

**Comparaisons
et Alignements
de structures 3D**

Comparaisons de structures 3D

Identiques

- Similarité/Différence de structures indépendantes:
 - Diffraction X contre RMN
 - Protéine native contre mutante

Similaires

- Prédire la fonction
- Histoire évolutive
- Domaines importants

Sans relation

- Identification de points communs entre protéines sans structure commune.
- Focaliser sur sites actifs, fixations de ligands sur sites actifs

Comparaisons de séquences

Similarité de séquences linéaires

Comparaisons de structures 3D

Relations dans l'espace 3D

Considérations

Quels atomes/régions seront comparé(e)s?

Structures rigides ou flexibles?

Tous les atomes ou $C\alpha$?

Maximisez le nombre d'atomes ou uniquement dans une région donnée?

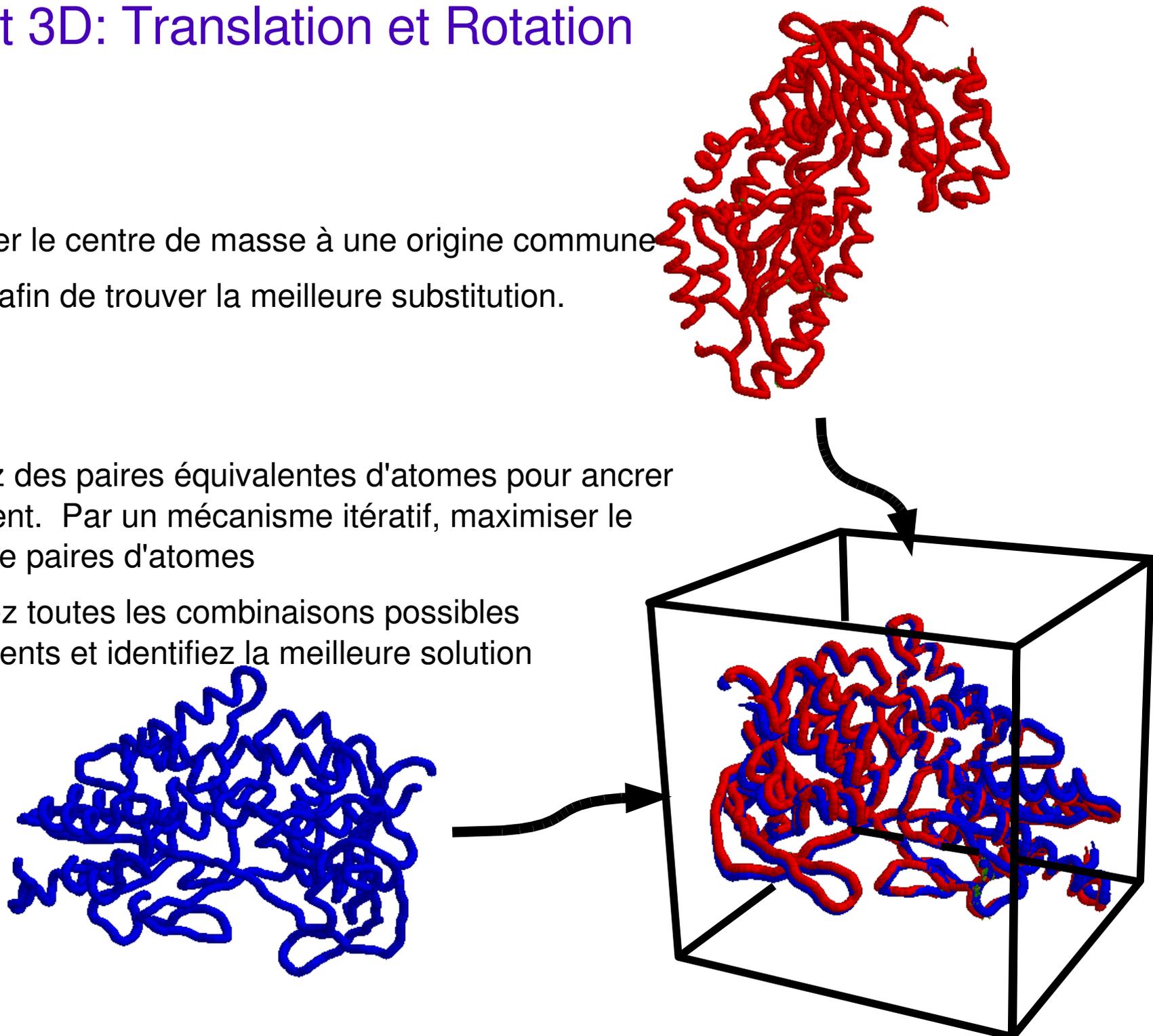
Alignement 3D: Translation et Rotation

Alignement

- Traduire le centre de masse à une origine commune
- Tourner afin de trouver la meilleure substitution.

Approches

- Identifiez des paires équivalentes d'atomes pour ancrer l'alignement. Par un mécanisme itératif, maximiser le nombre de paires d'atomes
- Examinez toutes les combinaisons possibles d'alignements et identifiez la meilleure solution



Alignement 3D: Algorithmes

Initiation

Examinez les structures secondaires et distances Ca-Cb pour identifier les repliements

Pénalités de « gap »

Pour les structures à régions discontinues qui ne s'alignent pas

Anticipez le fait que deux régions différentes peuvent s'aligner séparément mais pas dans le même alignement

Algorithmes d'alignement

Programmation dynamique

Matrice de distances

Fondé sur les structures secondaires

Alignement 3D: Programmation dynamique

Environnement local défini en termes de

- Distances interatomiques
- Angles dièdres
- Identité de chaînes latérales
- Exposition/Enfouissement des chaînes latérales

