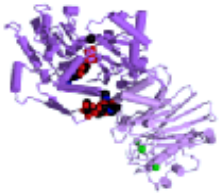
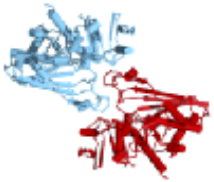
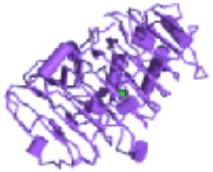
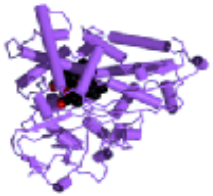
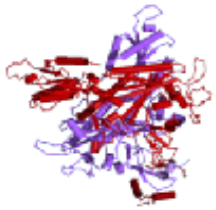
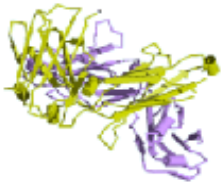
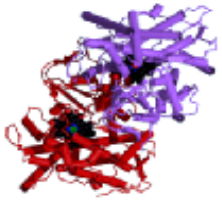
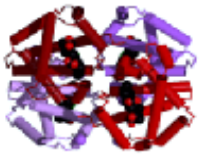







Protein structure and evolution

GT MASIM
16 novembre 2017

Mathilde Carpentier
Maître de conférences UPMC
Atelier de BioInformatique
Institut de Systématique, Evolution, Biodiversité
MNHN – CNRS – UMPC - EPHE




