

Meet-U

PROJET COLLABORATIF
PARIS 6 – PARIS 7 – PARIS 11

**Evolution des
interactions et
des interfaces**

les protéines évoluent . . .

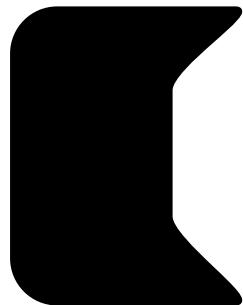
Au cours de l'évolution, les protéines sont soumises à des contraintes:

- structurales – pour maintenir leur stabilité en solution
- fonctionnelles – pour maintenir leur rôle dans la cellule

Les protéines évoluent lentement. La plupart des substitutions (98%) sont interdites, car délétères pour la structure, la fonction ou l'expression de la protéine.

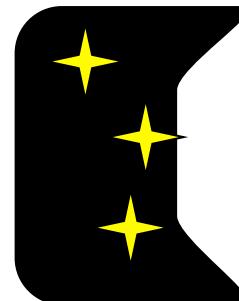
L'effet d'une mutation, à un instant donné, dépend du contexte de séquence.

P
de
S1



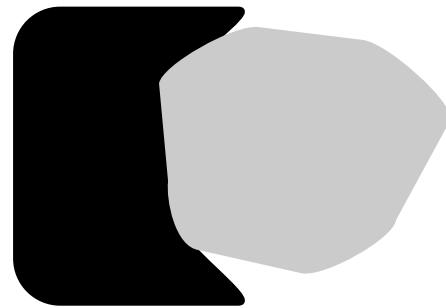
P et P' sont homologues,
elles descendent d'un
ancêtre commun

P'
de
S2

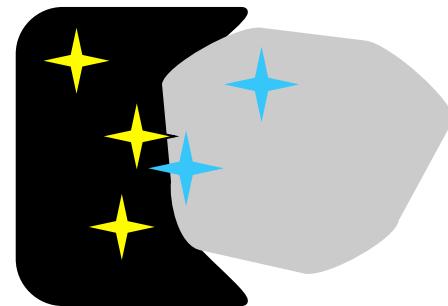


. . . leurs interactions aussi

R-L
de
S1



R'-L'
de
S2



R-L et R'-L' sont **interologues** (Vidal *et al.*, Science, 2000) : R et R' sont homologues, L et L' sont homologues. [InterEvol: <http://biodev.cea.fr/interevol/>]

Les interactions ne sont pas systématiquement conservées d'une espèce à l'autre, certaines sont gagnées ou perdues au cours de l'évolution (re-wiring).

Combien ? c'est difficile à estimer.

On ne sait pas exactement qui interagit avec qui dans chaque espèce.

Quelques définitions

homo-oligomères vs hétéro-oligomères

Les homo-oligomères sont formés par plusieurs copies d'une même chaîne protéique (ex: cytochrome c). Les hétéro-oligomères sont formés par des chaînes protéiques différentes (ex: ATPase). Les chaînes protéiques d'un oligomère sont appelées **sous-unités**.

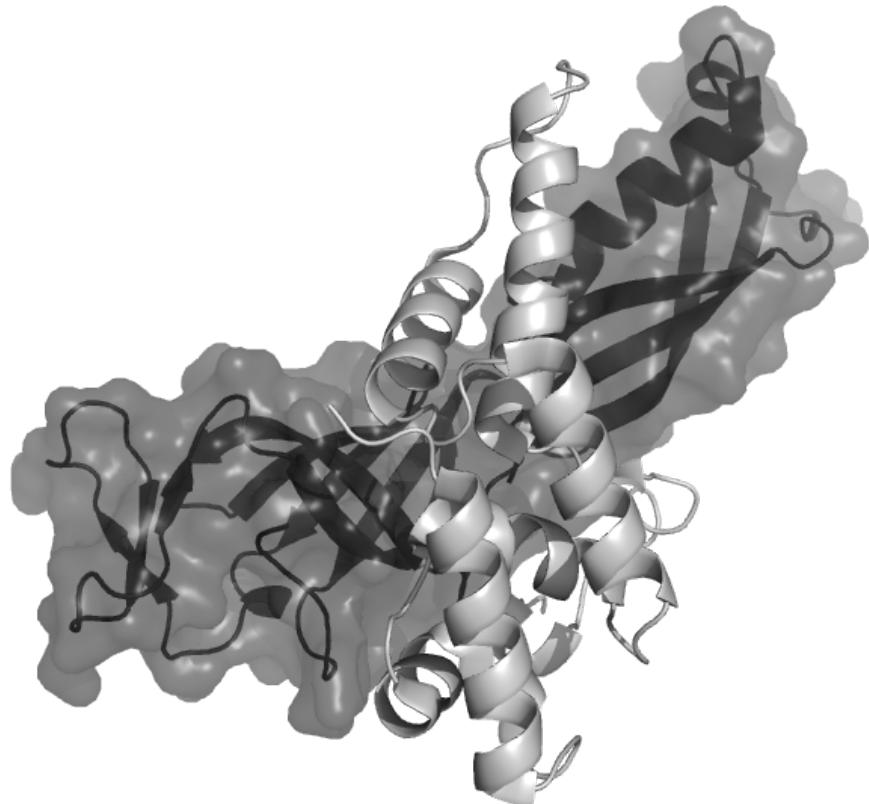
Complexes permanents vs complexes transients

Les complexes permanents sont caractérisés par des interactions très stables entre leurs constituants (ex ci-dessus). Les complexes transients ont une durée de vie plus courte, l'interaction est plus faibles (ex: récepteurs couplés à des protéines G).

Ces différents types de complexes n'évoluent pas à la même vitesse, et ne sont pas conservés de la même façon dans des espèces distinctes !

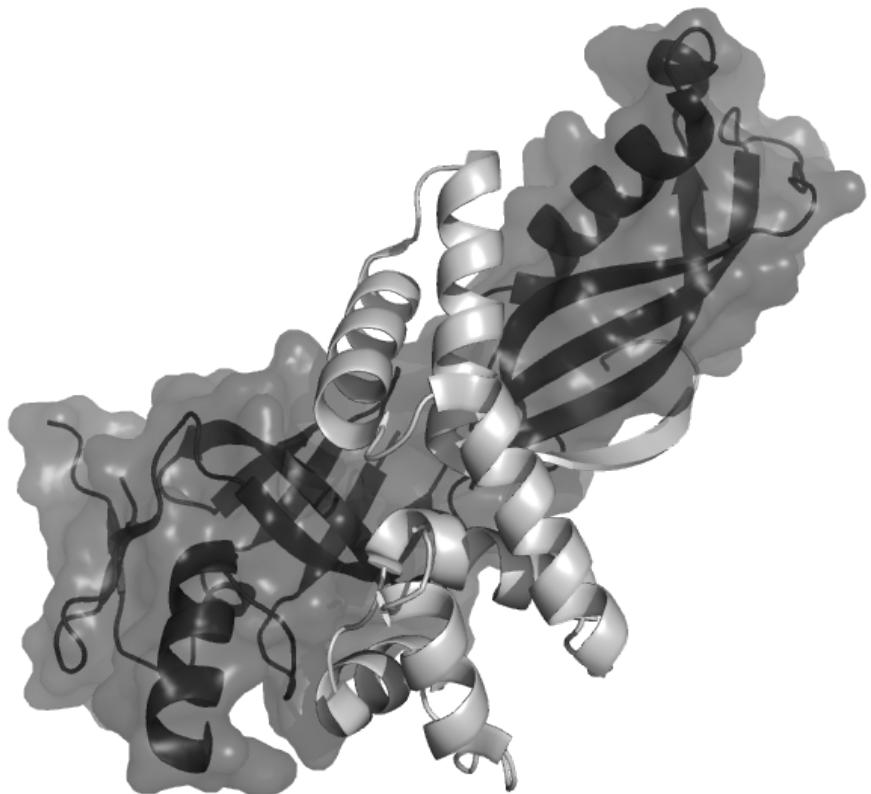
Interaction conservée signifie-t-
elle interface conservée ?