Alignement des structures en faisant coïncider les environnements locaux

Par exemple, dessiner des vecteurs représentant les liaisons $C\alpha-C\beta$ et superposer ces vecteurs

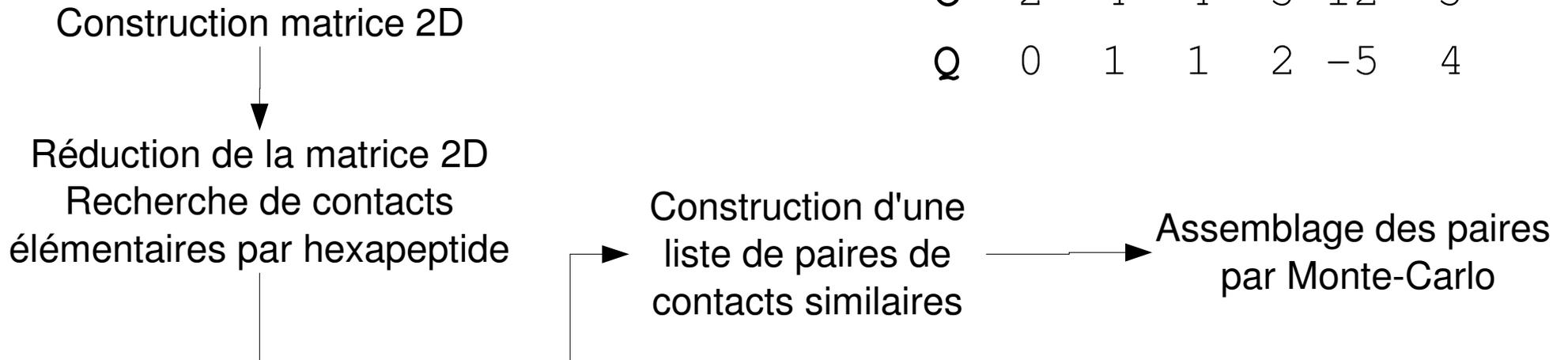
Alignement 3D: Matrice de distances

Approche

- Alignement de deux structures
- Conversion 3D en matrice 2D
- Recherche des atomes les plus proches entre les deux structures à aligner
- Fondée sur les distances $C\alpha-C\alpha$
- Principe de DALI (Distance ALIgnment tool)
 - Utilisé pour la banque FSSP

	A	R	N	D	C	Q
A	2	-2	0	0	-2	0
R	-2	6	0	-1	-4	1
N	0	0	2	2	-4	-1
D	0	-1	2	4	-5	2
C	-2	-4	-4	-5	12	-5
Q	0	1	1	2	-5	4

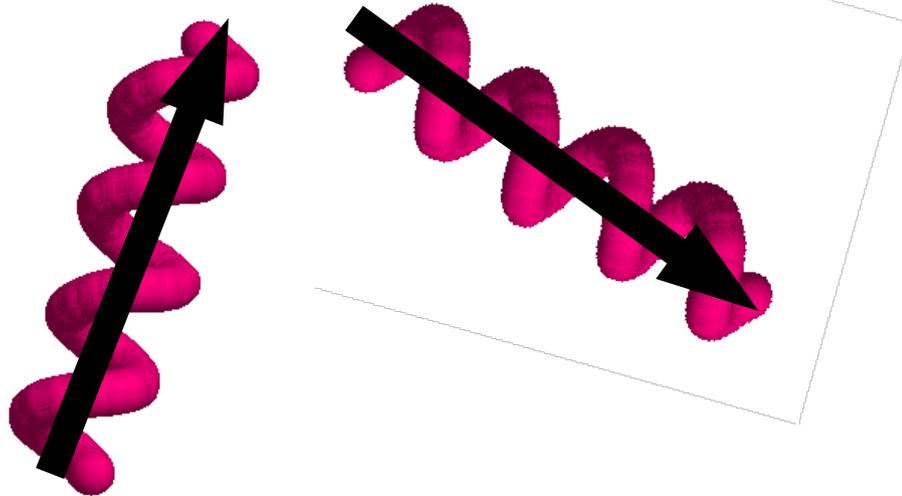
DALI (Distance ALIgnment tool)



Alignement 3D: Rapide par structures secondaires

Approche

- Les éléments de structures II sont représentés par un vecteur V (origine,direction)
- Comparaison des groupes de vecteurs pour identifier les repliements communs
- **Optionnel**: Utilisation de vecteurs supplémentaires
 - Arrangement des chaînes latérales (exposées/ enterrées)
- Probabilité qu'un groupe de structures secondaires soit présent entre des structures sans relation



Alignement 3D: Rapide par structures secondaires

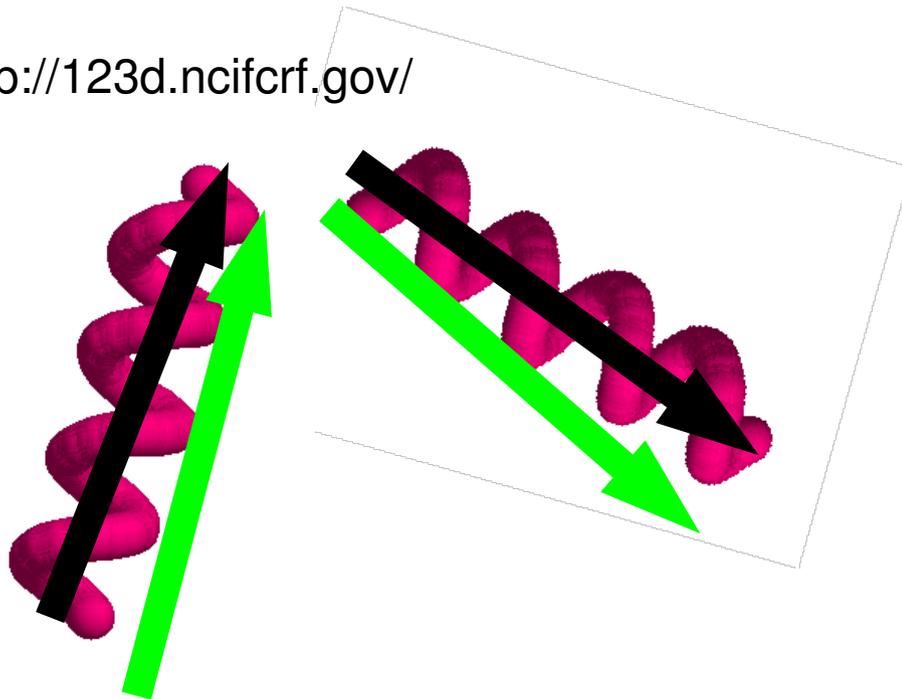
Logiciels automatiques d'alignements de structures secondaires