Atelier de BioInformatique (ABI)
l'ISYEB (MNHN) since October 2015

Permanent membres

G Achaz

S Brouillet

C Bertrand

M Carpentier

S Pasek

J Pothier

M Boccara

Associated members

B Billoud

E Duchaud

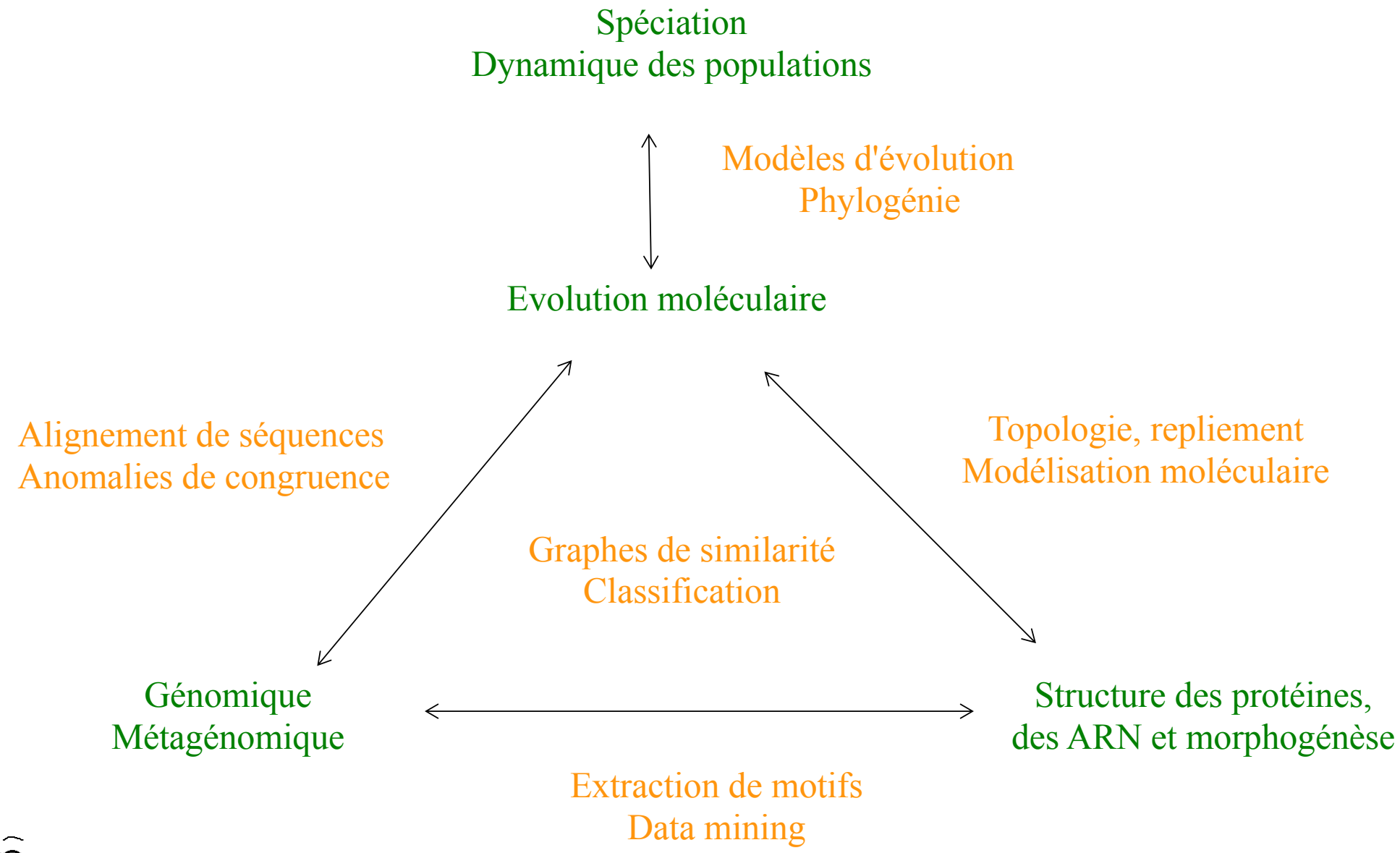
G Sapriel

H Soldano

I Lafontaine

P Brezellec

Atelier de BioInformatique (ABI)
ISYEB (MNHN) since October 2015



Protein structure comparison

Structural database scanning

- Yakusa[1,2]
  www.rpsb.jussieu.fr/Yakusa/

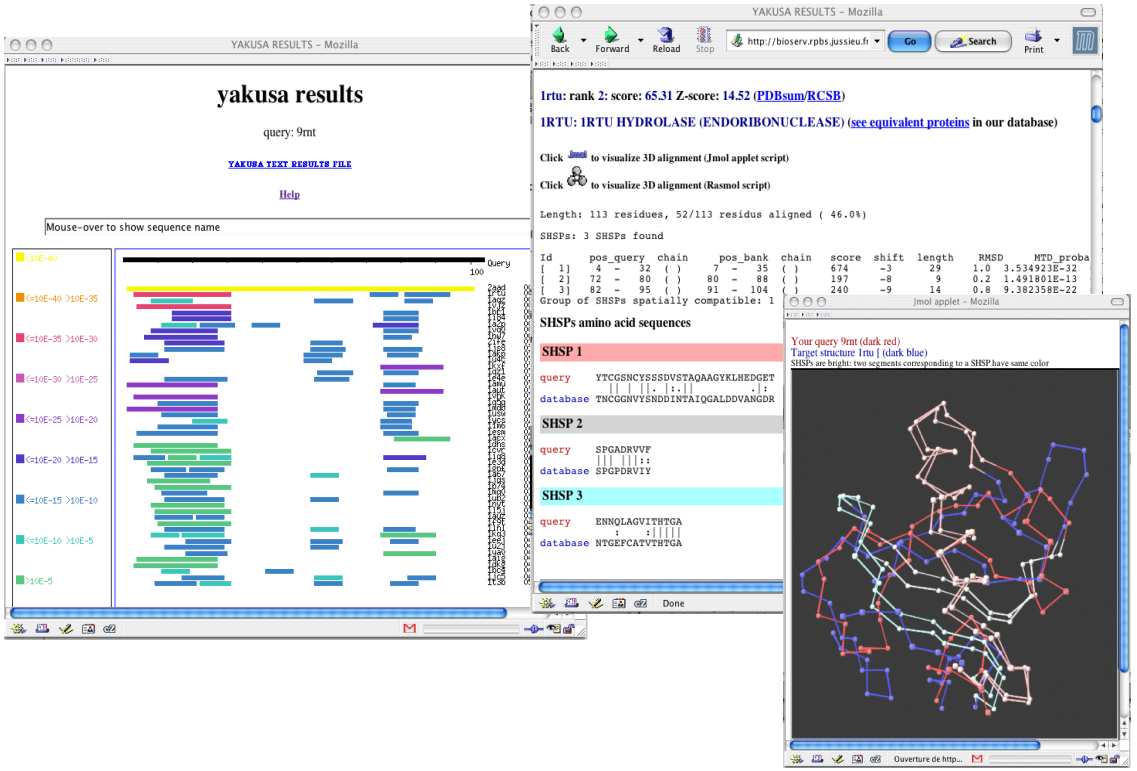
Multiple structural alignment

- Gok (KMR + alpha angles)

- « m-diagonals » methods

- Gibbs sampler method

- Relational motifs (Triades)[3,4]



[1] M. Carpentier, S. Brouillet, J. Pothier, YAKUSA: a fast structural databases scanning method, Proteins: Structure, Function, and Bioinformatics, volume 61, issue 1, pages 137-51.

[2] C. Alland, F. Moreews, D. Boens, M. Carpentier, S. Chiusa, M. Lonquety, N. Renault, Y. Wong, H. Cantalloube, J. Chomilier, J. Hochez, J. Pothier, B.O. Villoutreix, J.-F. Zagury, P. Tuffery, ;
RPBS: a web resource for structural bioinformatics, Nucleic Acid Research, 2005, 33: W44–W49

[3] N. Pisanti, H. Soldano, M. Carpentier, J. Pothier. **A relational extension of the notion of motifs: application to the common 3D protein substructures searching problem**. J Comput Biol (2009) N. Pisanti, H. Soldano, M. Carpentier, **Incremental Inference of Relational Motifs with a Degenerate Alphabet** , Lecture Note in Computer Science (2005).

[4] N. Pisanti, H. Soldano, M. Carpentier, J. Pothier, **Implicit and Explicit Representation of Approximated Motifs** KCL series book, edited by C. Iliopoulos, K. Park and K. Steinhfel (2005)

Comparison of sequence and structure alignment methods

- Do structure aligment methods detect homology?
- Are they better than sequence alignment methods?
- Is structure really more conserved than sequence?

Comparison of sequence and structure alignment methods

Are structural alignments really better than sequence alignments ?

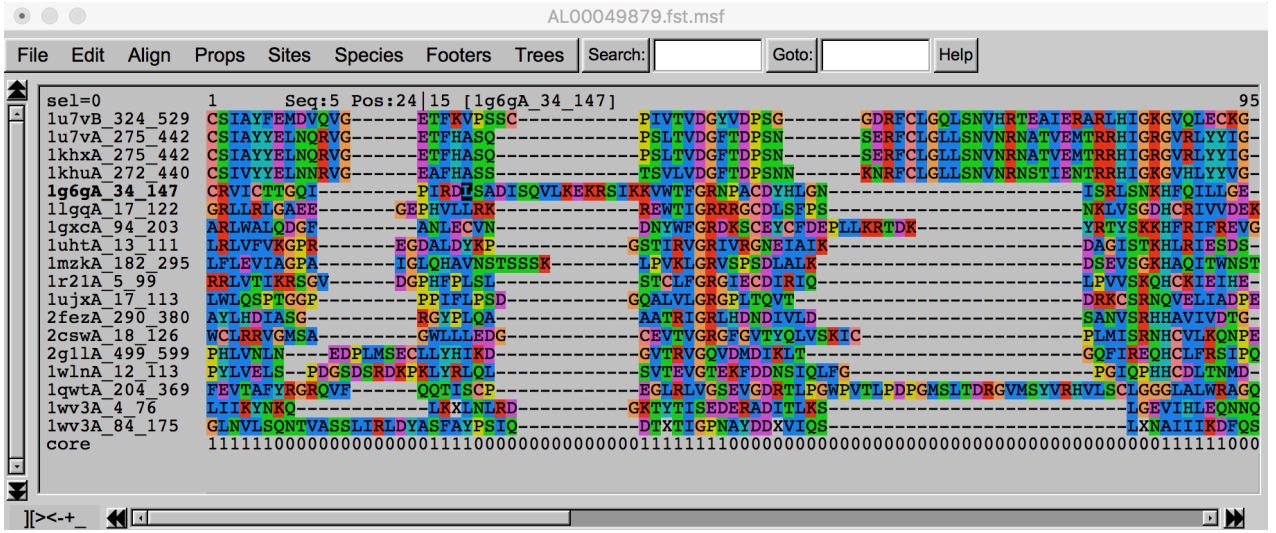
Data

Reference dataset = Manually curated protein multiple alignments with resolved structures

- Balibase 2[1] : 29 alignments
- Balibase 3[2] : 38 alignments
- Sisyphus[4] : 94 alignments

} 161 alignements

with a "core" alignment.



