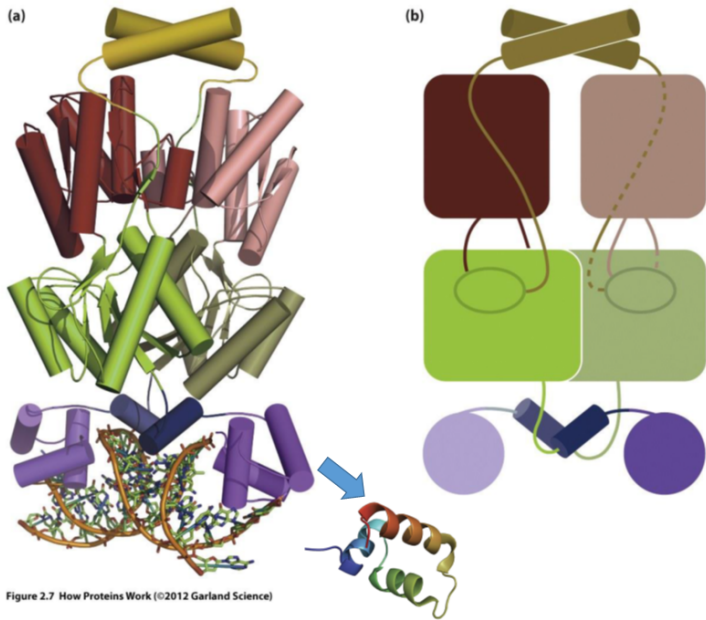
Protéines multidomaines



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

Exemple: lac repressor

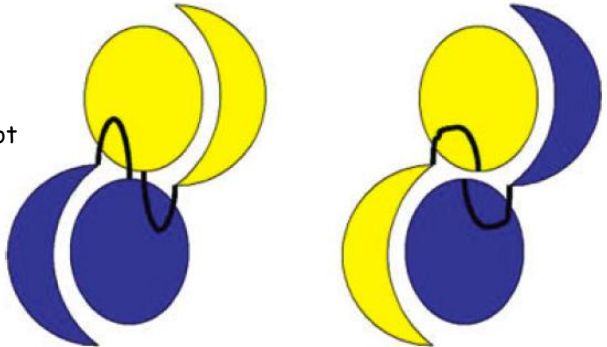




Domain Swapping

- 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.

Search

Go!

-  PDB Code
-  CATH Code
-  General Text

Navigation

[Home](#)
[Top of hierarchy](#)

[Home](#) > [Top](#) > [Class1](#) > [10](#) > [210](#) > [10](#) > [1](#) > [1](#) > [1](#) > [1utg00](#) [View this page as XML](#)

1utg00

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 Code	1utg
PDB Header	Uteroglobulin (oxidized)
PDB Source	Rabbit (<i>Oryctolagus cuniculus</i>) female genital tract

Domain Sequence	GICPRFAHVLENLLGLTPSSYETSLKEFEPDDTMKDAGM QMKKVLDSLPTTRENNIMKLEKIVKSPCLM
-----------------	---

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

Class
3 major classes

Mainly alpha

Alpha-beta

Mainly beta

Architecture
32 architectures

Barre

3-layer sandwich

2-layer sandwich

T

Topology or fold
~820 fold groups



~70 other
fold groups

H

**Homologous
superfamily**
-1400 superfamilies



Lactate Dehydrogenase
(9ldtA, domain 1)

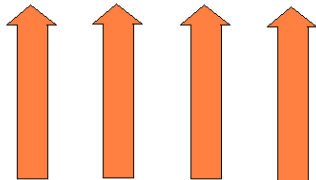
Flavodoxin
(2fox0, domain 1)

Sequence family
(35%)
~4000 families

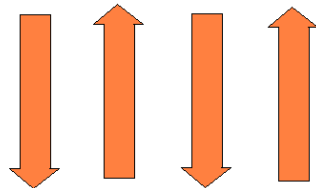


Examples of protein architecture

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



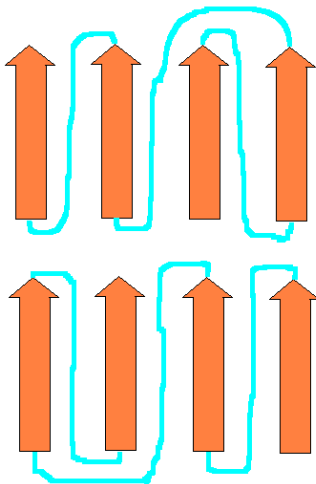
β -sheet with all pairs of strands parallel



