

# Représentations gros-grain pour la modélisation des protéines : Propriétés mécaniques et interactions

*Exposé présenté en vue de l'obtention de l'Habilitation à Diriger les Recherches, par*

**Sophie Sacquin-Mora**

Laboratoire de Biochimie Théorique-CNRS UPR 9080

Institut de Biologie Physico-Chimique, Paris

**I. Introduction générale sur les modèles gros-grain**

**II. Propriétés mécaniques des protéines :**

1. Le programme ProPHET
2. Résultats généraux (**utilisation autonome de ProPHET**)
3. Noyau mécanique dans les globines (**utilisation couplée à des simulations tout-atome**)
4. Réponse mécanique à une contrainte externe (**utilisation couplée à des approches expérimentales**)

**III. Interactions protéiques :**

1. Le programme MAXDo
2. Prédiction des partenaires d'interaction protéiques
3. Prédiction des sites d'interaction protéiques

**IV. Conclusions et Perspectives**

**I. Introduction générale sur les modèles gros-grain**

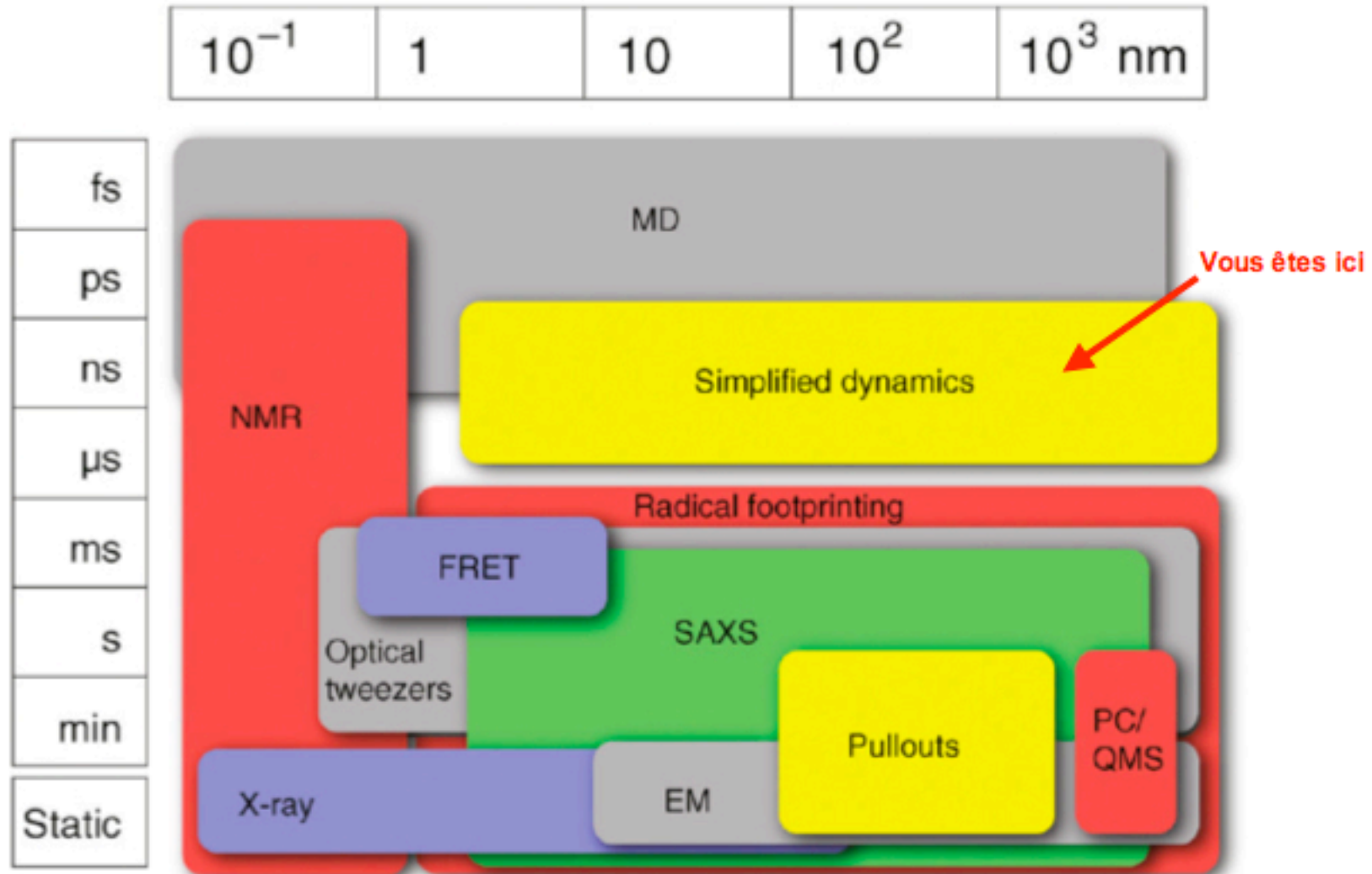
**II. Propriétés mécaniques des protéines :**

1. Le programme ProPHET
2. Résultats généraux (**utilisation autonome de ProPHET**)
3. Noyau mécanique dans les globines (**utilisation couplée à des simulations tout-atome**)
4. Réponse mécanique à une contrainte externe (**utilisation couplée à des approches expérimentales**)

**III. Interactions protéiques :**

1. Le programme MAXDo
2. Prédiction des partenaires d'interaction protéiques
3. Prédiction des sites d'interaction protéiques

**IV. Conclusions et Perspectives**



(Russel et al., *Curr. Opin. Cell Biol.*, 2009)

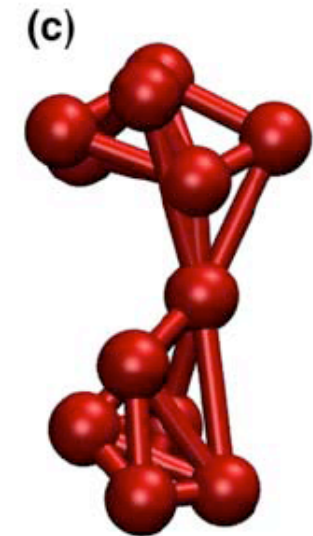
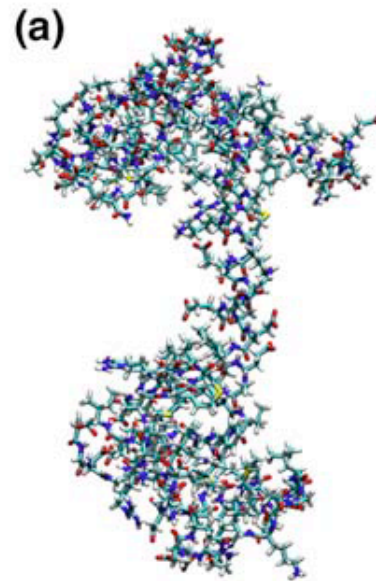
Calmoduline ( $\approx 150$  a.a.)

a) Modèle tout-atomes

( $\approx 1100$  atomes lourds)

b) Un grain par résidu

c) 12 noeuds

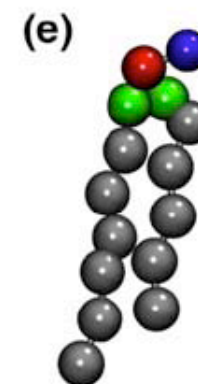
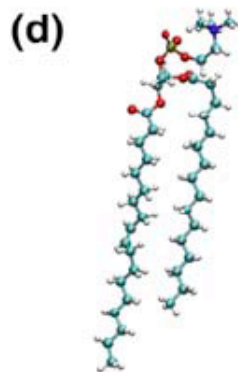


Lipide POPC

d) Modèle tout-atomes

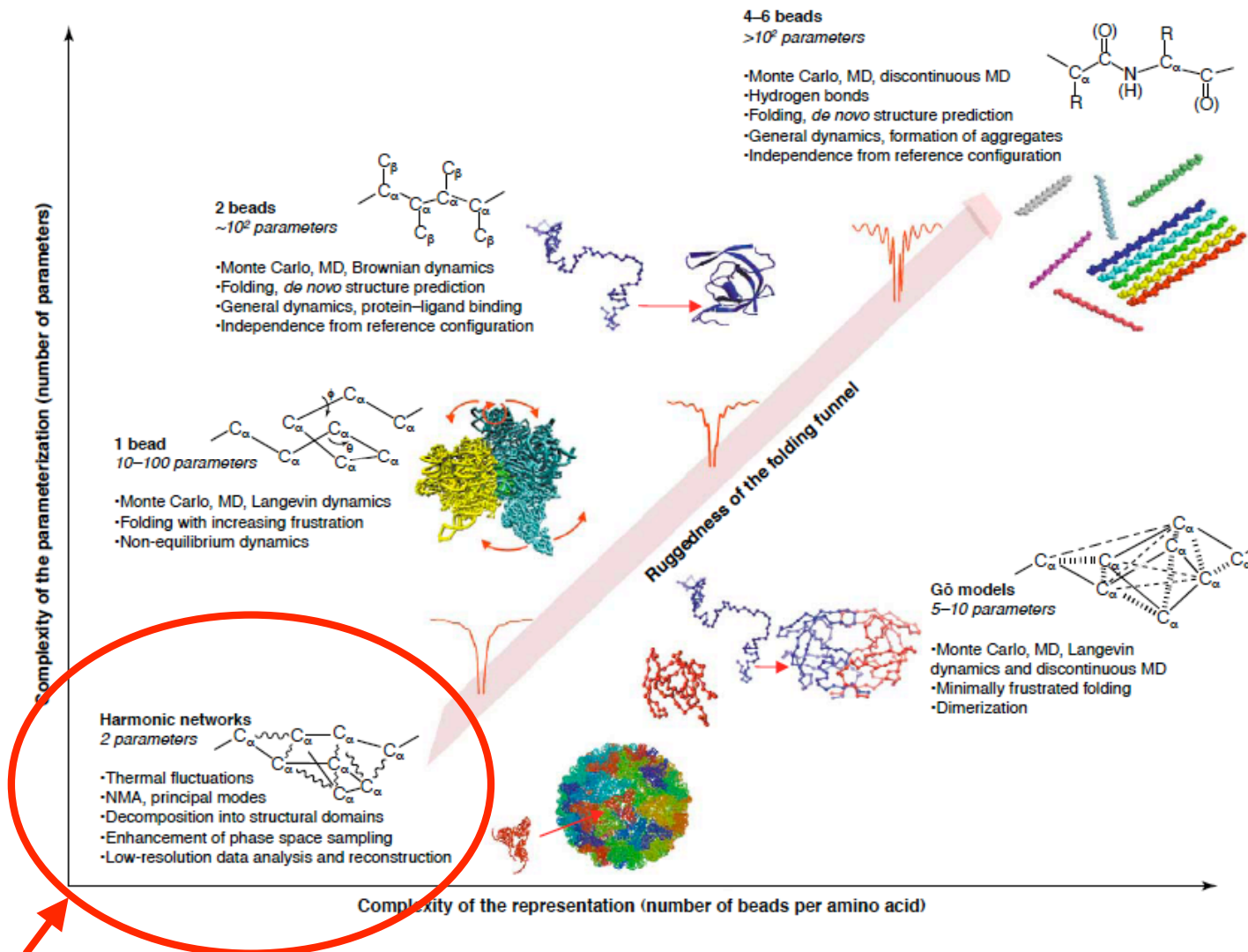
e) Modèle MARTINI

f) Ellipsoïde de Gay Berne



(Rader., *Curr. Opin. Pharmacol*, 2010)

# Introduction : Une (très) brève histoire des représentations réduites



Et maintenant on va là (Tirion, *Phys. Rev. Lett.*, 1996)

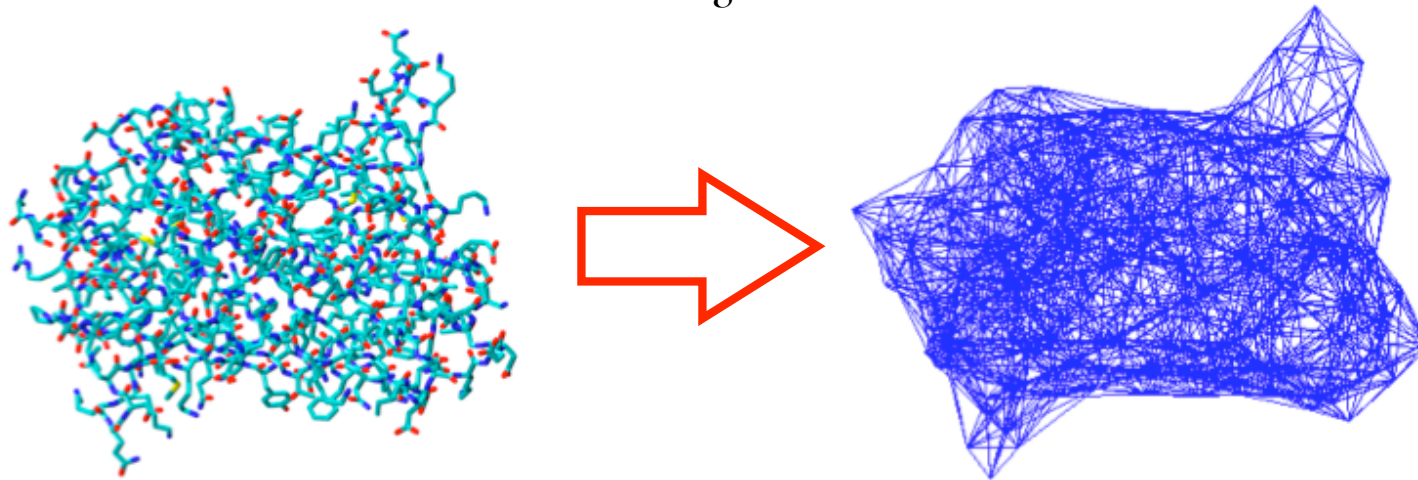
(Tozzini, *Curr. Opin. Struct. Biol.*, 2005)

## Le modèle en réseau de ressorts

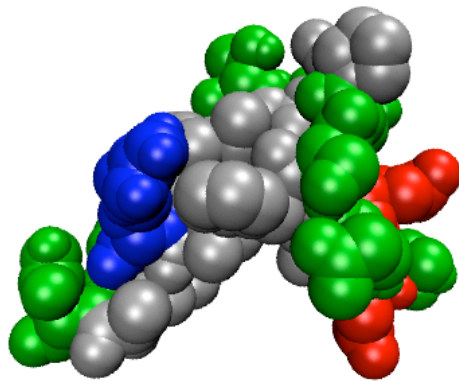
- Dans son article fondateur de 1996, **M. Tirion** propose de donner une forme quadratique à l'énergie intramoléculaire d'un système protéique selon l'équation suivante :

$$E = \frac{1}{2} \sum_{r_{ij}^0 < R_{cut}} \gamma (r_{ij} - r_{ij}^0)^2$$

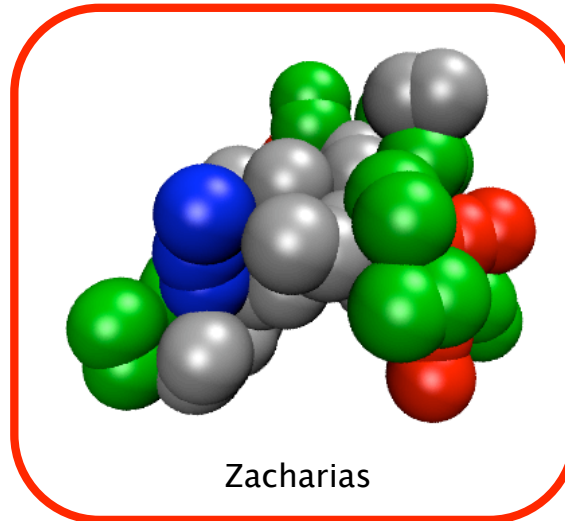
- Ce modèle simplifié à l'extrême permet de reproduire les modes de vibration basse fréquence et des facteurs de température obtenus par cristallographie.
- Le **Gaussian Network Model** (GNM) est d'une grande **robustesse**, ce qui permet d'utiliser des représentations réduites à différentes échelles permet de rendre compte de nombreux phénomènes protéiques : Dynamique fonctionnelle, structuration en domaines, changements conformationnels de grande amplitude...



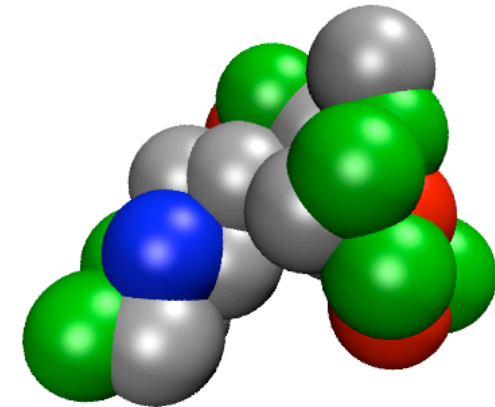
- Ce modèle a été développé pour modéliser les interactions inter-protéiques dans le cadre de simulations d'amarrage protéine-protéine ou **docking**.
- Chaque acide aminé est représenté par 1 à 3 pseudo-atomes en fonction de sa chaîne latérale. (Zacharias, *Prot. Sci.* 2003).



Représentation: Tout atome



Zacharias



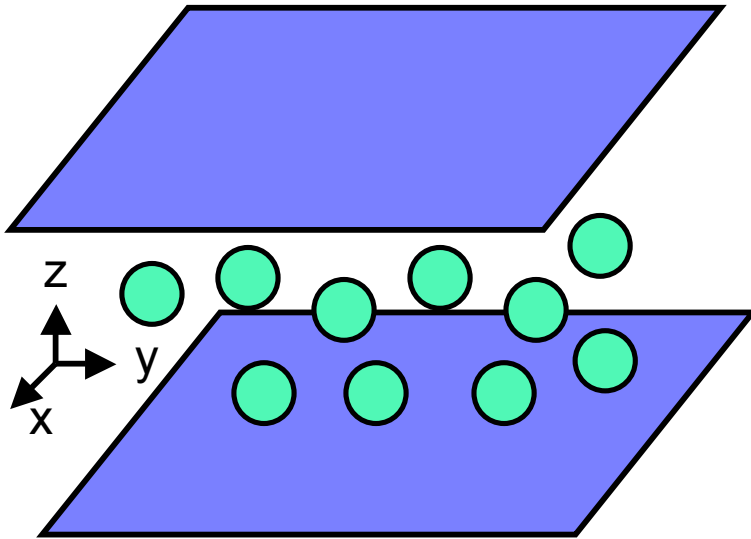
1 pseudo-atome par résidu

- Les interactions entre pseudo-atomes (appartenant à deux protéines distinctes) suivent un potentiel simplifié de la forme :

$$E_{ij} = \left( \frac{B_{ij}}{r_{ij}^8} - \frac{C_{ij}}{r_{ij}^6} \right) + \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}}, \text{ with } \epsilon(r_{ij}) = 15 r_{ij}$$



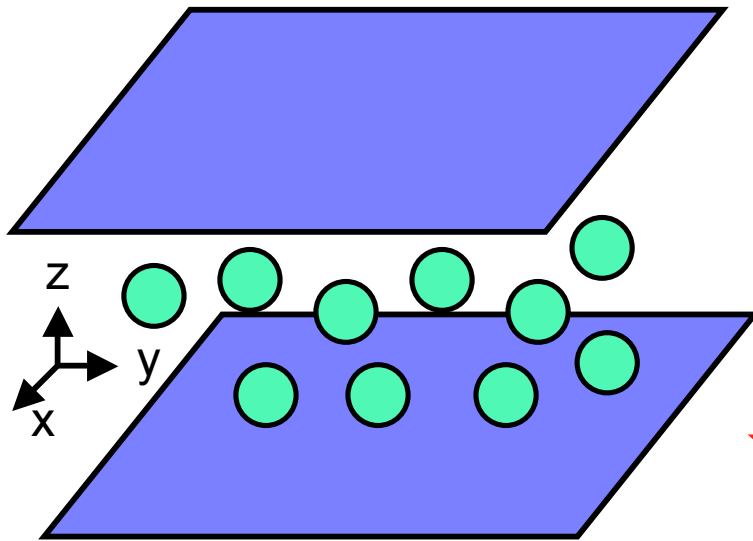
## *Pourquoi des modèles réduits ?*



$$\text{😊} \longleftrightarrow \text{😊} \quad u_{lj} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]$$

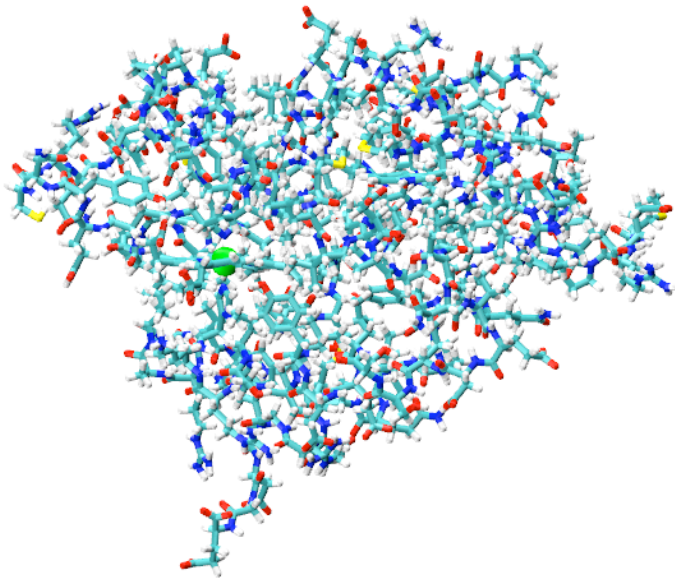
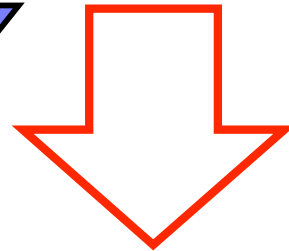
$$\Phi^{[k]}(z) = \varphi_{rep}^{[k]}(z) - \varphi_{att}^{[k]}(z)$$

# Pourquoi des modèles réduits ?



$$\text{😊} \longleftrightarrow \text{😊} \quad u_{ij} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]$$

$$\Phi^{[k]}(z) = \varphi_{rep}^{[k]}(z) - \varphi_{att}^{[k]}(z)$$



$$V(R, \theta, \phi) = \sum V_{intra}(R, \theta, \phi) + \sum V_{inter}(R)$$

$$V_{intra}(R, \theta, \phi) = V_{liaison}(R) + V_{pliage}(\theta) + V_{torsion}(\phi)$$

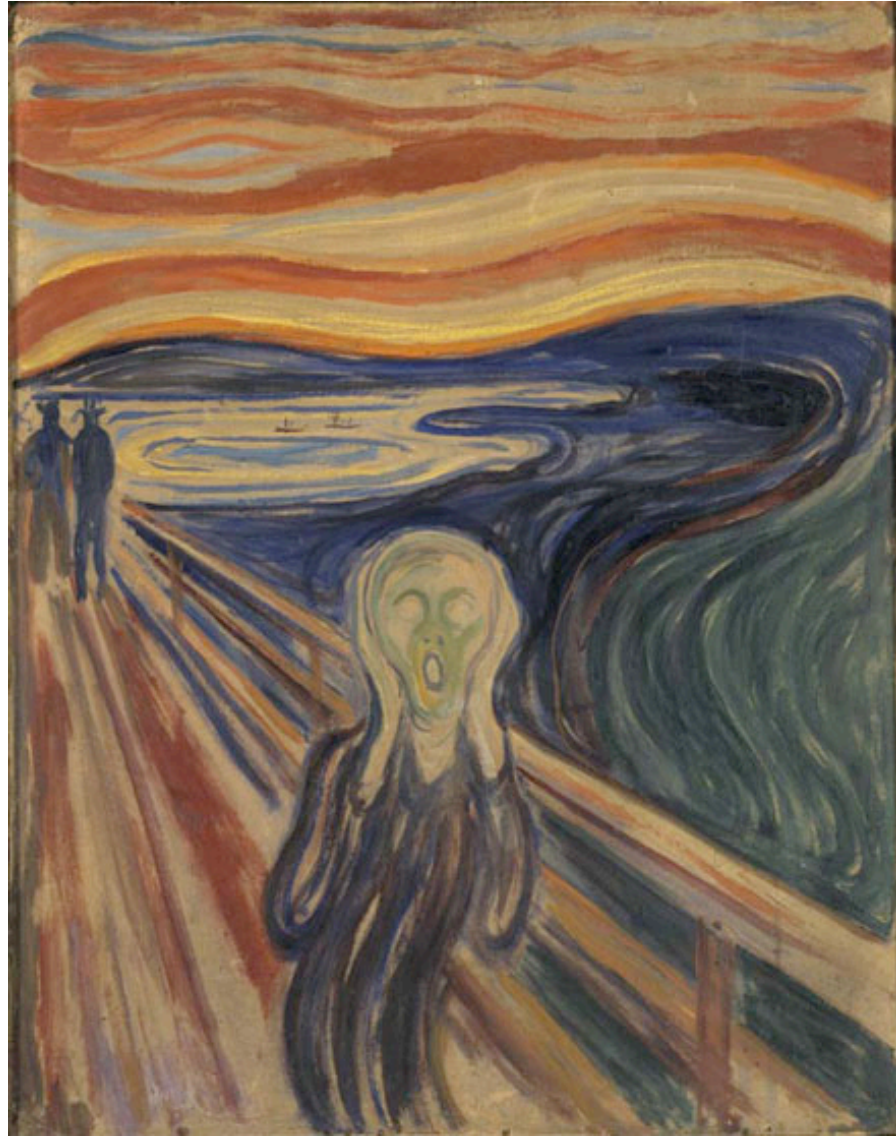
$$V_{intra}(R, \theta, \phi) = \frac{K_R}{2}(R - R_0)^2 + \frac{K_B}{2}(\cos\theta - \cos\theta_0) + \sum_{i=0}^8 a_i \cos^i(\phi)$$

$$V_{inter}(R) = V_{electrostatique}(R) + V_{polarisation}(R) + V_{repulsion-dispersion}(R)$$

$$V_{inter}(R) = \frac{1}{4\pi\epsilon_0} \sum_{i \neq j} \frac{q_i q_j}{R_{ij}} - \frac{1}{2} \mu_i^{ind} E_i^0 + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{R_{ij}} \right)^6 \right]$$

## *Pourquoi des modèles réduits ?*

---



(E. Munch, *Le Cri*, 1893).