Homstrad[3] : 365 alignments

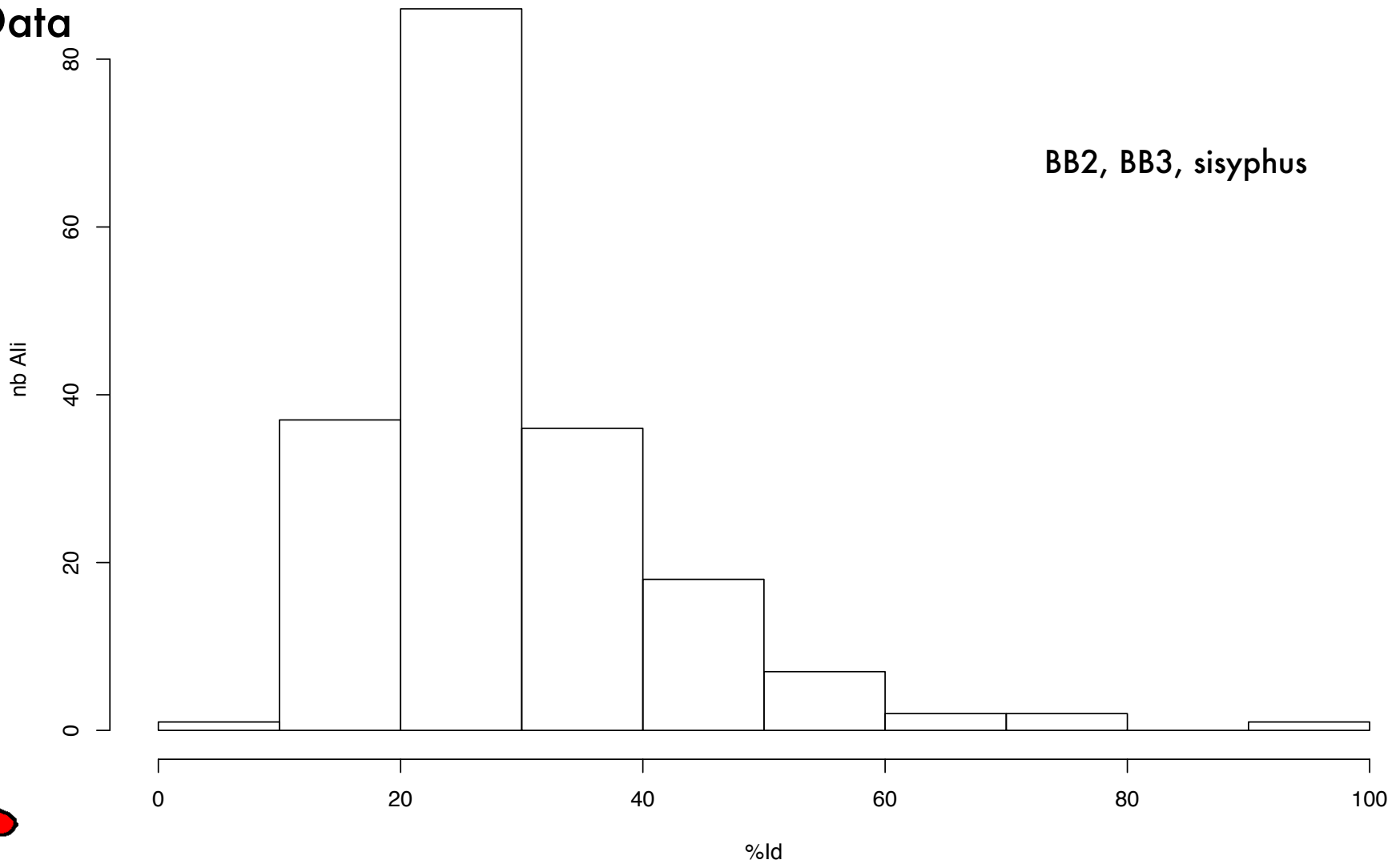
Problems:

- alignment by CE,
- no manual curation,
- core=SSE

1 Thompson et al. 1999
2 Thompson et al. 2005
3 Mizuguchi et al 1998
4 Andreeva et al 2007

Comparison of sequence and structure alignment methods

Data



BB2, BB3, sisyphus

Comparison of sequence and structure alignment methods

Scores [1]

- Sum of pairs (SP) : proportion of correctly aligned pairs

$$SP = \frac{\sum_{i=1}^{M} S_i}{\sum_{i=1}^{M_r} S_{ri}} \qquad S_i = \sum_{j=1, j\neq}^{N} \sum_{k=1}^{N} p_{ijk}$$

- Total Column (TC) : proportion of correctly aligned columns

$$CS = \frac{\sum_{i=1}^{M} C_i}{M}$$

1 Thompson et al. 1999

Comparison of sequence and structure alignment methods

Sequence alignment methods

DIALIGN	Morgenstern et al.	1998
CLUSTALW	Thompson et al	1994
TCOFFEE	Notredame et al	2000
MAFFT	Katoh et al	2002
MUSCLE	Edgar	2004
PRANK	Loytnoja et al	2005
PROBCONS	Mhabhashyam et al	2005
CLUSTALO	Sivers et al.	2011