VAST

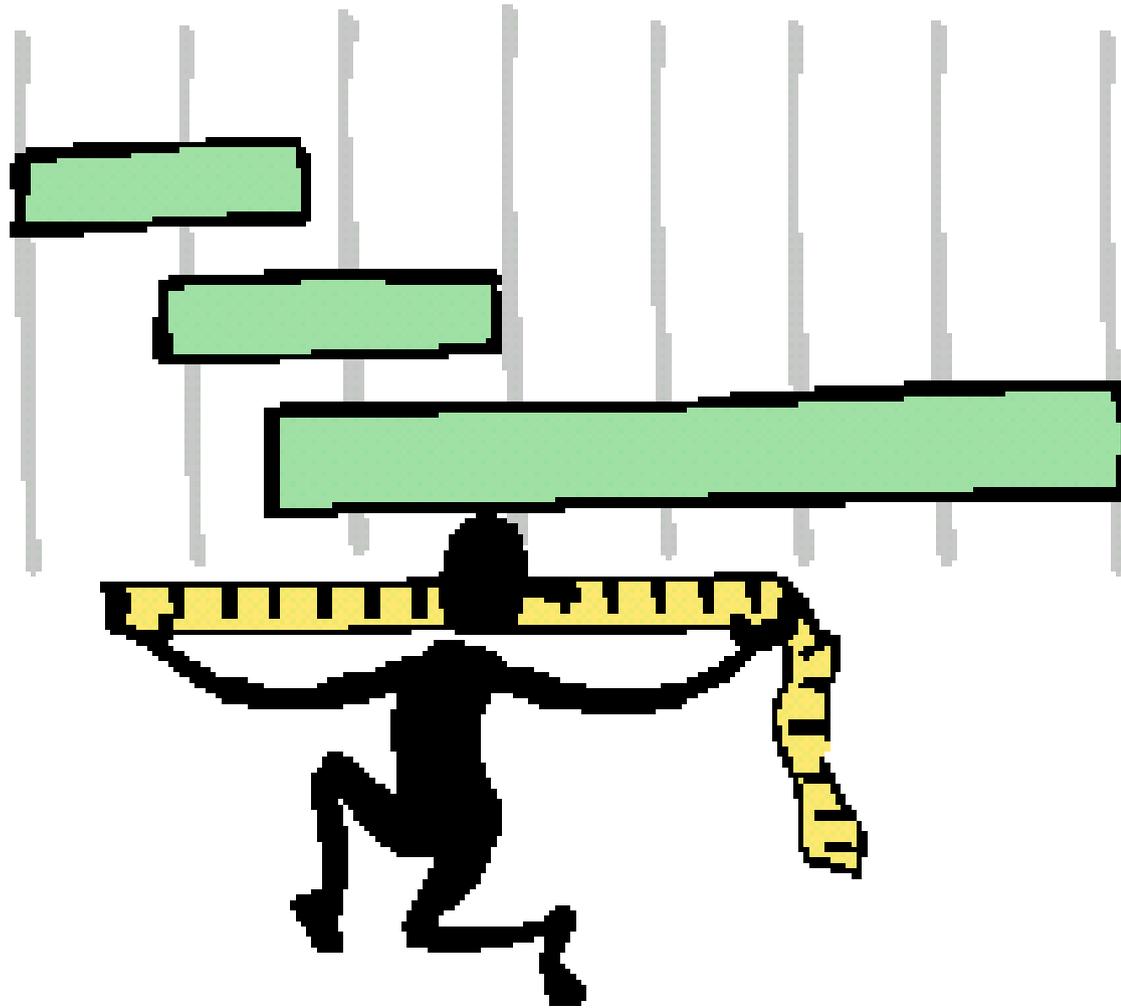
- <http://www.ncbi.nlm.nih.gov:80/Structure/VAST/vastsearch.html>