Interface conservée

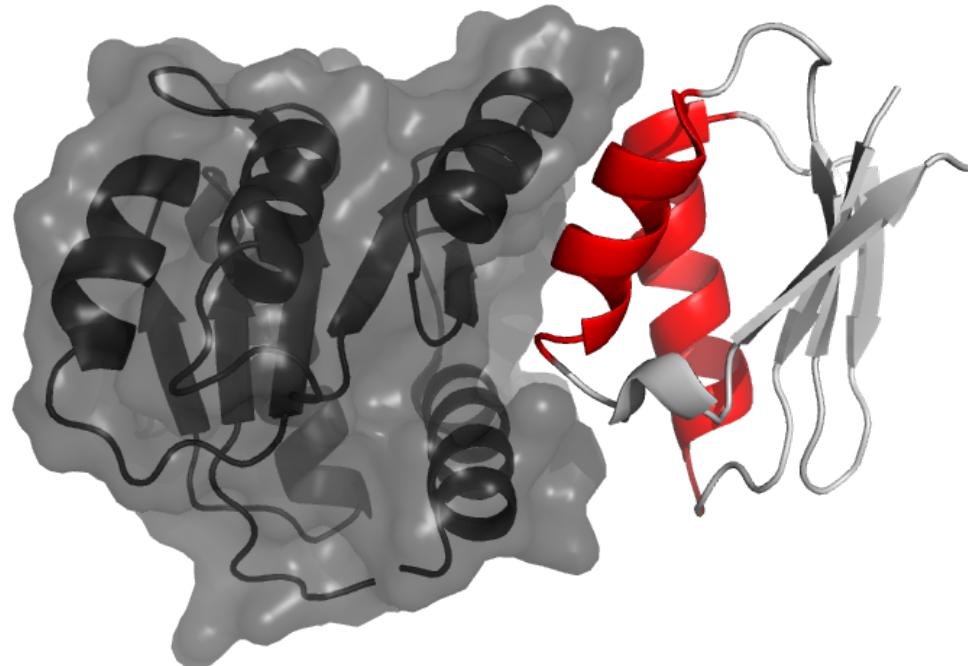


Sous-unités E et F de la
polymérase d'ARN II
Methanococcus jannaschii

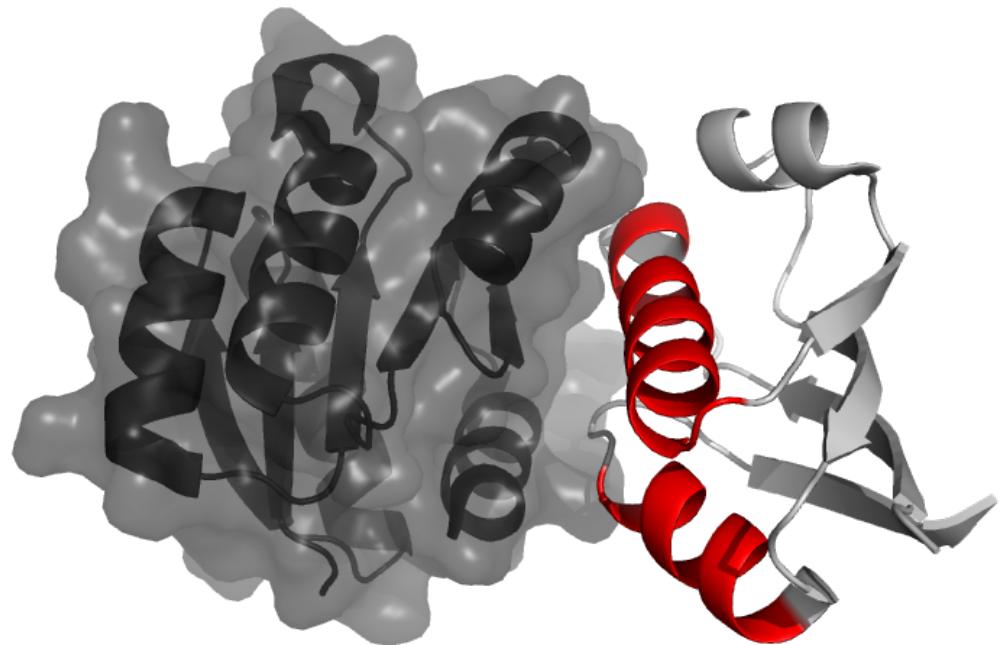
Rpb4 et Rpb7 de la
polymérase d'ARN II
Homo sapiens



Interface différente



CheY et CheA-P2
Escherichia coli



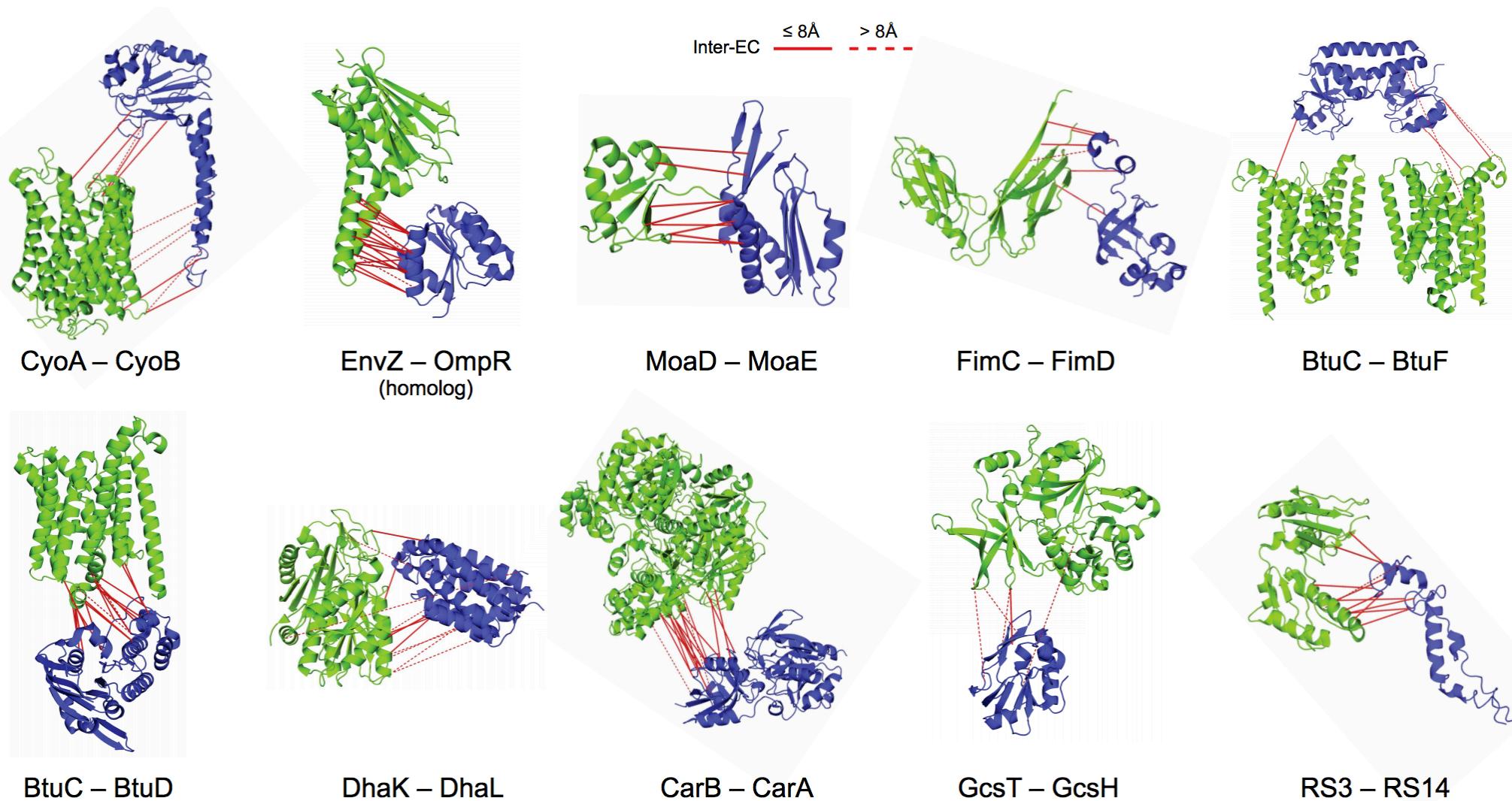
CheY et CheA-P2
Thermotaga maritima

Comment des protéines partenaires
évoluent-elles "ensemble" ?