**I. Introduction générale sur les modèles gros-grain**

**II. Propriétés mécaniques des protéines :**

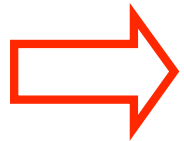
1. Le programme ProPHET
2. Résultats généraux (**utilisation autonome de ProPHET**)
3. Noyau mécanique dans les globines (**utilisation couplée à des simulations tout-atome**)
4. Réponse mécanique à une contrainte externe (**utilisation couplée à des approches expérimentales**)

**III. Interactions protéiques :**

1. Le programme MAXDo
2. Prédiction des partenaires d'interaction protéiques
3. Prédiction des sites d'interaction protéiques

**IV. Conclusions et Perspectives**

- Les seules données structurales sont insuffisantes pour comprendre le fonctionnement biologique des protéines :



**Prendre en compte la mécanique et la dynamique locale**

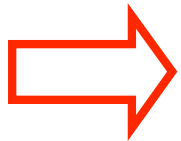
- Approches expérimentales : Facteurs de température (cristallographie), diffusion des neutrons (dynamique globale de la protéine), RMN (paramètres d'ordre)

⇒ Cas des résidus situés près du **site actif** :

Mise en évidence de faibles facteurs de température (et donc d'une flexibilité réduite).

- Approches théoriques : Simulations par Dynamique Moléculaire, étude des Modes Normaux

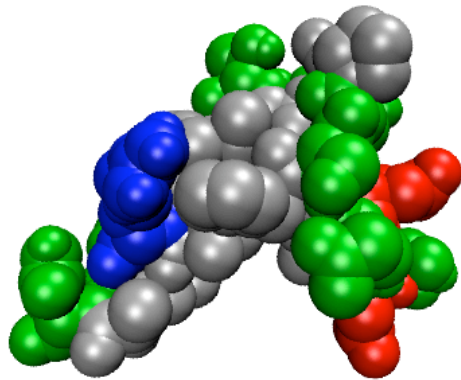
- Développement d'une approche permettant d'attribuer une constante de force à chaque résidu de la protéine.



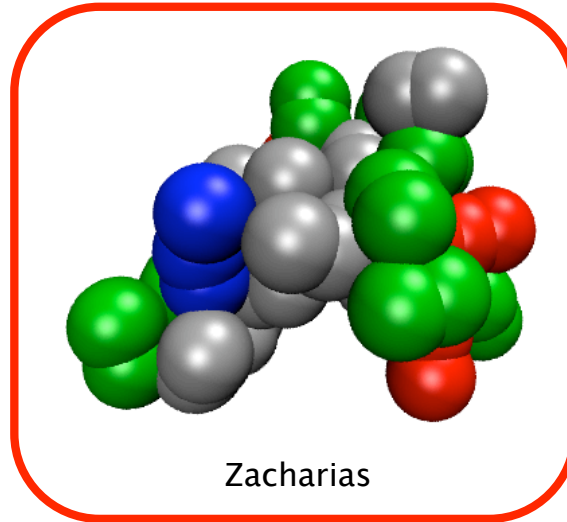
**Logiciel ProPHet (Probing Protein Heterogeneity)**

## *ProPHet : Représentation réduite des protéines:*

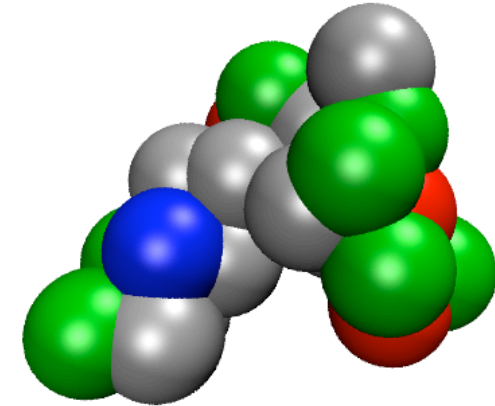
- Chaque acide aminé est représenté par 1 à 3 pseudo-atomes en fonction de sa chaîne latérale. (Zacharias, *Prot. Sci.* 2003).



Représentation: Tout atome

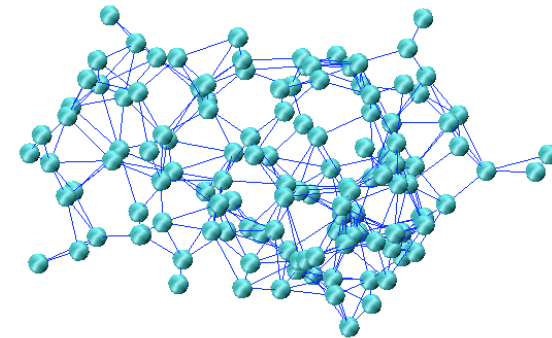
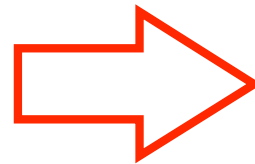
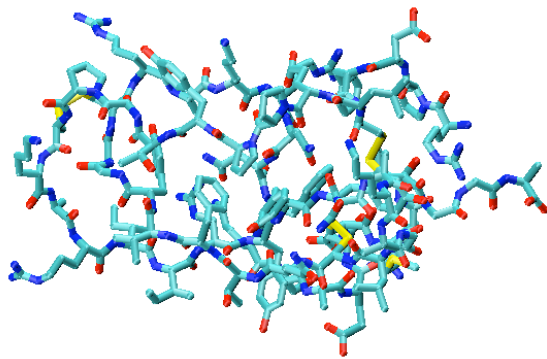


Zacharias



1 pseudo-atome par résidu

- Les interactions entre pseudo-atomes sont représentées par un potentiel harmonique (avec  $r_{\text{cut}}=9 \text{ \AA}$  et  $\gamma=0.6 \text{ kcal mol}^{-1}\text{\AA}^{-2}$ ).



## *ProPHet : Simulations de dynamique brownienne :*

---

$$r_i = r_i^0 + \sum_j \frac{D_{ij}^0 F_{ij}^0}{kT} \Delta t + R_i(\Delta t)$$

$$\langle R_i(\Delta t) \rangle = 0, \quad \langle R_i(\Delta t) R_j(\Delta t) \rangle = 2D_{ij}^0 \Delta t, \quad \Delta t = 10 \text{ fs}$$

- Les simulations entraînent des déformations de la molécule avec un rmsd d'environ 1,5 Å par rapport à la conformation de départ .
- La constante de force associée à chaque particule est calculée à partir des fluctuations autour de sa position d'origine :

Soit  $d_i$  la distance moyenne de la particule  $i$  à toutes les autres particules de la protéine

$$k_i = \frac{3k_B T}{\left\langle \left( d_i - \langle d_i \rangle_{simulation} \right)^2 \right\rangle_{simulation}}$$

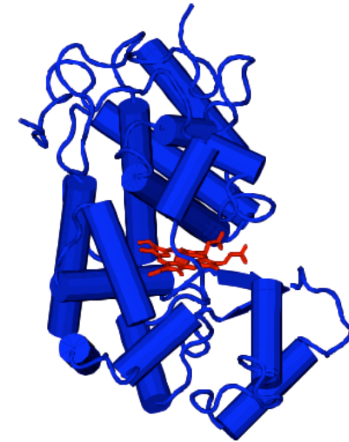
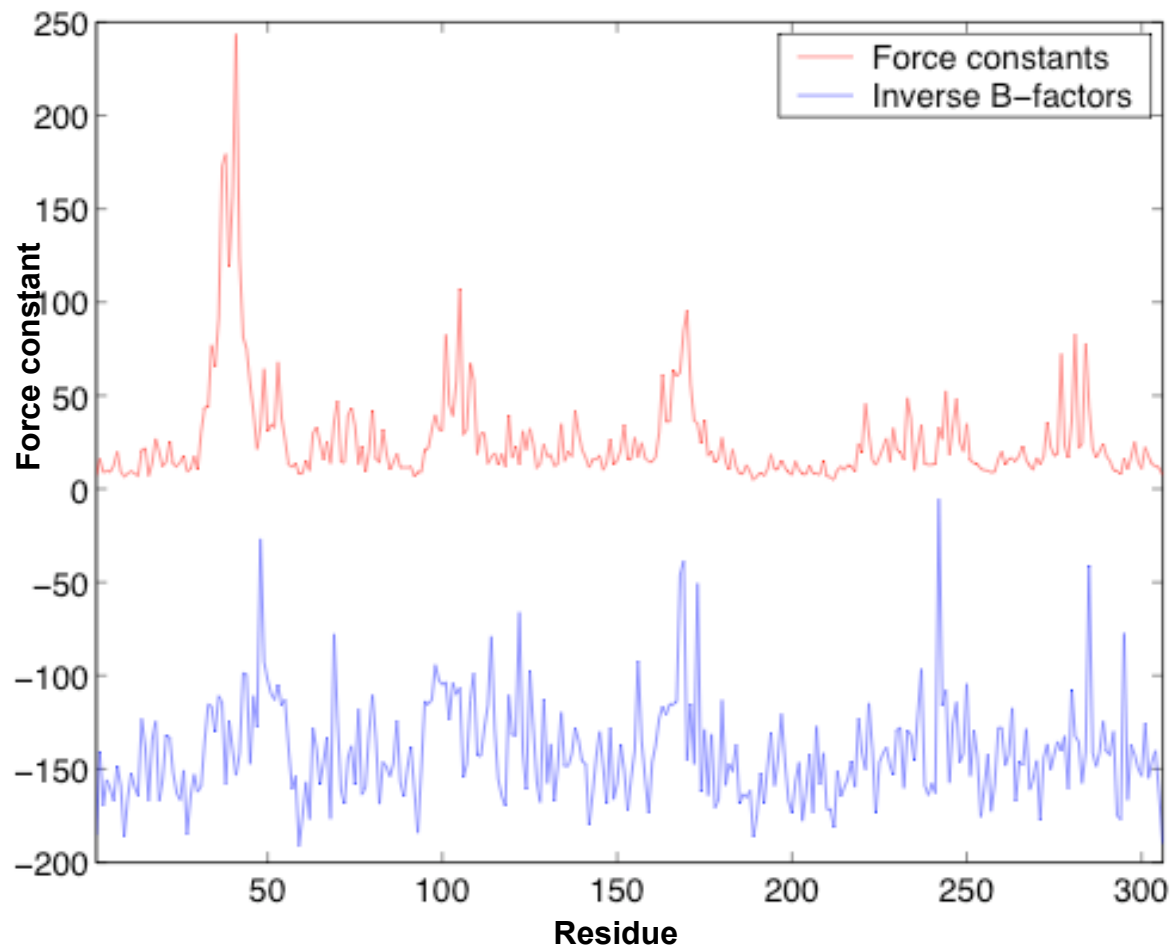
, pour la particule  $i$

$$K_k = \left\langle k_i \right\rangle_{i \in k}$$

, pour le résidu  $k$

# *Profil de rigidité protéique :*

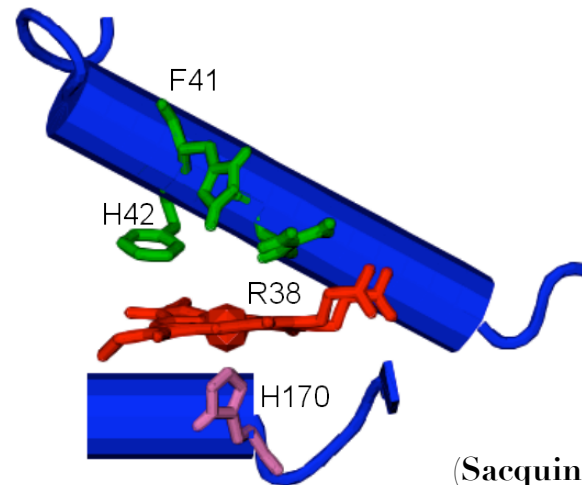
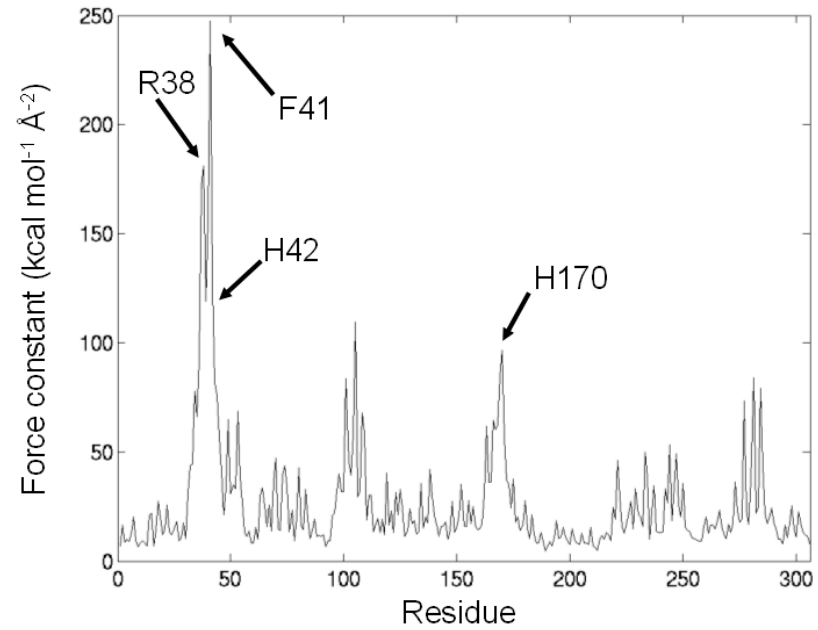
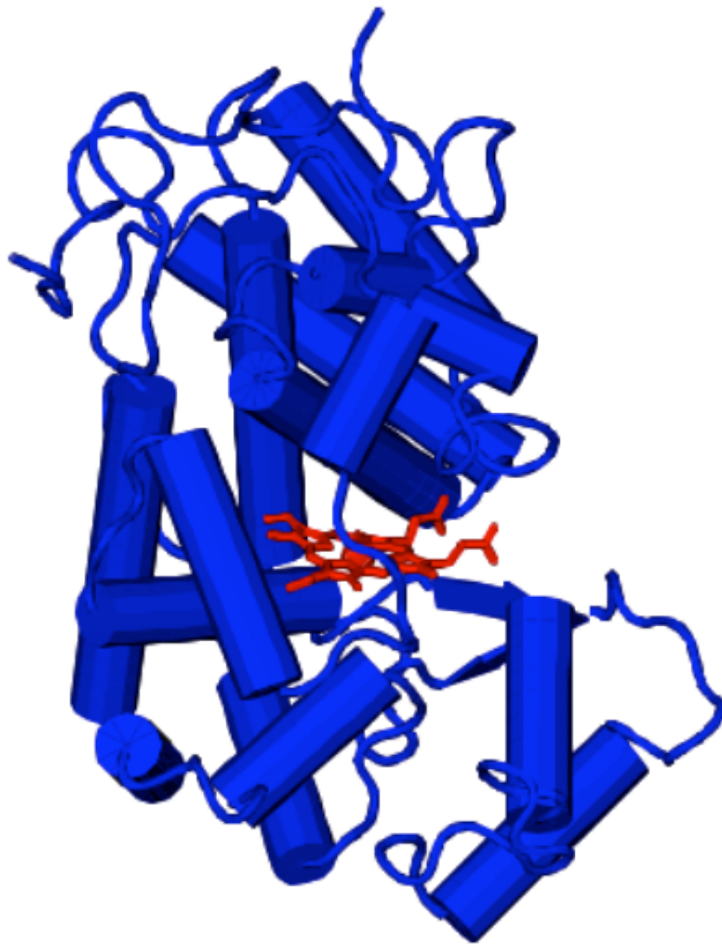
Horseradish peroxidase (1ATJ)





# Mise en évidence des résidus catalytiques :

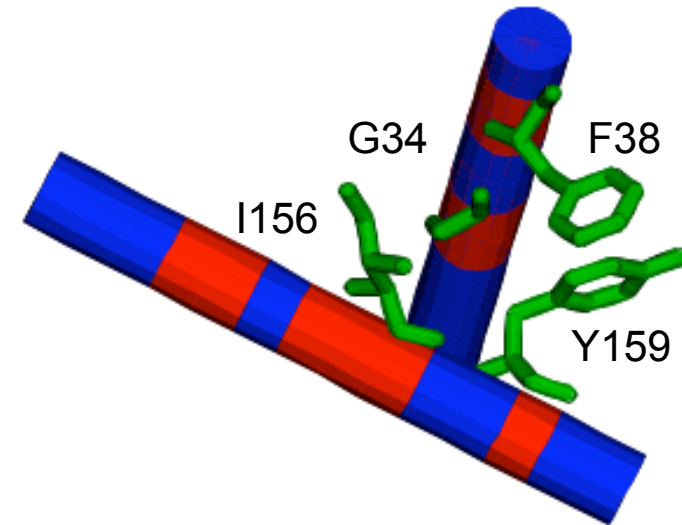
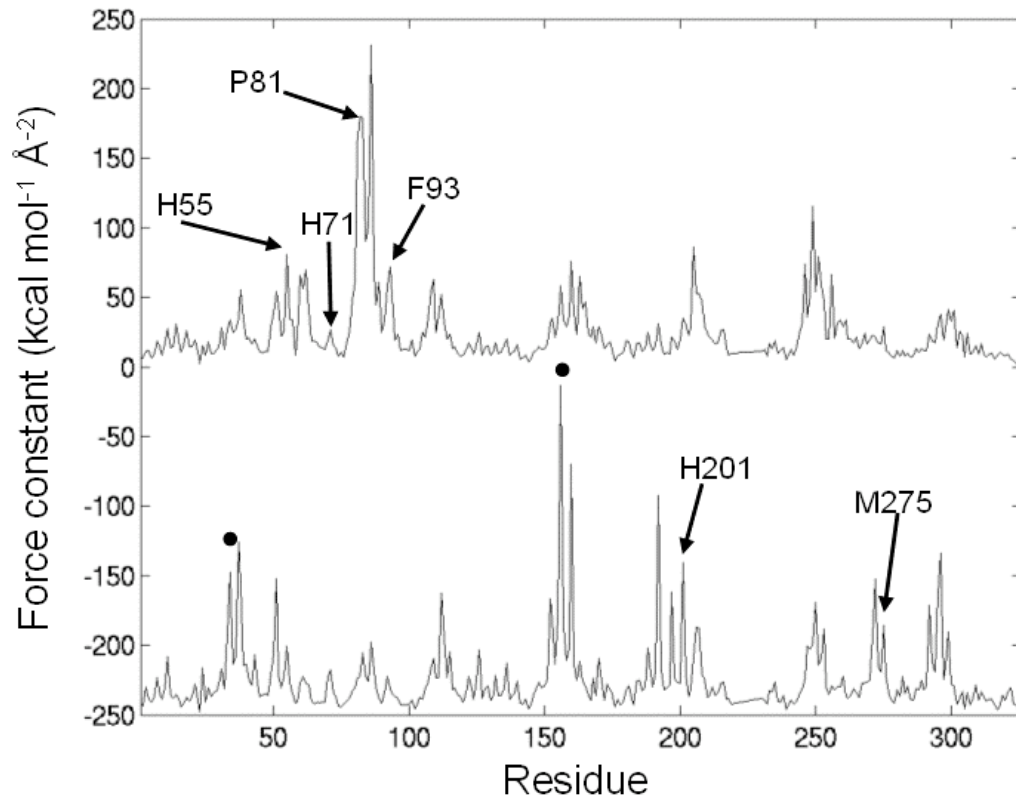
Horseradish peroxidase (1ATJ)



(Sacquin-Mora et al., 2006).