β -sheet with all pairs of strands anti-parallel

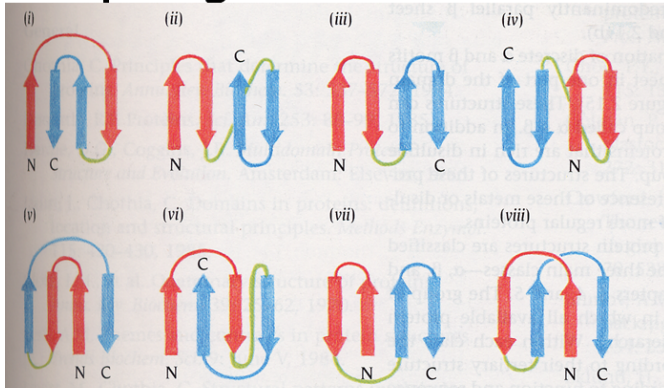
Examples of protein topology

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



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

Topologies des feuilletts 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

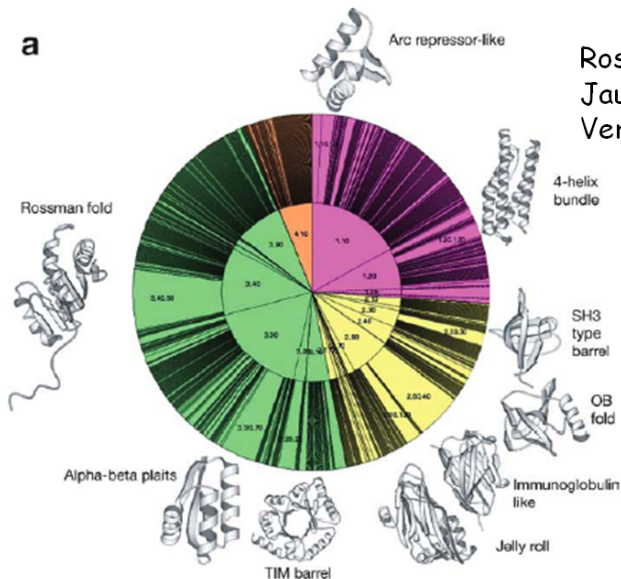
CLASSES

- **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

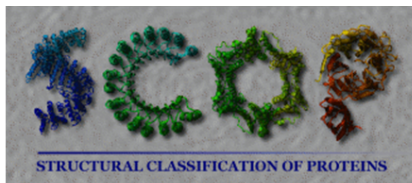
Rose : Alpha
Jaune : Beta
Vert : Alpha-beta



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.

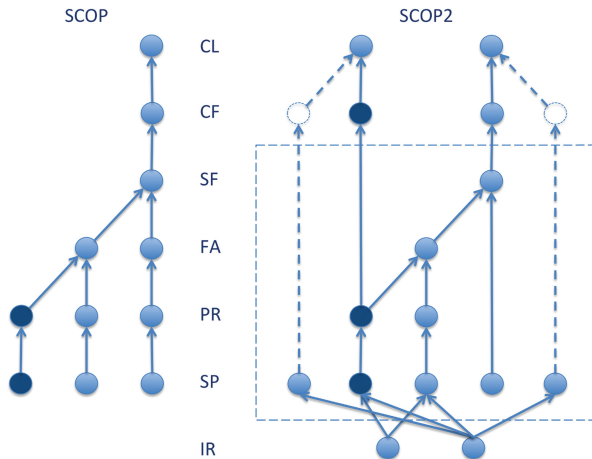
SCOP



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

SCOP 2

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



SCOP 2

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

The screenshot shows the SCOP 2 website interface with several annotations:

- Alternative ways for database access:** Points to the top navigation bar containing links: About, Browser, Graph, Download, Support.
- Main navigation menu:** Points to the same top navigation bar.
- Action buttons:** Points to the 'Configure', 'Share', 'Print', and 'Help' buttons located below the navigation bar.
- Navigation tabs:** Points to the tabs in the pop-up window: Structure, Parents, Children, Comments.
- Links to external sequence and structural resources:** Points to links in the pop-up window: View in PDB: SCRD, View in PDB: SCRD, View in UniProt: P03349.
- Click to bring up the node pop up window:** Points to the 'Fold change in the lambda C1 repressor DNA binding domain-like superfamily' node in the main graph.

The main graph displays a hierarchical structure of protein families, including nodes like SCOP2, Protein relationships, Protein type, Structural relationships, Structural class, Evolutionary events, Evolutionary relationships, common fold, All alpha proteins, Globular proteins, Alpha and beta proteins (a+b), Fold change in the lambda C1 repressor DNA binding domain-like superfamily, Family, Superfamily, POU domain-like, Lambda cro protein-like, Species, Protein, XRE-like regulatory proteins, Regulatory protein cro of the p52 fold-type, Regulatory protein cro of the lambda fold-type, motif, foldogen, BPP22_cro, Lambda_cro, Helix-Turn-Helix (HTH) motif, and Common subfold of the different types of the cro folds.

The pop-up window for the 'Fold change in the lambda C1 repressor DNA binding domain-like superfamily' node shows the following details:

- Domain:** IR-8033976-SCRD
- Comment:** complexed with p04
- Segments:** 6-36.A
- View in PDB:** SCRD
- View in PDB:** SCRD
- View in UniProt:** P03349
- Representative Sequence:** 66 residues. SCOP domains: 33 residues.

`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.

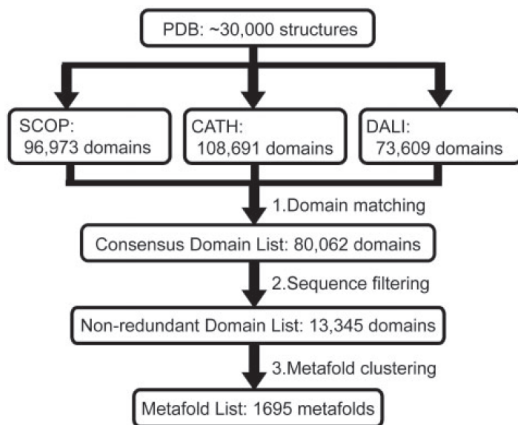
- Please note that PDB structures released after the last update will not be in the database! If you wish to find structural neighbours of these proteins, you are advised to submit the structure to the [Dali Server](#) instead.
- If you want to superimpose two particular structures, you can do it in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

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.¹, Stayrook, S.², Rosenberg, A.³, Lewis, M.³