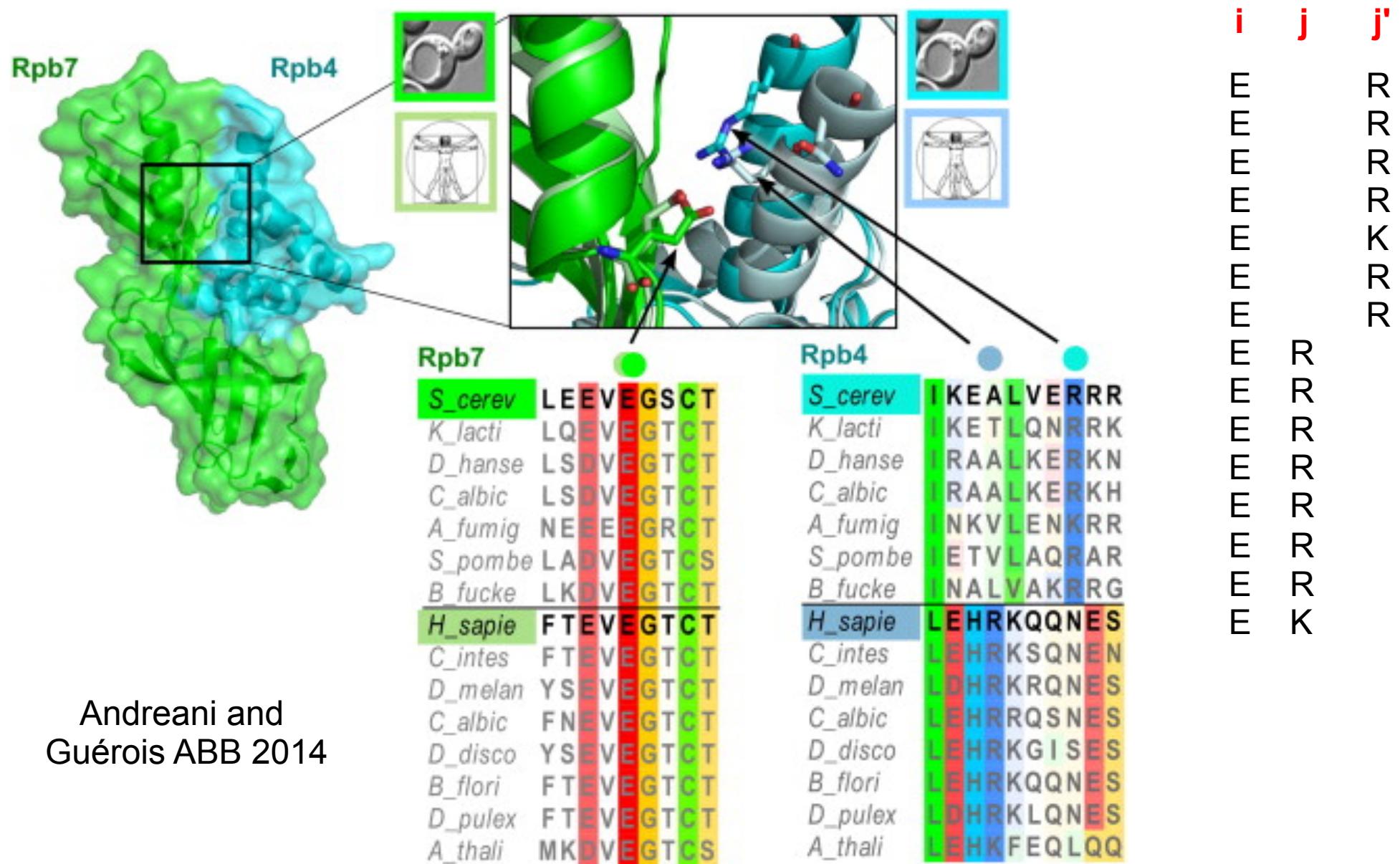
Coévolution

Une mutation dans la séquence d'un partenaire est compensée ou contre-balancée par une mutation dans la séquence de l'autre partenaire.

Points de contact



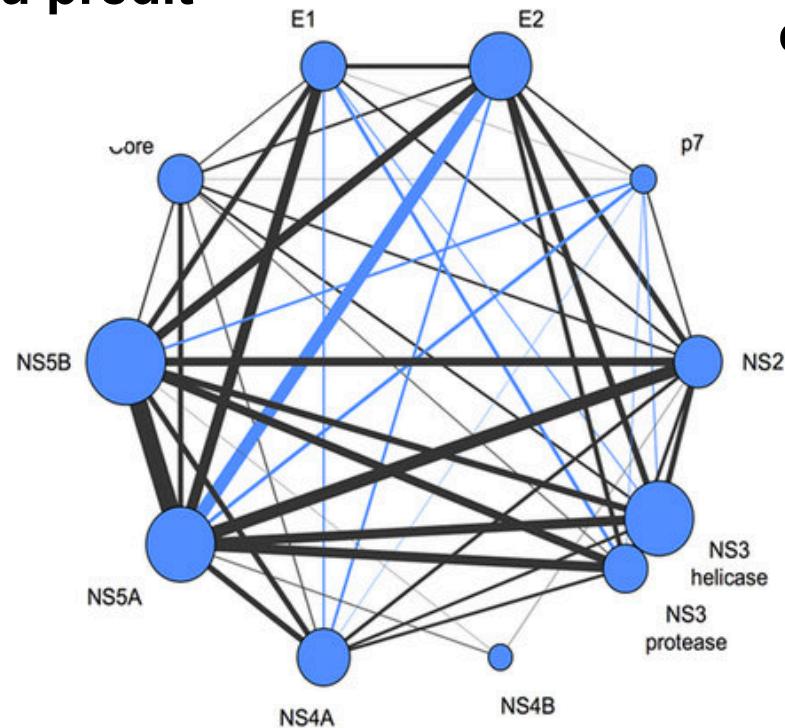
Coévolution entre triplets



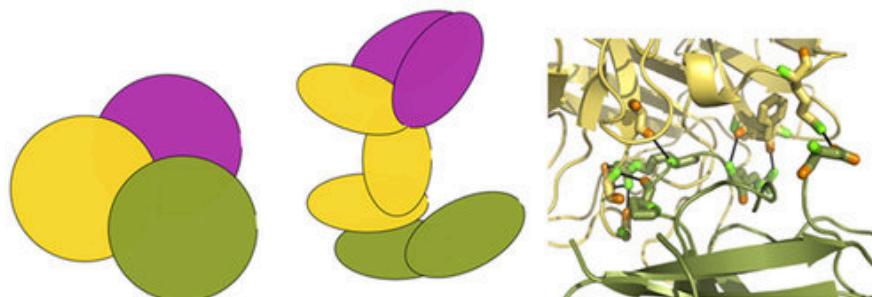
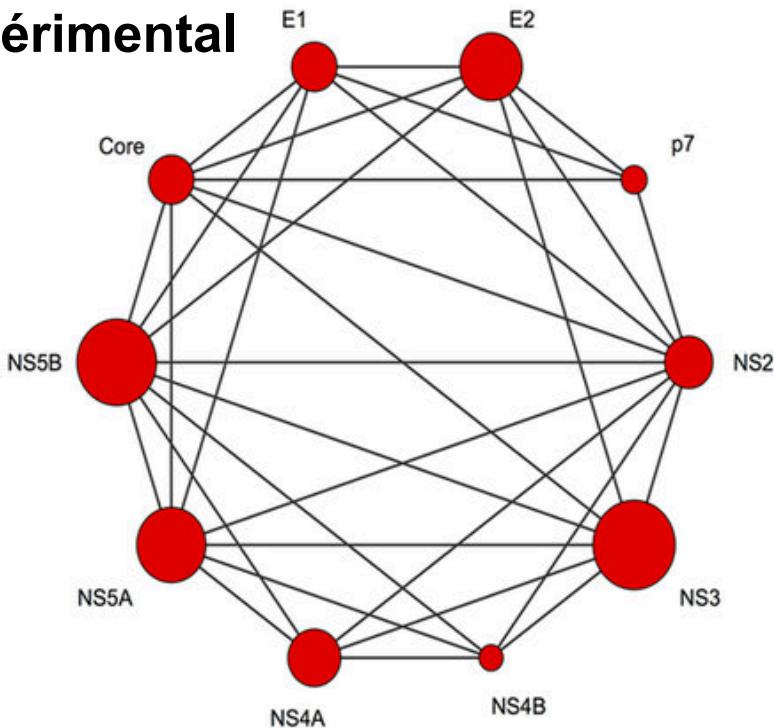
Andreani and
Guéris ABB 2014

Interactome du virus HCV

Réseau prédit



Réseau expérimental



resolution



Carbone *et al.* Sci. Reports 2016

Not only for interactions

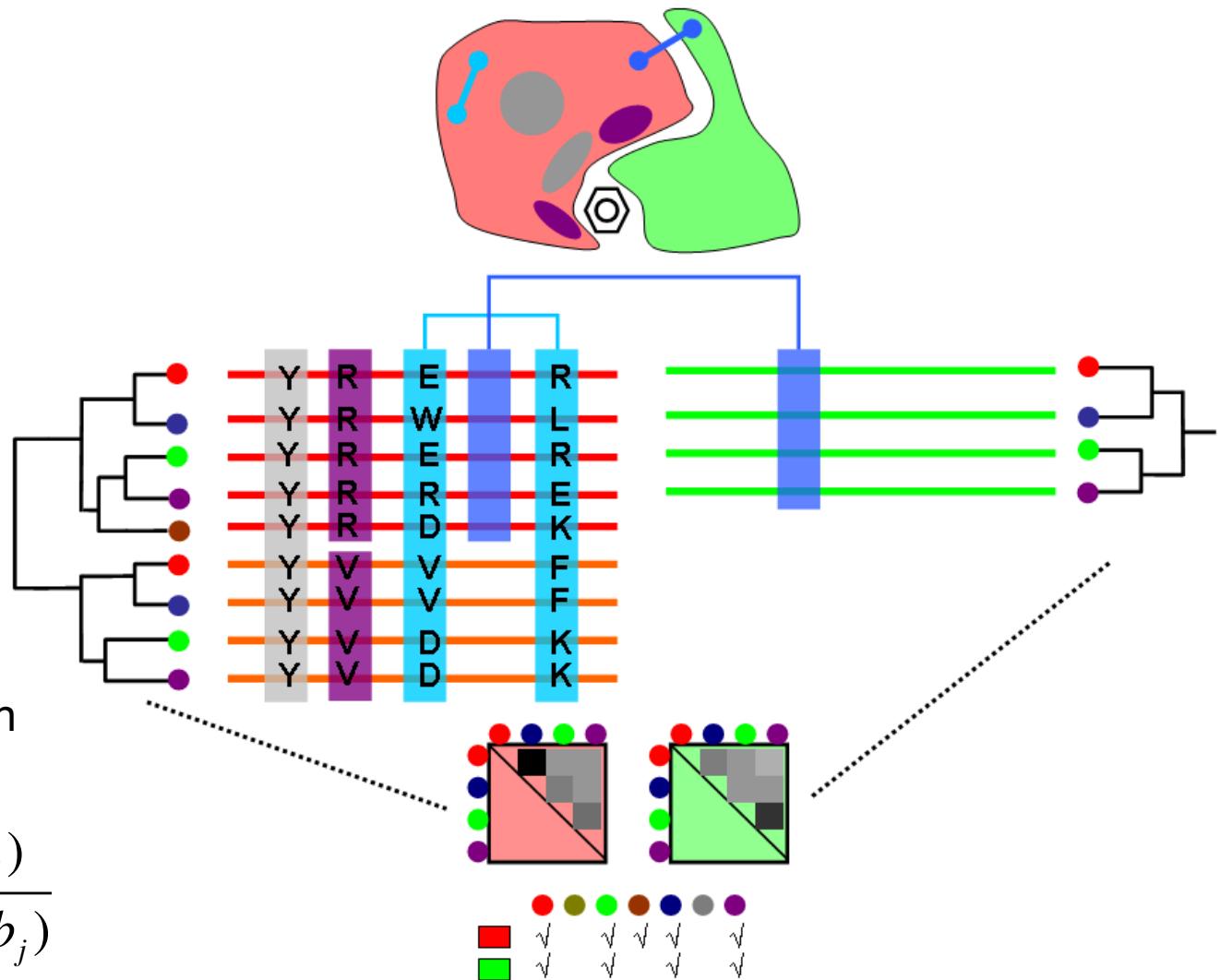
Un signal de coévolution est la marque d'une **dépendance fonctionnelle** entre 2 ou plusieurs positions, pas nécessairement un contact !