# Détection du noyau de repliement :

Cytochrome *c* peroxidase (1NML)



Résidus clés sur le plan structural

- Hélices- $\alpha$  orthogonales
- Positions (i, i+4) et (j, j+3)
- Groupes aromatiques en i+4 et j+3



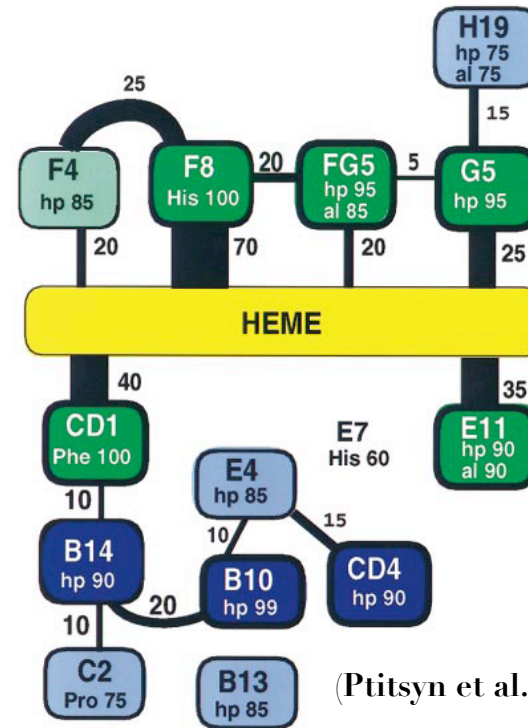
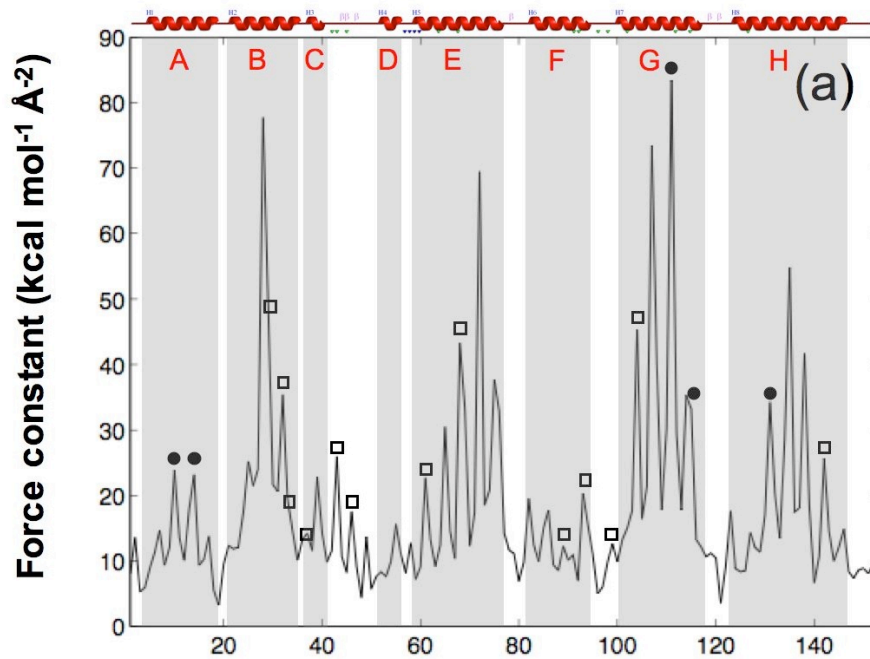
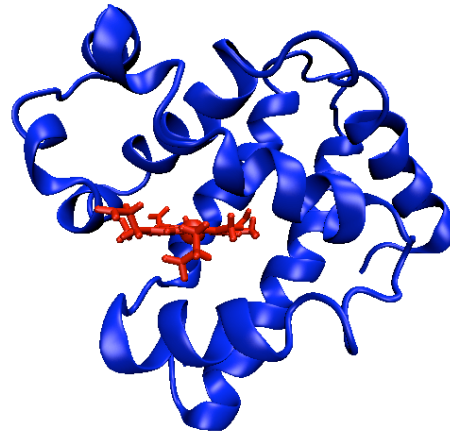
“Noyau de repliement” des  
domaines cytochrome *c*

(Ptitsyn et al., *J. Mol. Biol.*, 1998)

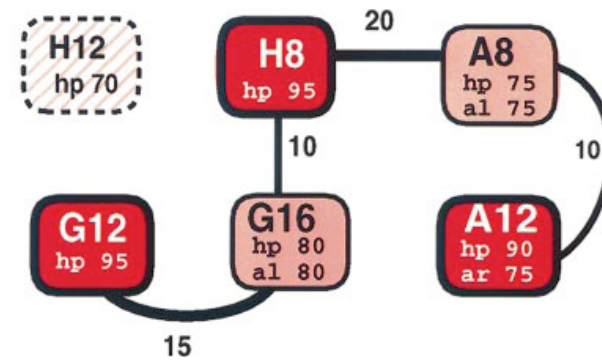
# Profil de rigidité dans les globines :

Myoglobine

(1YMB)



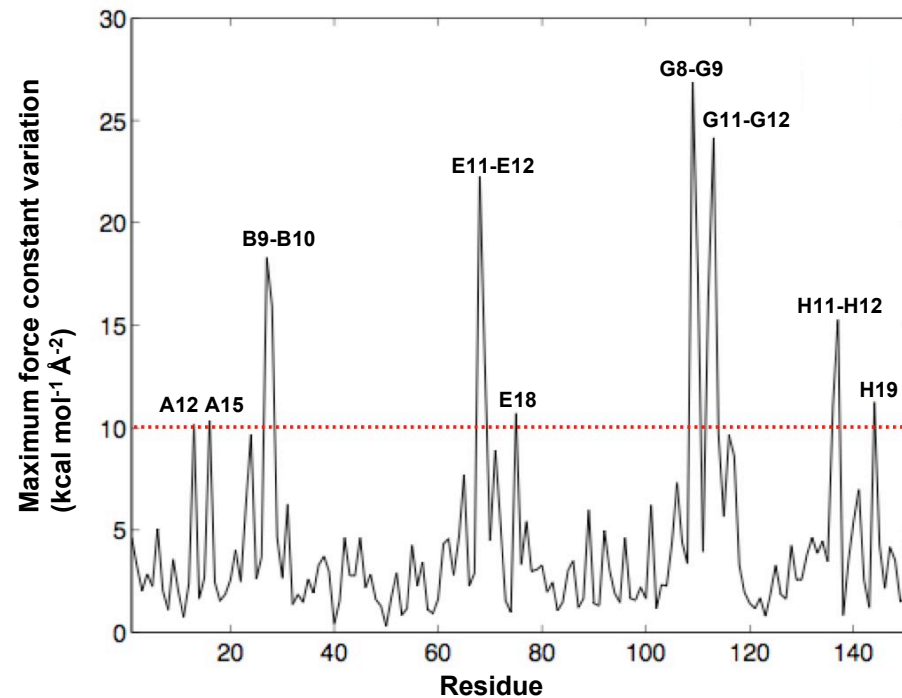
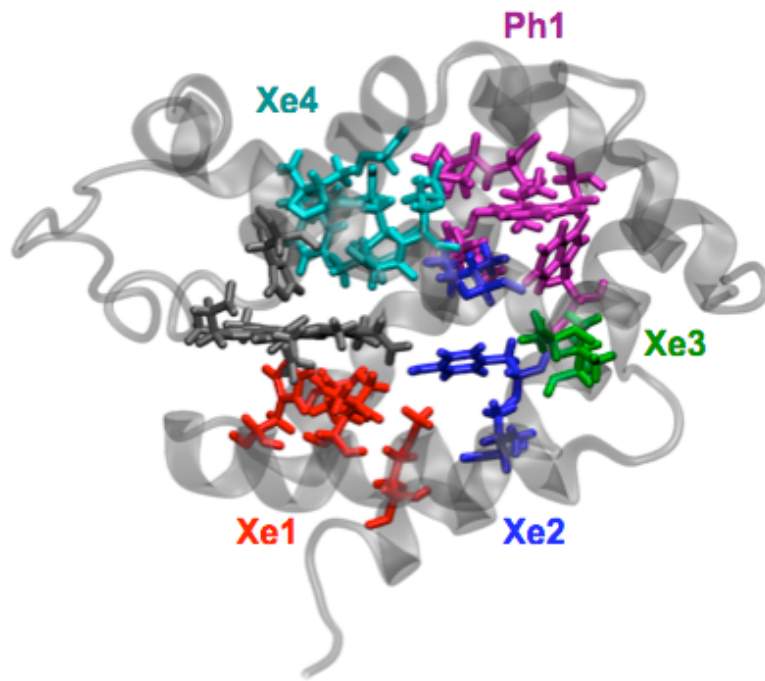
(Ptitsyn et al., *J. Mol. Biol.*, 1999)



## *Résidus frontière et résidus “mécaniquement sensibles” :*

Neuroglobine  
(1OJ6)

- On compare les profils de rigidités de 5 structures caractéristiques produites par des simulations MD tout atome.
- Les fluctuations du profil mettent en évidence un jeu de résidus **mécaniquement sensibles** qui sont systématiquement situés aux frontières entre des cavités internes.



## *Résidus frontière et résidus “mécaniquement sensibles” :*

- L'application de ce protocole à 6 chaînes distinctes (Ngb, Mgb, Cytoglobine, Hémoglobine tronquée, Hémoglobines  $\alpha$  et  $\beta$ ) fait apparaître 4 positions mécaniquement sensibles conservées au sein de la famille des globines : **E11, E15, G8 et G12**

```

1234567890123456 12345678901234567123456          1234561
AAAAAAAAAAAAAAAAA BBBBCCCCCCCCCCCCCCCCCCCC          DDDDDDE
1YMB_A  -GLSDGEWQQVLNVWGKVEADIAGHGQEVLRIRLFTGHPETLEKFDKFK-HLKTEAEMKAS 58
1OJ6_B  --MERPEPELIRQSWRAVSRSPLEHGTVLFARLFALEPDLLPLFQYNCRQFSSPEDCLSS 58
2DC3_A  EELSEAERKAVQAMWARLYANCEDVGVAILVRFVNFPSAKQYFSQFK-HMEDPLEMERS 59
1IDR_B  GLLSRLRKREPISYDKIG--GHEAIEVVVEDFYVRVLADDQLSAFFS-----G 47
2HHB_A  -VLSPADKTNVKAAWGKVGHAHAGEYGAEALERMFLSFPTTKTYFPHF--DLSH-----GS 52
2HHB_B  VHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFG-DLSTPDAVMGN 57

```

```

23456789012345678901          1234567890123456  1234567890123456
EEEEEEEEEEEEEEEEEEEE          FFFFFFFFFFFFFFFFFF  GGGGGGGGGGGGGGGG
1YMB_A  EDLKKHGTVVLTALGGILKKKGH---HEAELKPLAQSHATKHKIPIKYLEFISDAI IHVL 115
1OJ6_B  PEFLDHIRKVMLVIDAAVTNVEDLSSLEEYLASLGRKHR-AVGVKLSSFSTVGESLLYML 117
2DC3_A  PQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVV 119
1IDR_B  TNMSRLKKGQVEFFAALGGPEP-----YTGAPMKQVHQ-GRGITMHHFSLVAGHLADAL 101
2HHB_A  AQVKGHGKKVADALTNAVAHVDD---MPNALSALSDLHAHKL RVD PVNFKL LSHCLLVTL 109
2HHB_B  PKVKAHGKKVLGAFSDGLAHLDN---LKGTFATLSELHCDKLHVDPENFRLLGNVLCVVL 114

```

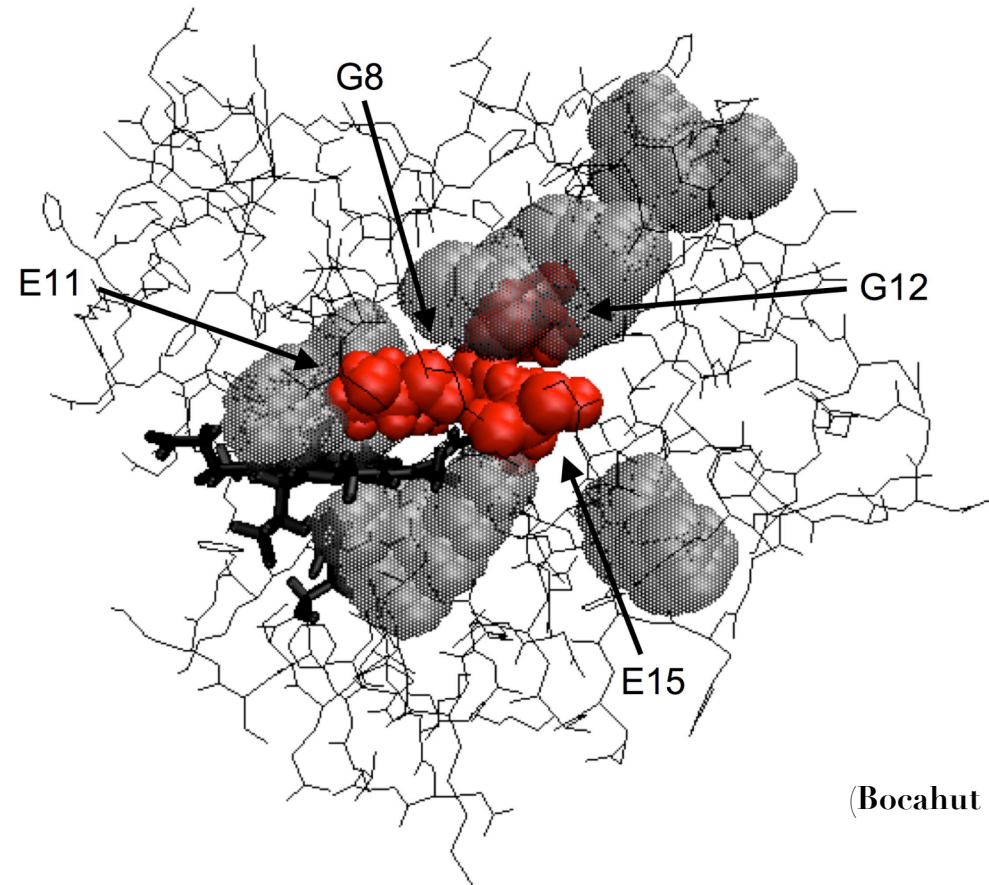
```

7890          1234567890123456789012345678
GGGG          HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
1YMB_A  HSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG 153
1OJ6_B  EKCLGPAFTPATRAAWSQLYGAVVQAMSRGWDGE---- 151
2DC3_A  AEEFASDFPPEQRAWAKLRGLIYSHVTAAYKEVGW-- 155
1IDR_B  T---AAGVPSETITEILGVIAPLAVDVT----- 126
2HHB_A  AAHLPAEFTPAVHASLDFKFLASVSTVLT SKYR----- 141
2HHB_B  AHHFGKEFTPPVQAAYQKVAGVANALAHKYH----- 146

```

## *Noyau mécanique des globines :*

---



(Bocahut et al., *JACS*, 2011)

- Les résidus constitutifs du **noyau mécanique** des globines jouent un rôle dans le contrôle de la diffusion de petits ligands au sein du réseau de cavités internes et ont tous été déjà signalés comme des résidus clés lors d'études expérimentales sur les globines.

## *Conclusion et perspectives :*

---

- Le logiciel ProPHet combine une représentation réduite des protéines et un algorithme de dynamique brownienne pour **sonder les propriétés mécaniques des protéines à l'échelle du résidu**.
- Les profils ainsi obtenus permettent de pointer des **résidus clés sur les plans structural et fonctionnel**, via leur rigidité accrue.
- Le profil de rigidité d'une protéine présente des fluctuations dynamiques, qui permettent de cerner un petit nombre de **résidus mécaniquement sensibles**, dont la rigidité varie de manière importante au cours du temps.
- Dans le cas des globines, ces résidus sont conservés et correspondent aux **résidus frontières** situés aux interfaces entre cavités internes. Les positions E11-E15-G8-G12 forment notamment un **noyau mécanique** essentiel pour réguler la diffusion de ligands internes.
- Application à d'autres familles de protéines globulaires ?

## *Réponse enzymatique à une contrainte externe :*

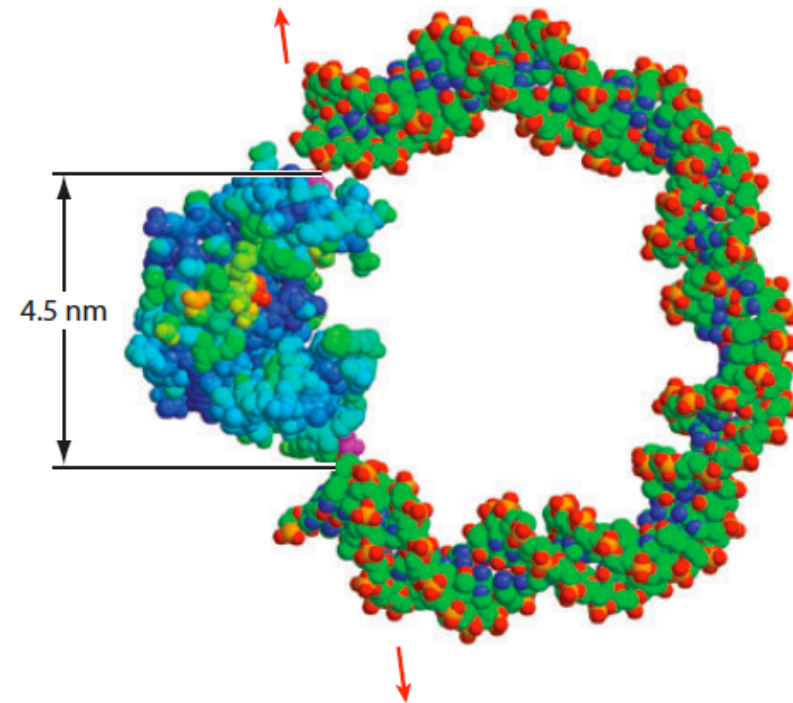
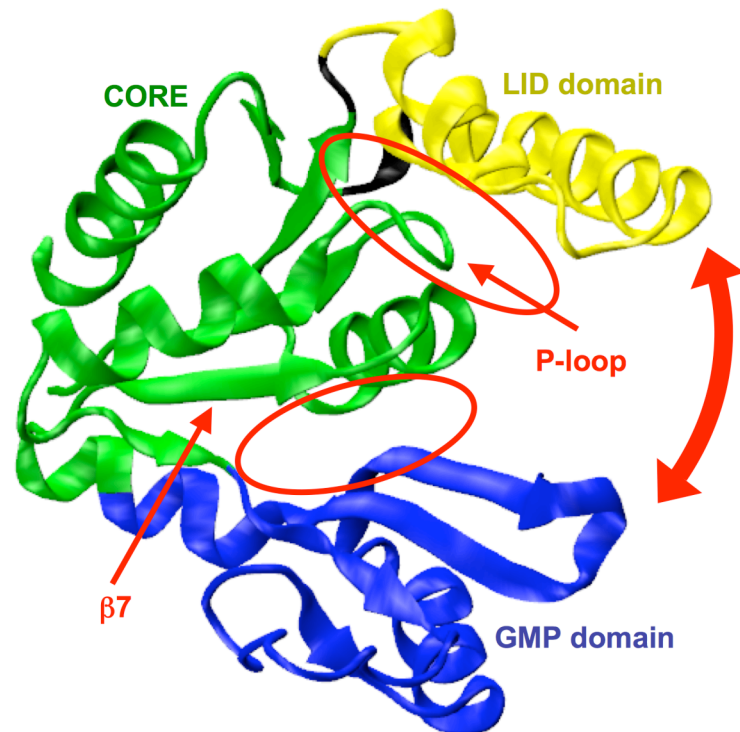
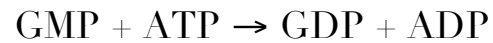
---

- Développement récent des approches de type **molécule unique**, microscopie à force atomique, pinces optiques ou magnétiques, pour l'étude de la mécanique des protéines (repliement, anisotropie des éléments de structure secondaire).
- Jusqu'ici les perturbations de l'activité enzymatique étaient liées à **des changements structuraux importants**.
- La conception de « chimères » protéine-ADN permet d'évaluer la réponse fonctionnelle de la protéine soumise à **une contrainte externe de faible amplitude**.



# Réponse enzymatique à une contrainte externe :

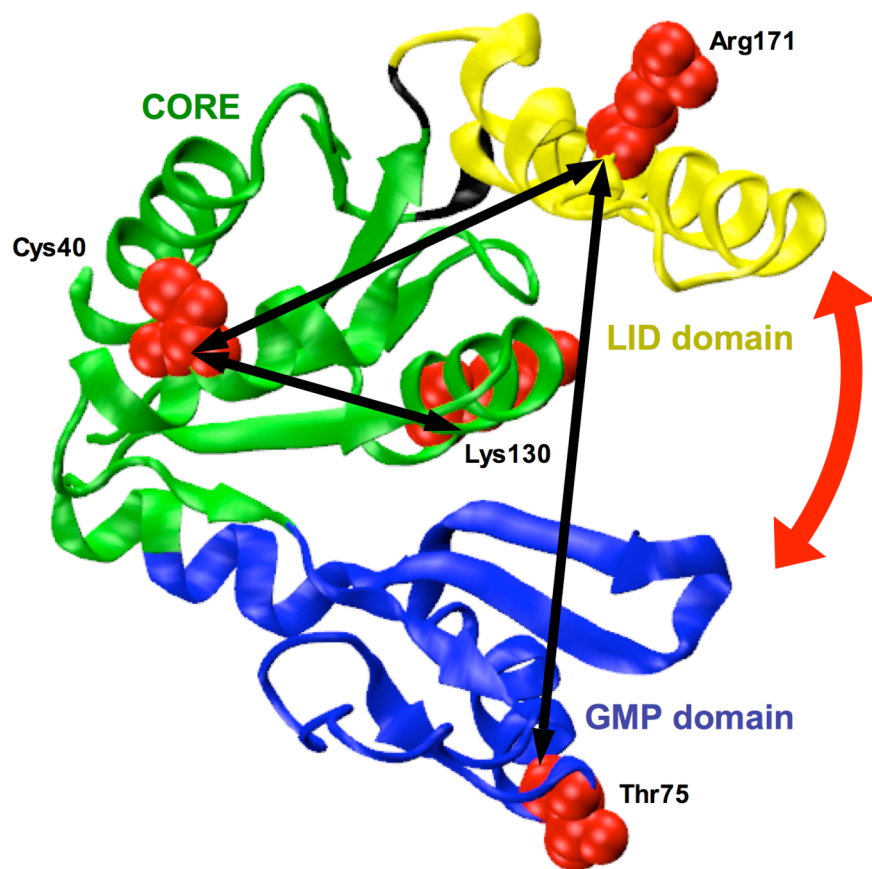
## *M.T* Guanylate Kinase :



(Zocchi, *Annu. Rev. Biophys.*, 2009)

- Trois directions ont été testées expérimentalement, et on suit l'évolution de trois paramètres par spectroscopie:  $K_G$  (GMP binding),  $K_A$  (ATP binding) et  $k_{cat}$  (taux catalytique).