SARF2

- <http://123d.ncifcrf.gov/>



Qualité de l'alignement



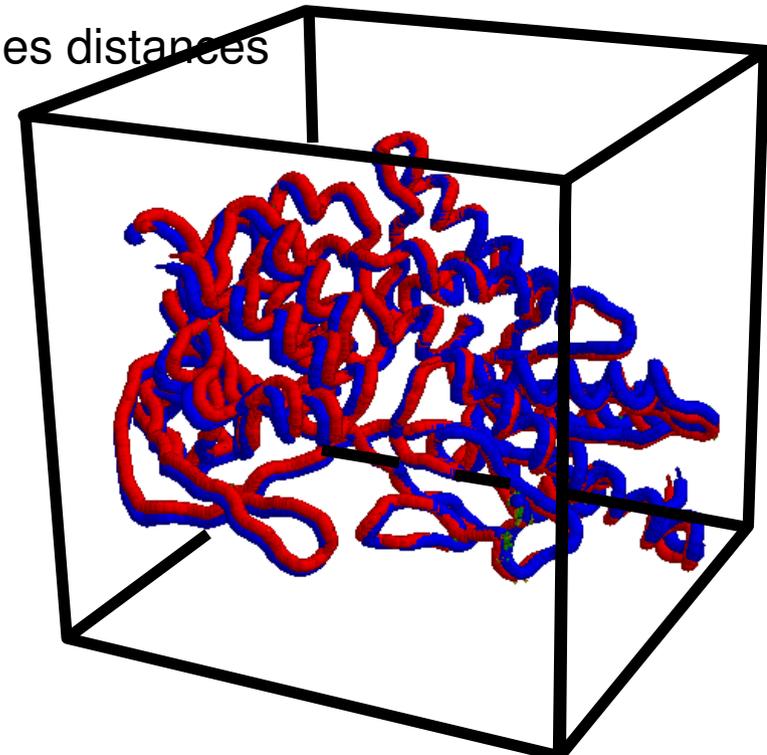
Qualité des alignements 3D

Calcul de la déviation entre les deux structures alignées

RMSD : Root Mean Square Deviation

- Justesse entre les deux ensembles de coordonnées
- Meilleur < 3 angstroems
- Calculez les distances C α -C α , summez les carrés des distances et divisez par le nombre de paires; le tout à la racine carrée:

$$\text{RMSD} = \sqrt{\sum_{i=1}^N D_i^2 / N}$$





**Prédiction
de
structures
3D**

Prédiction de la structure 2D à partir de la structure 1D

Par des méthodes statistiques

- Les premières datent de 1974.
- A partir de la connaissance des structures 3D d'un échantillon de protéines modèles, on établit une table d'occurrences comptabilisant les proportions observées de chacun des 20 acides aminés dans un état structural donné.
- La prédiction est établie à partir de cette table.

Grâce aux propriétés physico-chimiques des acides aminés

- La charge, l'hydrophobicité et l'hydrophilie.
- Celles-ci influent en effet directement sur la position des résidus à l'intérieur (coeur hydrophobe: les hélices α et feuilletts β dominant) ou à la surface de la protéine (boucles).

Par la méthode des plus proches voisins (Nearest neighbours method)

- Utilisation de la similarité entre sous-séquences de protéines

Par les méthodes d'apprentissage par réseaux de neurones

Par les chaînes de Markov

L'influence de la structure 3D sur la 2D n'est pas prise en compte par ces méthodes.

Prédiction de structures spécialisées

Leucine zippers

- Hélices α antiparallèles maintenues par des interactions entre Leu toutes les 7

Structures torsadées (coiled coils)

- 2 ou 3 hélices α enroulées les unes aux autres en torsade "gauche"
- **Multicoil** <http://gaiberg.wi.mit.edu/cgi-bin/multicoil.pl>
- **COILS2** http://www.ch.embnet.org/software/COILS_form.html

Régions transmembranaires

- Domaines de 20-30 aa avec une forte hydrophobicité
- PHDhtm
- PHDtopology
- Tmpred (TMbase)

Prédiction de la structure 3D à partir de la structure 1D

Pourquoi?

- Les méthodes expérimentales comme diffraction X ou RMN sont:
 - techniquement difficiles
 - coûteuses
- Plus de **500 repliements** structuraux parmi **12500 structures tertiaires** connues
- Différentes séquences adoptent le même repliement.

Forte probabilité
pour qu'une
séquence protéique
ait déjà des types de
repliements connus

Prédiction de la structure 3D à partir de la structure 1D

Modèle d'homologie (homology modelling)

La séquence protéique cible est alignée sur toutes les séquences protéiques de la PDB

Si le degré d'homologie est significatif, un modèle est construit.

Construction du modèle

Utilisation de la protéine homologue comme "empreinte" (template) pour déduire la structure tertiaire de la protéine inconnue

Méthode la plus utilisée

- Si la protéine étudiée partage plus de **30% d'acides aminés, après alignement**, avec une protéine connue, on peut supposer que ces deux séquences possèdent des **structures 3D proches**.
- La protéine connue peut alors servir **d'empreinte** pour construire un modèle structural de la protéine étudiée (modélisation moléculaire).
- Plus la similarité sera grande, meilleure sera la modélisation.

Cas No1: Il existe une similarité entre la séquence de la protéine étudiée et une séquence de protéine répertoriée

Méthodes substitutives

- Elles consistent à **échanger les chaînes latérales** de l'empreinte par celles de la séquence à modéliser (réorientation manuelle grâce à un logiciel graphique)
- Recherche à minimiser
 - l'encombrement stérique
 - les mauvais contacts
 - l'énergie qu'il est nécessaire d'appliquer à la molécule pour lui donner la conformation induite.

Méthodes géométriques

- Elles consistent à mesurer des distances sur la protéine empreinte puis à les appliquer à la molécule à modéliser.
- Elles cherchent ensuite à déterminer les **conformations possibles** qui présentent
 - une optimisation énergétique mécanique et dynamique (liaisons covalentes , angles de valence, rayons de Van der Waals)
 - un respect des contraintes de distance déduite de l'empreinte.

Prédiction de la structure 3D: le threading

Reconnaissance de repliements (threading)

- Utilisée lorsque la modélisation par homologie a échoué.
- Alignement de la séquence cible sur des modèles de repliements
- Fondée sur les propriétés hydrophobes des résidus

Méthode moyennement utilisée

Cas No2: Les banques de données de séquences ne recèlent aucune séquence similaire avec la protéine étudiée :

La séquence est comparée en vue de sa "compatibilité" (similarité structurale) avec des repliements connus

- A partir d'un échantillon de toutes les unités de repliements ("folds") possibles
- On "enfile" (*to thread*) la séquence sur la structure 3D en cherchant à minimiser l'énergie d'interaction de l'ensemble des résidus
- Tout en optimisant
 - la répartition spatiale des acides aminés hydrophiles et hydrophobes
 - en prenant en compte le rapport surface/volume, le rayon de giration, etc.