Journal: (2007) J.Mol.Biol. 370: 609-619

PubMed: 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

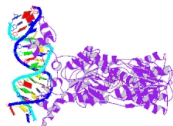
Classification: Transcription/dna Structure Weight: 125535.28 

Molecule: DNA
 (5'-D>(*DAP*DAP*DTP*DTP*DGP*DGP*DAP*DGP*DCP*DGP*DCP*DTP*DCP*DAP*DCP*DAP*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

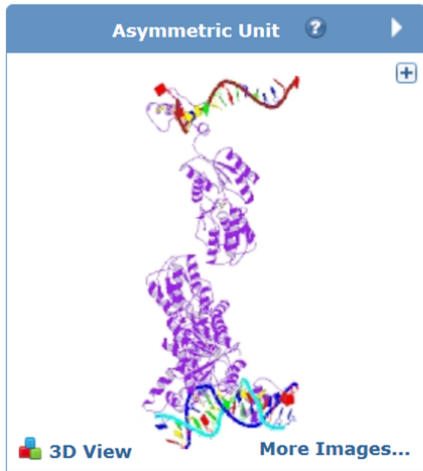
Biological assembly 1 assigned by authors and generated by PISA (software)

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:

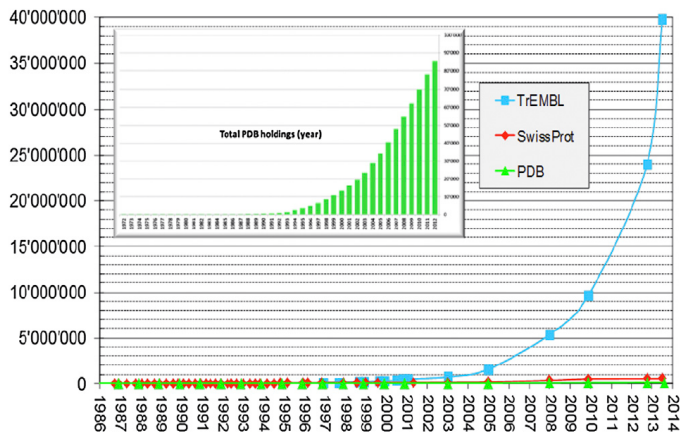
http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

Corrigé manuellement (que homomères):

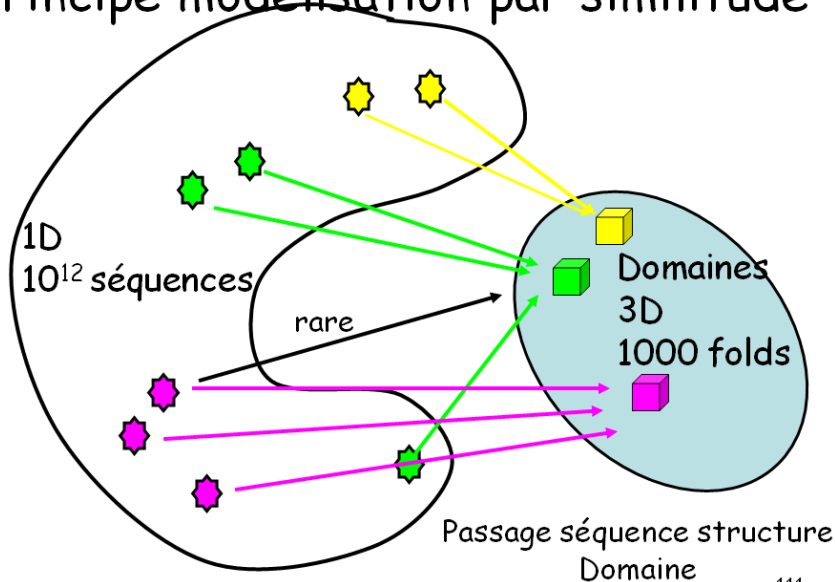
PiQSi: http://supfam.mrc-lmb.cam.ac.uk/elevy/piqsi/piqsi_home.cgi

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 displaying the CMBI Bioinformatics Seminars website. The page has a header with the CMBI logo and the title 'Bioinformatics Seminars'. A left sidebar contains a navigation menu with categories like 'Lay-out', 'Introduction', 'Homology Modelling', 'Validation', 'Force Field', 'Drug Design', 'Enzyme calculations', and 'Miscellaneous'. The main content area is titled 'Homology Modelling: Intro' and contains text about the course objectives, the technique of homology modelling, and a note about Force Field computations. At the bottom right of the main content area, there is a small icon of a house with arrows pointing left and right.

LAY-OUT
Introduction
[The course](#)
[Dependencies](#)
[Goals](#)
[Logistics](#)
[Intro practicals](#)
[Molecular graphics](#)
Homology Modelling
[Intro](#)
[Video](#)
[Practical](#)
[Article](#)
Validation
[Intro](#)
[Video](#)
[Practical](#)
[Article](#)
Force Field
[Intro](#)
[Seminar](#)
[Practical 1](#)
[Practical 2](#)
Drug Design
[Intro](#)
[Seminar](#)
[Practical](#)
Enzyme calculations
Miscellaneous
[HELP](#)
[Exercise files](#)
[Seminars](#)
[Wiki](#)

Bioinformatics Seminars

Homology Modelling: Intro

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.