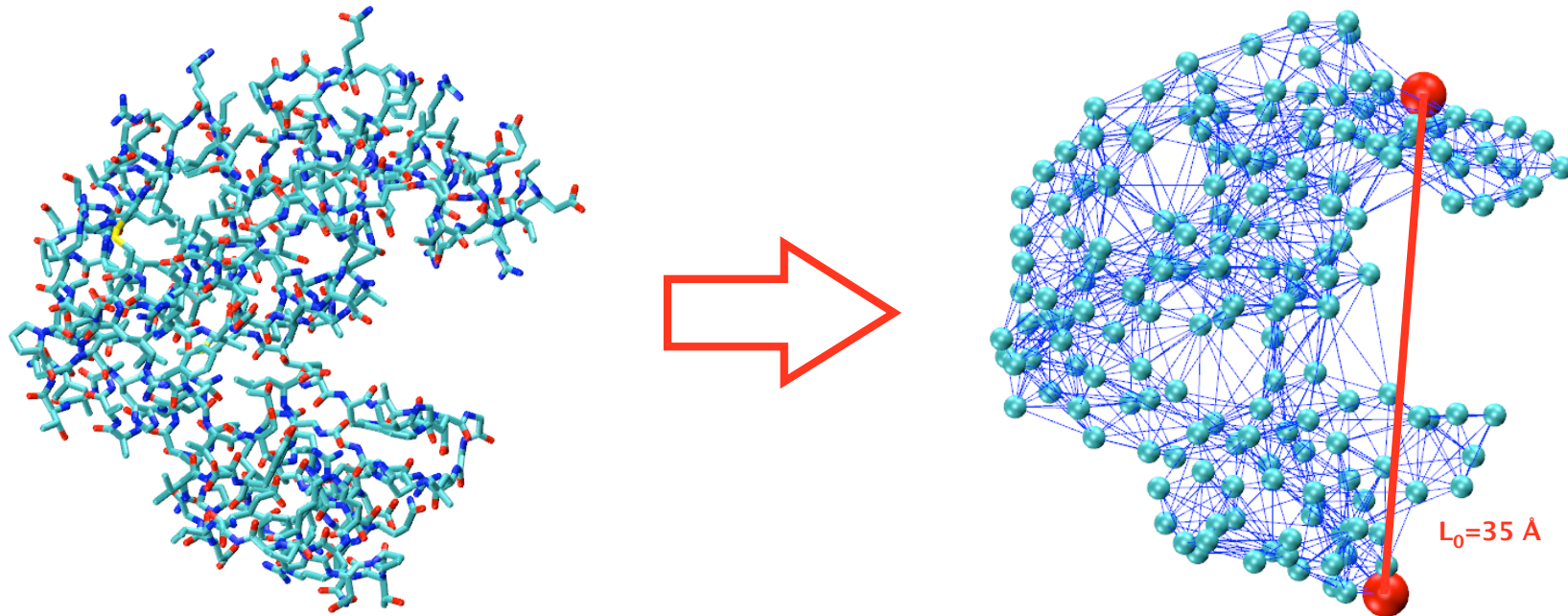
# Réponse enzymatique à une contrainte externe :



Direction	$K_G$	$K_A$	$k_{cat}$
75/171	$K_G$ augmente (affinité pour GMP diminue)	-	-
40/171	-	-	$k_{cat}$ diminue
30/140	-	-	-

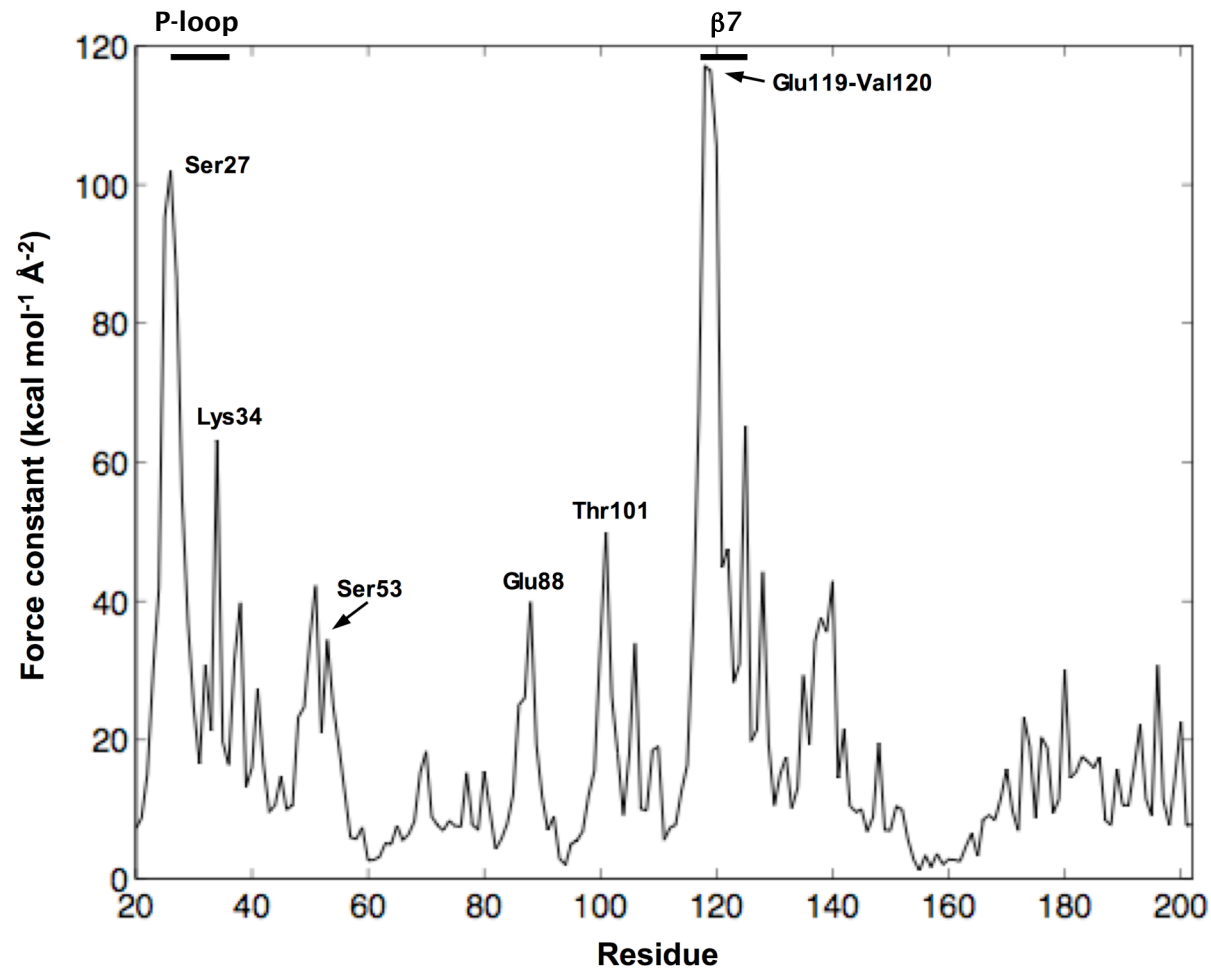
## *Modification de ProPHEt :*

- La contrainte externe est modélisée par un ressort supplémentaire dont les paramètres sont déterminés par les données expérimentales ( $L_C=150 \text{ \AA}$  et  $k_C=1,2 \text{ kcal mol}^{-1}\text{\AA}^{-2}$ )



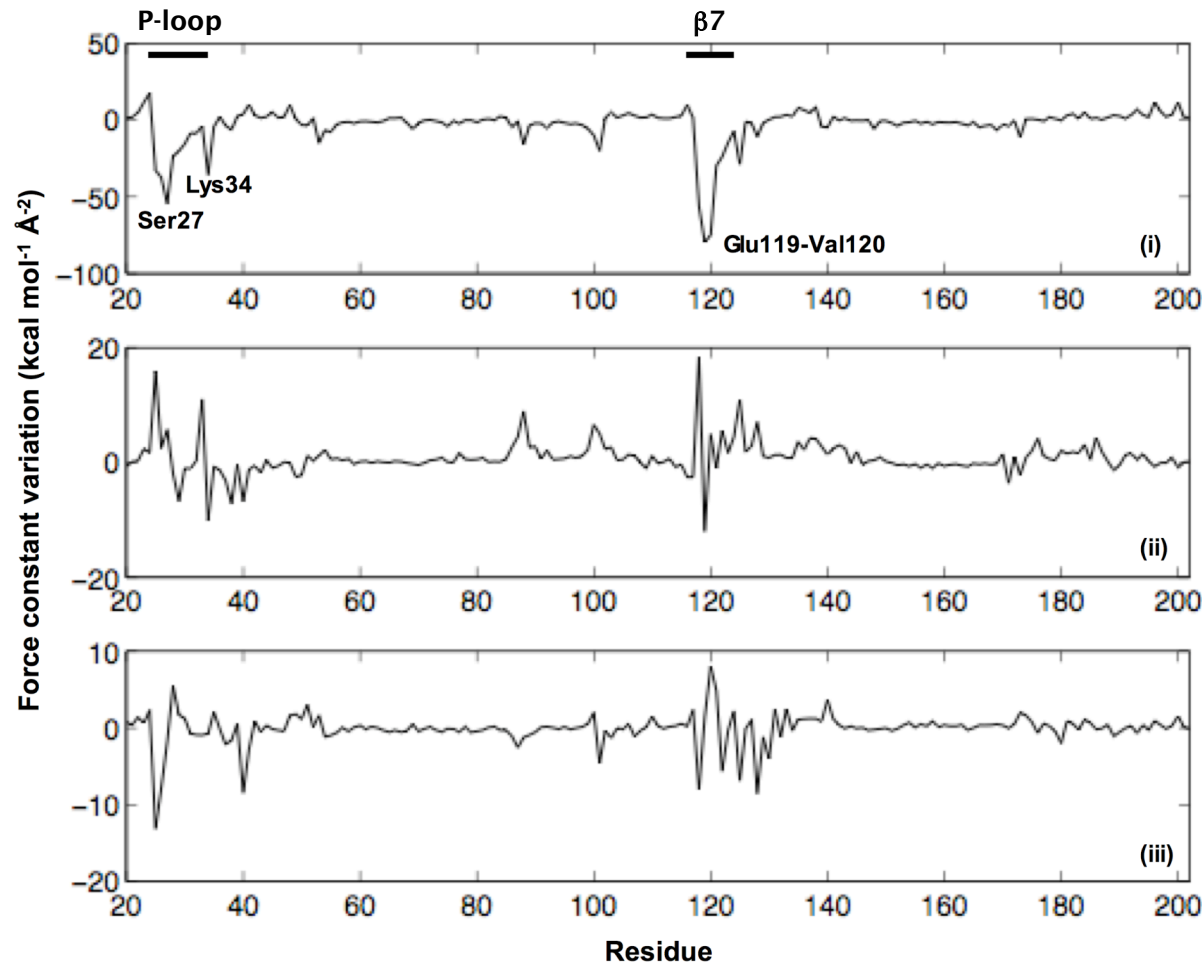
## *Propriétés mécaniques de GK :*

- Profil de rigidité :



# Propriétés mécaniques de GK :

- Variations du profil de rigidité:



75/171

Distorsion importante du site de fixation de GMP

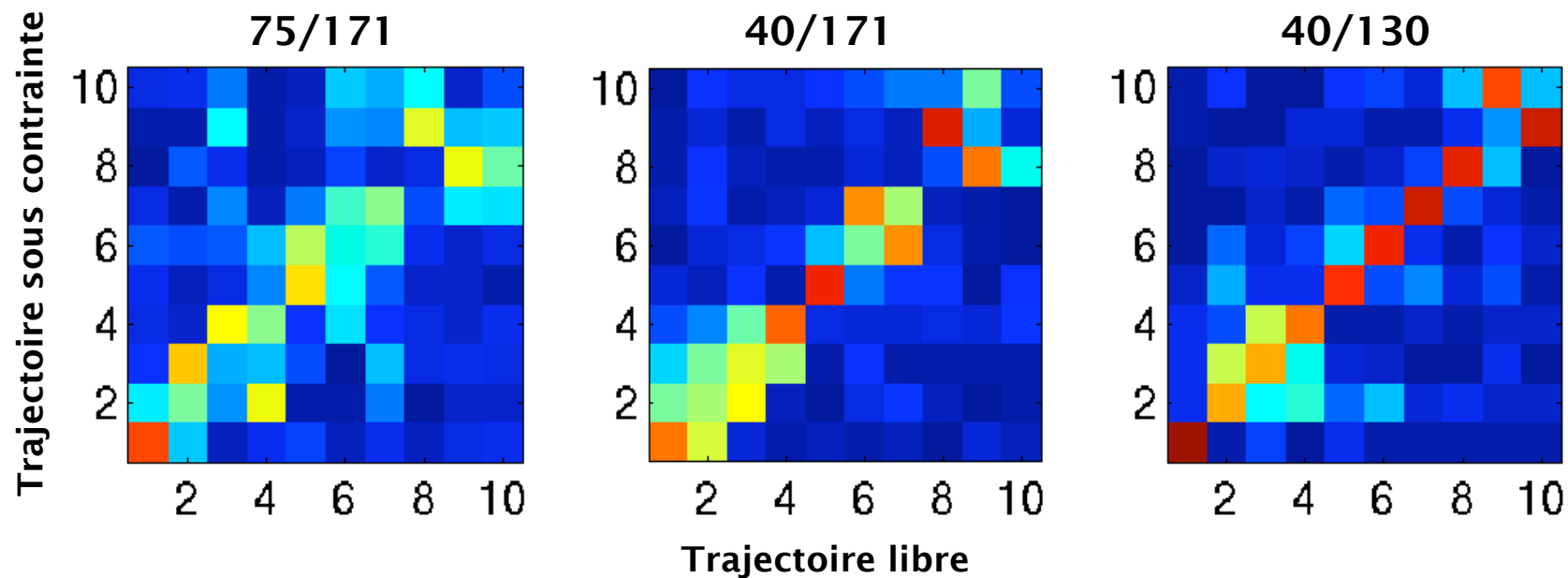
⇒ Effet sur  $K_G$

40/171

40/130

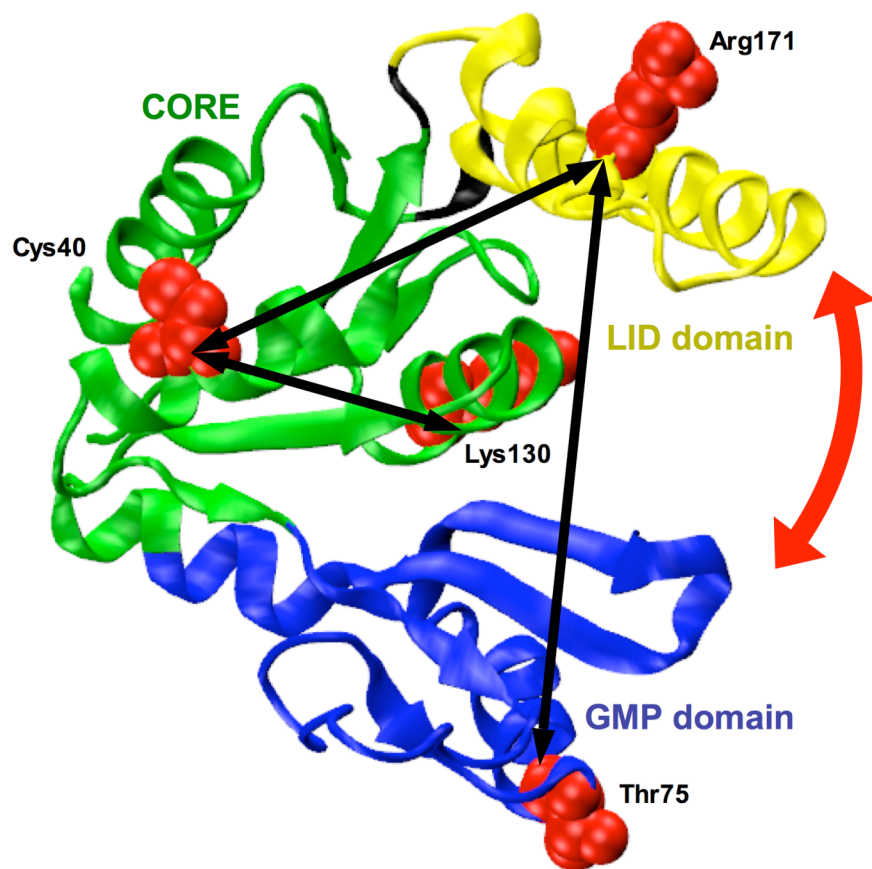
## *Dynamique sous contrainte :*

- On effectue une analyse par composantes principales (PCA) des dynamiques sous contrainte que l'on compare à la trajectoire de la protéine libre.
- On s'intéresse plus particulièrement à la conservations du **premier mode normal** de vibration de la protéine (mouvement d'ouverture/fermeture de l'enzyme)



CMO	0.655	0.803	0.829
P <sub>1</sub>	0.828	0.771 ⇒ Effet sur k <sub>cat</sub>	0.972

# Réponse enzymatique à une contrainte externe :



Direction	$K_G$	$K_A$	$k_{cat}$
75/171	$K_G$ augmente (affinité pour GMP diminue)	-	-
40/171	-	-	$k_{cat}$ diminue
30/140	-	-	-

## *Conclusion et perspectives :*

---

- Via une approche simple associant un **modèle réduit** pour la représentation des protéines et des simulations par **dynamique brownienne** on peut faire le lien entre les propriétés **mécaniques** et **dynamiques** de GK et son **activité enzymatique**.
- Il semble que la variation de  $K_G$  soit liée à la perturbation **mécanique locale** du site de fixation de GMP.
- La variation de  $k_{cat}$  viendrait elle d'une perturbation **globale de la dynamique** de la protéine (mouvement d'ouverture/fermeture).
- Via un scan de la surface protéique on peut proposer une nouvelle direction de contrainte Asp65-Leu122 qui entraînerait des variations de  $K_G$  **et**  $k_{cat}$ .
- Par contre on ne peut isoler de direction perturbant le seul site de fixation de l'ATP.



Contrôle de l'activité enzymatique par des « mutants mécaniques » ?



**I. Introduction générale sur les modèles gros-grain**

**II. Propriétés mécaniques des protéines :**

1. Le programme ProPHET
2. Résultats généraux (**utilisation autonome de ProPHET**)
3. Noyau mécanique dans les globines (**utilisation couplée à des simulations tout-atome**)
4. Réponse mécanique à une contrainte externe (**utilisation couplée à des approches expérimentales**)

**III. Interactions protéiques :**

1. Le programme MAXDo
2. Prédiction des partenaires d'interaction protéiques
3. Prédiction des sites d'interaction protéiques

**IV. Conclusions et Perspectives**

- **L'identification de partenaires protéiques potentiels** et la **compréhension des interactions protéiques** constituent un problème fondamental en biologie avec des conséquences immédiates dans le domaine du « drug design ».
- Dans le cadre du programme Décrypthon (mis en place par le CNRS, l'AFM et IBM), nous avons développé un projet autour trois axes :
  - ✓ **Développement d'un algorithme de docking** pour la prédiction des interactions protéine-protéine (S. Sacquin-Mora, R.Lavery).
  - ✓ Identification des sites d'interaction protéiques avec la méthode **Evolutionary Trace** (A. Carbone, S. Engelen).
  - ✓ **Mise en place d'une grille de calcul parallèle** permettant le traitement simultané d'un très grand nombre de données (R. Bolze, World Community Grid).

- **L'identification de partenaires protéiques potentiels** et la **compréhension des interactions protéiques** constituent un problème fondamental en biologie avec des conséquences immédiates dans le domaine du « drug design ».
- Dans le cadre du programme Décrypthon (mis en place par le CNRS, l'AFM et IBM), nous avons développé un projet autour trois axes :

✓ **Développement d'un algorithme de docking** pour la prédiction des interactions protéine-protéine (S. Sacquin-Mora, R.Lavery).

✓ Identification des sites d'interaction protéiques avec la méthode **Evolutionary Trace** (A. Carbone, S. Engelen).

✓ **Mise en place d'une grille de calcul parallèle** permettant le traitement simultané d'un très grand nombre de données (R. Bolze, World Community Grid).

# « Ceci n'est pas du Docking »

On s'intéresse à la détermination des partenaires protéiques potentiels au sein d'une base de données (**pré-docking**),

**PAS** à leur agencement détaillé dans le complexe (organisation des chaînes latérales...).

Pour cela on va « détourner » des outils de docking rigide usuels dans le cadre du programme **MAXDo** (Macromolecular Association via Cross-Docking)

# Une introduction très générale au docking :

**Objectif** : Retrouver la structure tridimensionnelle d'un assemblage macromoléculaire (ici des complexes protéine-protéine) à partir des composants séparés.

Les structures de départ proviennent de :

- Protein Data Bank : rayons X, RMN
- Modèle de reconstruction (*par homologie*)

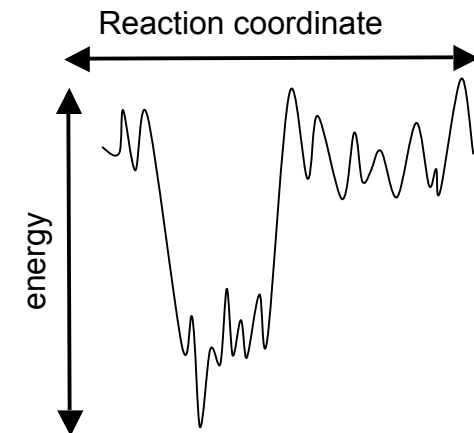
Ces structures non liées peuvent être **légèrement** ou **TRÈS** différentes de la structure protéique au sein du complexe:



On se limite ici aux problèmes de docking **rigide**

Afin de trouver la conformation ligand/récepteur optimale (d'énergie minimale), on doit explorer un **espace conformationnel complexe** (6 variables)

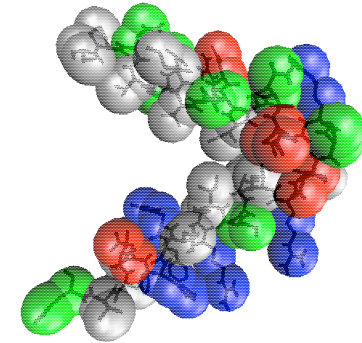
**Comment réduire la complexité du problème ?**  
(et les temps de calcul)



## MAXDo : Représentation réduite

---

- Chaque acide aminé est représenté par 1 à 3 pseudo-atomes  
(Zacharias, 2003)



- Potentiel d'interaction **intermoléculaire** :

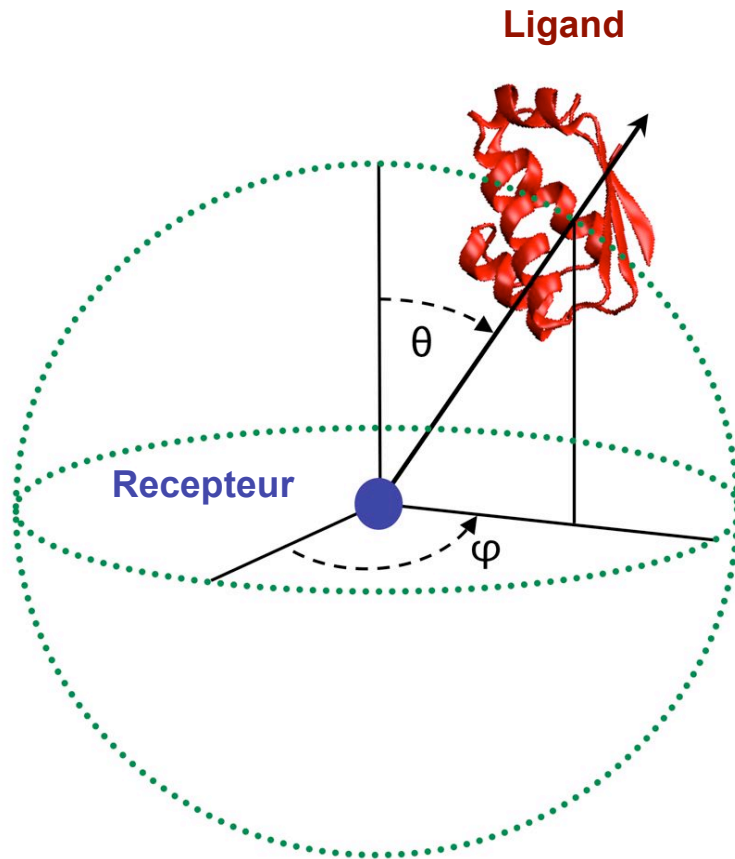
$$E_{ij} = \left( \frac{B_{ij}}{r_{ij}^8} - \frac{C_{ij}}{r_{ij}^6} \right) + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}, \text{ with } \varepsilon(r_{ij}) = 15 r_{ij}$$

Où  $i$  et  $j$  sont deux pseudo atomes appartenant à deux protéines **distinctes**,

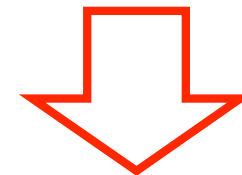
$r_{ij}$  est la distance entre  $i$  et  $j$ ,

$B_{ij}$  et  $C_{ij}$  dépendent des propriétés physico chimiques des résidus auxquels appartiennent  $i$  et  $j$ .