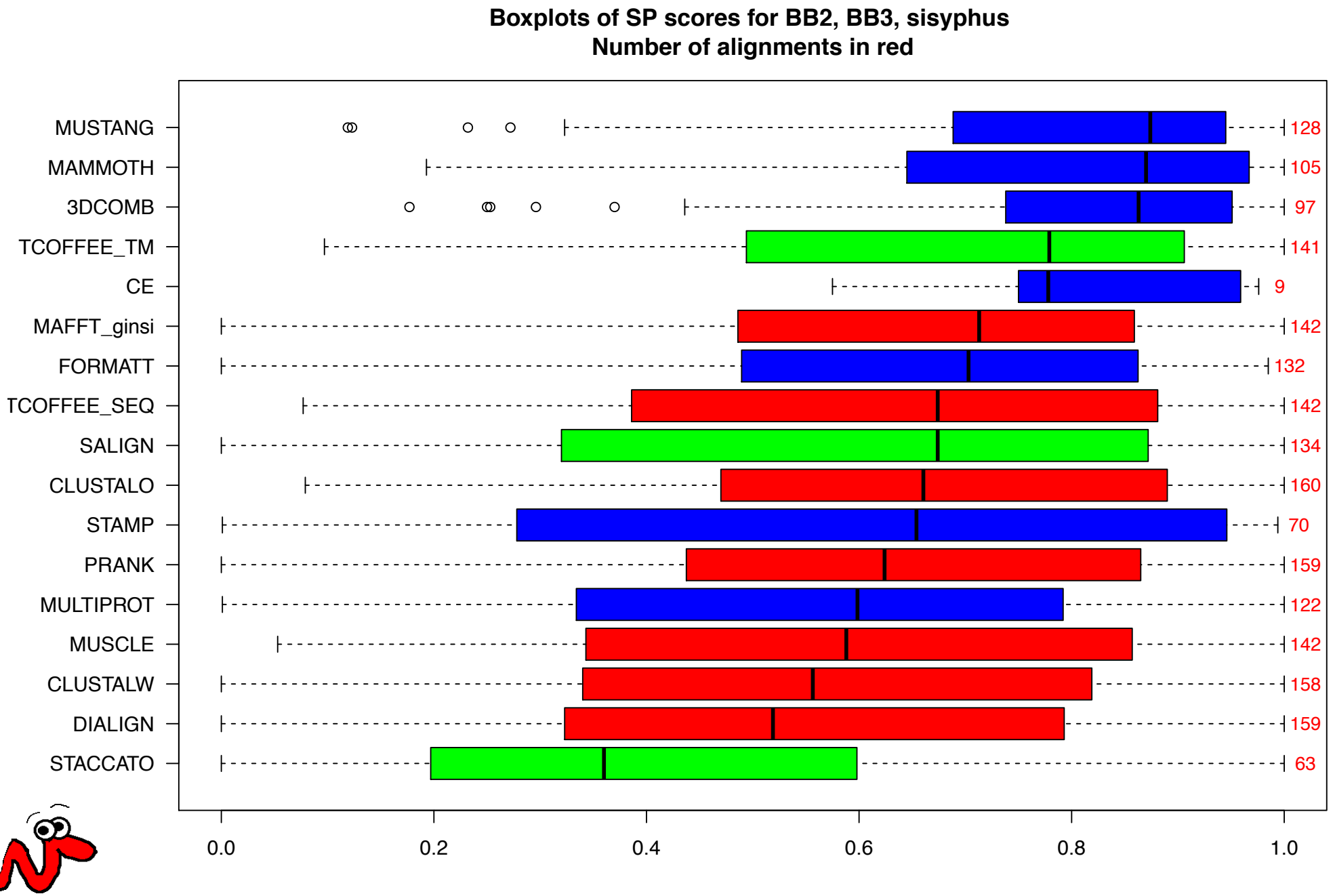
Structural alignment methods

Structure+Sequence alignment methods

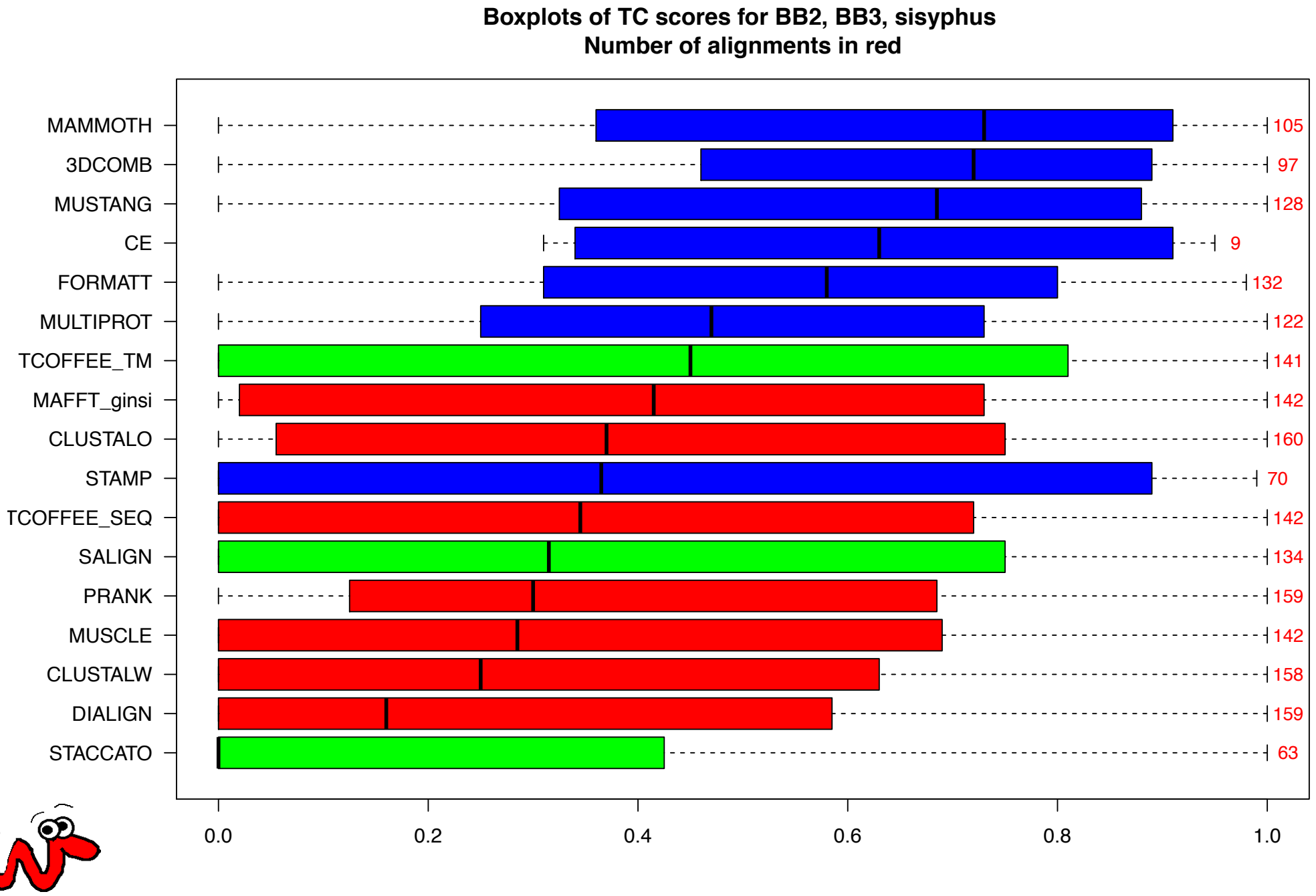
SSAP	C. Orengo & W. Taylor	1989
STAMP	**R. Russell and G. Barton**	**1992**
multal	Taylor, Flores et Orengo	1994
ProFit	ACR. Martin	1996
CE/CE-MC	**I. Shindyalov**	**2000**
Matras	**K. Nishikawa**	**2000**
PrISM	B. Honig	2000
MASS	O. Dror and H. Wolfson	2003
MolCom	S.D. O'Hearn	2003
SSM	E. Krissinel	2003
MALECON	S. Wodak	2004
MultiProt	**M. Shatsky and H. Wolfs**	**2004**
SWAPSC	Mario A. Fares	2004
C-BOP	E. Sandelin	2005
MAMMOTH-mult	**D. Lupyan**	**2005**
MUSTANG	**A.S. Konagurthu et al.**	**2005**
POSA	Y. Ye and A. Godzik	2005
TetraDA	J. Roach	2005
CBA	J. Ebert	2006