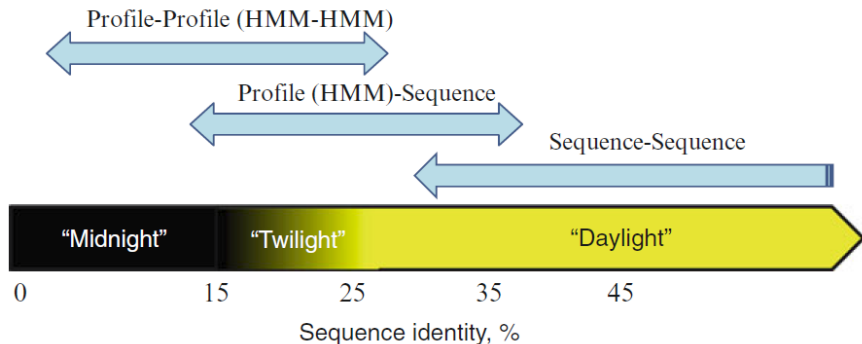
It is nowadays sometimes also possible to predict the structure of a protein without the use of a homolog with known structure. This field is not yet developed far enough yet to teach about it because what we would teach today might be called old-stuff tomorrow.

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.

Article:

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

Identité de séquence

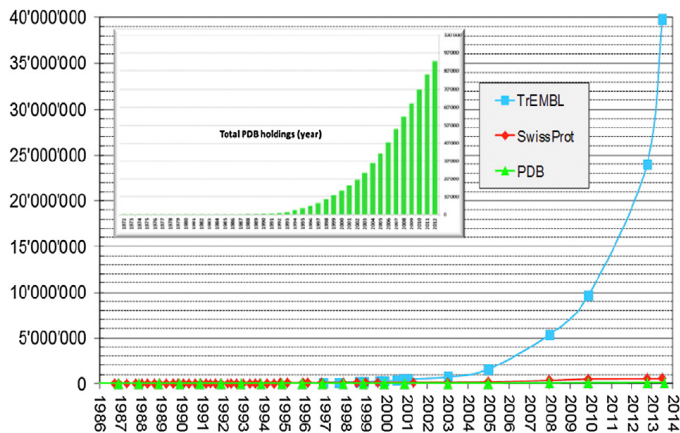


Venclovas, "Homology Modeling", ch. 3, Methods in Mol. Biol.(2012)

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)
- Comblent 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

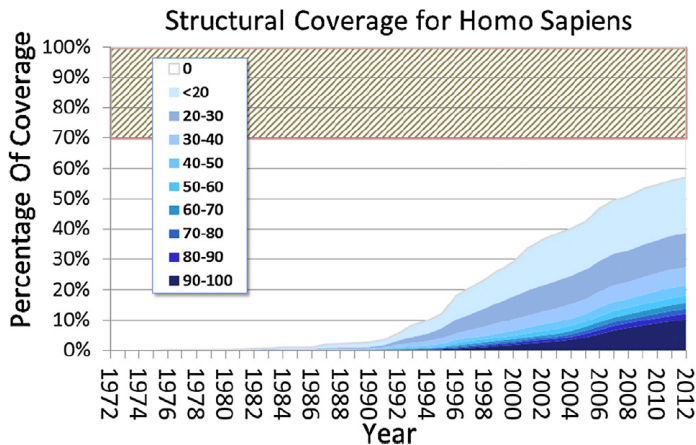


Table 1. Commonly Used Tools and Services for Protein Structure Modeling and Prediction

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.ccmb.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://robetta.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

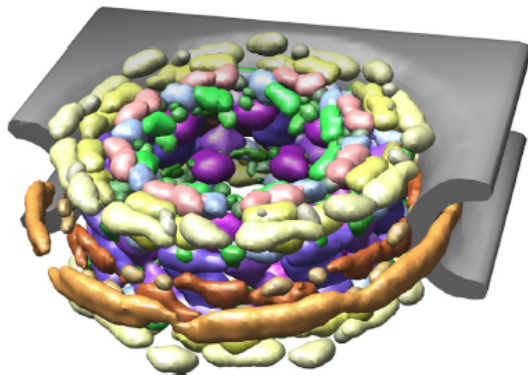


Figure 4. Integrative Structure Model of the NPC

The molecular architecture of the approximately 50 MDa transmembrane NPC

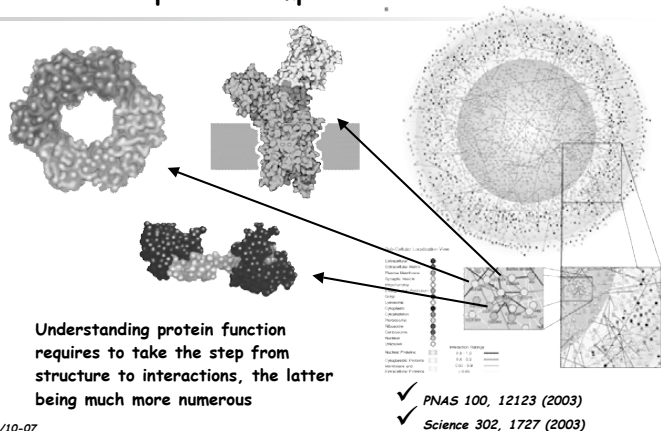
5 Docking protéine-protéine

- 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

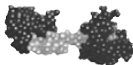
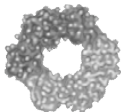
- 3D structure of a large number of unbound/free proteins solved => PDB
- Only about 1000 types of folds, almost all known.
- => Comparative modeling / Homology modeling

