Computational design of protein:

peptide recognition

David Mignon et Nicolas Panel

Laboratoire de Biochimie , **Ecole Polytechnique**

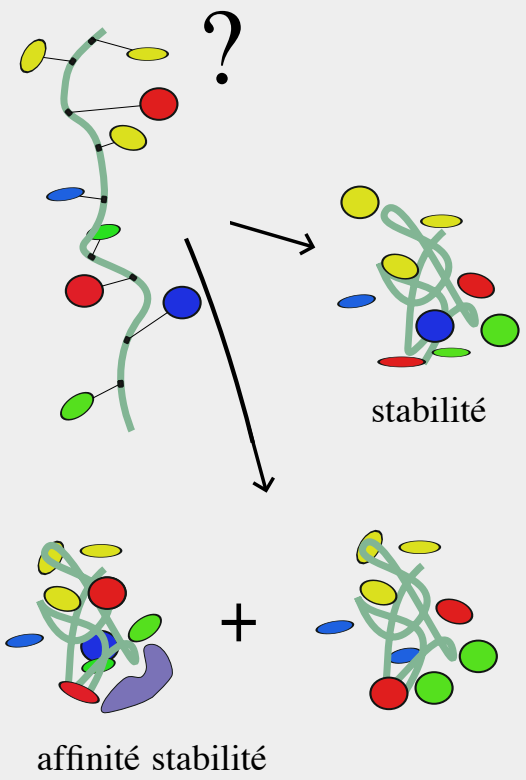
- Méthode CPD et performances
- Applications et simulations détaillées

Computational protein design (CPD)

Concevoir ou modifier des protéines par informatique
pour leur conférer de nouvelles propriétés

Applications:

- protéines redessinées
- enzymes, complexes
- insertion d'acides aminés non-naturels



Principaux programmes de CPD:

- ORBIT (Mayo,1996)
- Toulbar2 (Allouche,2014)
- Proteus (Simonson,2008)
- Rosetta (Baker,2003)

Protein design with Proteus

Conformational space

- Particular structure for backbone
- Side chains rotamers
- Simple model of unfolded state
  (important for stability)

Energy function

- intra-protein molecular mechanics
- Solvant: dielectric model = Generalized Born

Le Monte-Carlo

algorithme Metropolis-Hastings

Le but est de générer une collection d'états échantillonnés selon la distribution de Boltzmann.

$$p(etat) = \frac{1}{Z} e^{(-\frac{E}{RT})}$$

L'algorithme définit une chaîne de Markov pour laquelle:

1- La distribution de probabilité des états est stationnaire

C'est garantie par la balance détaillée.

2- Il n'y a qu'une seule distribution stationnaire.

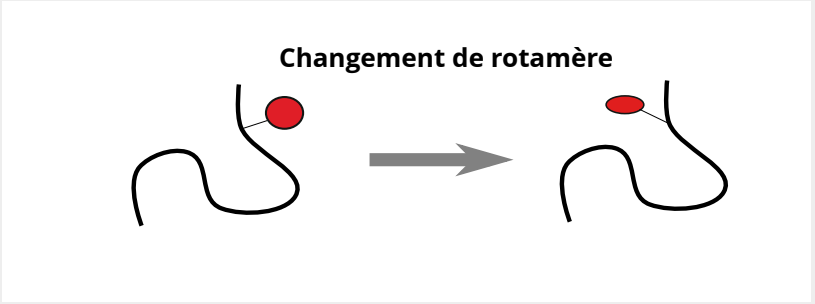
C'est garantie par le caractère ergodique de la chaine.