# MAXDo : Algorithme de docking



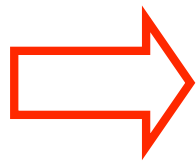
- Afin de trouver la meilleure configuration possible ligand/récepteur, on effectue une **minimisation multiple** en partant d'un très grand nombre de points de départ
- le nombre de positions initiales du ligand dépend de la taille du récepteur  
(1 point/ $10\text{\AA}^2$ )
- Pour chaque position initiale du ligand, on considère 210 orientations de départ



**Au total on a entre  $10^5$  et  $5 \cdot 10^5$  minimisations d'énergie à faire par couple protéique**

- La minimisation de la position du ligand se fait **sous contrainte** afin d'obtenir une **carte de potentiel de la surface du récepteur**

- Les calculs de docking sont effectués pour l'ensemble des protéines contenues dans la base de données de Mintseris et al. (*Proteins*, 2005) qui comprend 84 complexes protéiques (et 168 protéines distinctes).
- On effectue ces calculs pour les partenaires expérimentaux, mais aussi pour l'ensemble des couples de protéines possibles (soit environ  $150^2=22500$  couples)



### « Docking croisé à grande échelle »

- Une série de calculs préliminaires portant sur une base restreinte de six complexes a été effectuée sur la grille universitaire mise en place par IBM.
- Les calculs pour l'ensemble de la base sont faits sur une grille d'internautes :

≈7000 ans de calculs sur un PC



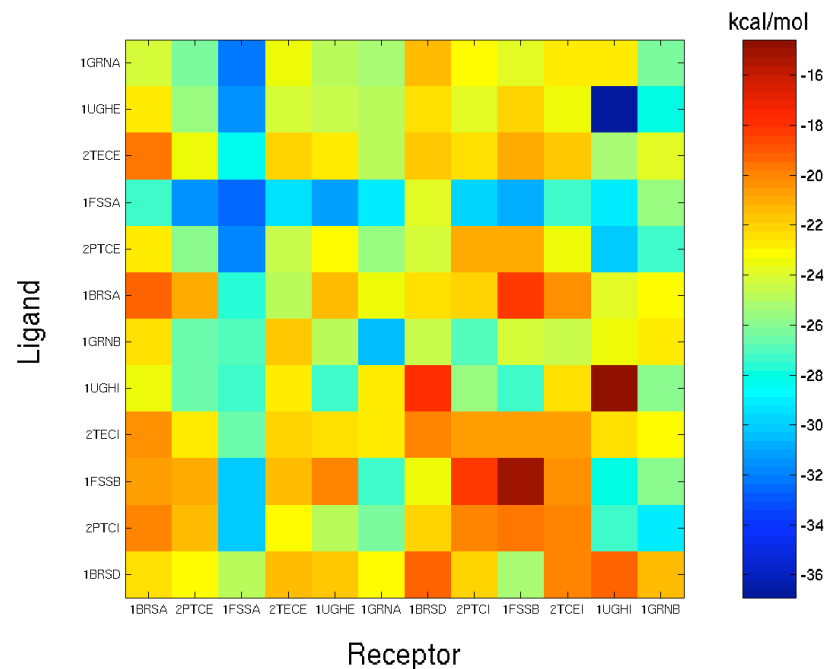
≈5 mois sur WCG (World Community Grid)



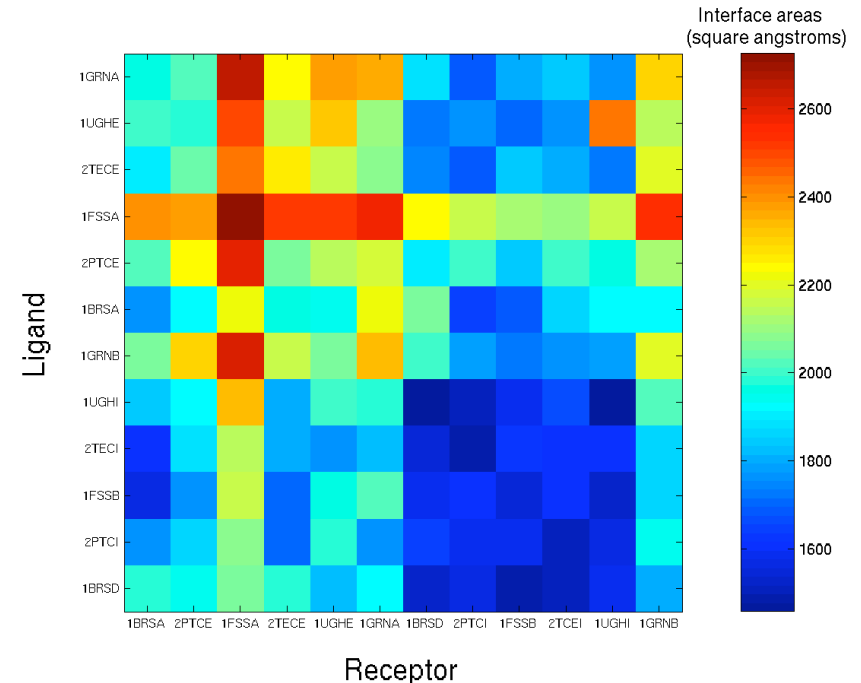
# *Prédiction des partenaires d'interaction (base restreinte) :*

Peut on distinguer interfaces “expérimentales” et “artificielles” ?

(on part des structures liées )



Énergies d'interaction (kcal.mol<sup>-1</sup>)



Aires d'interface (Å<sup>2</sup>)

# Prédiction des partenaires d'interaction (base restreinte) :

Peut on distinguer interfaces “expérimentales” et “artificielles” ?

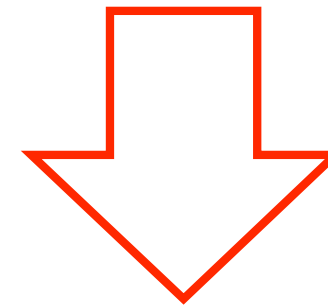
(on part des structures liées)

Indice d'interaction :

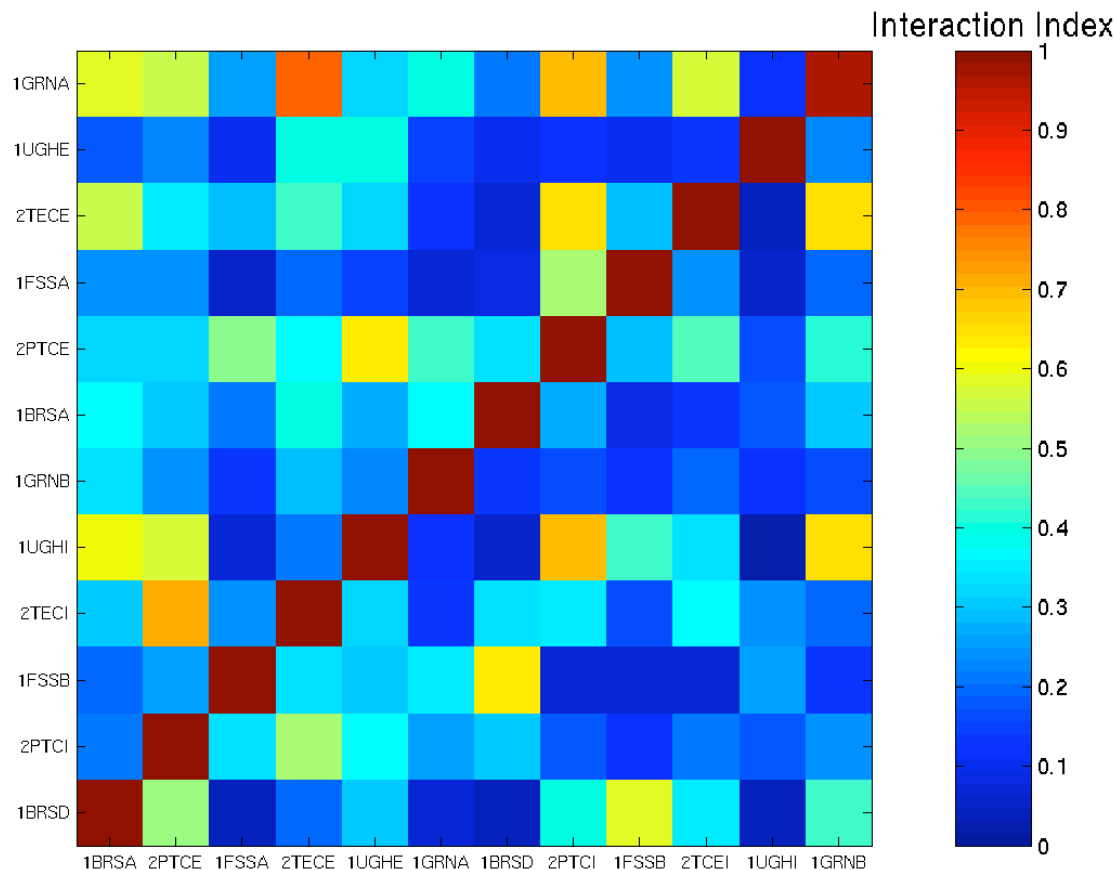
$$I = f_{\text{rec}} * f_{\text{lig}} * E_{\text{int}}$$

$f_{\text{rec}}$  = Fraction de résidus d'interfaces (récepteur)

$f_{\text{lig}}$  = Fraction de résidus d'interfaces (ligand)



On a besoin de connaître les résidus d'interface



## Prédiction des sites d'interaction protéiques :

Pour une couple protéique donné  $P_1/P_2$ , on analyse l'ensemble des interfaces générées lors du docking et pour chaque résidu  $i$  appartenant au ligand  $P_2$  on calcul deux indices

- **Interface Propensity,**

Les résidus appartenant

aux **“bonnes interfaces”**

(basses en énergie) sont favorisés

$$IP_{P_1P_2}(i) = \frac{\sum_{j \in N_{\text{int}, P_1P_2}(i)} \exp\left(-\frac{E(j) - E_{\min}}{RT}\right)}{\sum_{j \in N_{\text{pos}, P_1P_2}} \exp\left(-\frac{E(j) - E_{\min}}{RT}\right)},$$

- **Normalized Interface Propensity,**

$$NIP_{P_1P_2}(i) = \frac{IP_{P_1P_2}(i) - \langle IP_{P_1P_2} \rangle_{i \in \text{Surf } 2}}{\max(IP_{P_1P_2})_{i \in \text{Surf } 2} - \langle IP_{P_1P_2} \rangle_{i \in \text{Surf } 2}}.$$

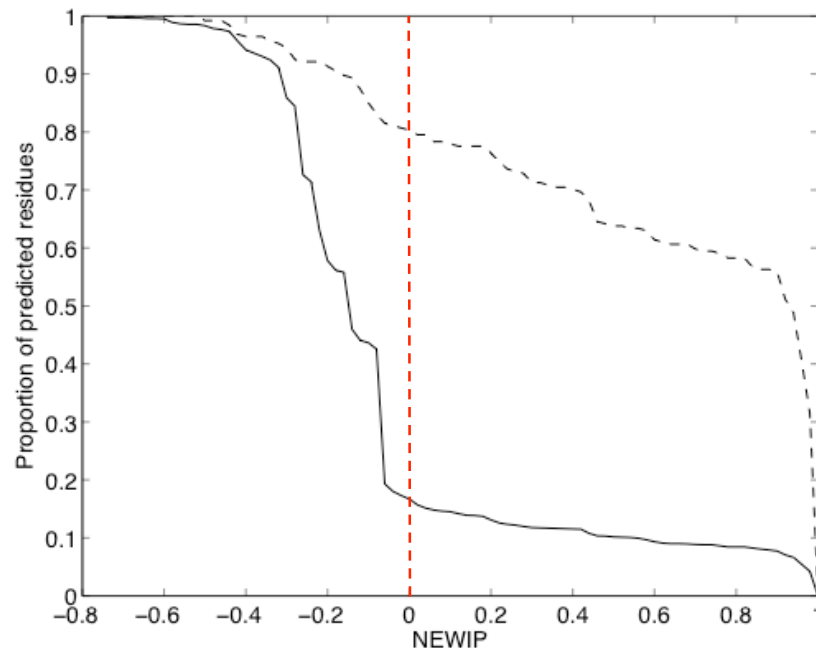
**Les Résidus avec un NIP > 0 ont plus de probabilité d'appartenir à l'interface protéique.**

# *Prédiction des sites d'interaction protéiques (base restreinte) :*

Docking simple :

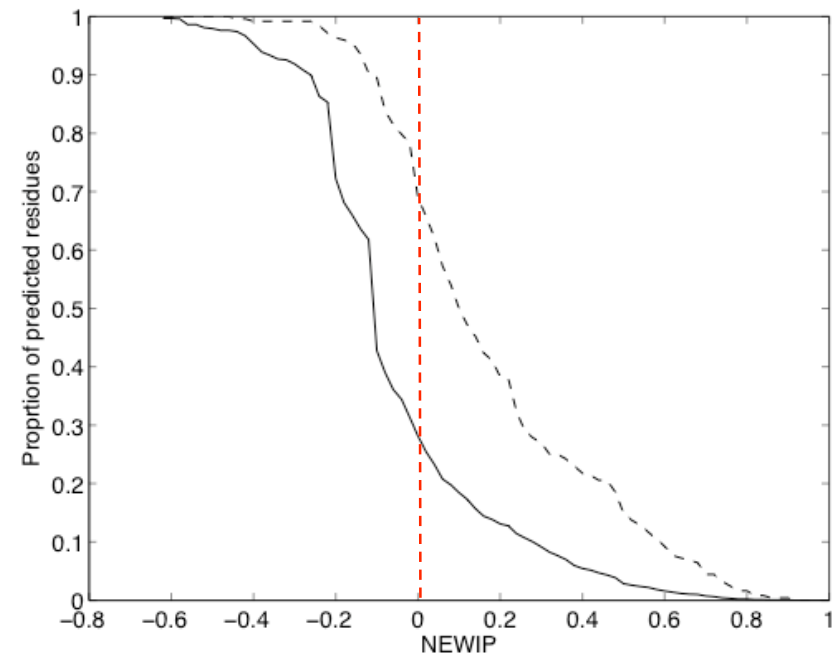
(interfaces entre partenaires expérimentaux)

NIP=0 : Couv. =18%, Sens. =81%



Docking croisé (toutes les interfaces) :

Couv. =31%, Sens. =78%

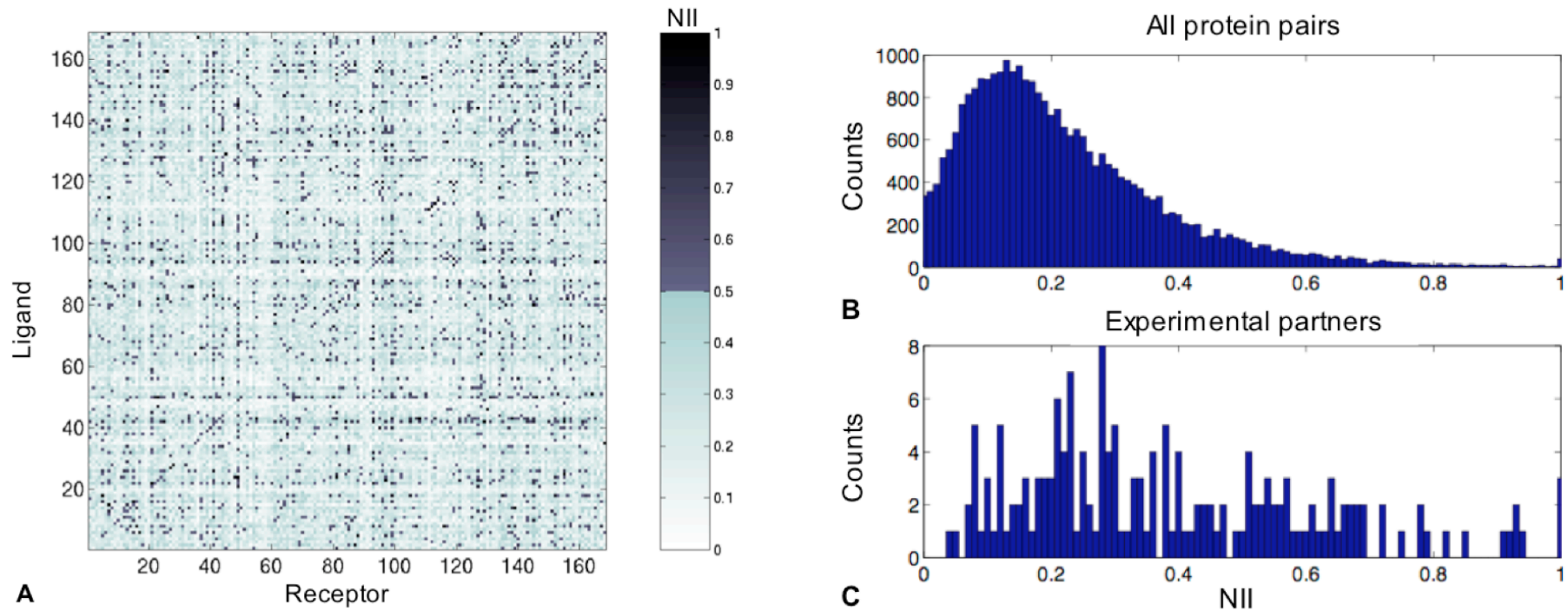


Couverture : ———

Sensibilité : - - - - -

# *Docking croisé sur le benchmark complet :*

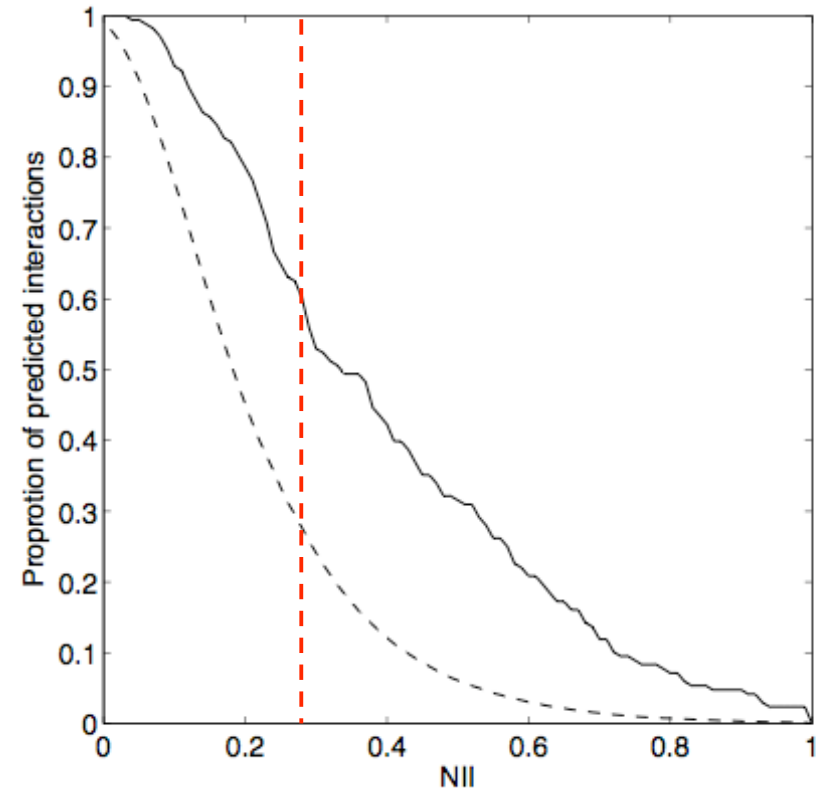
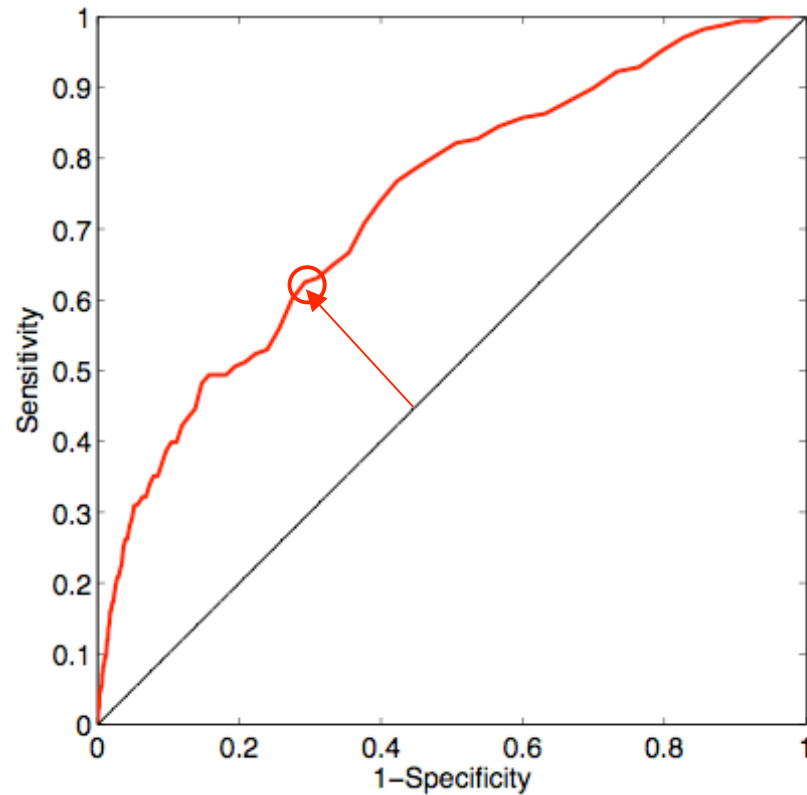
On travaille avec des structures **non-liées**.



$\langle \text{NII} \rangle = 0.22$  pour toutes les paires protéiques

$\langle \text{NII} \rangle = 0.40$  pour les partenaires expérimentaux

## *Docking croisé sur le benchmark complet :*



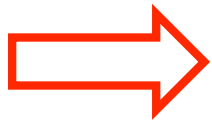
Prédiction des bons partenaires :

**NII=0.27: Couv. =29%, Sens. =63%**

## *Conclusions :*

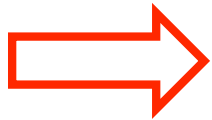
---

- La description réduite des protéines associée à un algorithme de docking par minimisation multiple permet de retrouver les bonnes configurations pour les complexes expérimentaux.
- L'analyse des résultats de docking croisé nécessite l'emploi de critères supplémentaires pour différencier interfaces spécifiques ou non-spécifiques :



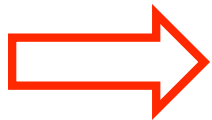
**La connaissance préalable des résidus appartenant à l'interface protéique est nécessaire**

- Comparaison des interfaces obtenues par docking-croisé avec les résidus détectés par des approches évolutives (Joint Evolutionary Trees)



**Exploiter les informations apportées par JET pour restreindre l'espace de recherche lors du docking (et donc accélérer les calculs)**

- Objectif à long terme:



**Développement d'un outil de prédiction des partenaires protéiques**

**Application de la méthode à des protéines impliquées dans des maladies neuromusculaires**

- Mécanisme d'ouverture d'un canal ionique ⇒ **Projet ANR Nicochimera**
- Biocatalyseur d'oxydation de l'hydrogène pour les piles à combustible  
⇒ **Projet ANR Biopac**
- Adaptation de ProPHET à d'autres représentations réduites ⇒ **Lecture des fichiers MARTINI**
- Vers une cartographie des interactions protéiques à l'échelle du génome  
⇒ **Projet Mapping**



- Mécanisme d'ouverture d'un canal ionique ⇒ **Projet ANR Nicochimera**

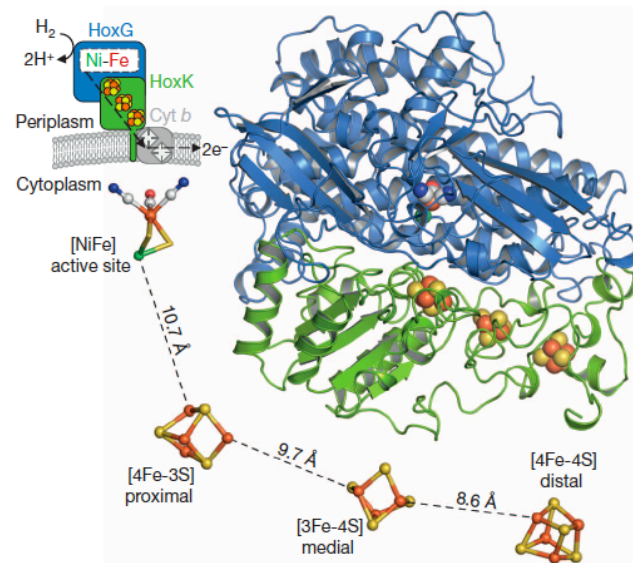
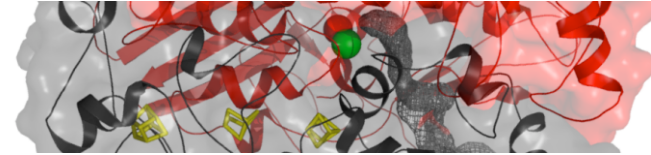
- Biocatalyseur d'oxydation de l'hydrogène pour les piles à combustible  
⇒ **Projet ANR Biopac**

- Adaptation de ProPHET à d'autres représentations réduites ⇒ **Lecture des fichiers MARTINI**