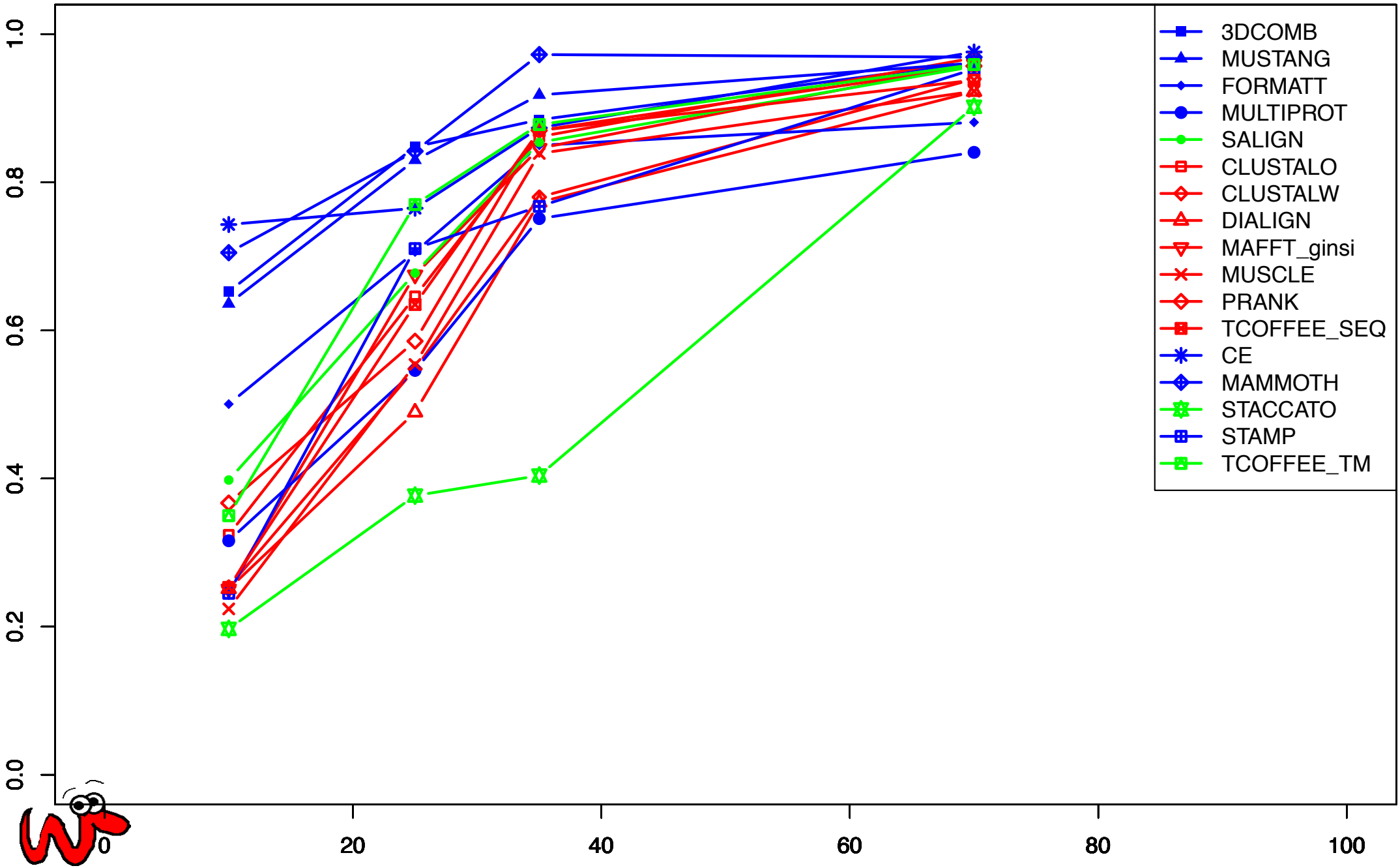
STACCATO	**Shatsky et al.**	**2006**
STRAP	C. Gille	2006
UCSF Chimera	E. Meng *et al.*	2006
CURVE	D. Zhi	2006
CAALIGN	T.J. Oldfield	2007
CLEMAPS	W-M. Zheng	2007
3DCOFFEE	**Notredame et al.**	**2007**
PyMOL	W. L. DeLano	2007
SALIGN	**M.S. Madhusudhan *et al***	**2007**
Vorolign	Birzele F, Gewehr J E, Csa	2007
BLOMAPS	W-M. Zheng & S. Wang	2008
Matt/Formatt	**M. Menke**	**2008**
mistral	micheletti et orland	2009
SMOLIGN	H. Sun et al	2010
EpitopeMatch	S. Jakuschev	2011
3DCOMB	**S. Wang and J. Xu**	**2012**
msTALI	P. Shealy & H. Valafar	2012
mulPBA	A.P. Joseph *et. al.*	2012
Fit3D[9]	F. Kaiser *et al.*	2015