L'exploration Monte-Carlo

Déplacement MC dans l'espace des conformations:

modification du rotamère à une position i sur la protéine repliée

$$\Delta E = E(.., rot_i^{new}, ..) - E(.., rot_i^{old}, ..)$$

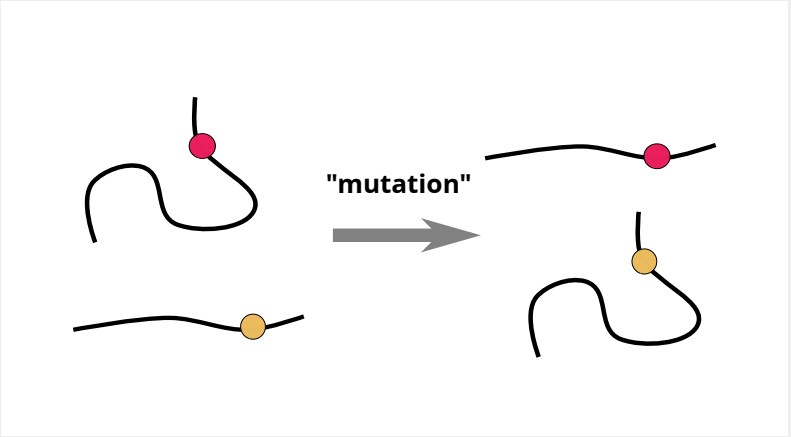


Déplacement MC dans l'espace des séquences :

modification du type de chaîne latérale à une position i sur la protéine repliée.

En même temps, une mutation inverse sur la protéine dépliée, en i.

$$\Delta E = \Delta E_f - \Delta E_{uf}$$



Etat déplié: les énergies de référence

- **Définition**: Pour une séquence **s**, l'énergie de l'état déplié est de la forme:

$$E_s^u = \sum_{i \subset s} E_{t_i}^r$$

Ce sont des paramètres ajustables.

- **L'objectif** est de déterminer les E_t^r pour obtenir les bonnes fréquences d'acide aminé.

- **La méthode** (maximum de vraisemblance):

  Soit $\mathcal{S}$ un ensemble de séquences de Swissprot , p($\mathcal{S}$) sa probabilité de Boltzmann est une fonction des E_t^r .

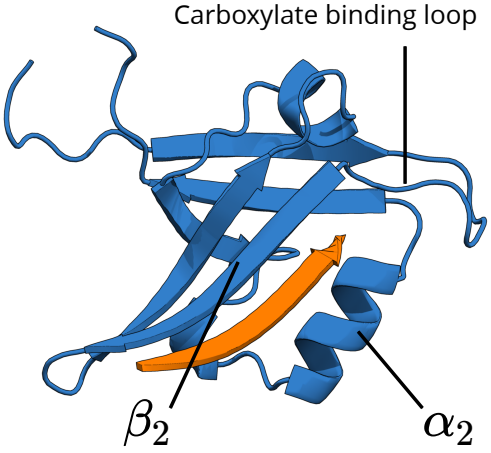
  Nous cherchons les E_t^r qui maximisent p($\mathcal{S}$), elles réalisent notre objectif.

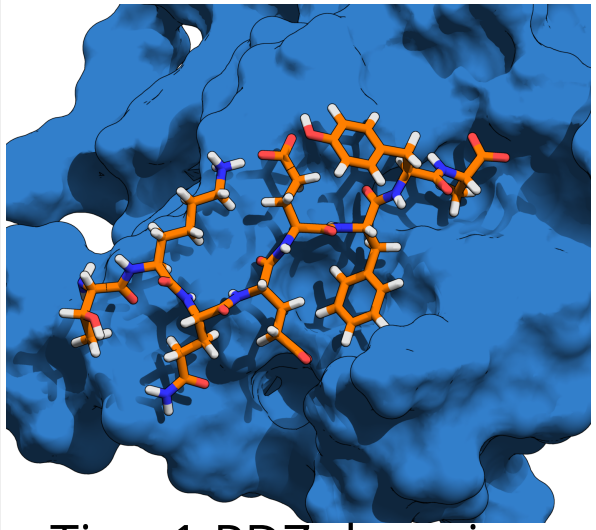
- **Un algorithme itératif:**

$$E_t^r(n+1) = E_t^r(n) + \delta E \times (freq_t^{exp} - freq_t^{proteus_n})$$

PDZ domains as test systems

- Participate in PPIs
- Around 80-100 residues
- Conserved structure:
  - 5/6 β-strands
  - 2 α-helices
- Recognize 4-7 C-terminal amino acids of their targets ($K_d = 10 - 80 \mu M$)
- β-sheet augmentation mechanism
- Binding groove by β_2 and α_2
- Carboxylate binding loop highly conserved