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?**

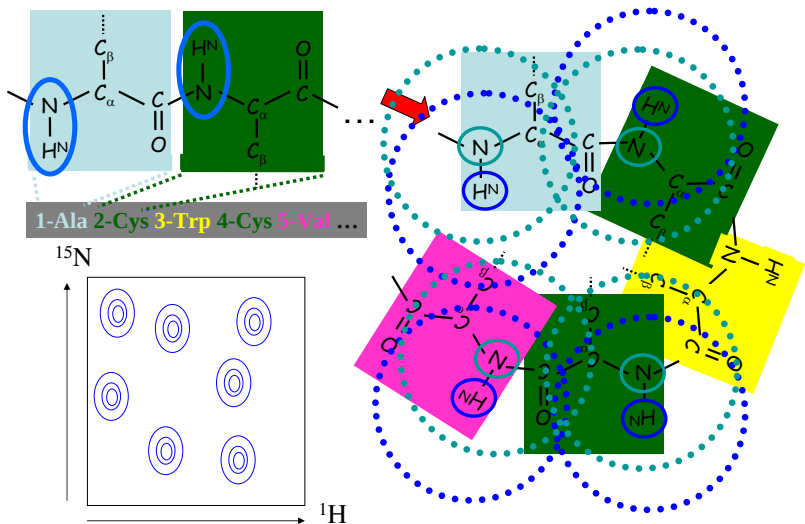


- Models provide structural insight into function and mechanism of action
- Models can drive and guide experimental studies
- Models can help understand and rationalize the effect of disease-related mutations
- Models provide a starting point for drug design

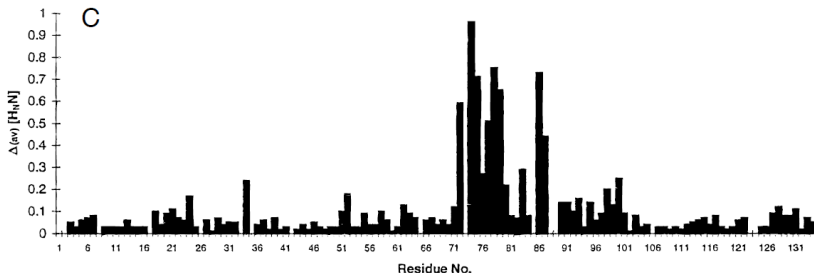
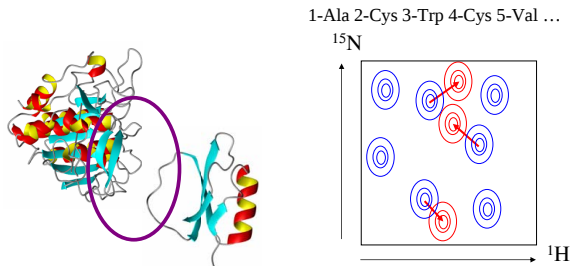
AB/10-07

Docking with experimental data

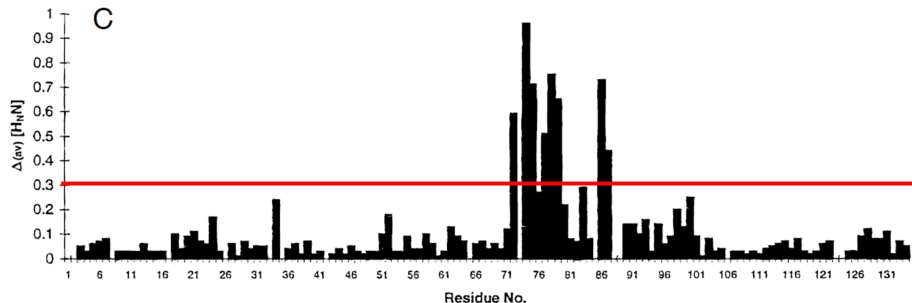
Chemical shift



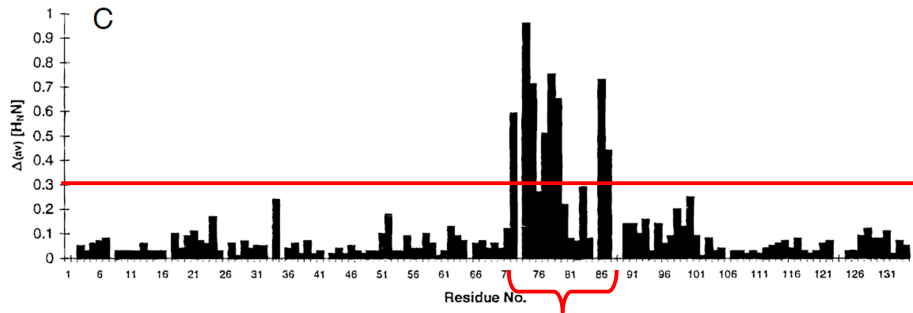
Chemical Shift Perturbation (CSP)



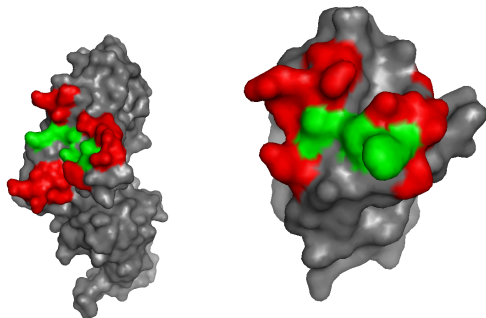
Chemical Shift Perturbation (CSP)



Chemical Shift Perturbation (CSP)