Beaucoup de mesures ont été développées pour détecter ces signaux.

liste non-exhaustive :
[http://csbg.cnb.csic.es/
rev_coevol_nrg/](http://csbg.cnb.csic.es/rev_coevol_nrg/)

La plus "simple" est l'information mutuelle entre 2 positions

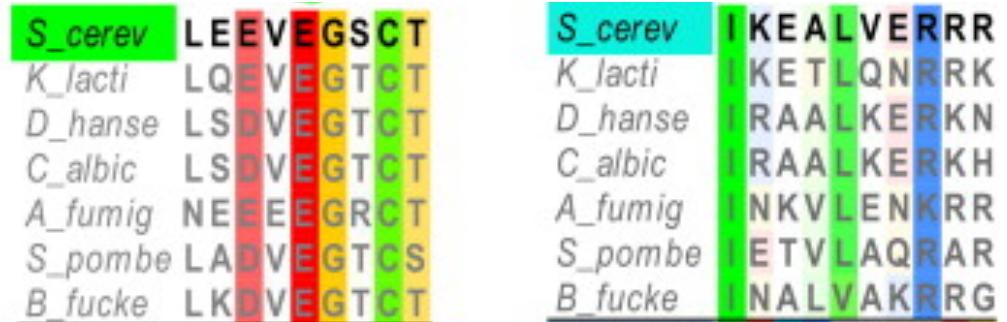
$$MI_{i,j} = \sum_{a,b} f(a_i, b_j) \log \frac{f(a_i, b_j)}{f(a_i)f(b_j)}$$



Mesures de coévolution

Une mesure classique et simple, qui requiert un grand nombre de séquences différentes et qui ne prend pas en compte les relations évolutives entre ces séquences est l'information mutuelle (MI). Elle repose sur l'entropie de Shannon.

$$H_i = -\sum_{k=1}^n p(a_k) \log p(a_k)$$



$$MI_{i,j} = -\sum_{k=1}^n \sum_{l=1}^m p(a_k, a_l) \log p(a_k, a_l)$$

Plusieurs corrections/modifications de cette mesure ont été introduites pour s'affranchir du bruit inhérent aux données.

MI_p: Dunn et al. 2008 Bioinformatics Mutual information without the influence...

MI_c: Lee & Kim 2009 Bioinformatics A new method for revealing correlated...

DCA: Weigt et al 2009 PNAS Identification of direct residue contacts...

Coévolution

Une mutation dans la séquence d'un partenaire est compensée ou contre-balancée par une mutation dans la séquence de l'autre partenaire.

Avantages

Les signaux de coévolution peuvent être très précis, résolution au niveau du résidu, et indiquer des points de contacts.

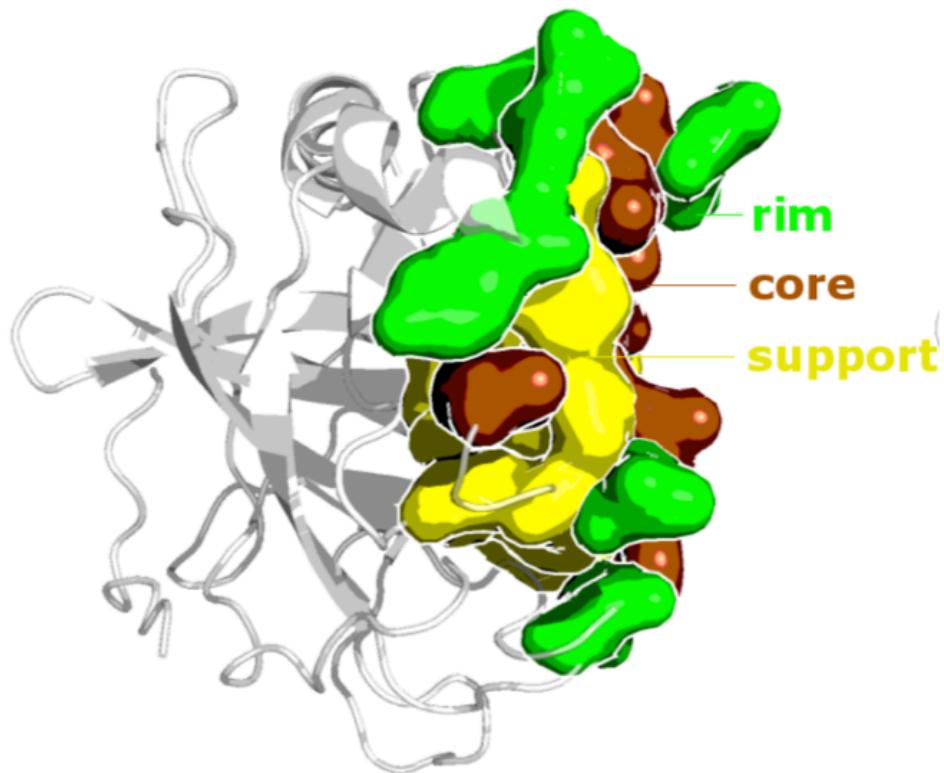
Inconvénients

- la notion de dépendance fonctionnelle est plus large que celle de contact
- la plupart des méthodes requièrent un grand nombre de séquences
- les signaux de coévolution entre partenaires non-permanents sont généralement faibles et difficiles à détecter.

Que peut-on dire à partir d'une protéine seule sur ses interactions ?

Structure de l'interface

Les résidus à l'interface peuvent être classés selon leur degré d'enfouissement
(Levy, *J. Mol Biol.* 2010).

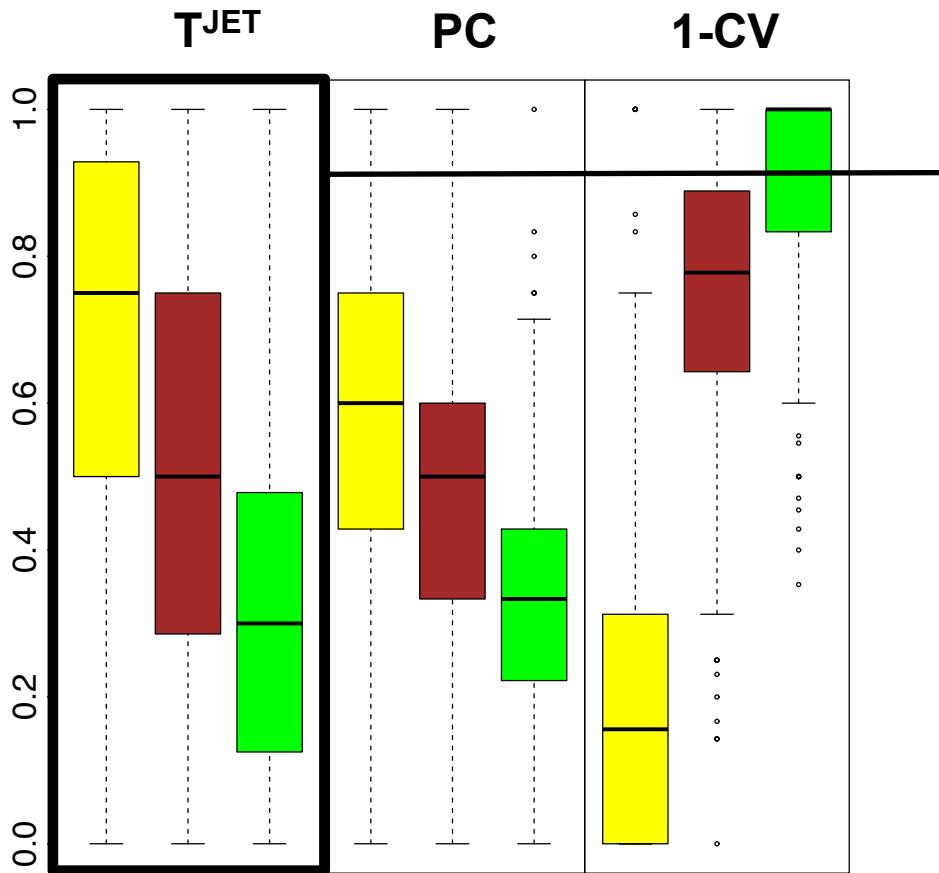


- **support**: $\text{rsa}_u \leq 25\%$, $\text{rsa}_b \leq 25\%$
- **core** : $\text{rsa}_u > 25\%$, $\text{rsa}_b \leq 25\%$
- **rim** : $\text{rsa}_u > 25\%$, $\text{rsa}_b > 25\%$