La technique du *threading* ne retourne pas une solution mais un **ensemble de structures candidates potentielles** (elle permet de distinguer les solutions possibles des solutions incompatibles).

Cas No2: Les banques de données de séquences ne recèlent aucune séquence similaire avec la protéine étudiée :

Approches pour déterminer cette "compatibilité"

- Empreinte environnementale (*Environmental template*)
 - Transformation de la séquence sous forme d'une matrice dans laquelle on a la probabilité qu'un résidu soit présent dans une certaine position du repliement
 - Bowie *et al.* (1991) a déterminé 18 types d'environnement.
- Méthode des potentiels de contact
 - Analyse la proximité des contacts entre résidus dans la structure
 - Fondé sur des calculs de distances entre toutes les paires d'acide aminées de la séquence positionnées sur la structure 3D (Sippl & Flockner 1996).
 - Détermine si les positions à l'intérieur de la séquence cible peuvent produire des interactions similaires (les plus favorables énergétiquement)

Prédiction de la structure 3D: Threading

Méthodologie

- Séquence cible déplacée position par position sur la structure
- Le repliement est modelé par calcul de paires interatomiques pour aligner la séquence sur le backbone de l'empreinte
- Optimisation du modèle par minimisation pseudo-énergétiques
- L'alignement le plus stable énergétiquement est considéré comme le plus favorable
- Nécessite des calculs intensifs

Prédiction de la structure 3D: Threading

Les outils de threading

- 123D: <http://123d.ncifcrf.gov/123D+.html>

Prédiction de la structure 3D à partir de la structure 1D

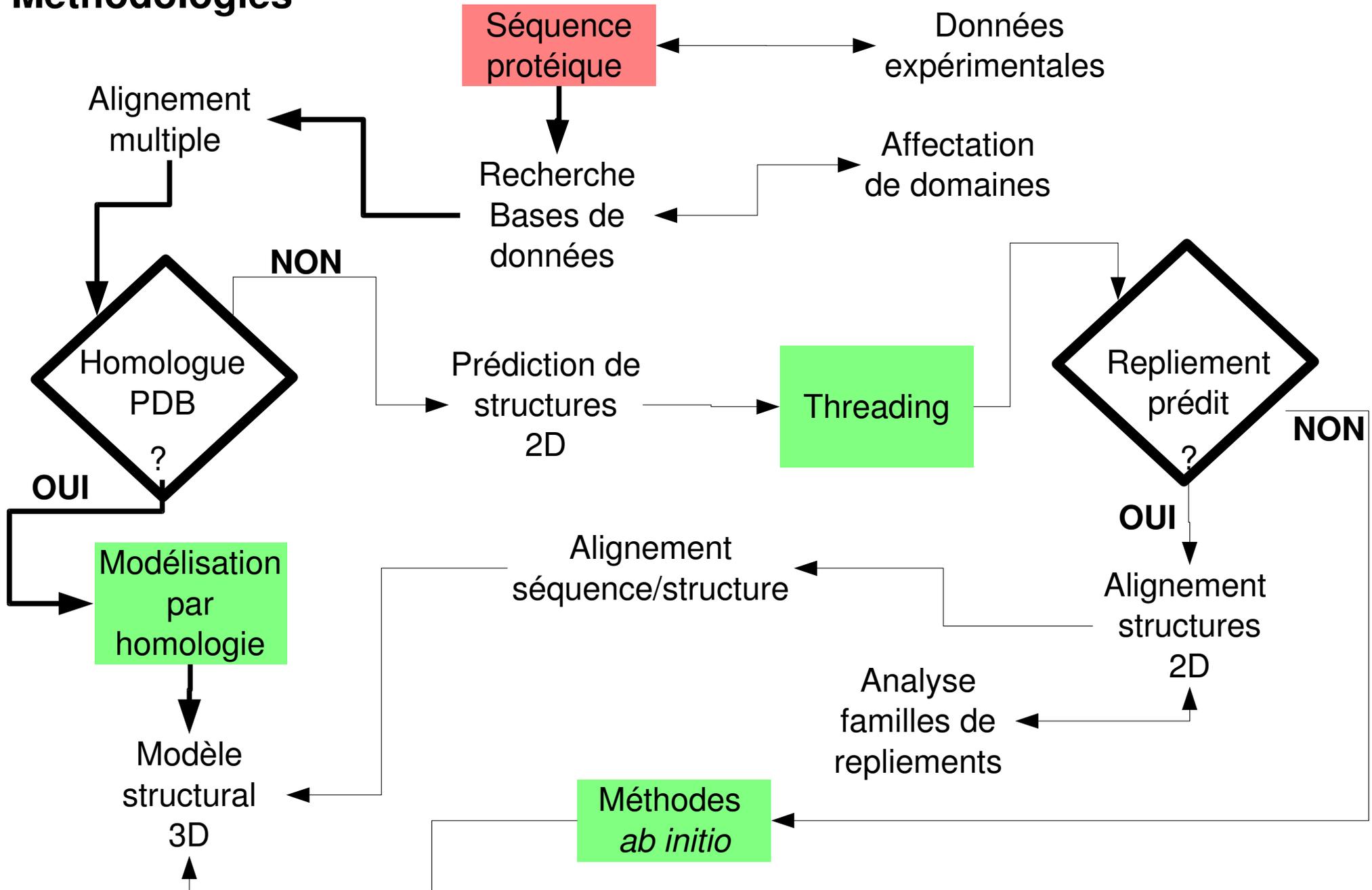
Méthodes *ab initio*

- Modélisation de toutes les énergies mises en jeu dans les repliements
- Trouver un modèle dont l'énergie est la plus faible.