Interface localization on 3D structures

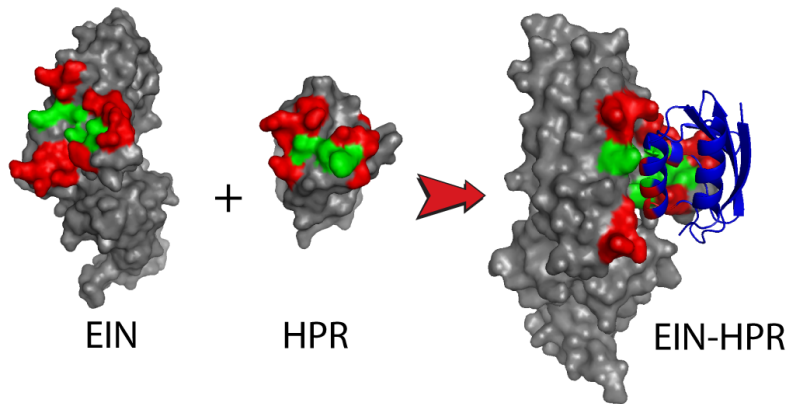


EIN

HPR

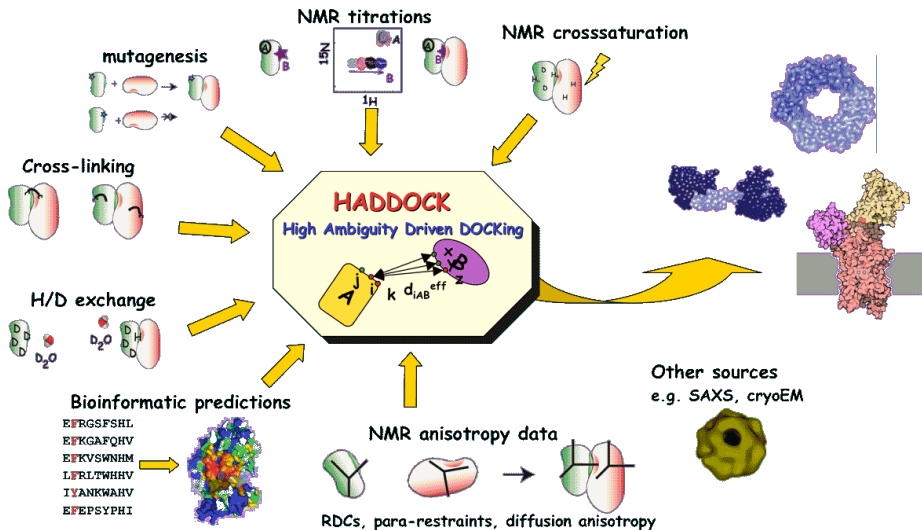
red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

Docking

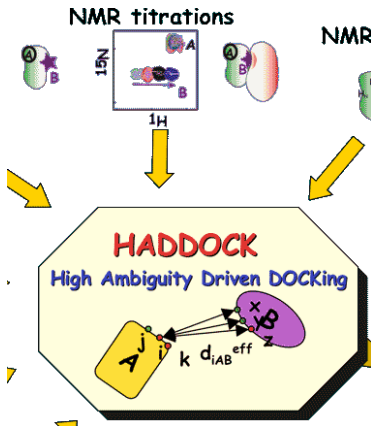


red = active residues derived from CSP data and surface accessibility
green = passive residues, i.e. the surface neighbors of the active residues