Tiam1 PDZ domain

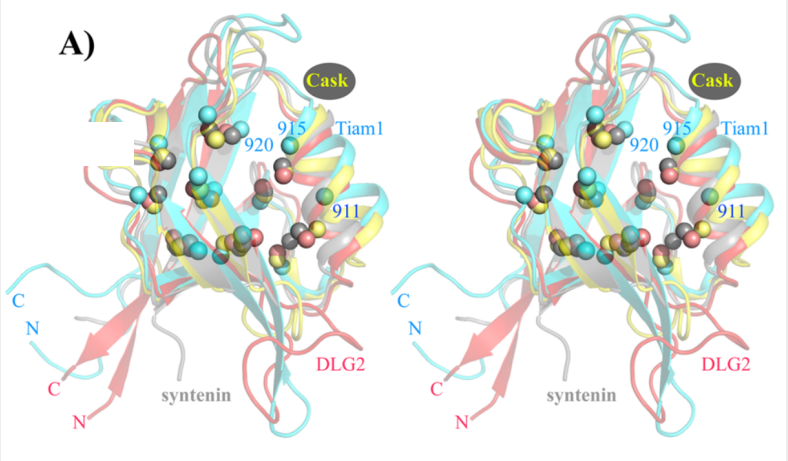
Paramétrisation, mesure de performance de notre modèle CPD

- 8 protéines PDZ

1G9O,1R6J,2BYG,1IHJ,1N7E,3K82,Tiam1,Cask (entre 72 et 84 résidus)

- Design de toutes les positions sauf Gly et Pro

- Optimisation des énergies de référence sur des homologues proches des protéines dans Swissprot.



- Une simulation finale est effectuée avec les énergies de référence optimisées: REMC ,8 marcheurs, 750 millions de pas chacun.

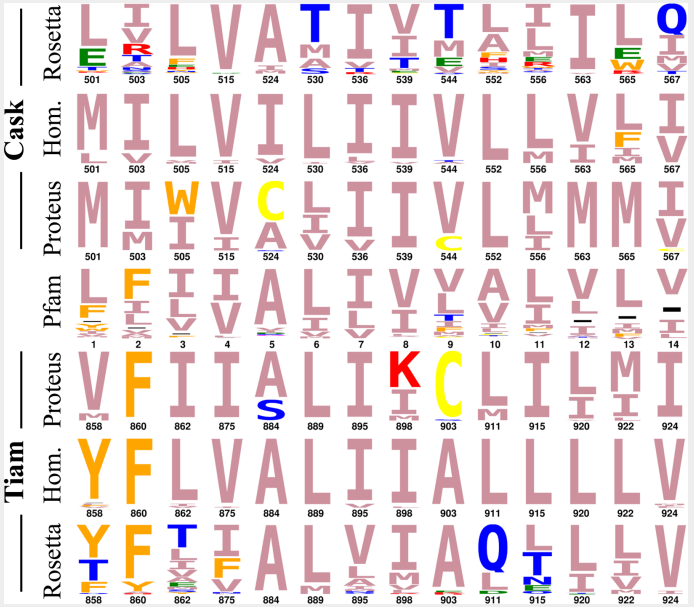
- Comparaisons des séquences calculées à la base de données Pfam-RP55

- Comparaisons au Séquences calculées par Rosetta

- Test de nos séquences avec les outils Superfamily de reconnaissance de repliement.