Méthode peu utilisée

Prédiction de la structure 3D à partir de la structure 1D

Méthodologies



Boite à outils



Construction de modèles

Programmes permettant construction automatique de modèles

- **GENO3D**

- PSI-BLAST pour identifier les homologues possédant des structures 3D utilisables comme empreintes
- <http://geno3d-pbil.ibcp.fr>

- **SWISS-MODEL**

- Compare la séquence à ExPDB pour trouver un homologue
- Définir ses propres empreintes (à partir de threading)

Construction de modèles

Programmes permettant construction automatique de modèles

- **GENO3D**

- PSI-BLAST pour identifier les homologues possédant des structures 3D utilisables comme empreintes
- <http://geno3d-pbil.ibcp.fr>

- **SWISS-MODEL**

- Compare la séquence à ExPDB pour trouver un homologue
- Définir ses propres empreintes (à partir de threading)

The ExPDB database

The structure database used by **SWISS-MODEL** is derived from the Protein Data Bank (PDB). In order to allow rapid retrieval of the relevant structural information, each protein chain of the database was placed in an individual file.

The ExPDB codes are constructed according to the following rule:

PDBCODE+ChainID



Pôle Bio-Informatique Lyonnais

Geno3D

Geno3D is the [IBCP](#) contribution to [PBIL](#) in Lyon, France

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

Thursday, February 17th 2003 : **GENO3D** is now running on Linux cluster ([see news](#))

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

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

Database :

Sequence 1 name (optional) :

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

Number of PSI-BLAST run

GENO3D 2 : AUTOMATIC MODELING OF PROTEINS THREE-DIMENSIONAL STRUCTURE

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

Database :

Sequence 1

Paste protein sequence or ID :

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

Number of PSI-BLAST run

Job GENO3D2 (ID: 26787) is running on GENO3D server (started on Mon Jun 21 09:58:04 CET 2007)

Results will be shown below. Please wait and don't go back.

Run GENO3D2

FIRST STEP : Segment to modelling

Select the template(s) to use
 Template selection

% identity

Link to alignment

Link to NPSA tools

PSI-BLAST run 3 for Monomer_0 sample

TEMPLATE	E	FIRST	LAST	ID	ALIGNEMENT	COMMENT	NPSA link
<input type="checkbox"/> pdb1b82A-0	2e-72	1	236	100	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1b82B-0	3e-72	1	236	100	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1a20-0	3e-72	1	236	58	see alignment	PEROXIDASE	NPSA
<input type="checkbox"/> pdb1b80A-0	1e-71	1	236	99	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1b80B-0	2e-71	1	236	99	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1b85A-0	3e-71	1	236	99	see alignment	OXIDOREDUCTASE	NPSA

TEMPLATE	E	FIRST	LAST	ID	ALIGNEMENT	COMMENT	NPSA link
<input checked="" type="checkbox"/> pdb1bep-0	2e-45	2	236	18	see alignment	PEROXIDASE	NPSA
<input type="checkbox"/> pdb1bj9-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input checked="" type="checkbox"/> pdb1ocf-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1a2g-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1ocb-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input checked="" type="checkbox"/> pdb1bek-0	2e-45	2	236	18	see alignment	PEROXIDASE	NPSA
<input type="checkbox"/> pdbecop-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE(H2O2(A))	NPSA
<input type="checkbox"/> pdb1i-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1j-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1k-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE(H2O2(A))	NPSA
<input type="checkbox"/> pdb1l-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA
<input type="checkbox"/> pdb1m-0	2e-45	2	236	18	see alignment	OXIDOREDUCTASE	NPSA

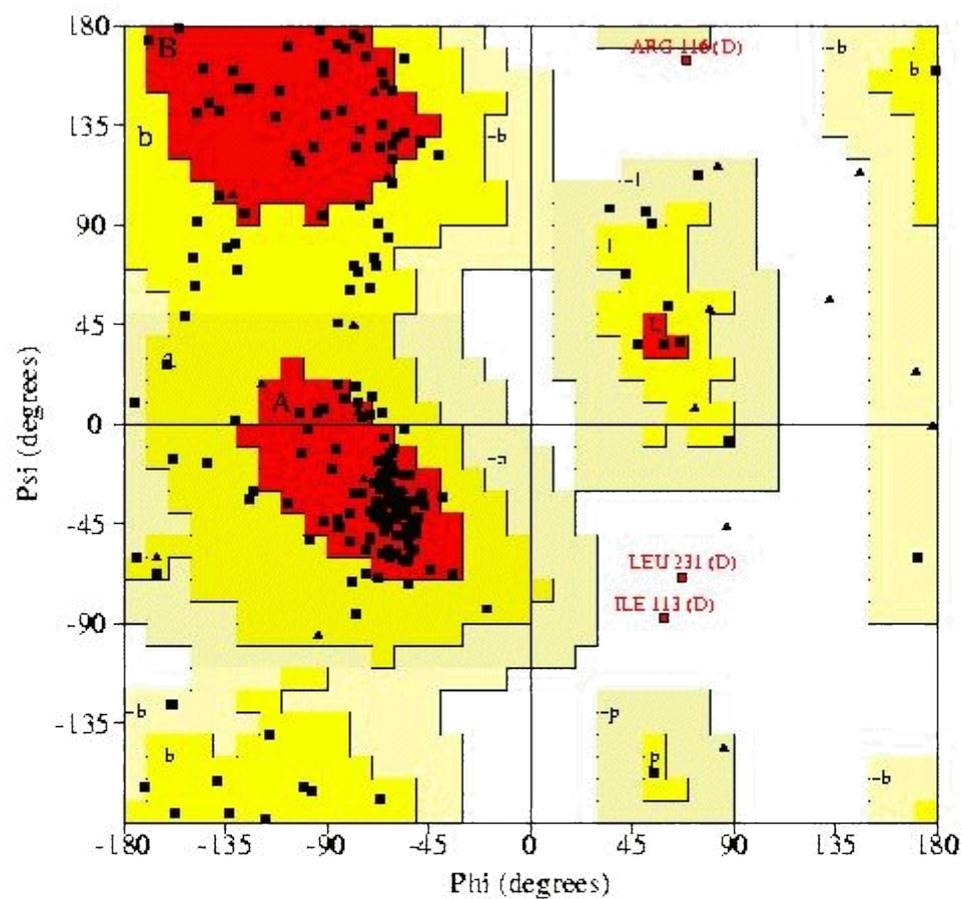
- Sequence of this chain :

```
IRLVFHDSIA ISPAMEAQGK FGGGGADGSI MIFDDIETAF HPNIGLDEIV
KLQKPFVQKH GVTPGDFIAF AGAVALSNCP GAPQMNFFTG RAPATQPAPD
GLVPEPFHTV DQIINRVNDA GEFDELELVW MLSAHSVAAV NDVDPTVQGL
PFDSTPGIFD SQFFVETQLR GTAFPGSGGN QGEVESPLPG EIRIQSDHTI
ARDSRTACEW QSFVNNQSKL VDDFQFIFLA LTQLG
```