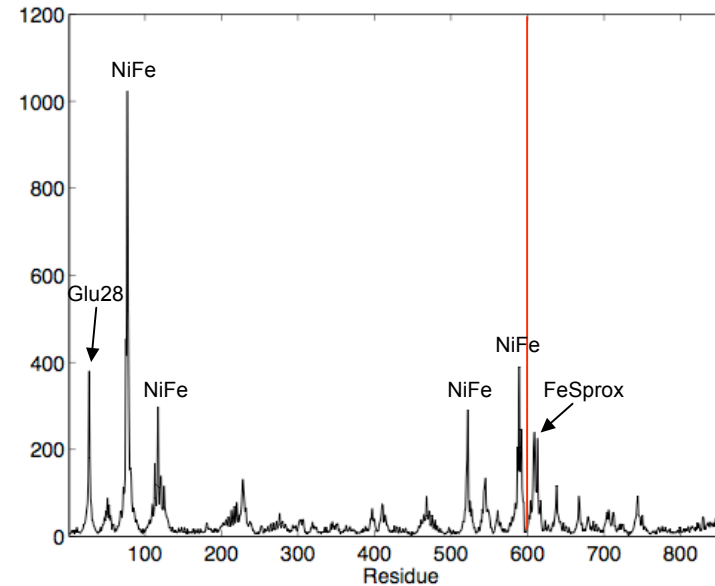
- Vers une cartographie des interactions protéiques à l'échelle du génome  
⇒ **Projet Mapping**

• Biocatalyseur d'oxydation de l'hydrogène pour les piles à combustible

⇒ **Projet ANR Biopac**



(Fritsch et al., *Nature*, 2011)



- Existence d'un noyau mécanique ?
- Perturbation du profil mécanique lors du greffage sur une électrode (modification de ProPHet)

## **Richard Lavery**

### **Mécanique des protéines :**

- Emilie Laforet, Marc Baaden, Olivier Delalande (LBT, IBPC)
- Pierre Sebban, Christiane Alba-Simionesco (Laboratoire de Chimie Physique, Université Paris-11, Orsay)
- Giovanni Venturoli (Lab. Biophysique et Biochimie, U. Bologne)

### **Interactions protéiques :**

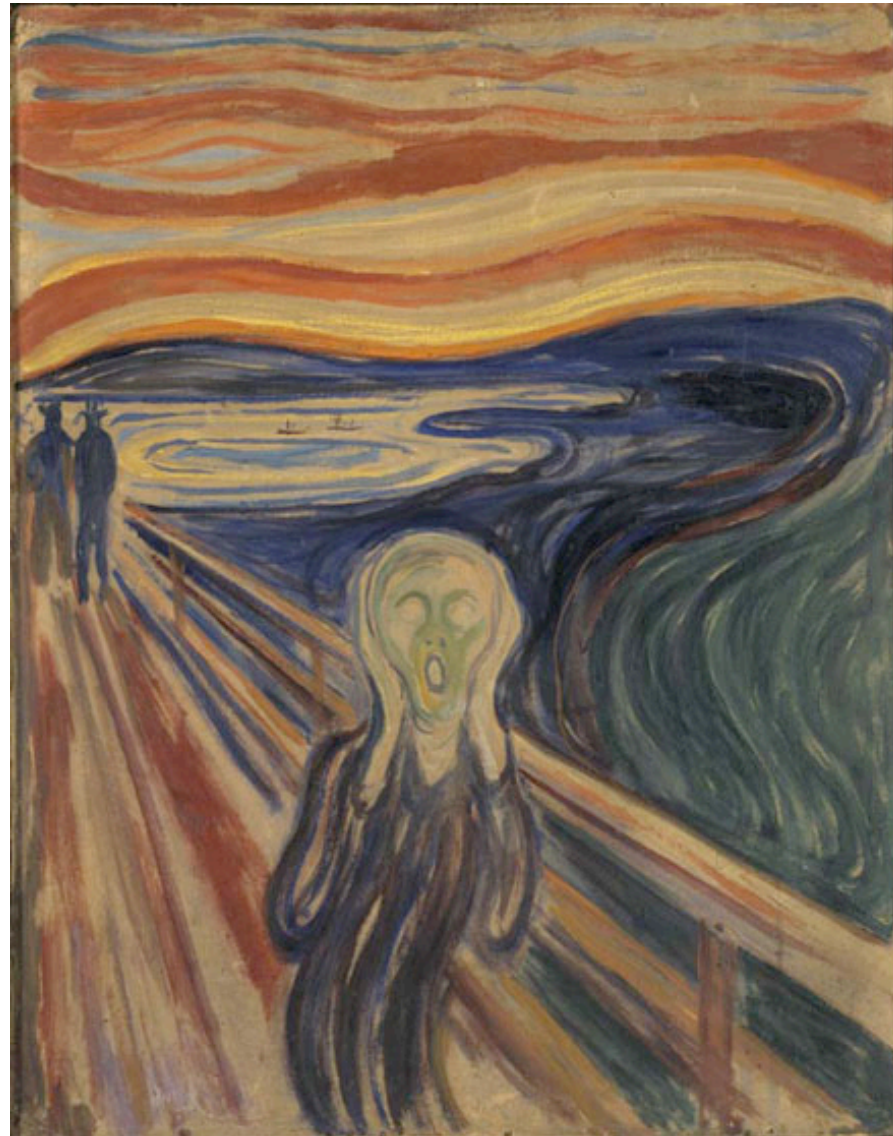
- Alessandra Carbone, Stefan Engelen (Génomique Analytique, Université Paris-6)

### **Projet Neuroglobine :**

- **Anthony Bocahut**
- Sophie Bernad, Valérie Derrien et Pierre Sebban (LCP, Orsay), spectroscopie UV-vis.
- Ewen Lescop (ICSN, Gif-surYvette), RMN
- Ron Clarke (Université de Sydney), ITC
- Chris Garvey (ANSTO, Sydney), neutrons

*Une autre Conclusion :*

---



(E. Munch, *Le Cri*, 1893).



G. Courbet, *Portrait de l'artiste dit Le Désespéré*, 1843-1845

*Portrait de chercheur mesurant l'ampleur de la tâche lui restant à accomplir pour comprendre comment fonctionne une protéine*

---

## Hydrodynamic interactions :

- Rotne-Prager tensor :

$$\mathbf{D}_{ii} = \frac{kT}{6\pi\eta a_i} \mathbf{I}_3,$$

$$\mathbf{D}_{ij} = \frac{kT}{8\pi\eta r_{ij}} \left[ \left( \mathbf{I}_3 + \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2} \right) + \frac{2a_{ij}^2}{r_{ij}^2} \left( \frac{1}{3} \mathbf{I}_3 - \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{r_{ij}^2} \right) \right]$$

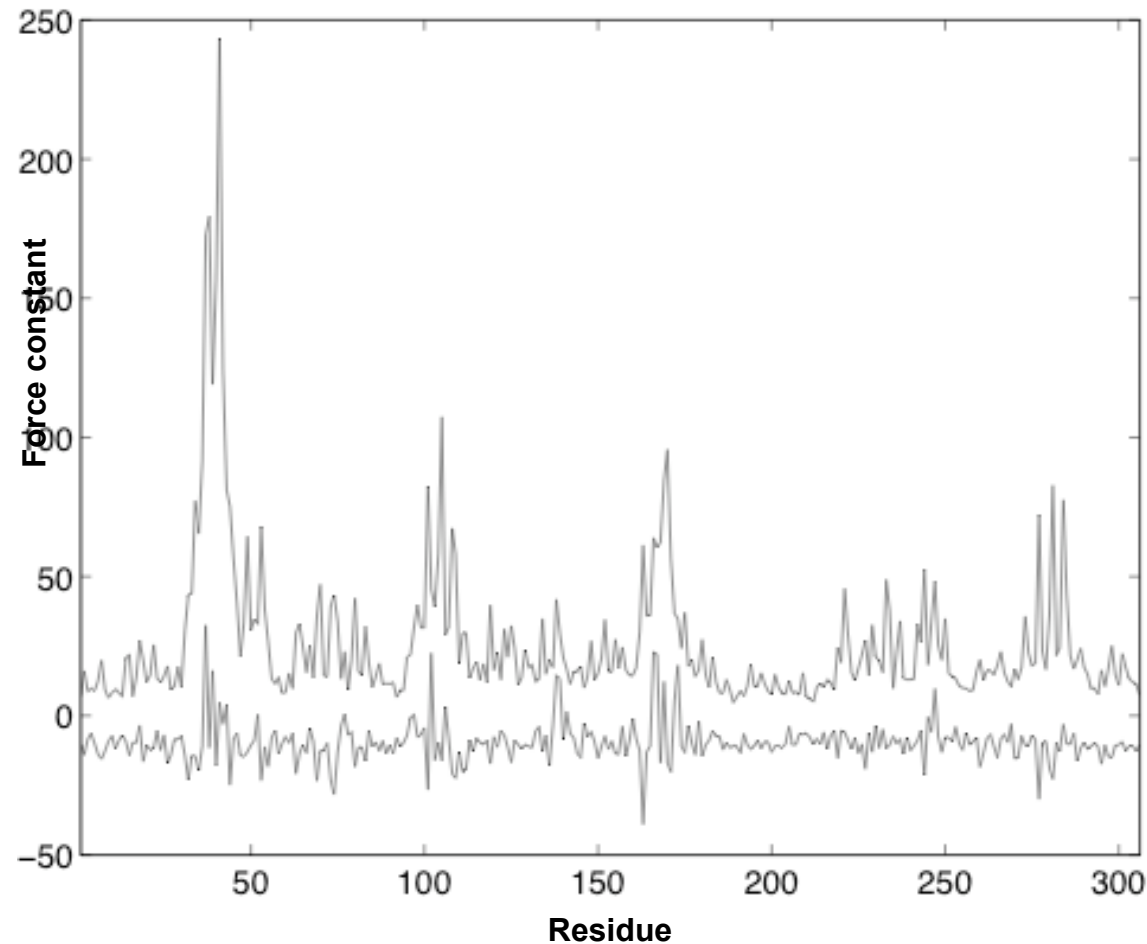
- Correction for overlapping pseudo-atoms ( $r_{ij} < 2a_{ij}$ ) :

$$\mathbf{D}_{ij} = \frac{kT}{6\pi\eta a_{ij}} \left[ \left( 1 + \frac{9}{32} \frac{r_{ij}}{a_{ij}} \right) \mathbf{I}_3 + \frac{3}{32} \frac{\mathbf{r}_{ij}\mathbf{r}_{ij}}{a_{ij} r_{ij}} \right]$$

- With no HI, we simply take  $\mathbf{D}_{ij}=0$ .

# *Prise en compte du groupe prosthétique*

Horseradish peroxidase (1ATJ)



Les profils de rigidités sont qualitativement conservés lorsqu'on introduit le cofacteur dans la représentation

DB (2-3p/r)

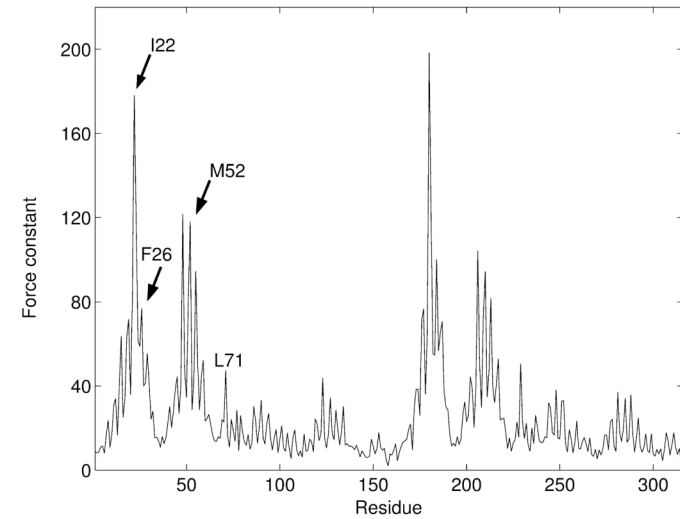
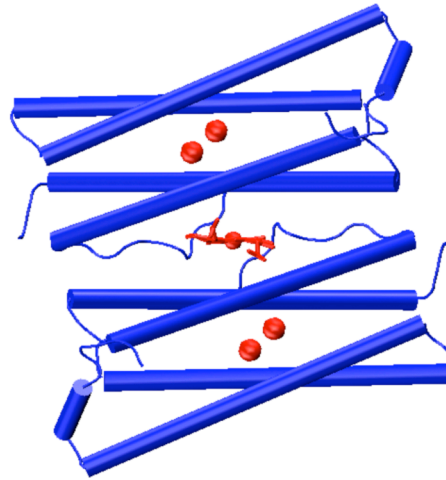
← Variation de k après introduction de l'hème

# Signatures structurales :

Bacterioferritin

(1BCF),

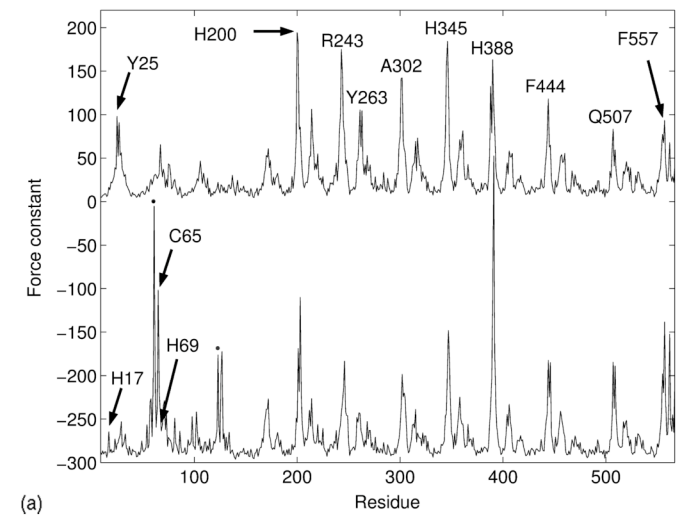
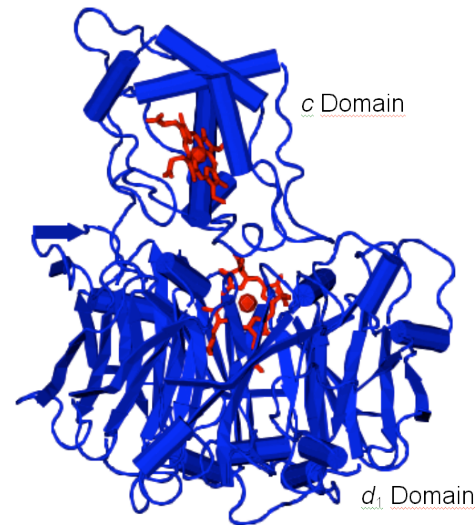
Hélices  $\alpha$



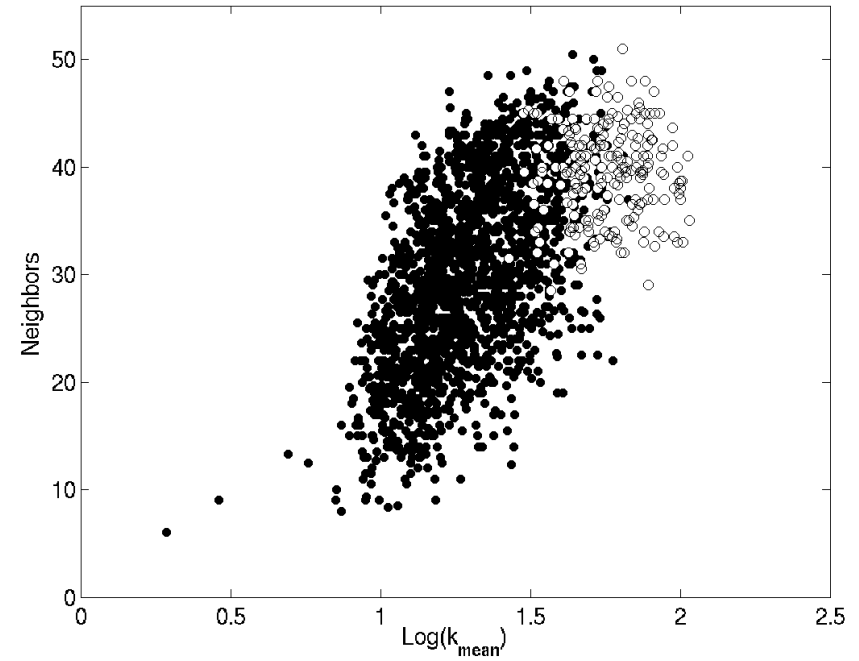
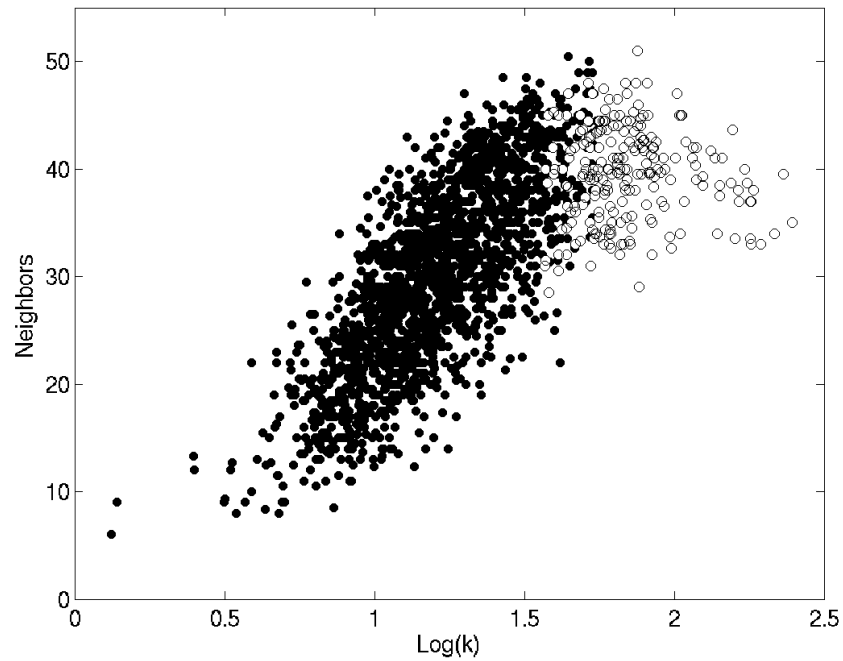
Cytochrome  $cd_1$

(1QKS),

Tonneau  $\beta$



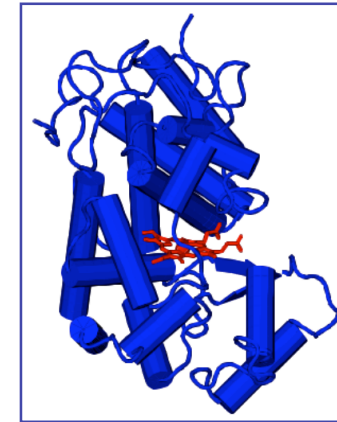
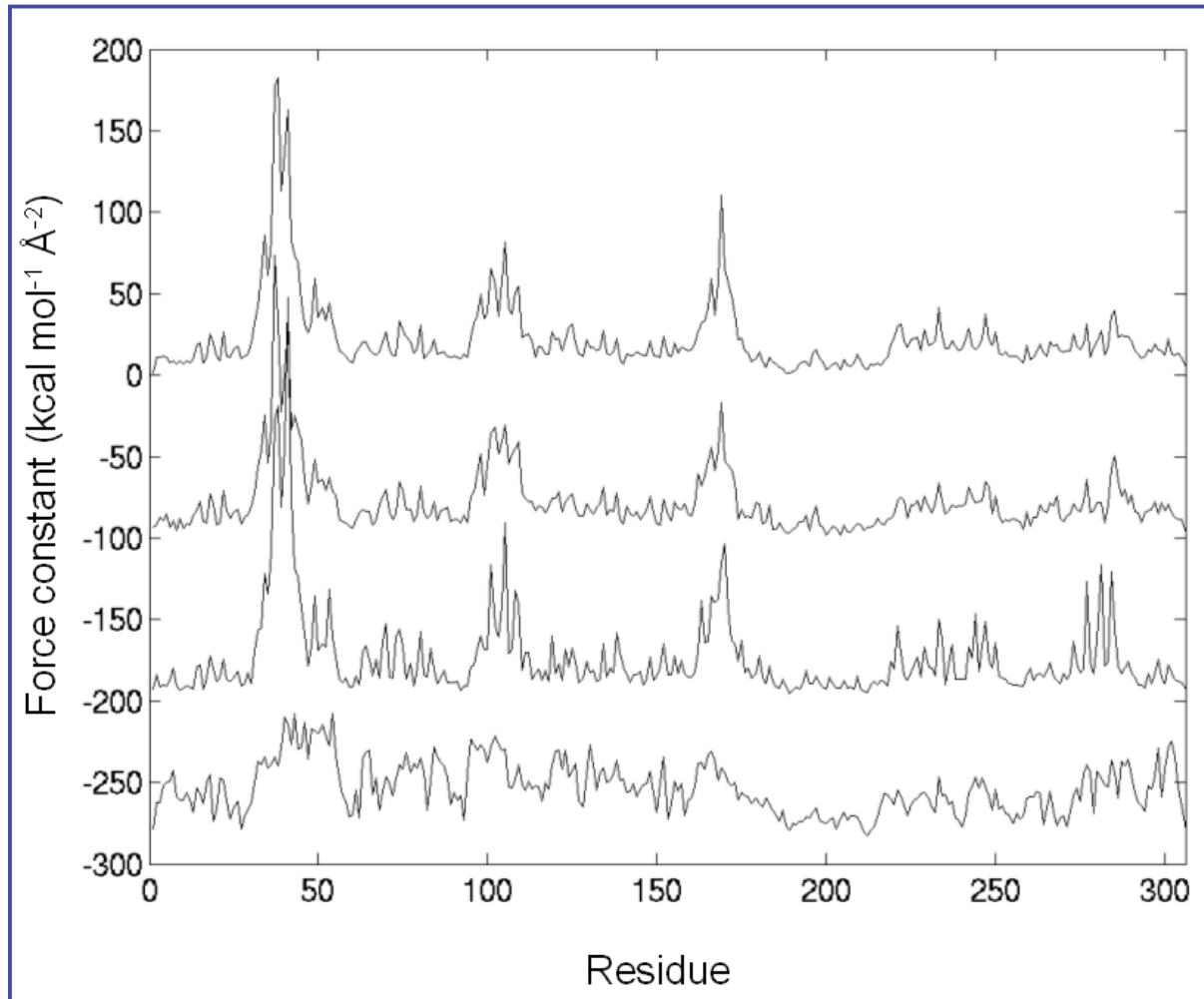




- Rigid residues,  $k_i > \langle k \rangle + \sigma(k)$
- Other residues.