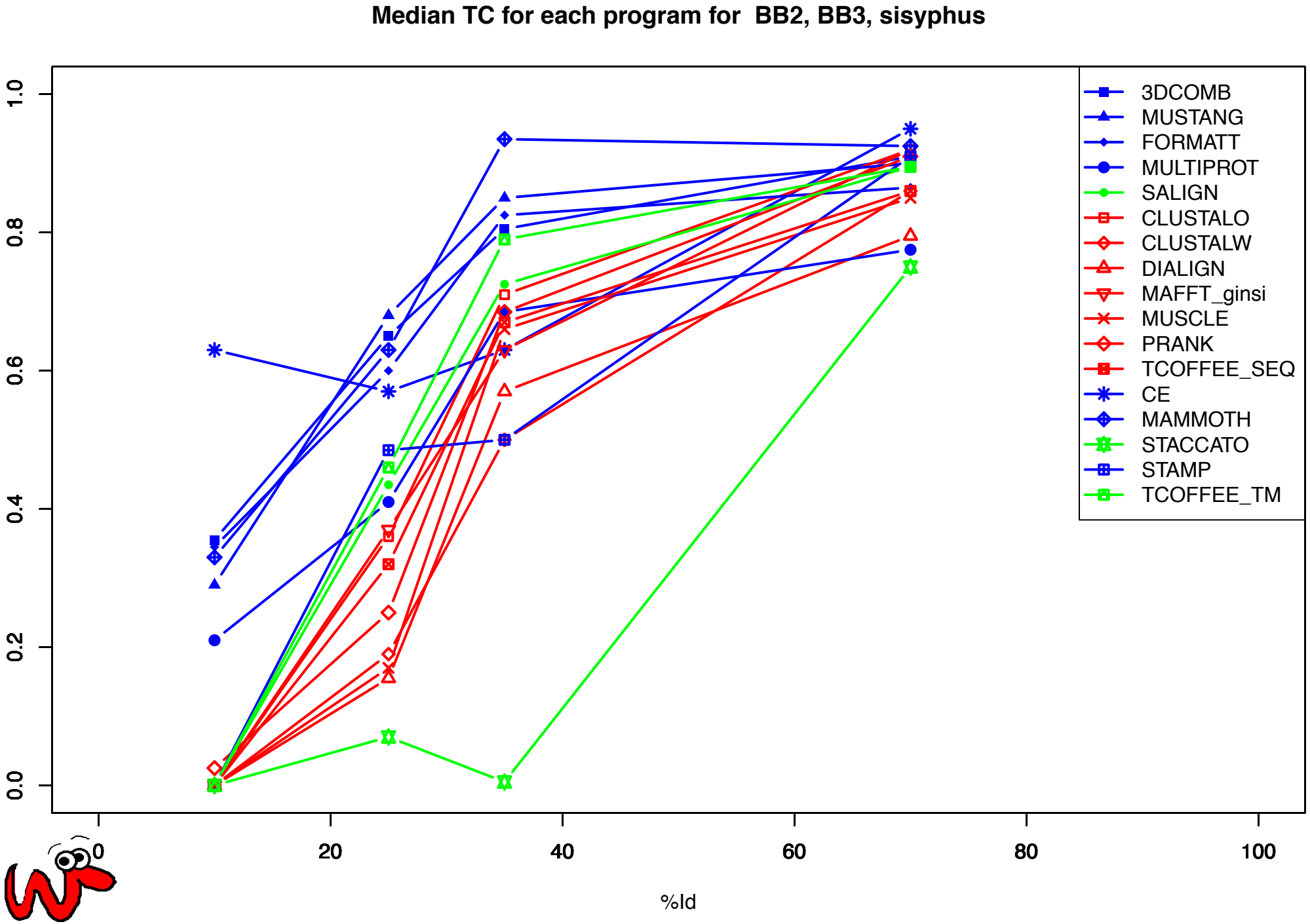
SP (sum of pairs)



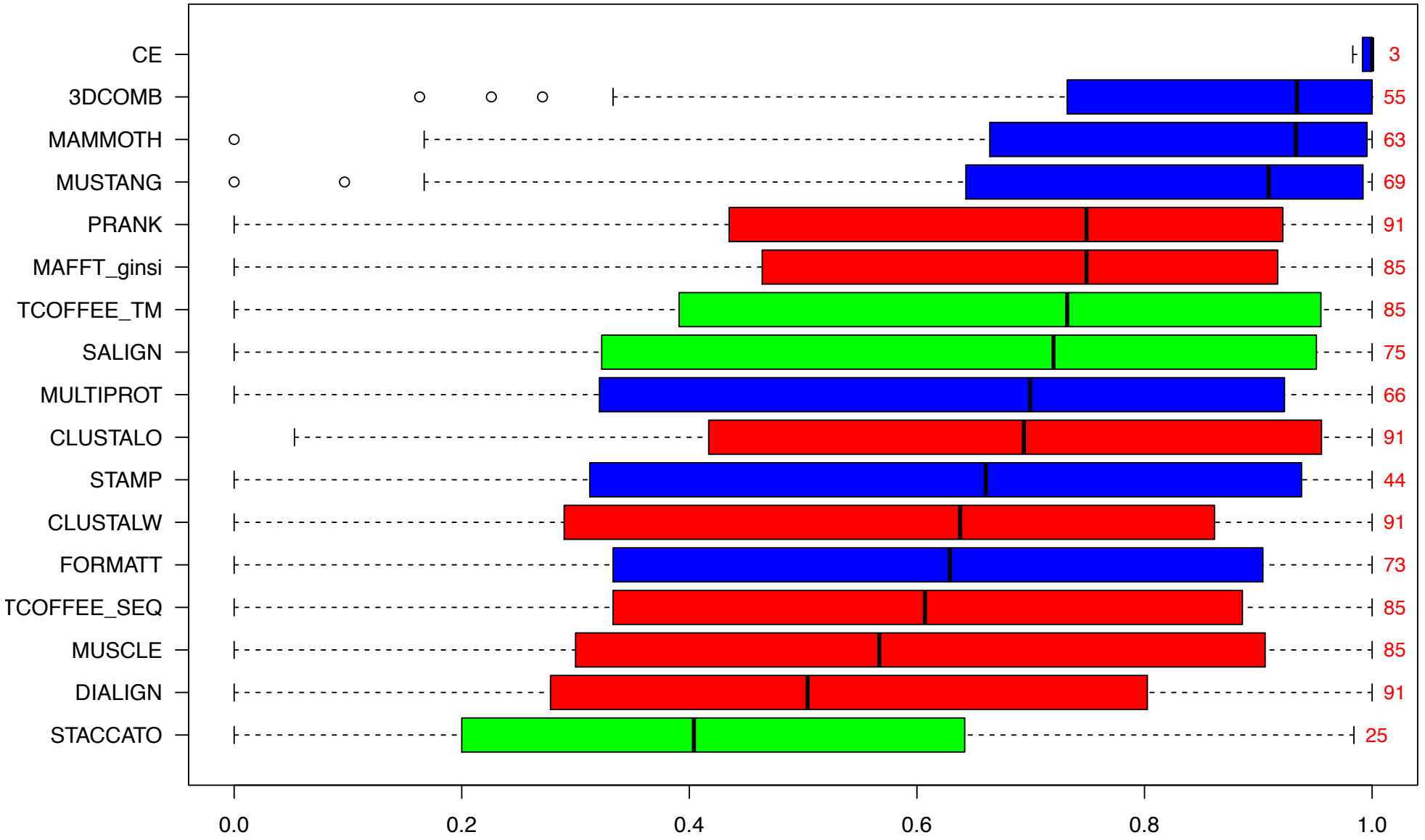
TC (Total Columns)