- This chain was modelled using 3 templates :

Template	Alignment (Clustalw format)	Secondary information (Sov)	
Identity			
pdb1bek_0	alignment_1_1.aln	71.0%	18.5%
pdb1bep_0	alignment_1_2.aln	71.0%	18.5%
pdb1ccl_0	alignment_1_3.aln	69.2%	18.1%

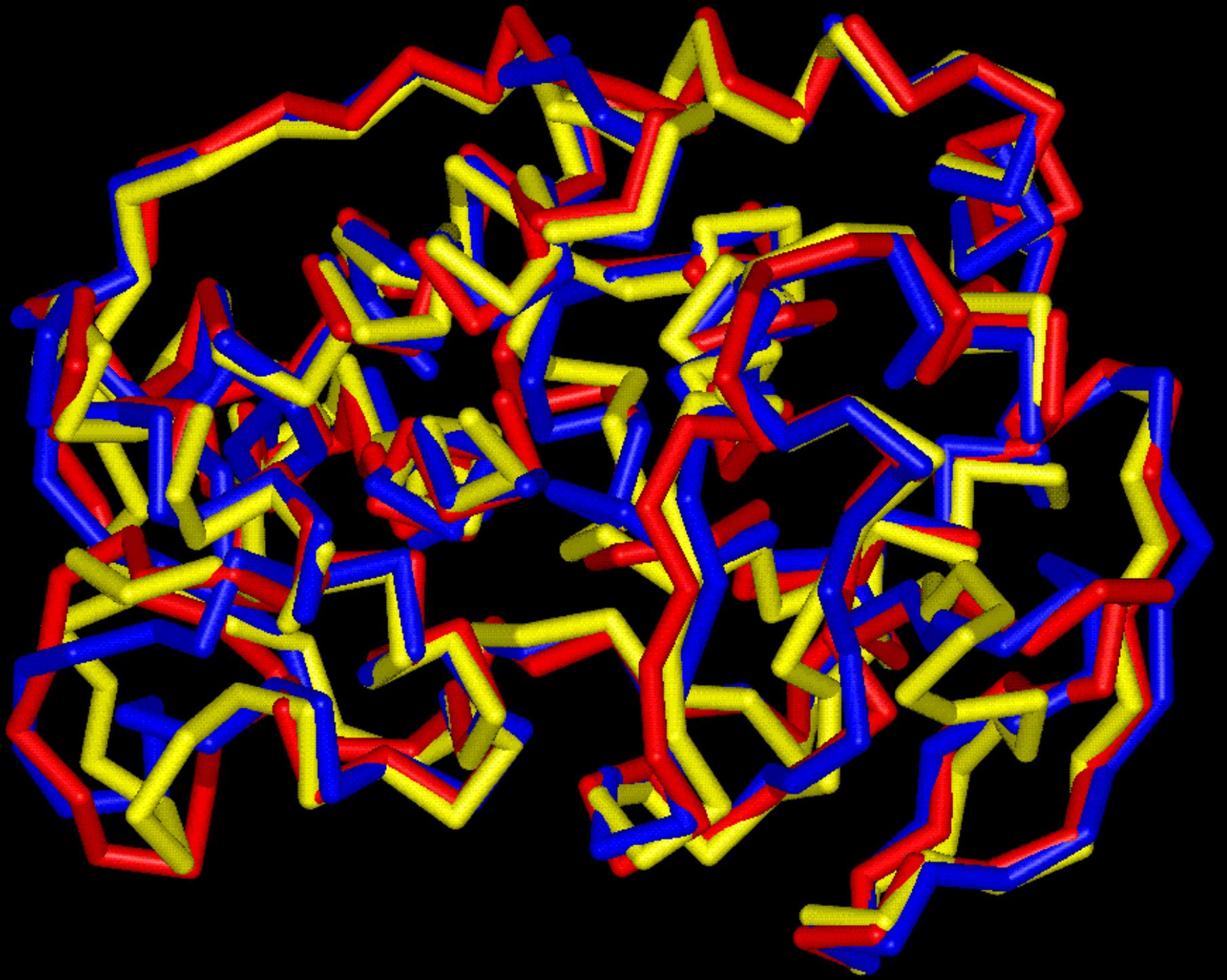
Ramachandran Plot



Plot statistics

Residues in most favoured regions [A,B,L]	138	71.1%
Residues in additional allowed regions [ab,l,p]	45	23.2%
Residues in generously allowed regions [-a,-b,-l,-p]	8	4.1%
Residues in disallowed regions	3	1.5%
Number of non-glycine and non-proline residues	194	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	23	
Number of proline residues	17	
Total number of residues	235	

Based on an analysis of 115 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.