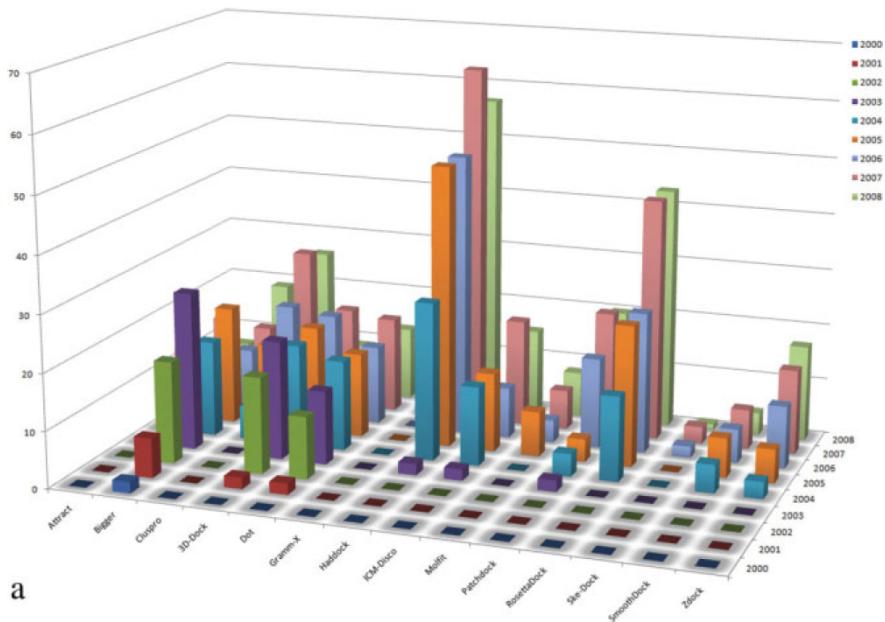
Haddock - <http://haddock.chem.uu.nl>



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

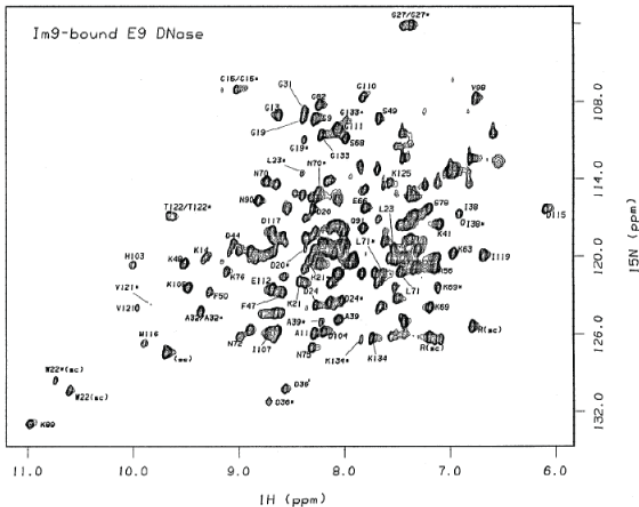


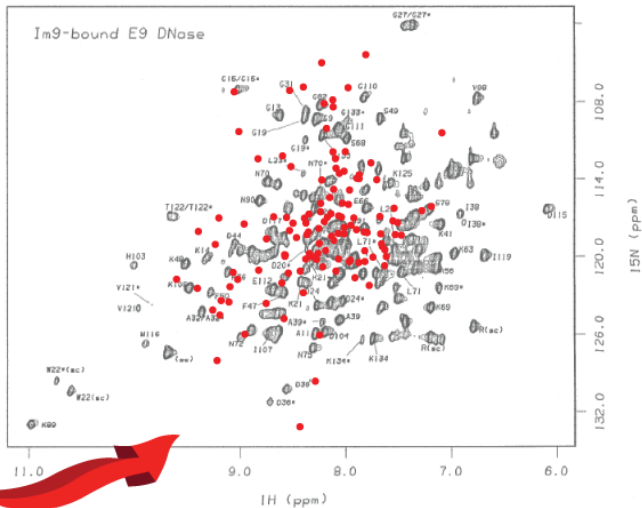
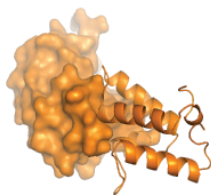
$$E_{Haddock} = E_{vdW} + E_{elec} + E_{AIR} + E_{desolv}$$



a

Docking with NMR data: CS-HADDOCK





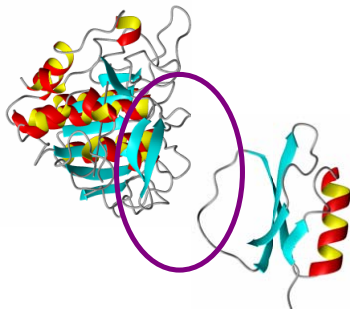
CS-predictor

3D to CS with ShiftX

Contributions to calculated CS δ_{calc} :

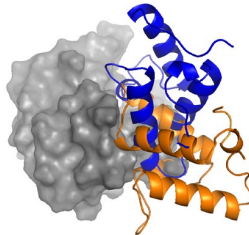
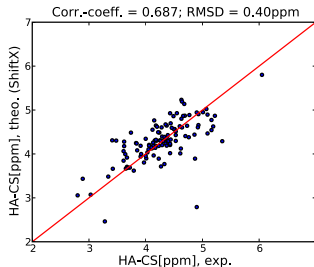
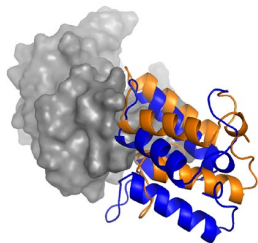
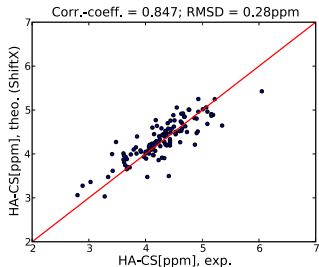
$$\delta_{calc} = \delta_{coil} + \delta_{RC} + \delta_{EF} + \delta_{HB} + \delta_{HS}$$

- 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)