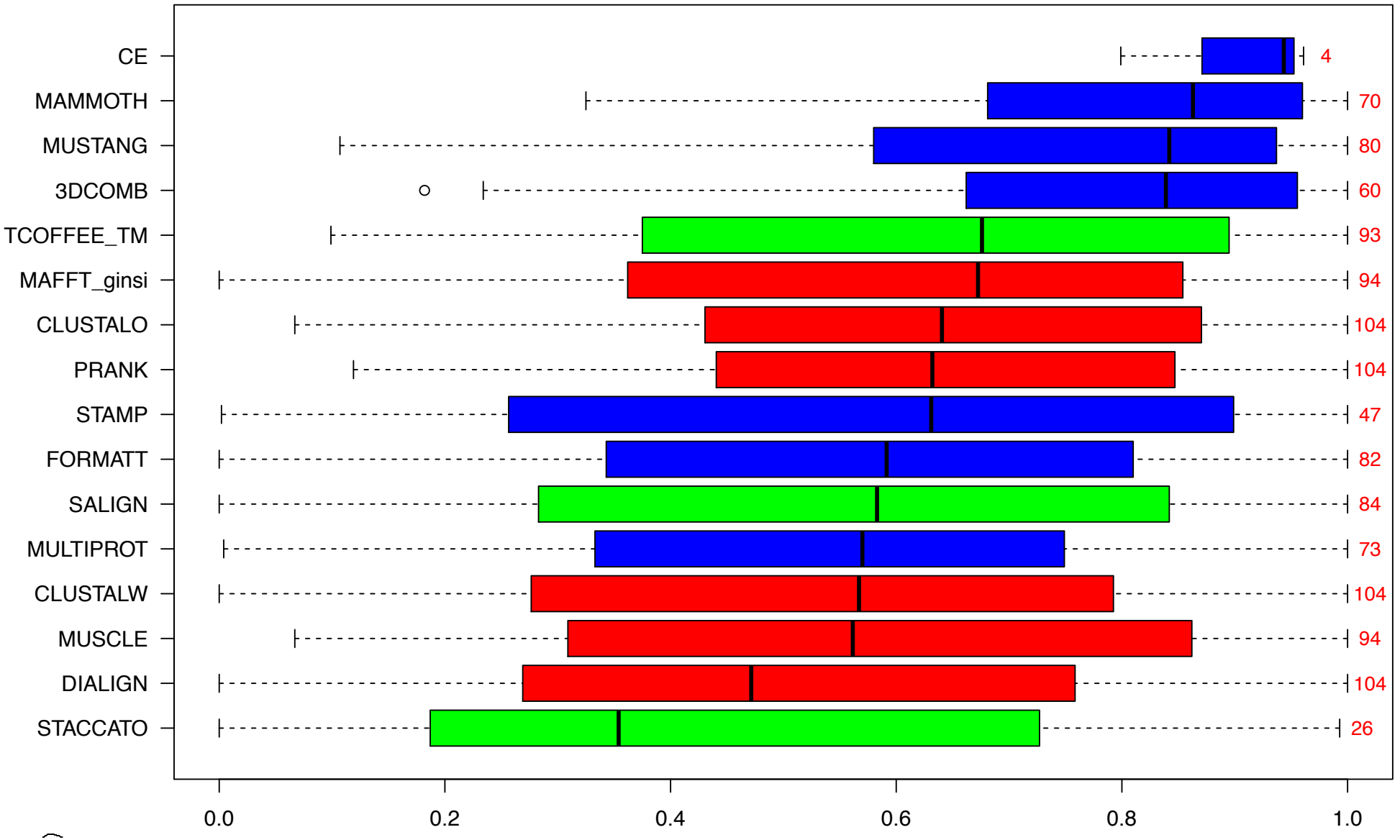

Median SP for each program for BB2, BB3, sisyphus
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
3DCOMB
MUSTANG
FORMATT
MULTIPROT
SALIGN
CLUSTALO
CLUSTALW
DIALIGN
MAFFT_ginsi
MUSCLE
PRANK
TCOFFEE_SEQ
CE
MAMMOTH
STACCATO
STAMP
TCOFFEE_TM


Median TC for each program for BB2, BB3, sisyphus
3DCOMB
MUSTANG
FORMATT
MULTIPROT
SALIGN
CLUSTALO
CLUSTALW
DIALIGN
MAFFT_ginsi
MUSCLE
PRANK
TCOFFEE_SEQ
CE
MAMMOTH
STACCATO
STAMP
TCOFFEE_TM
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
%Id