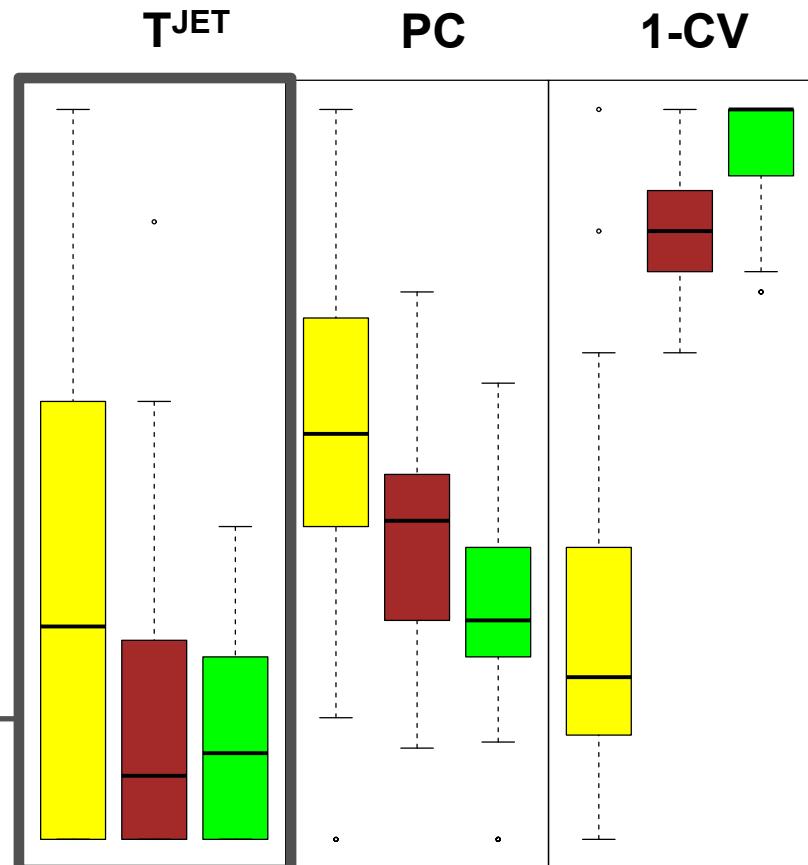
rsa: relative solvent accessible surface area (calculée avec NACCESS, MSMS...)
u/b: unbound/bound

rsa $\leq 25\%$: résidu enfoui
rsa $> 25\%$: résidu de surface

Propriétés de l'interface

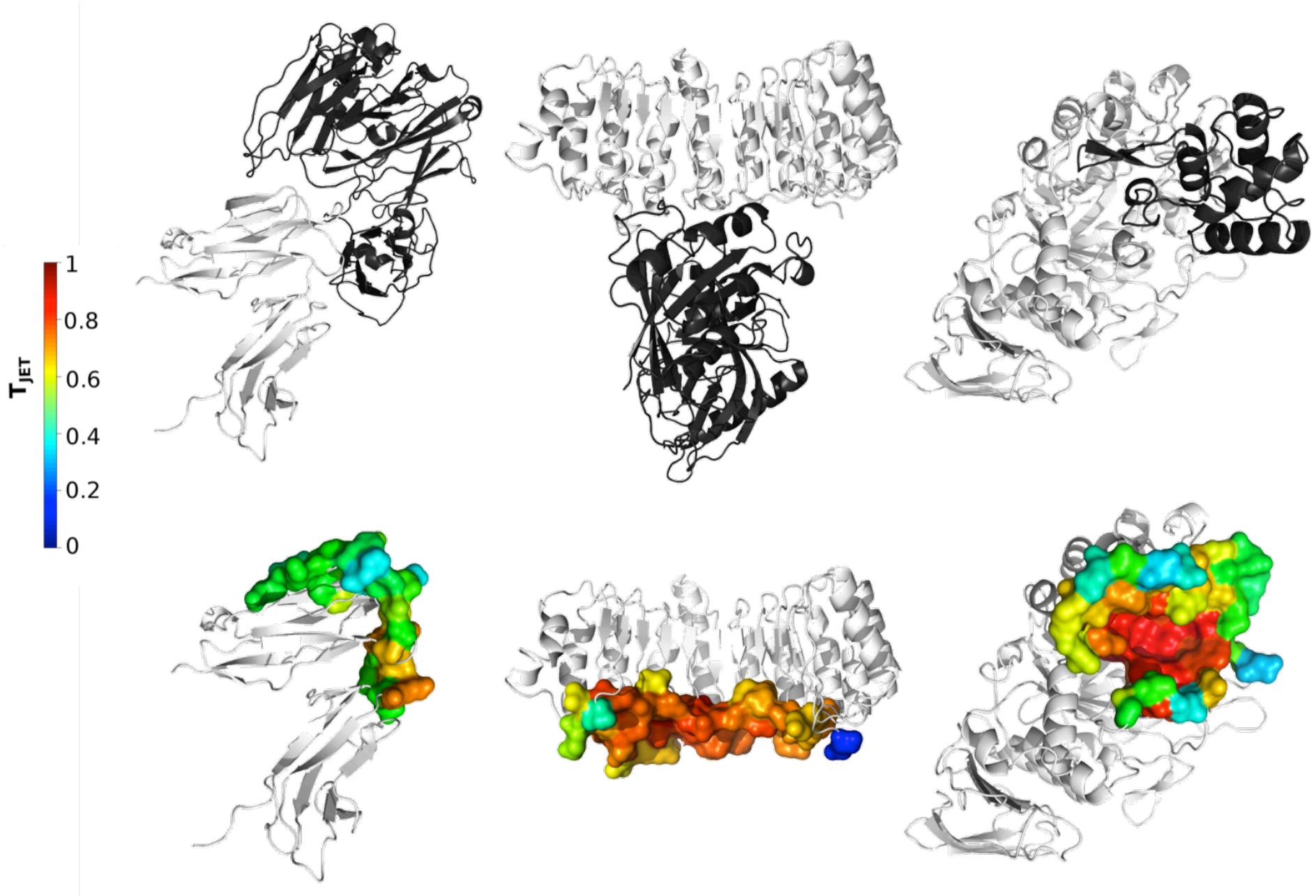


En général, les résidus du support, et certains du core, sont plus conservés que le reste de la protéine.
[stats sur 352 protéines, PPDBv4]