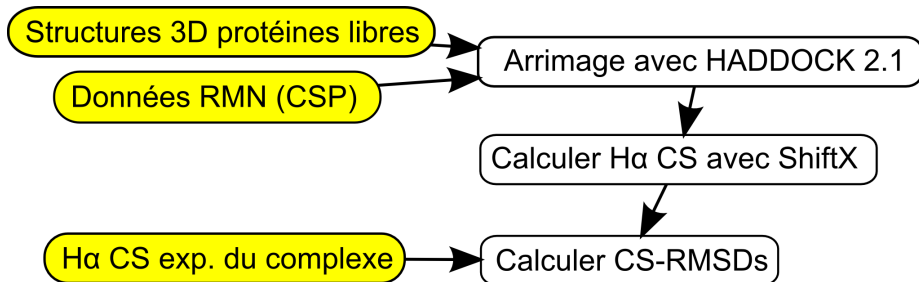


Neal et al., *J. Biomol. NMR* 26: 215-240, 2003

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



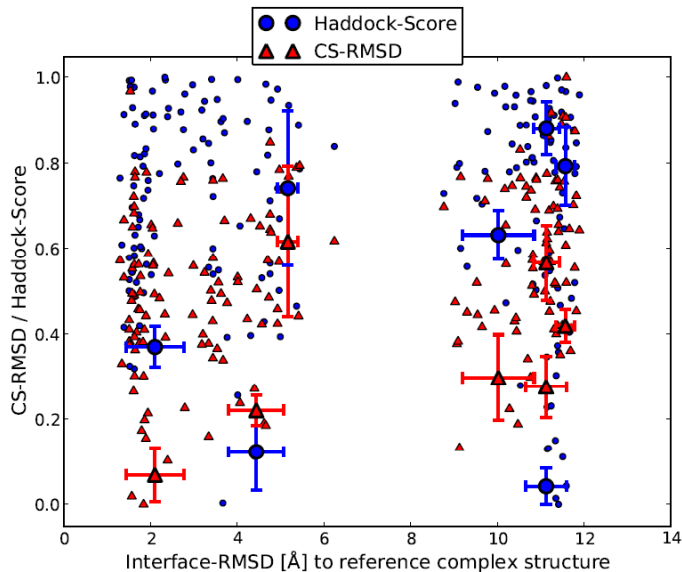
Protocole d'arrimage CS-HADDOCK



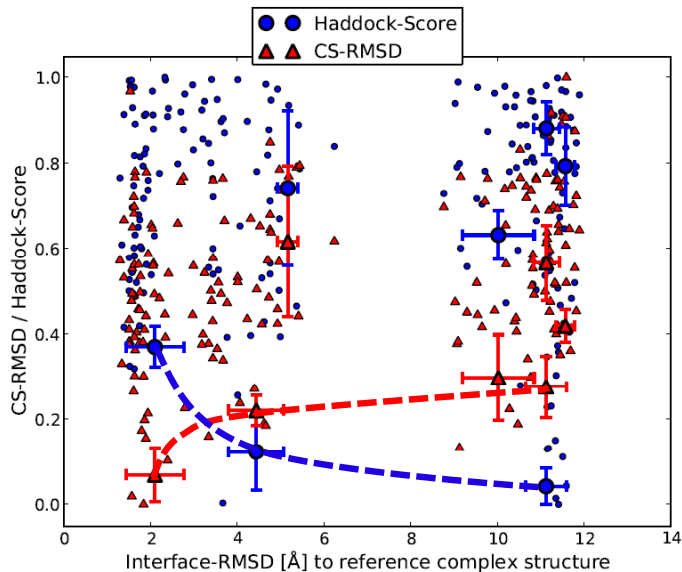
CS-RMSD =

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

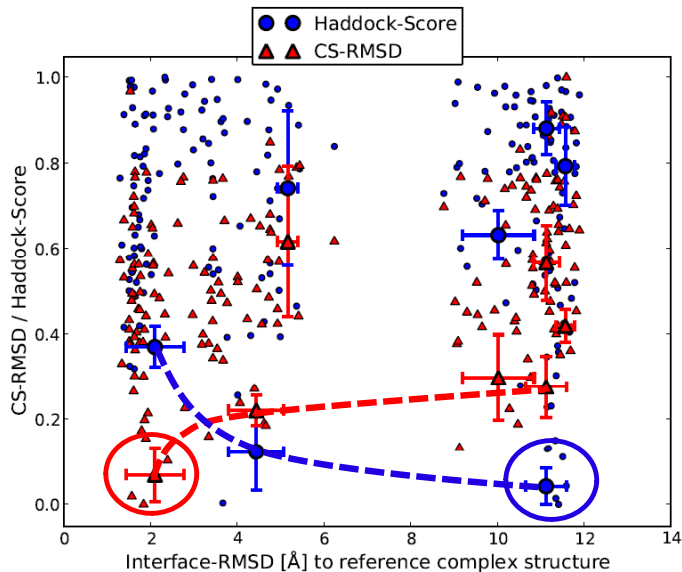
CS-HADDOCK vs HADDOCK



CS-HADDOCK vs HADDOCK

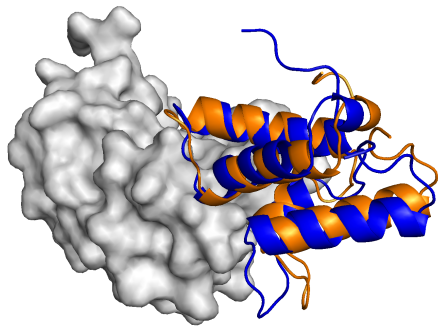


CS-HADDOCK vs HADDOCK

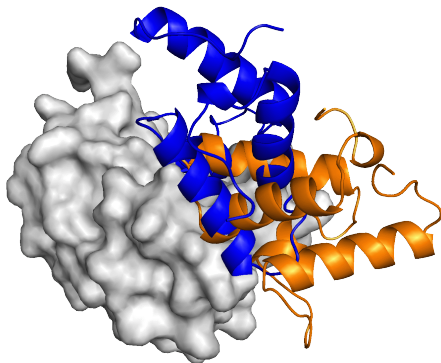


CS-HADDOCK vs HADDOCK

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



(a) CS-RMSD score



(b) HADDOCK score

The end

- MERCI pour votre attention!

6 Bibliography



Andrusier, N., Mashiach, E., Nussinov, R., and Wolfson, H. J. (2008). Principles of flexible protein-protein docking. *Proteins*, 73(2):271–289.
PMID: 18655061.



Bonvin, A. M. J. J. (2006). Flexible protein-protein docking. *Curr. Opin. Struct. Biol.*, 16(2):194–200.
PMID: 16488145.



Deupi, X. and Kobilka, B. K. (2010). Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology*, 25(5):293–303.



Lensink, M. F. and Wodak, S. J. (2010). Docking and scoring protein interactions: CAPRI 2009. *Proteins*, 78(15):3073–3084.
PMID: 20806235.



Sandak, B., Wolfson, H. J., and Nussinov, R. (1998). Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. *Proteins*, 32(2):159–174.
PMID: 9714156.



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.



Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). Geometry-based flexible and symmetric protein docking. *Proteins*, 60(2):224–231.
PMID: 15981269.



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



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.



Tuffery, P. and Derreumaux, P. (2012). Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J R Soc Interface*, 9(66):20–33.
PMID: 21993006.



Zacharias, M. (2010). Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.*, 20(2):180–186.
PMID: 20194014.