Séquence Proteus et Rosetta sous forme de logos

positions à la surface



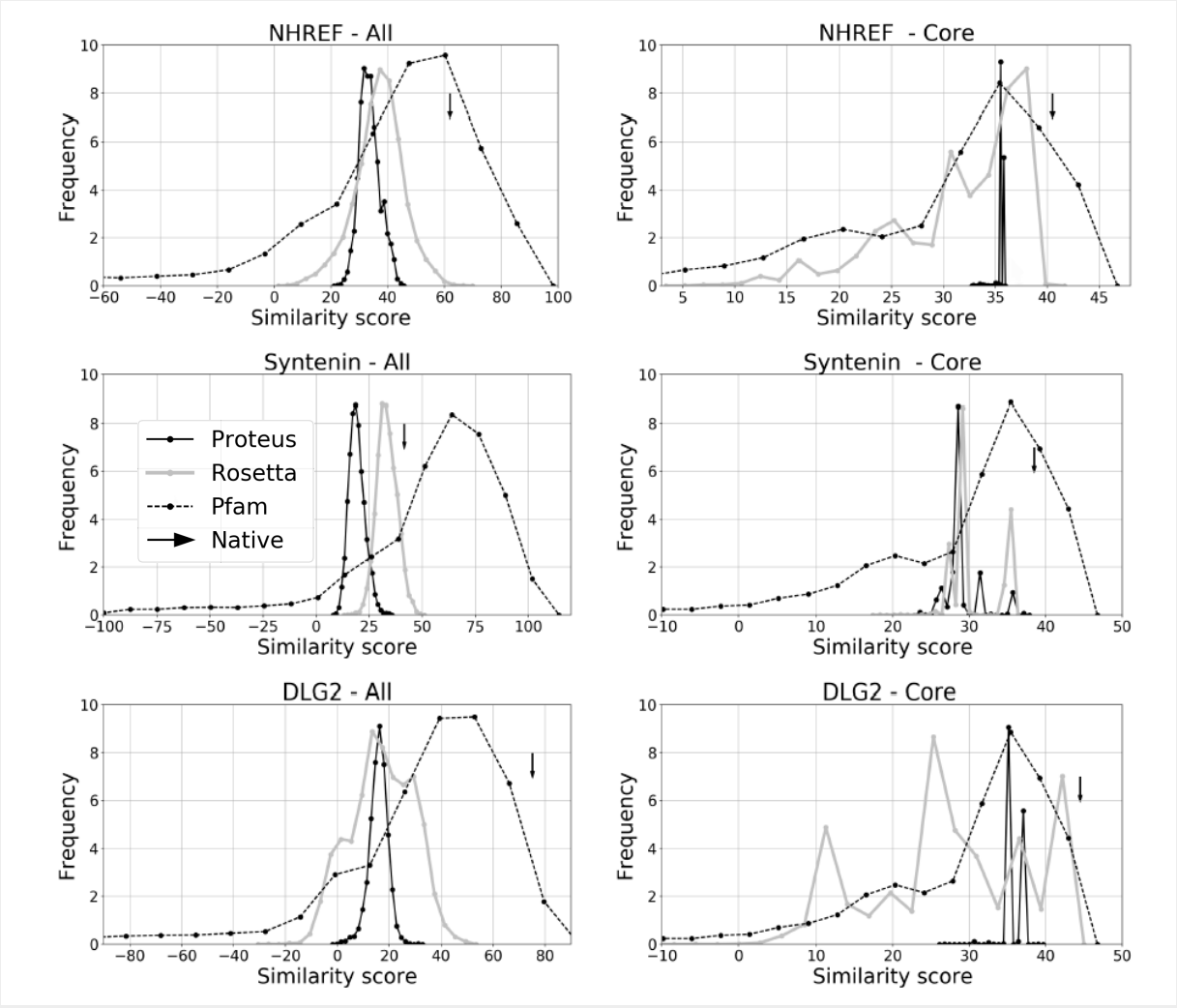
positions du cœur

Reconnaissance de pli avec Superfamily

	Proteus		Rosetta	
Protein	Family Evalue	Family success	Family Evalue	Family success
NHREF	$8.94\ 10^{-2}$	10000	$2.2\ 10^{-3}$	10000
Syntenin	$2.69\ 10^{-3}$	10000	$1.8\ 10^{-3}$	10000
DLG2	$1.96\ 10^{-3}$	10000	$9.6\ 10^{-4}$	10000
Tiam1	$1.96\ 10^{-3}$	10000	$2.8\ 10^{-2}$	9030
Cask	$1.96\ 10^{-3}$	10000	$7.5\ 10^{-3}$	9832

Comparaison aux séquences naturelles

Similarité des séquences



Et Nicolas Panel...