# Calculating the force constants:

## Horseradish peroxidase (1ATJ)



Minimization (1p/r)

BD (1p/r)

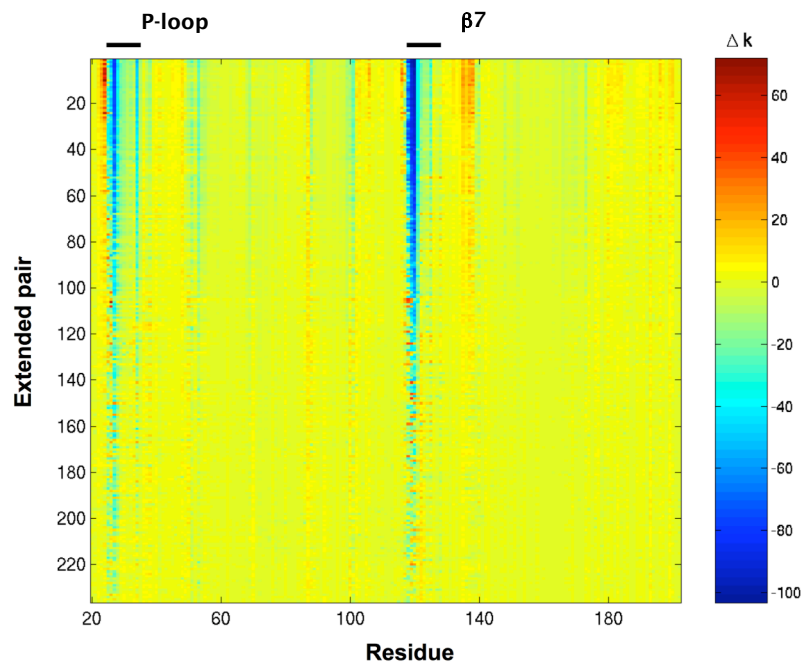
BD (2-3p/r)

B-factors (inverse)

## Scan de la surface protéique :

▪ On effectue une recherche systématique à la surface de GK en testant 236 paires de résidus présentant les caractéristiques suivantes :

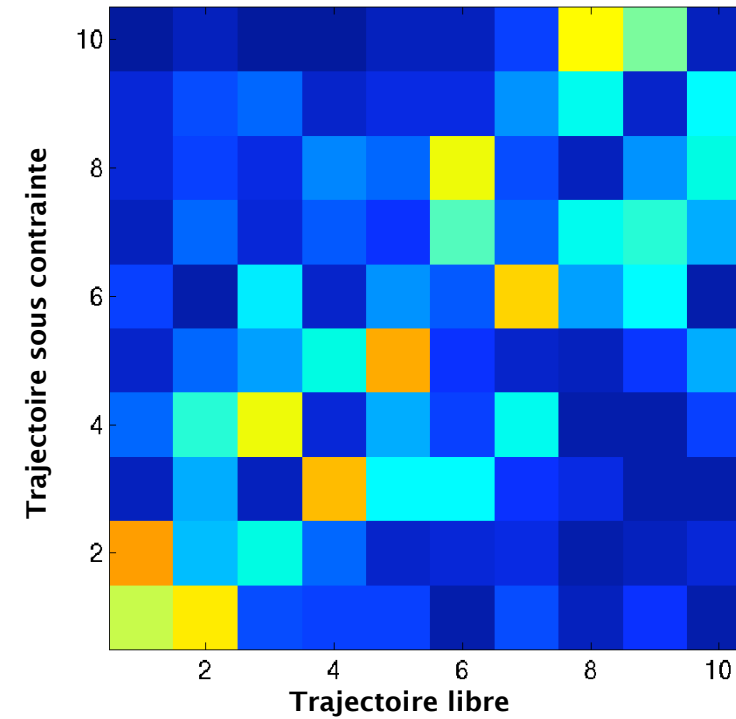
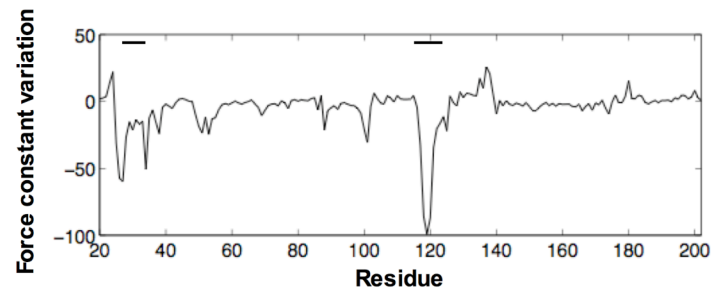
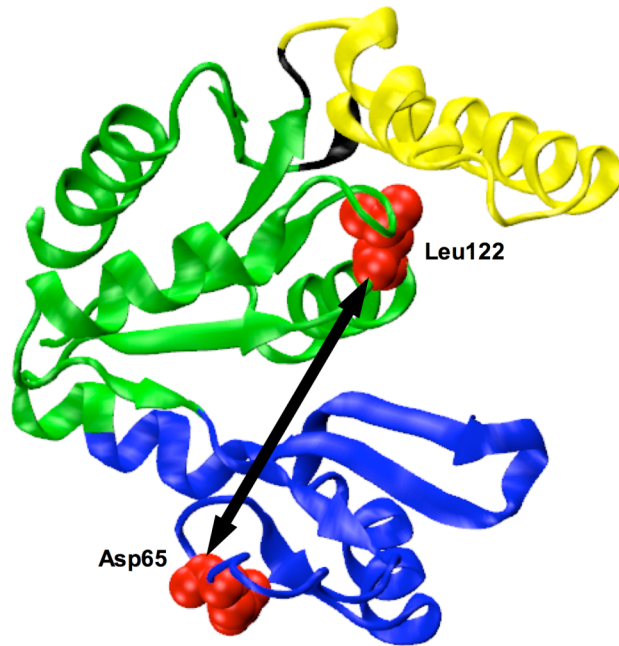
- Résidus de surface, **SASA  $\geq 5\%$**
- Résidus distants d'au moins **20 Å**
- Distance le long de la séquence d'au moins **30 résidus**
- On élimine les **paires redondantes** ( $|i-i'| \leq 5$  et  $|j-j'| \leq 5$  résidus)



La P-loop et le feuillet b7 sont les zones les plus sensibles sur le plan mécanique

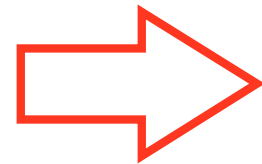
## Scan de la surface protéique :

- On isole la paire de résidus Asp65/Leu122, qui entraîne la perturbation mécanique la plus importante.



CMO 0.66

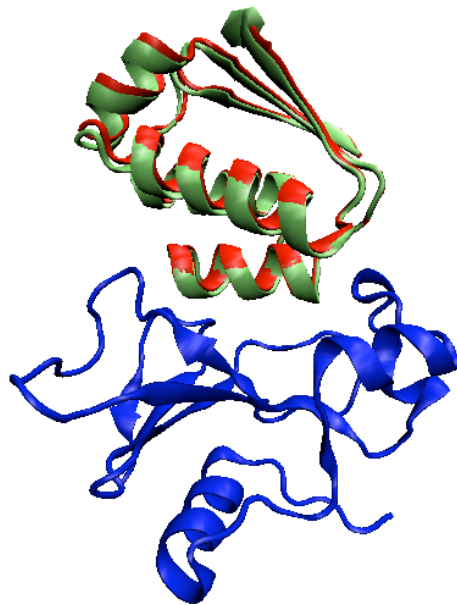
$P_1$  0.55



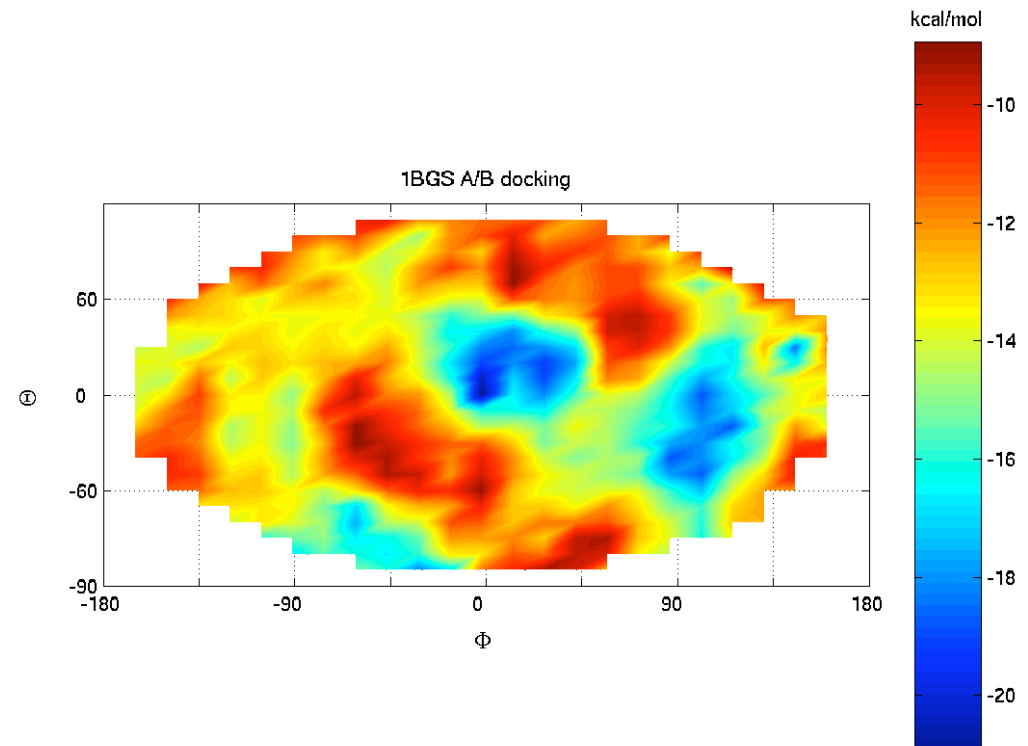
Effet sur  $K_G$  et  $k_{cat}$

## *Validation de l'algorithme et du champ de force :*

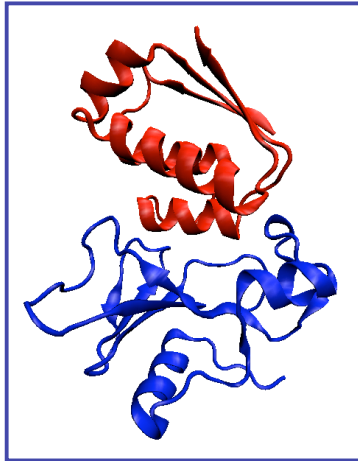
Une première série de calculs sur les partenaires expérimentaux donne des structures correctes pour les complexes et une bonne localisation du site d'interaction sur la surface protéique



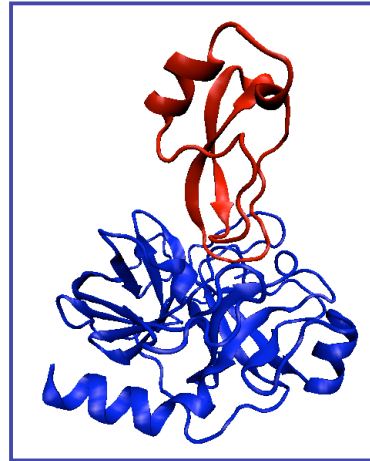
Barnase/barstar : **0.69 Å**



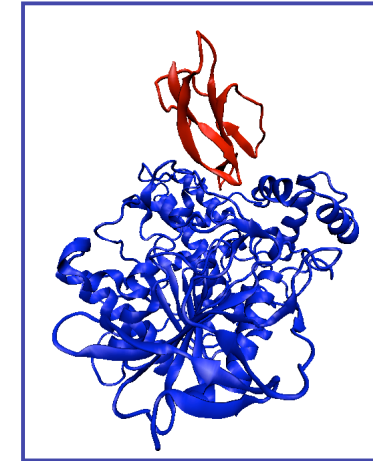
## *Results on a reduced dataset :*



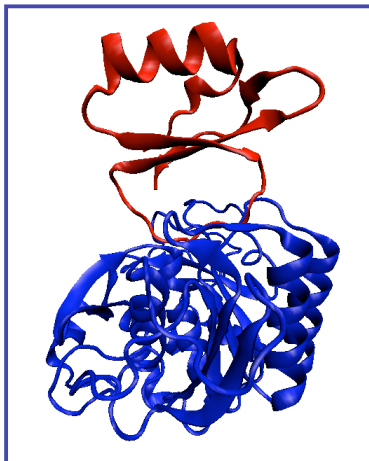
**Barnase/barstar (1BGS)**



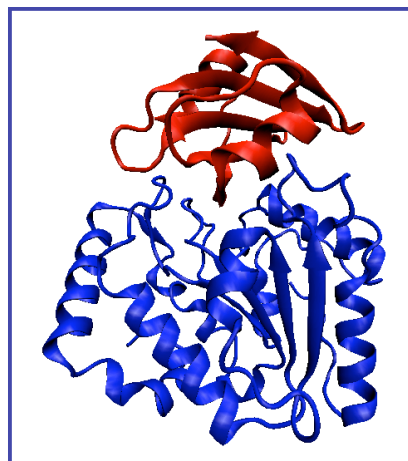
**$\beta$ -trypsin/inhibitor (2PTC)**



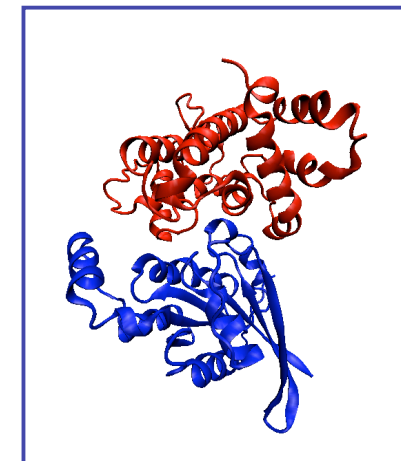
**Acetylcholinesterase/fasciculin II (1FSS)**



**Thermitase/eglin C (2TEC)**



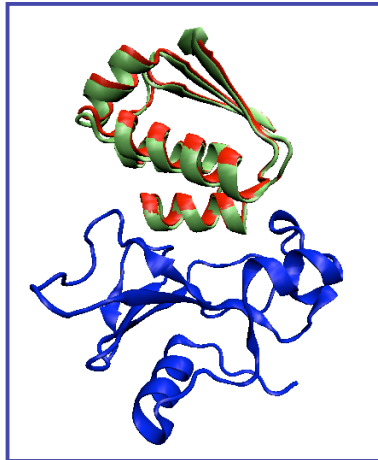
**Uracil-DNA glycosylase/inhibitor  
(1UGH)**



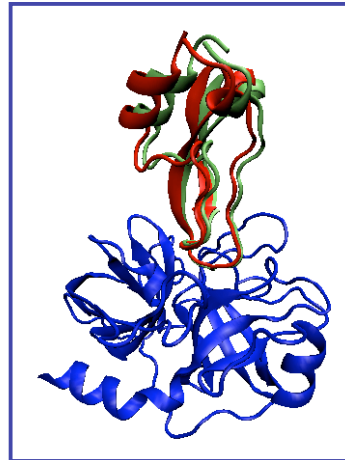
**CDC42 GTPase/CDC42 GAP (1GRN)**

# *Validation the interaction potential :*

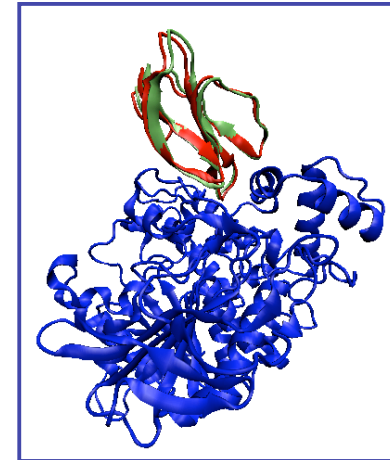
Best ligand rmsd obtained :



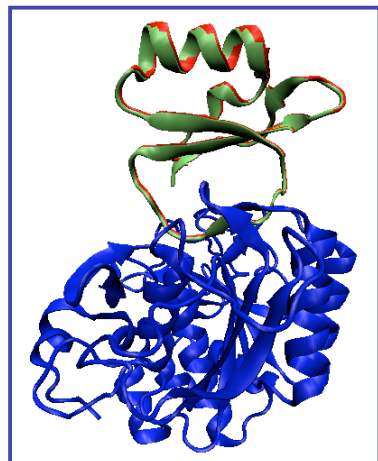
Barnase/barstar : **0.69 Å**



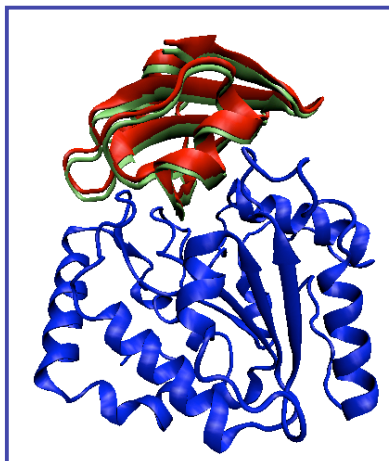
$\beta$ -trypsin/inhibitor : **2.19 Å**



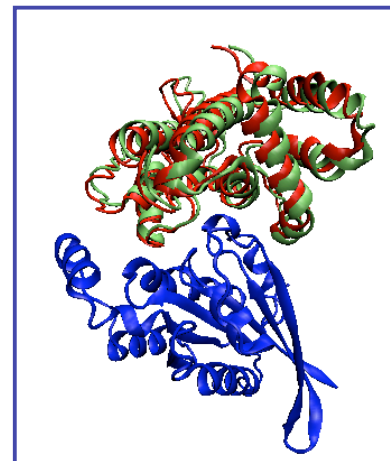
Acetylcholinesterase/fasciculin II : **1.03 Å**



Thermitase/eglin C : **0.43 Å**



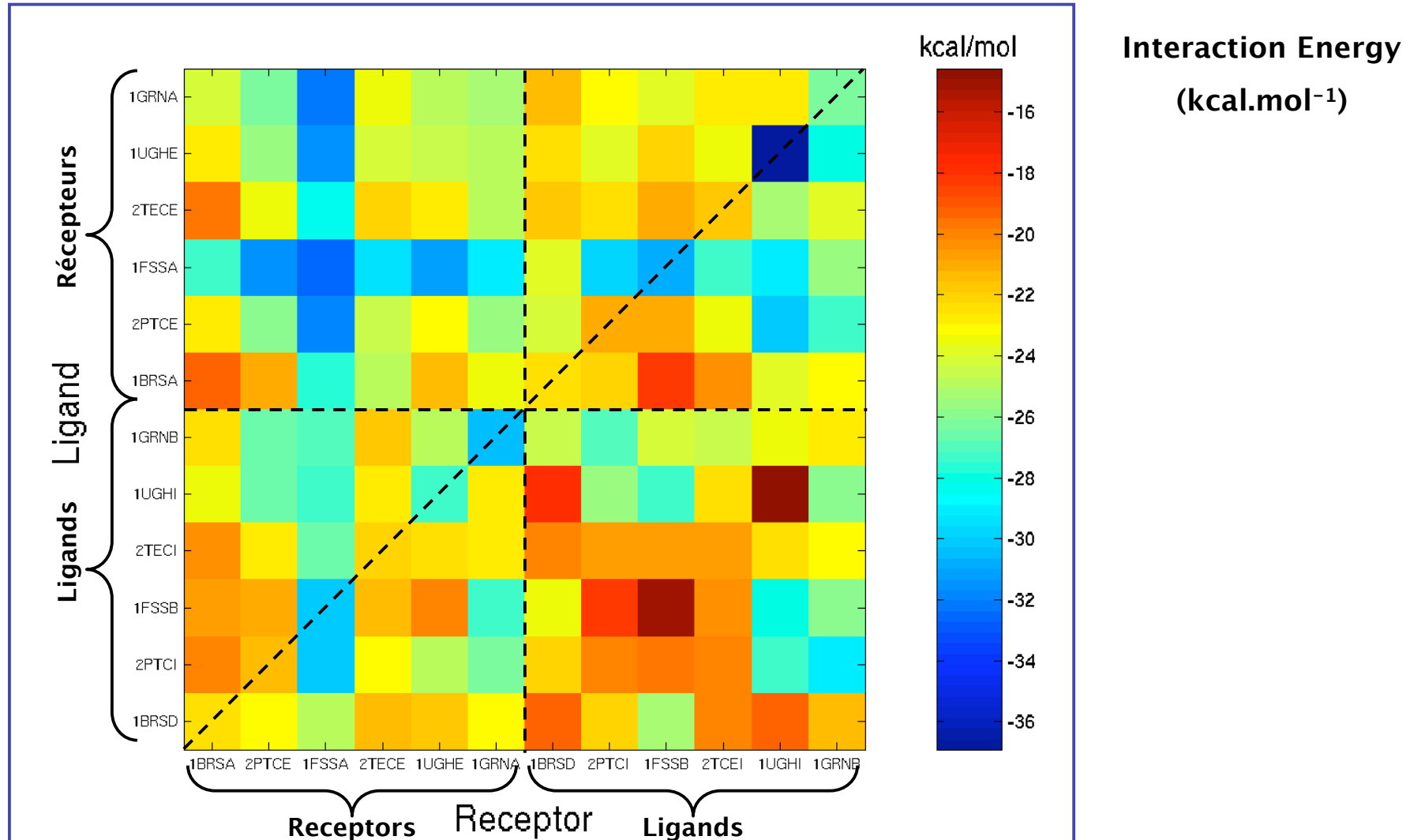
Uracil-DNA glycosylase/inhibitor : **0.99 Å**



CDC42 GTPase/CDC42 GAP : **1.83 Å**

# Cross-docking results :

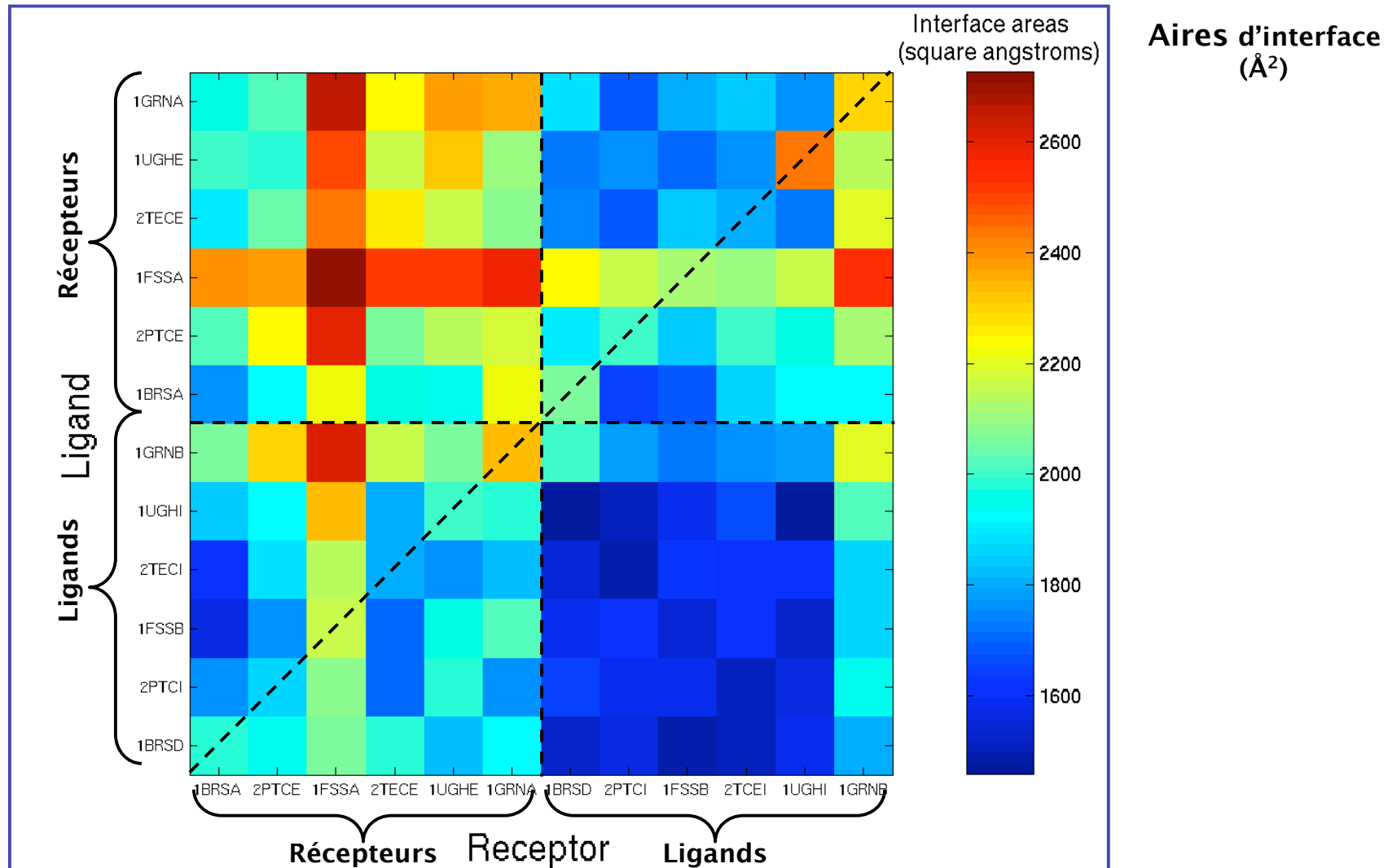
Can we distinguish between “experimental” and “artificial” protein interfaces ?