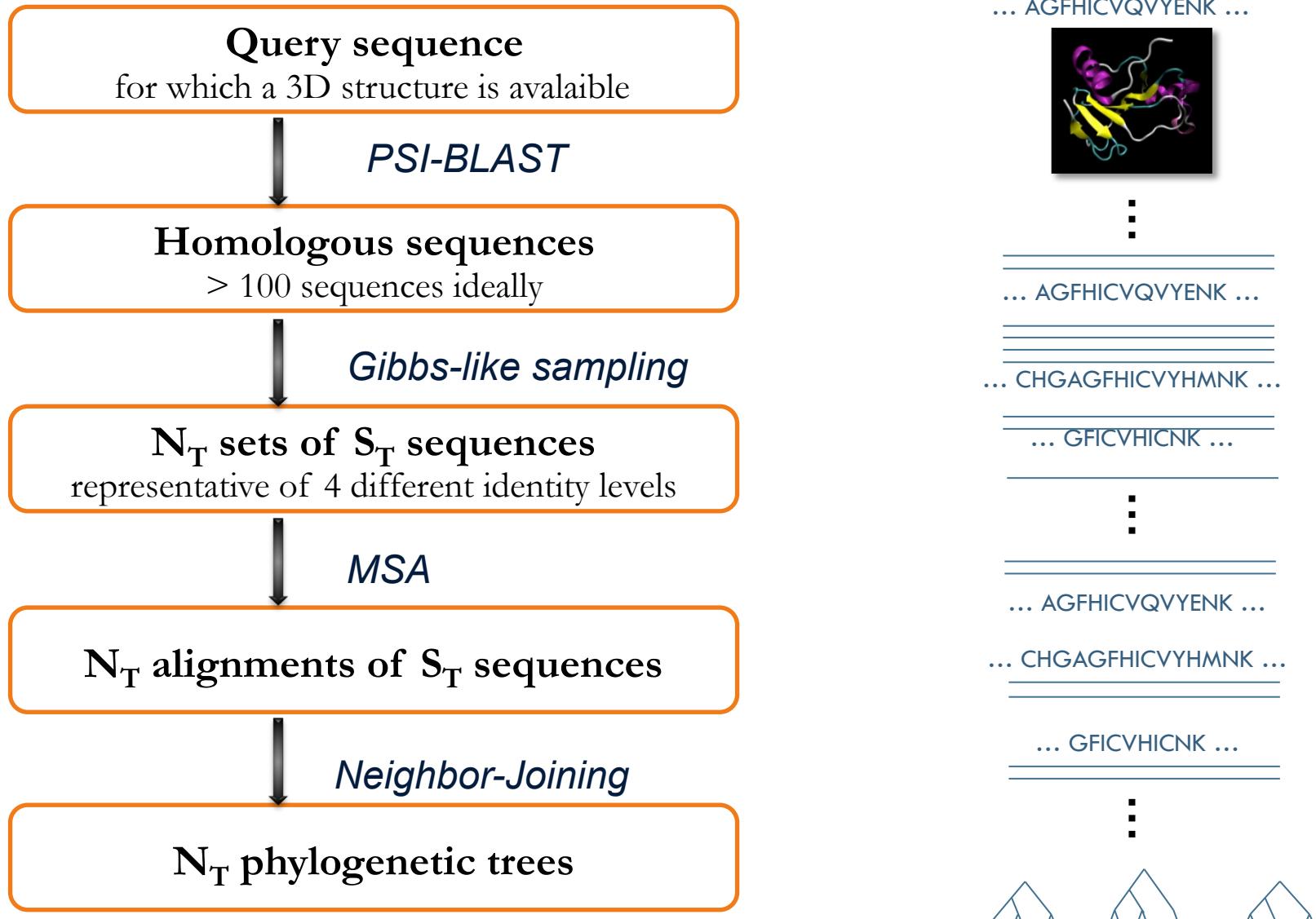


Mais ce n'est pas toujours le cas ! Les sites de fixation des antigènes sur les anticorps possèdent des propriétés très différentes !

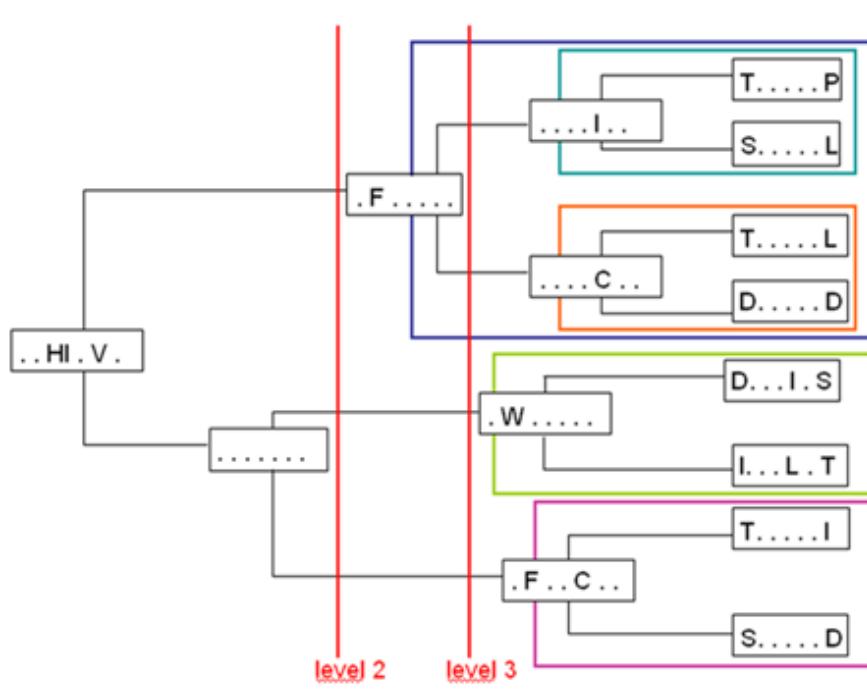
Quelques exemples



Trace évolutive jointe (T_{JET})



Trace évolutive jointe (T_{JET})



Level 2 :

.....
.F.....
.W.....
.F..C..

Level 3 :

.F.....
....I..
....C..
.W.....
.F..C..

.....
.3..1..
..HI.V.
.....
.XHI.V.

.....
.3..3..
.XHI.V.
.....
.XHI XV.

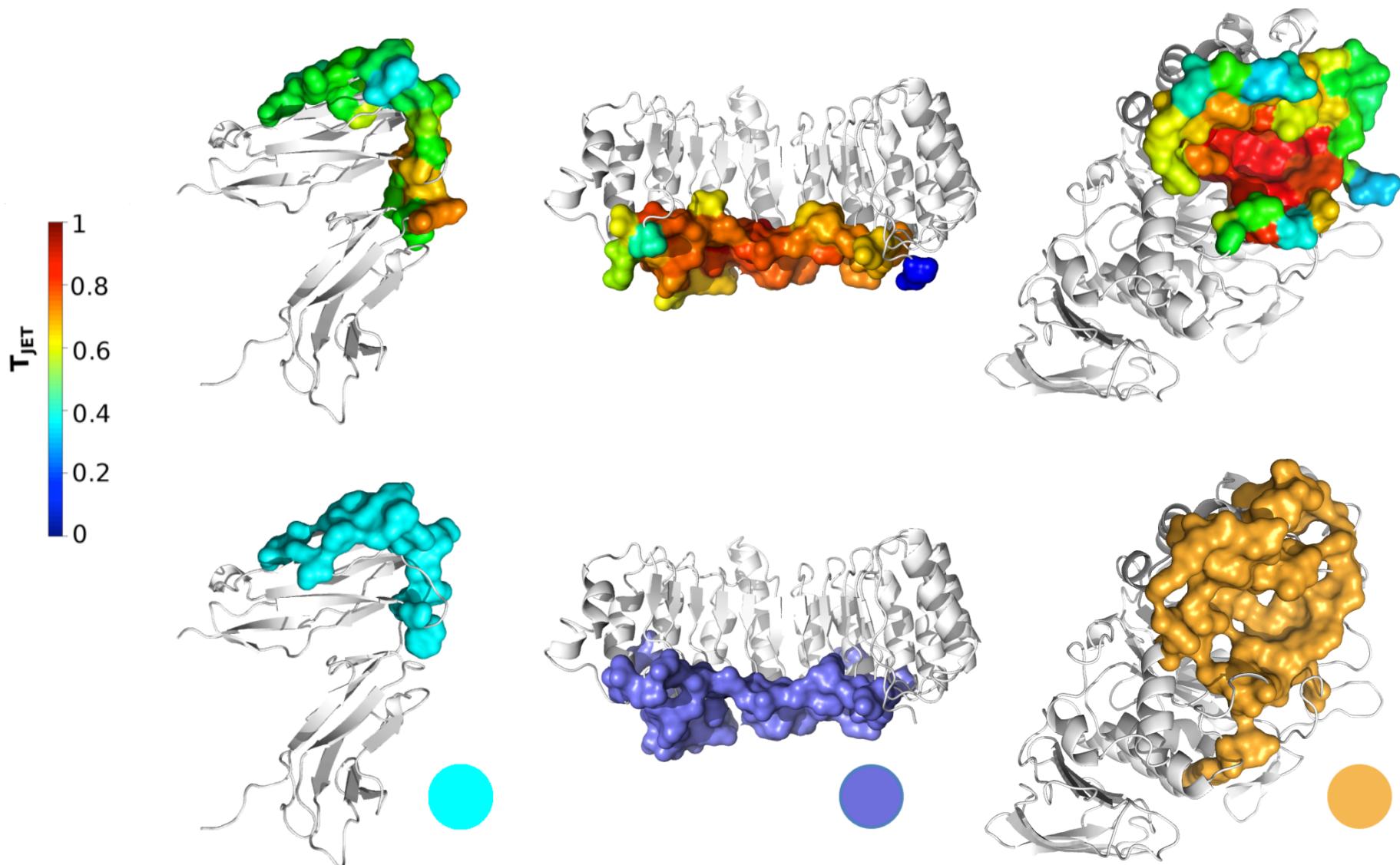
Pour chaque position dans la séquence de la protéine, la trace évolutive correspond au niveau (profondeur) dans l'arbre à partir duquel l'acide aminé à cette position a été fixé.

Les traces sont moyennées sur l'ensemble des arbres générées.
(résultats statistiquement robustes)

La notion de trace évolutive jointe est inspirée de celle **de trace évolutive** (Lichtarge, *J. Mol. Biol.* 1996).

T_{JET} est implémentée dans JET
(Engelen *et al.* PLoS CB 2009) et JET2
(Laine & Carbone, PLoS CB 2015)

Prédictions JET²



Comment intégrer de l'information
évolutive dans le docking

Evolution dans le docking

Type de données

- sites d'interaction prédis pour chaque partenaire
- points de contact prédis entre les partenaires
- coordonnées tridimensionnelles d'interologues.