SWISS-MODEL - Mozilla Firefox

Fichier Edition Affichage Aller à Marque-pages Outils ?

http://swissmodel.expasy.org//SWISS-MODEL.html

Hotmail Personnaliser les liens Windows Media Windows

Coyote Linux Web Admini... http://www.n...iucrabs.html BU sciences SWISS-MODEL MolScript v2.1: About the... Structure Explorer - 1MYF

MENU

Modeling requests:

- [First Approach mode](#)
- [Alignment Interface](#)
- [Project \(optimise\) mode](#)
- [Oligomer modeling](#)
- [GPCR mode](#)

Interactive tools

- [DeepView - Swiss-PdbViewer](#), a tool for viewing and manipulating protein structures and models.
- [Lookup](#) the ExpDB template codes accessible to SWISS-MODEL.
- [Search](#) the template sequences accessible to SWISS-MODEL.
- [Examples](#) using SWISS-MODEL and the Swiss-PdbViewer.
- [ANOLEA](#) Protein structure quality check (atomic non-local

HELP

- [Frequently Asked Questions.](#)
- [Visualising 3D models.](#)
- [Reliability of models.](#)
- [How SWISS-MODEL works.](#)
- [How ProModII works.](#)
- [Modelling of oligomeric proteins.](#)



SWISS-MODEL

An Automated Comparative Protein Modelling Server

SIB - Biozentrum Basel site provided by:



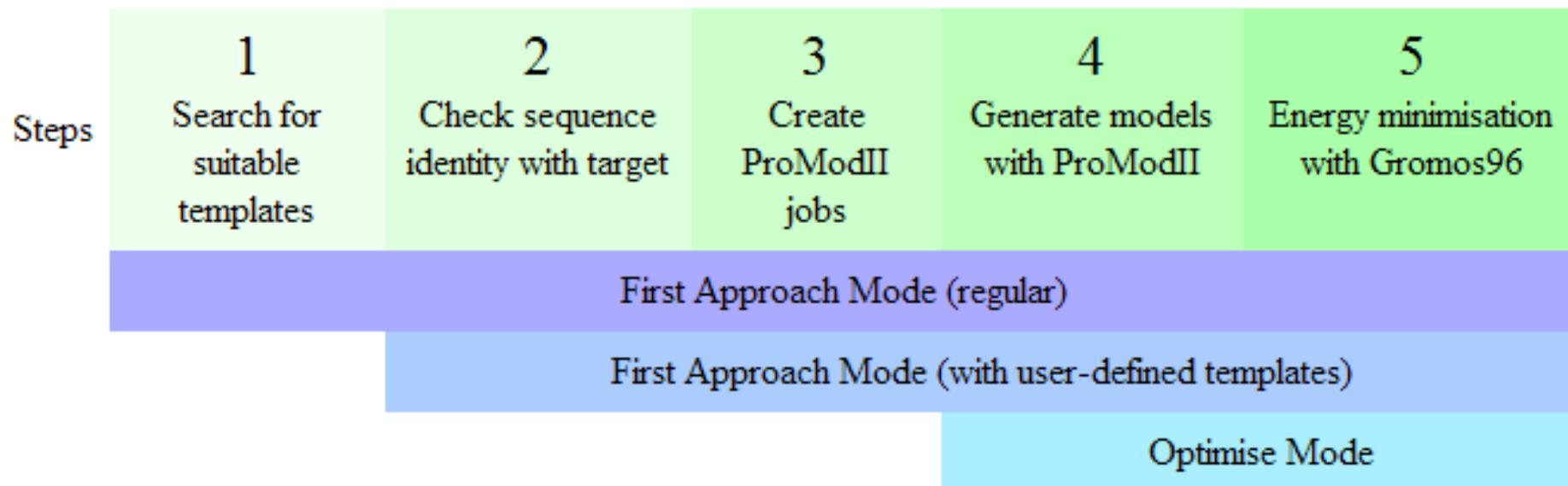
[SWISS-MODEL](#) is a fully automated protein structure homology-modeling server, accessible via the [ExpASY](#) web server, or from the program [DeepView](#) (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us refine the sequence analysis and modelling algorithms, please [report](#) of possible bugs and problems with the modelling procedure.

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and is now being further developed within the [SIB - Swiss Institute of Bioinformatics](#) in collaboration between Torsten Schwede at the [Structural Bioinformatics Group](#), Biozentrum (University of Basel) and Nicolas Guex at [GlaxoSmithKline](#).

The computational resources for the SWISS-MODEL server are provided in collaboration by the Biozentrum (University Basel) and the [Advanced Biomedical Computing Center](#) (NCI Frederick, USA).

Terminé



Step	Program/Method	Database	Action
1	BLASTP2	ExNRL-3D	Will find all similarities of target sequence with sequences of known structure.
2	SIM	-	Will select all templates with sequence identities above 25% and projected model size larger than 20 residues. Furthermore, this step will detect domains which can be modelled based on unrelated templates
3	-	-	Generate ProModII input files
4	ProModII	ExPDB	Generate all models
5	Gromos96	-	Energy minimisation of all models

Modélisation de la protéine par ProMod

Etapas

- * **Superposition** of related 3D-structures.
- * Generation of a **multiple a alignement** with the sequence to be modelled.
- * Generation of a **Framework** for the new sequence.
- * Rebuild lacking **Loops**.
- * Complete and correct **BackBone**.
- * Correct and rebuild **Side Chains**.
- * Verify model structure **quality** and check **Packing**.
- * Refine structure by **Energy minimisation** and **Molecular Dynamics**.