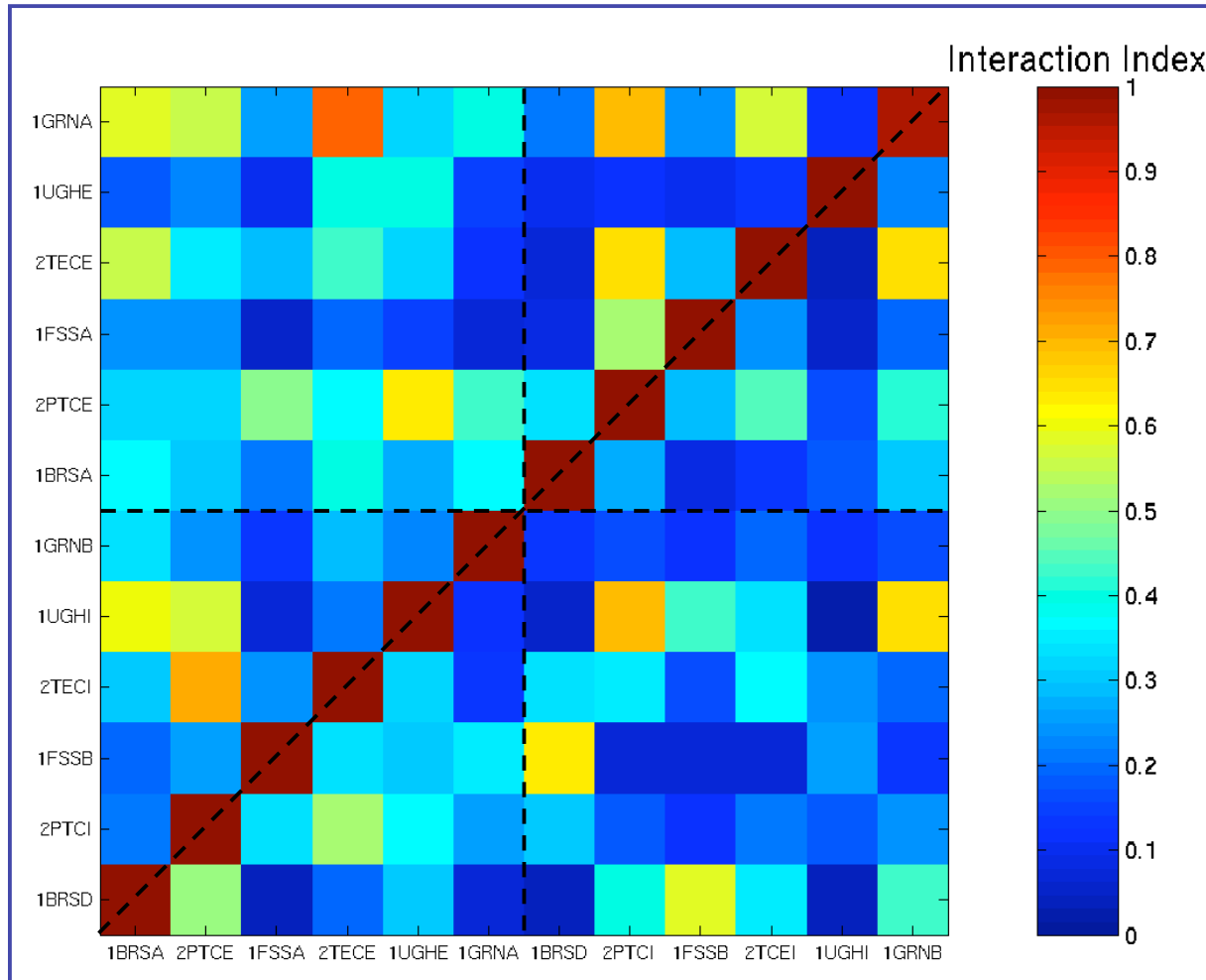
# Cross-docking results :

Can we distinguish between “experimental” and “artificial” protein interfaces ?



# Cross-docking results :

Can we distinguish between “experimental” and “artificial” protein interfaces ?

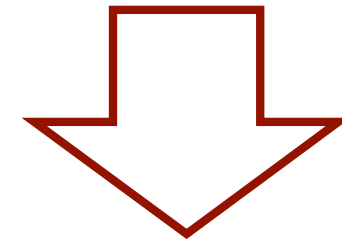


Indice d'interaction :

$$I = f_{\text{rec}} * f_{\text{lig}} * E_{\text{int}}$$

$f_{\text{rec}}$  = Fraction de résidus  
d'interface du récepteur

$f_{\text{lig}}$  = Fraction de résidus  
d'interface du ligand



**On a besoin de  
connaître les  
résidus d'interface**

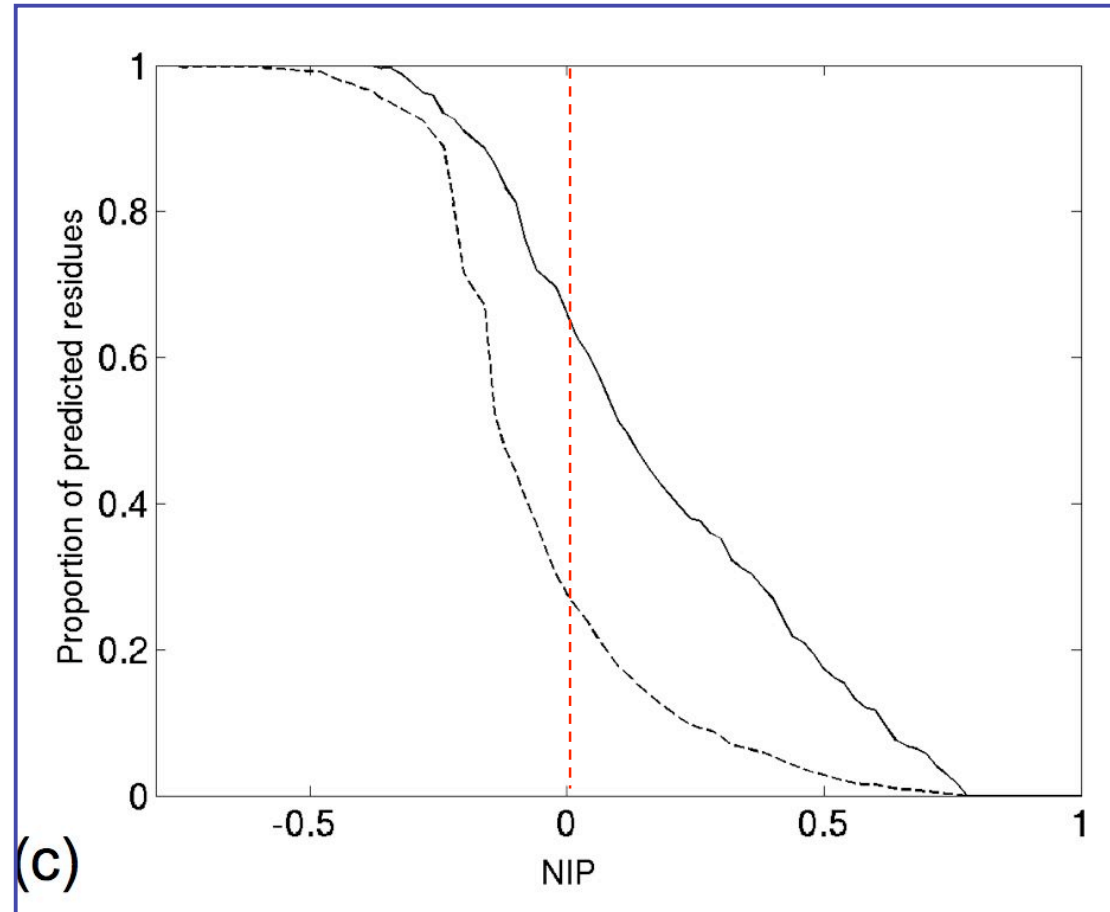
# Predicting interface residues:

Cross-docking predictions with unbound structures :

At NIP=0: Cov. =30%, Sens. =70%

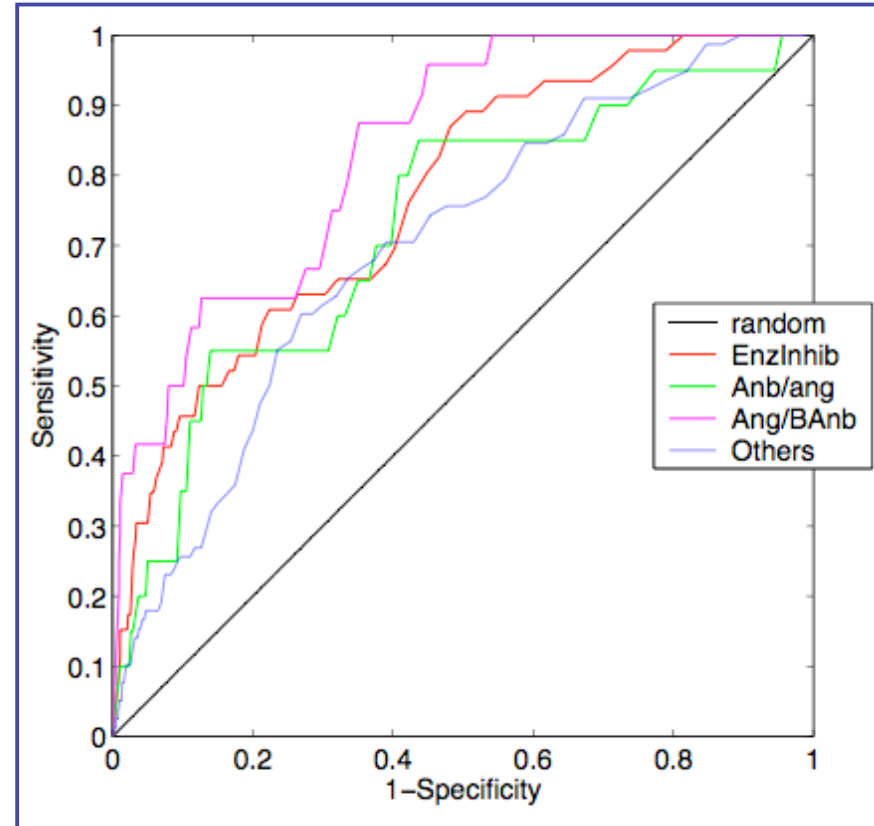
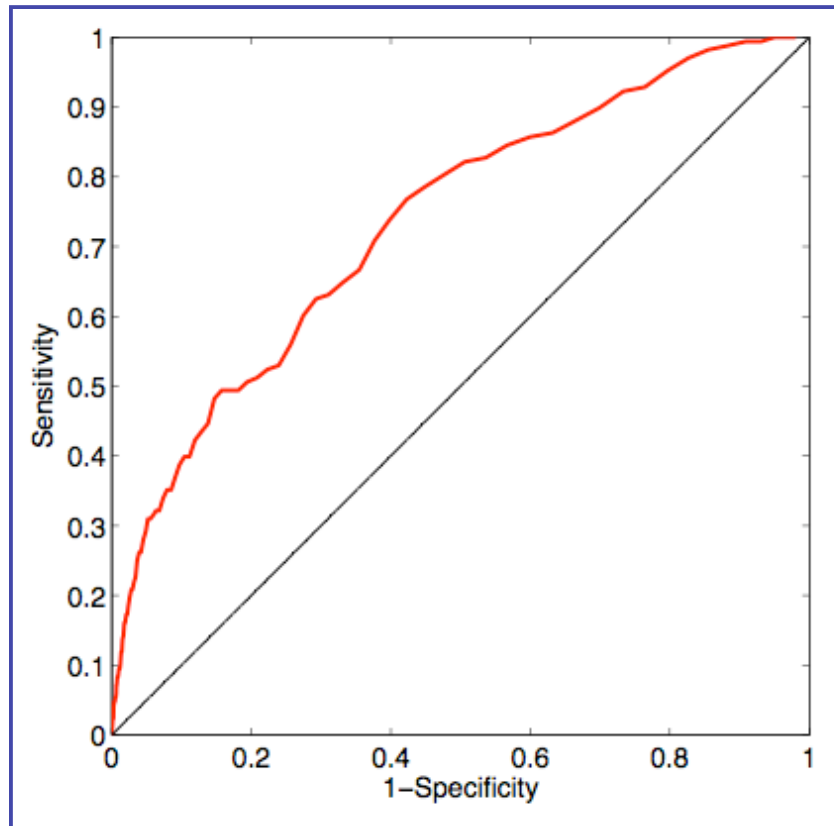
Coverage : ———

Sensibility : - - - - -



# Cross-docking results on the complete dataset:

Influence of the type of complex considered :

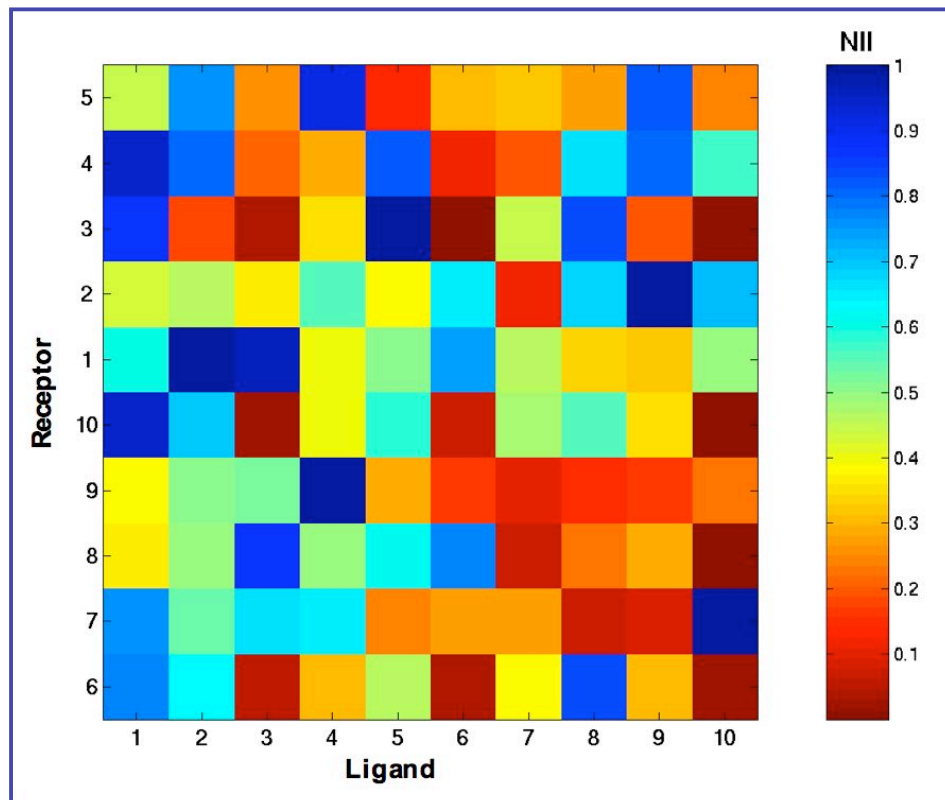


84 protein complexes:

- 23 Enzyme/Inhibitor
- 10 Antigen/Antibody
- 12 Antigen/Bound Antibody
- 39 Others

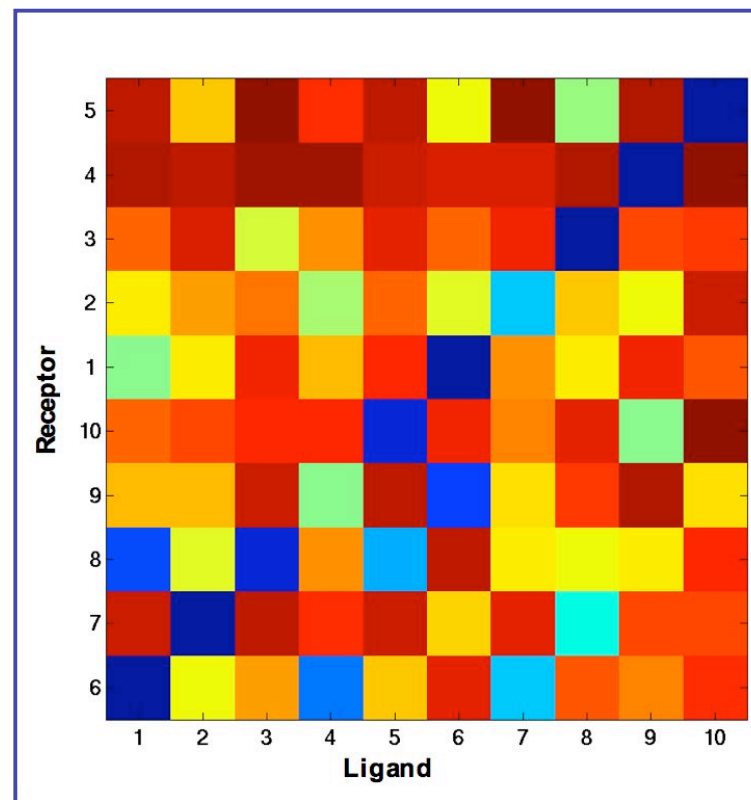
# Cross-docking results on a reduced dataset:

CD-matrices with unbound structures :



Enzyme/inhibitor test set :

$$\langle \text{NII} \rangle = 0.45$$
$$\langle \text{NII} \rangle_{\text{exp}} = 0.65$$



Antigen and others test set :

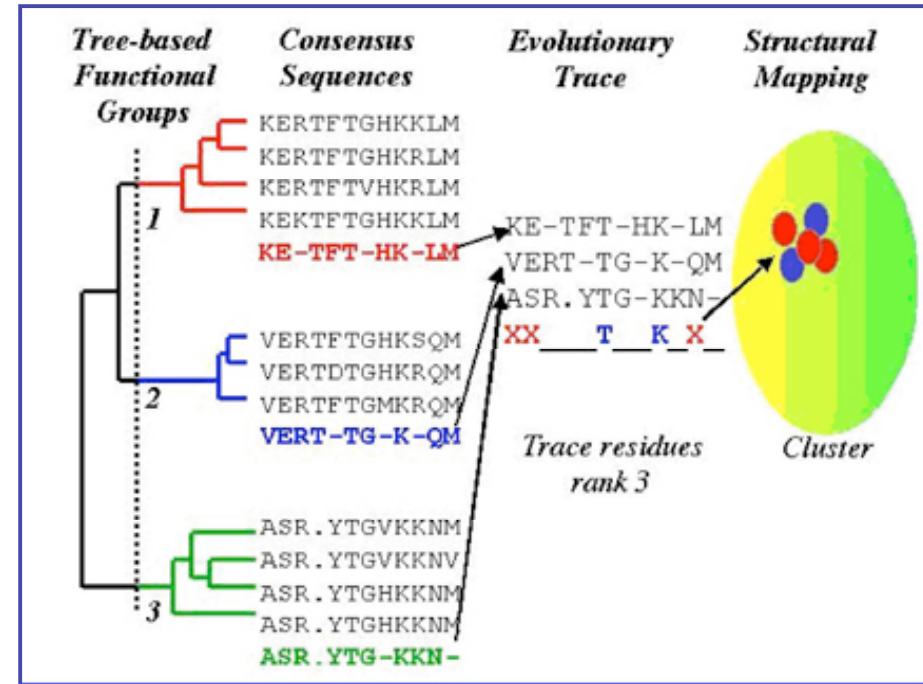
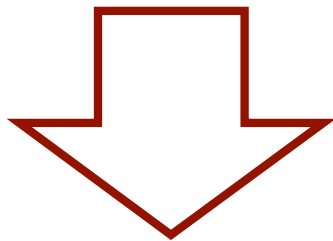
$$\langle \text{NII} \rangle = 0.32$$
$$\langle \text{NII} \rangle_{\text{exp}} = 0.91$$

# Approche Phylogénétique :

Développement du programme JET (Joint Evolutionary Trees), reprenant la méthode Evolutionary Trace :

Des études préalables avec ET ont montré que :

- Les résidus trace obtenus forment des cluster dans la structure 3D de la protéine.
- Ces cluster sont localisés au niveau des sites fonctionnels ou des interfaces moléculaires.

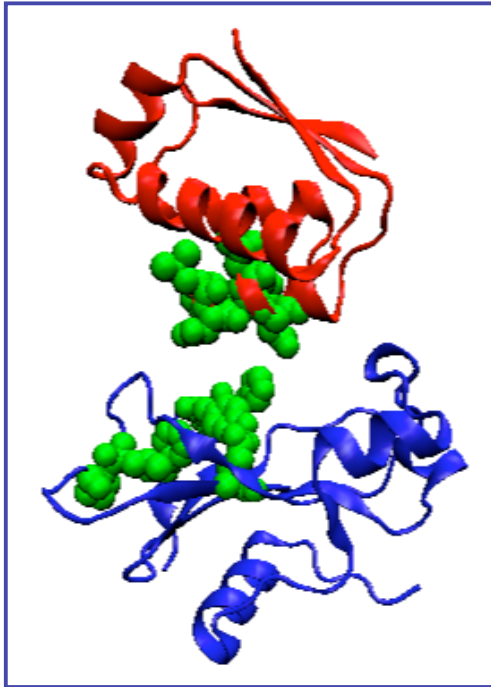


Lichtarge, 1996

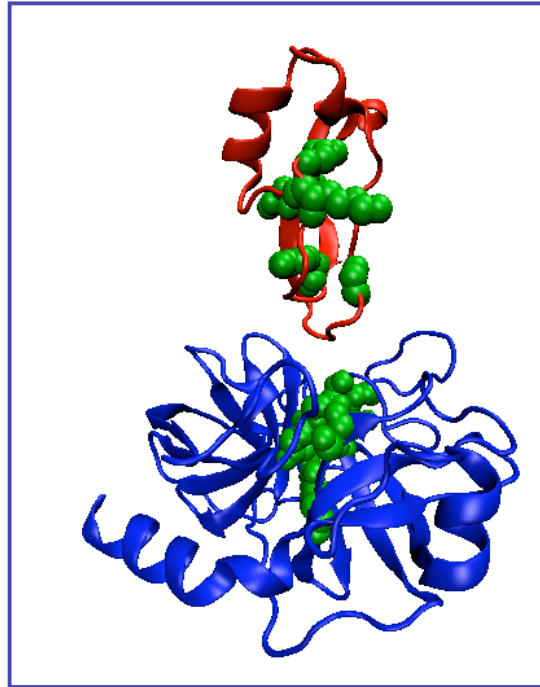
**Informations sur le site d'interaction protéique à partir de la seule séquence en acides aminés (sans connaître les partenaires potentiels)**

## Approche Phylogénétique :

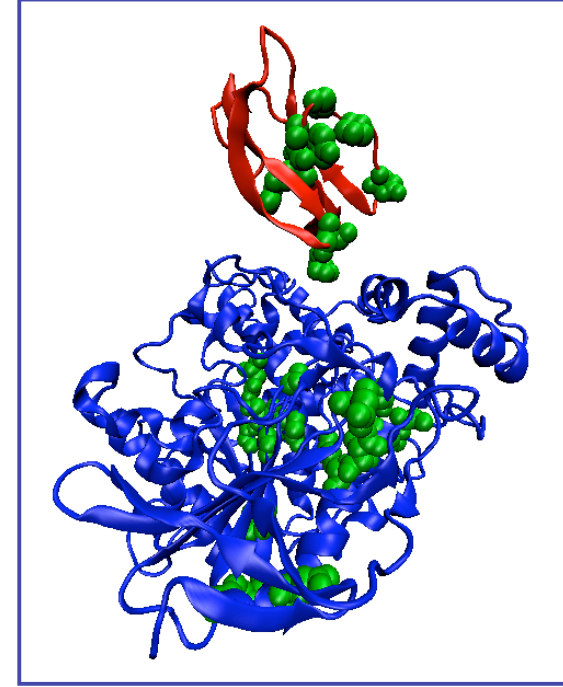
---



Barnase/barstar



$\beta$ -trypsin/inhibitor



Acetylcholinesterase/fasciculin II

- Problème de sélectivité pour les petits inhibiteurs par rapport aux gros récepteurs (2PTC)
- Cas des protéines comportant plus d'un site d'interaction (1FSS).