Type d'utilisations

- (1) Guider/restreindre la recherche de l'espace de docking
- (2) Améliorer l'évaluation des poses (fonction de score)
- (3) Calquer la prédiction du complexe sur des structures connues de "patrons"

(1) Restreindre le sampling

Les signaux de coévolution (corrélations entre paires de résidus) peuvent être utilisés comme contraintes de distance : on impose que les résidus co-évolués soient proches dans le complexe prédit.

Protocole utilisé dans HADDOCK :

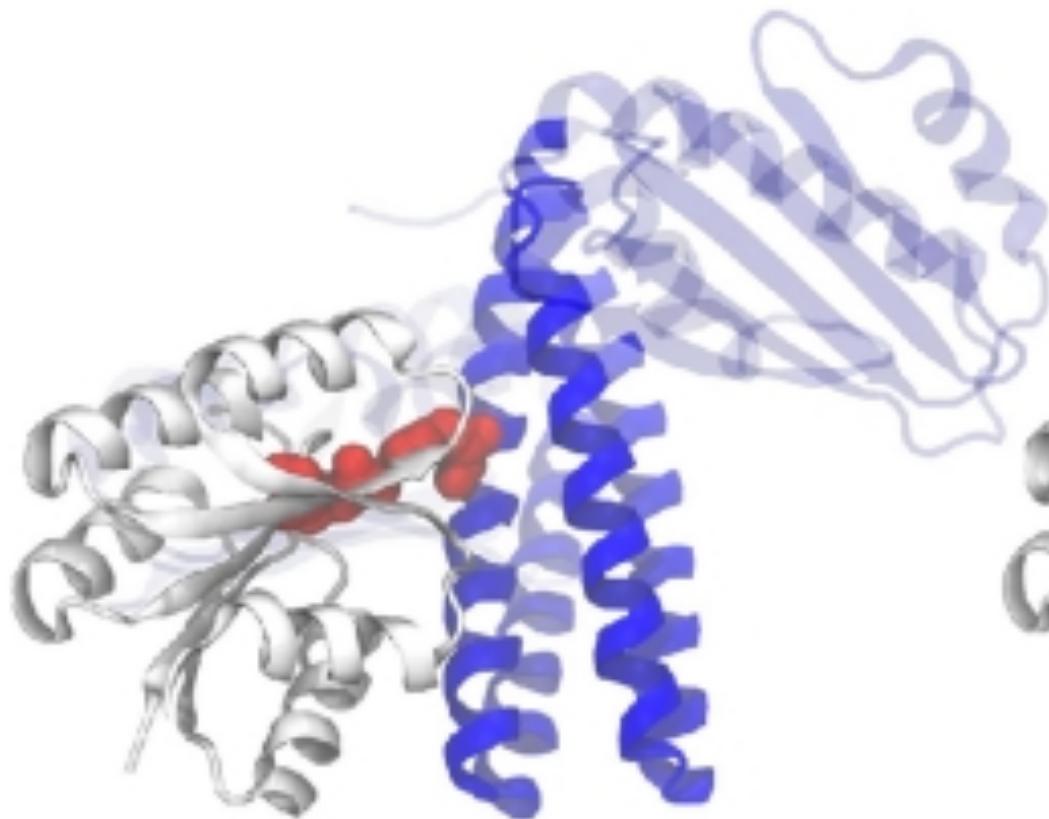
<http://www.bonvinlab.org/software/haddock2.2/>

Chaque contrainte de distance est traitée de manière ambiguë (distances effective, minimum et maximum)

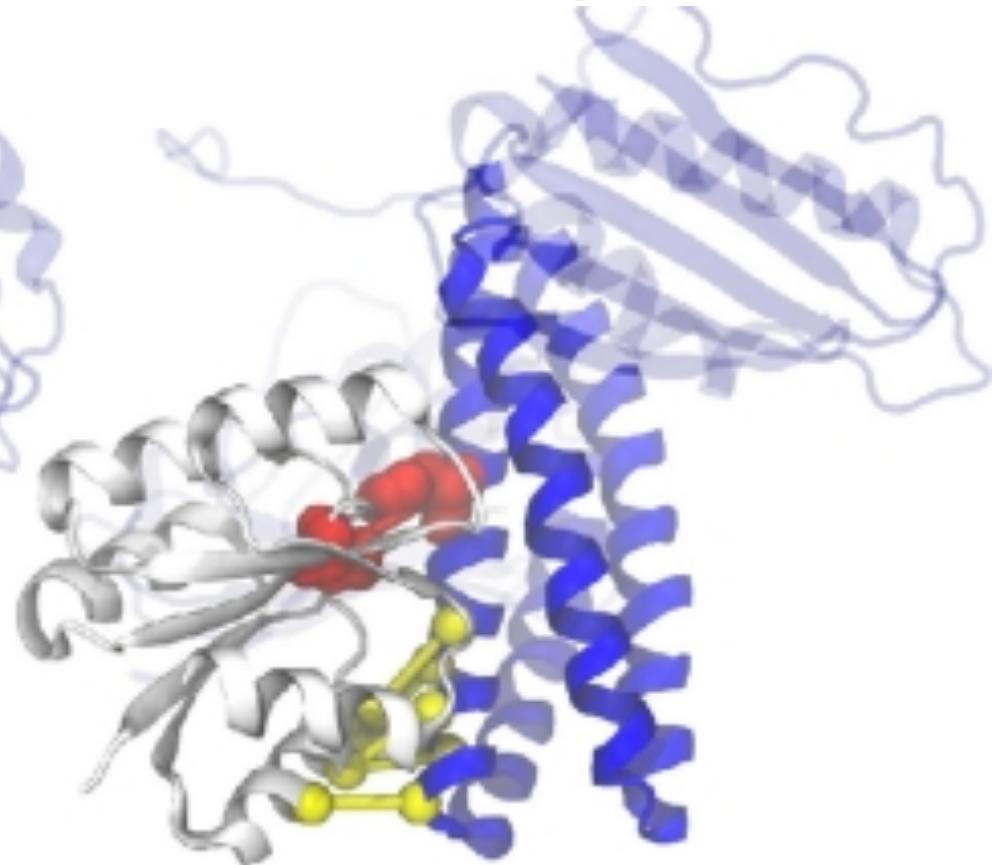
Historiquement développé pour inclure des mesures de RMN.

(1) Restreindre le sampling

Structure expérimentale



Structure prédictée



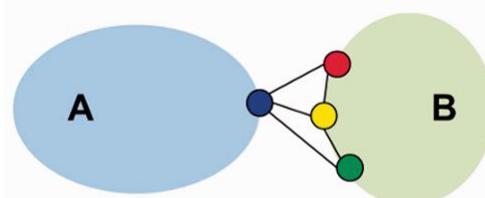
En jaune : paires de résidus identifiées comme co-évoluées par la méthode DCA.
Weigt et al. *PNAS* 2009

(2) Améliorer le scoring

potentiels statistiques entre paires et triplets

information extraite de l'alignement des séquences homologues

A Inter-molecular contacts from interface tessellation



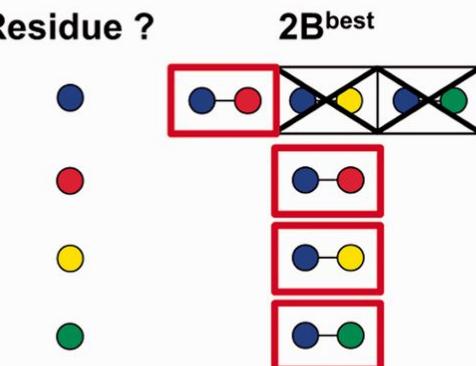
B Decomposition into 2- and 3-body scores

● ●	0.53	● ● ●	0.04
● ● ○	-0.05	● ○ ○	
● ○ ○	-0.11	○ ○ ○	-0.26

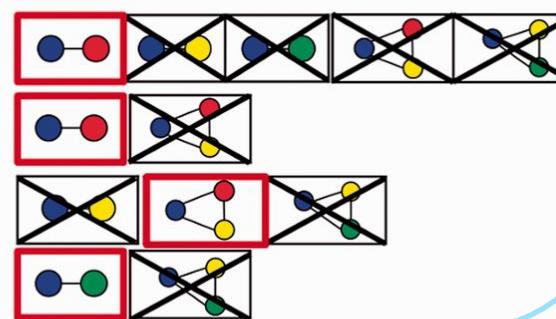
$$2B = (0.53 - 0.05 - 0.11) = 0.37 \quad 3B = (0.04 - 0.26) = -0.22$$

C Search for best score per residue in $2B^{\text{best}}$ and $2/3B^{\text{best}}$ schemes

Residue ?



$2/3B^{\text{best}}$



D Inclusion of evolutionary information in case of the $2B_{\text{evol}}$ scheme

Alignment for A

	i	j
<i>H. sapiens</i>	○ ○ ○	● ○ ○
<i>M. musculus</i>	○ ○ ○	○ ○ ○
<i>D. rerio</i>	○ ○ ○	○ ○ ○
:	○ ○ ○	○ ○ ○
<i>S. cerevisiae</i>	○ ○ ○	● ○ ○

Alignment for B

	i	j
<i>H. sapiens</i>	○ ○ ○	● ○ ○
<i>M. musculus</i>	○ ○ ○	○ ○ ○
<i>D. rerio</i>	○ ○ ○	○ ○ ○
:	○ ○ ○	○ ○ ○
<i>S. cerevisiae</i>	○ ○ ○	● ○ ○

$2B_{\text{evol}}$ score term for

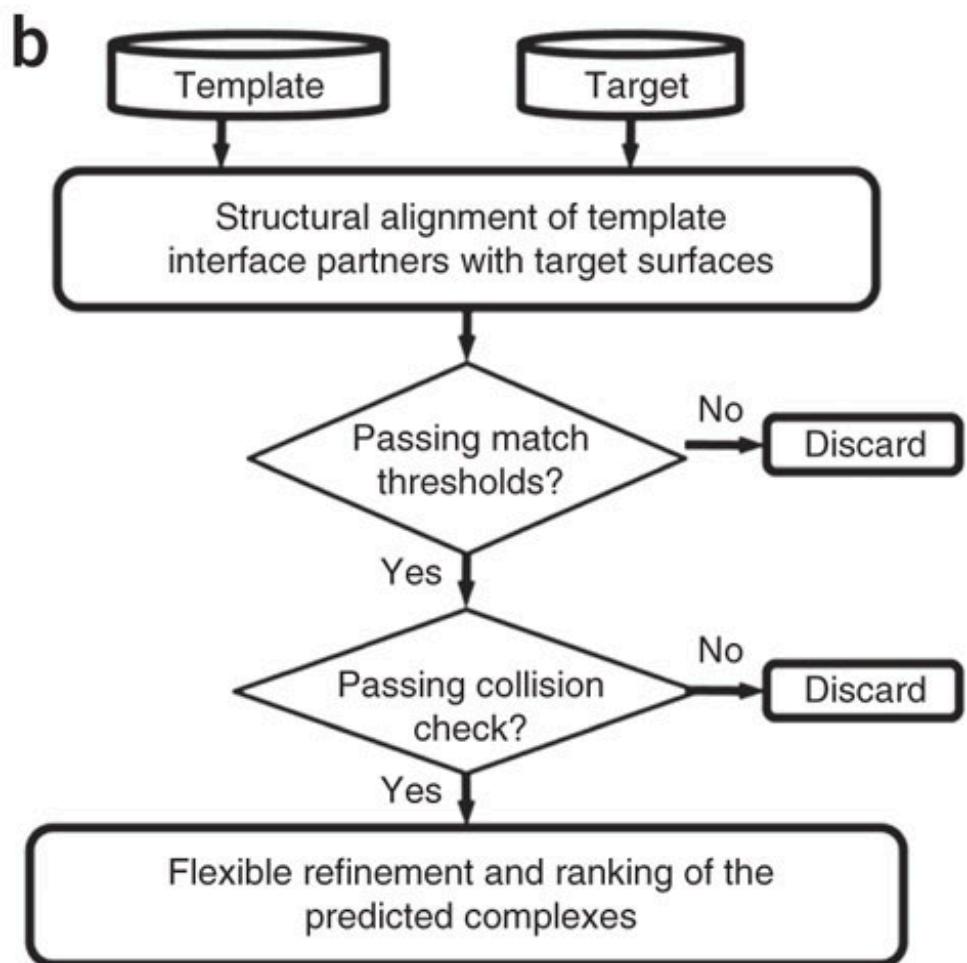
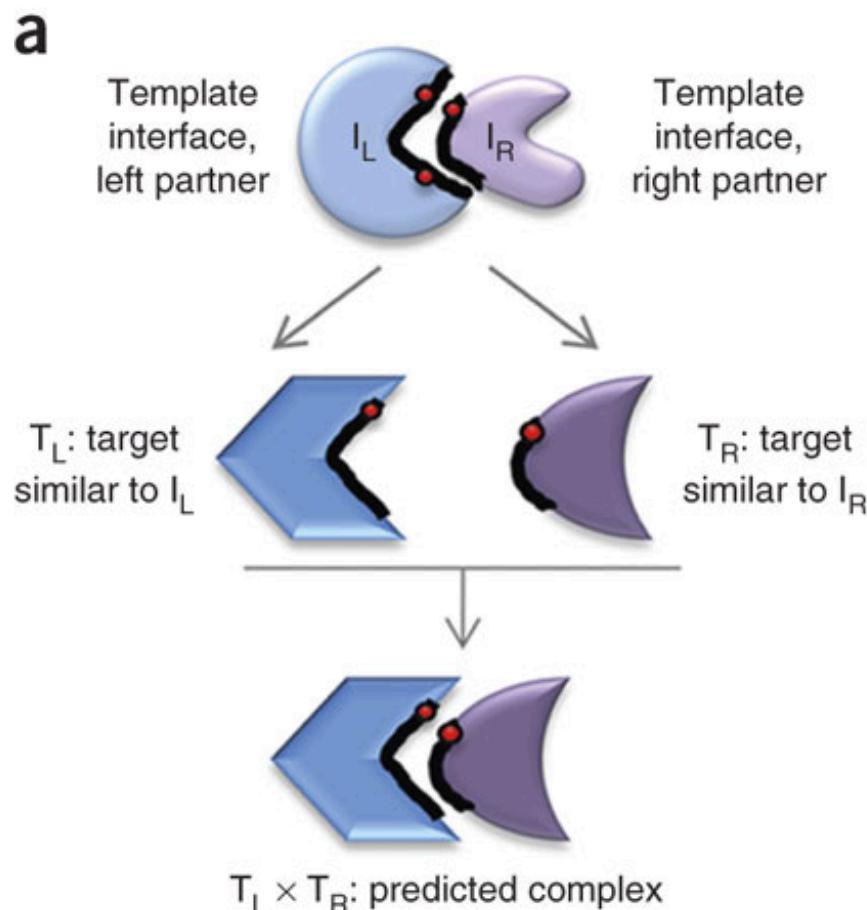
+	● ●	0.53
+	● ●	0.53
+	● ●	0.53
+	● ○	-0.05
+	○ ○	-0.18
+	○ ●	0.53

1.89

InterEvScore :
<http://biodev.cea.fr/interevol/interevscore/>
Andreani et al.
Bioinformatics 2013

(3) Template-based docking

PRISM (Tuncbag *et al.* Nat. Protocols 2011): <http://cosbi.ku.edu.tr/prism/index.php>



(3) Template-based docking

Avantages

Efficacité/rapidité du code, pas besoin d'échantillonner tout l'espace de recherche de docking.

Inconvénients

- limité aux structures résolues et disponibles dans la PDB
- tributaire de la qualité des structures des patrons.

Outils

Calculs de surface accessible au solvant :

SURFace

http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:SURFace_Algorithms

NACCESS

<http://www.bioinf.manchester.ac.uk/naccess/>

Python-based ASA

<http://boscoh.com/protein/calculating-the-solvent-accessible-surface-area-asa.html>

MSMS

<http://mgltools.scripps.edu/downloads>

Calc-surface (libproteingeometry package)

<http://www.molmovdb.org/geometry/>

GETAREA (web server)

<http://curie.utmb.edu/getarea.html>

Prédiction de sites fonctionnels à partir de la notion de conservation :

Consurf

http://bental.tau.ac.il/new_ConSurfDB/

Evolutionary Trace

<http://mammoth.bcm.tmc.edu/lab.html>

INTREPID

<https://github.com/berkeleyphylogenomics/intrepid>

Joint Evolutionary Trees (JET²)

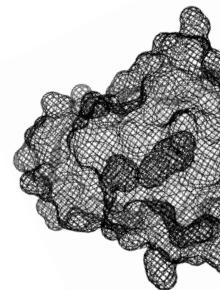
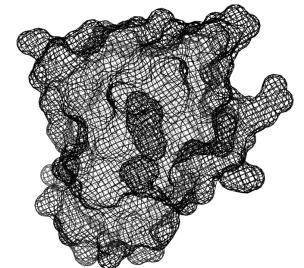
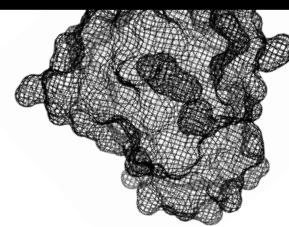
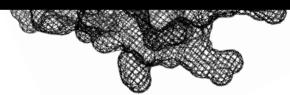
<http://www.lcqb.upmc.fr/JET2/JET2.html>

http://www.jet2viewer.upmc.fr/jet2_viewer/report/index.html

Conseil du jour



LENTEMENT, JE SUIS PRESSÉ(E)



A faire pour la prochaine fois

- Développer un prototype pour la partie sampling
- Regarder en détail les fichiers PDB tests et déterminer les différentes propriétés de ces complexes
- Implémenter une fonction de score et l'évaluer sur ces PDBs
- Nous présenter votre avancement sous forme de diaporama (intra-équipe)