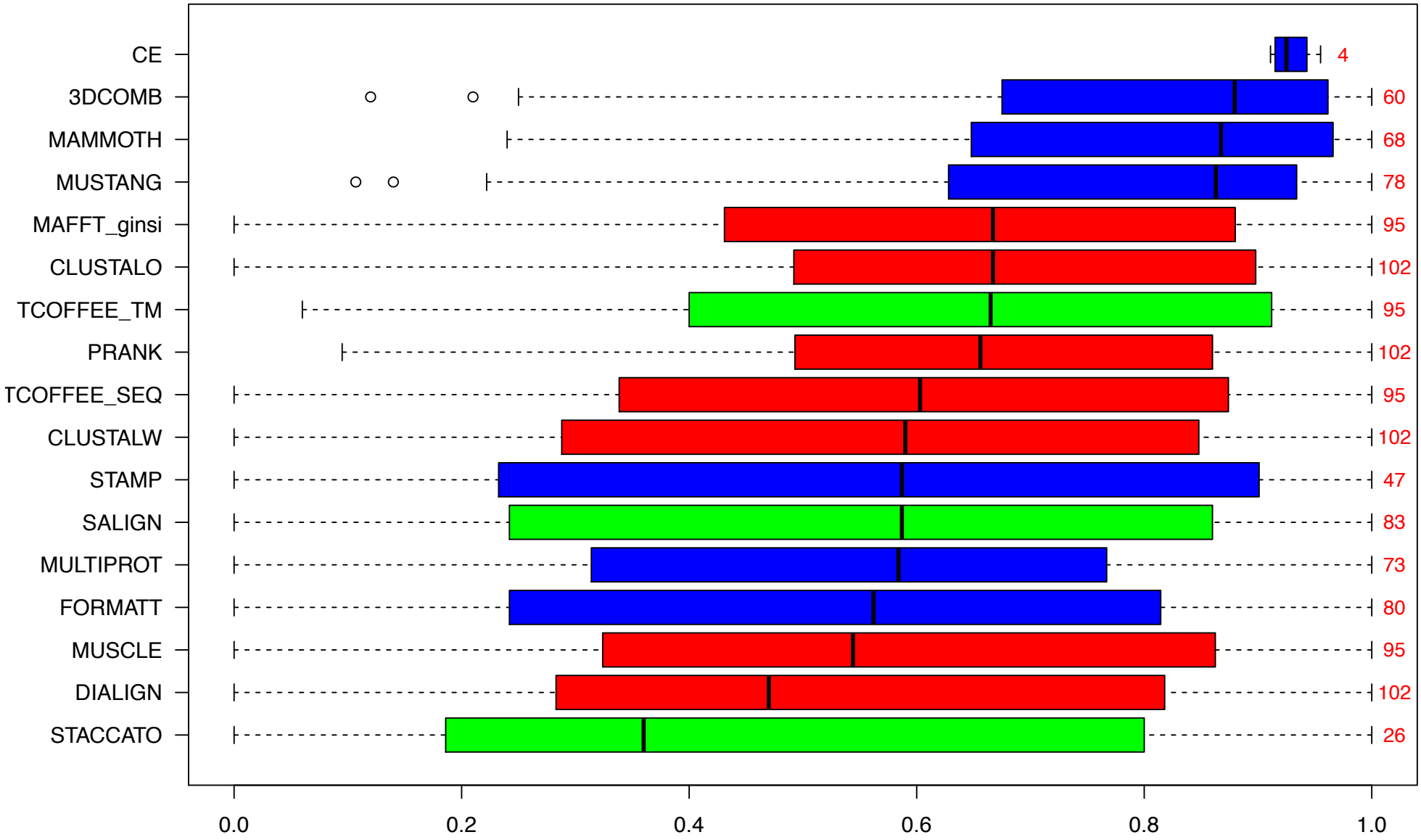
SP Residues in helices



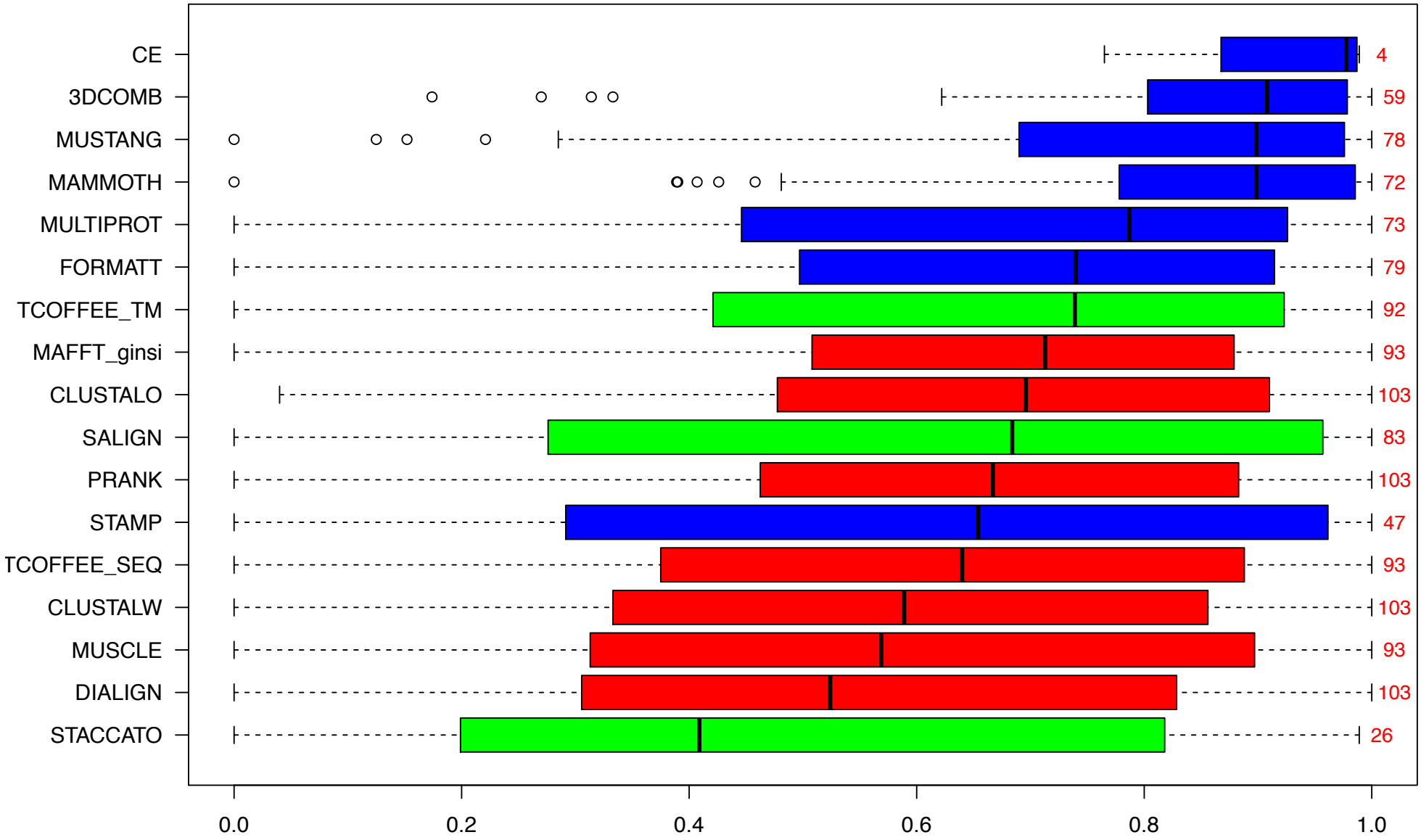
SP Residues in strands



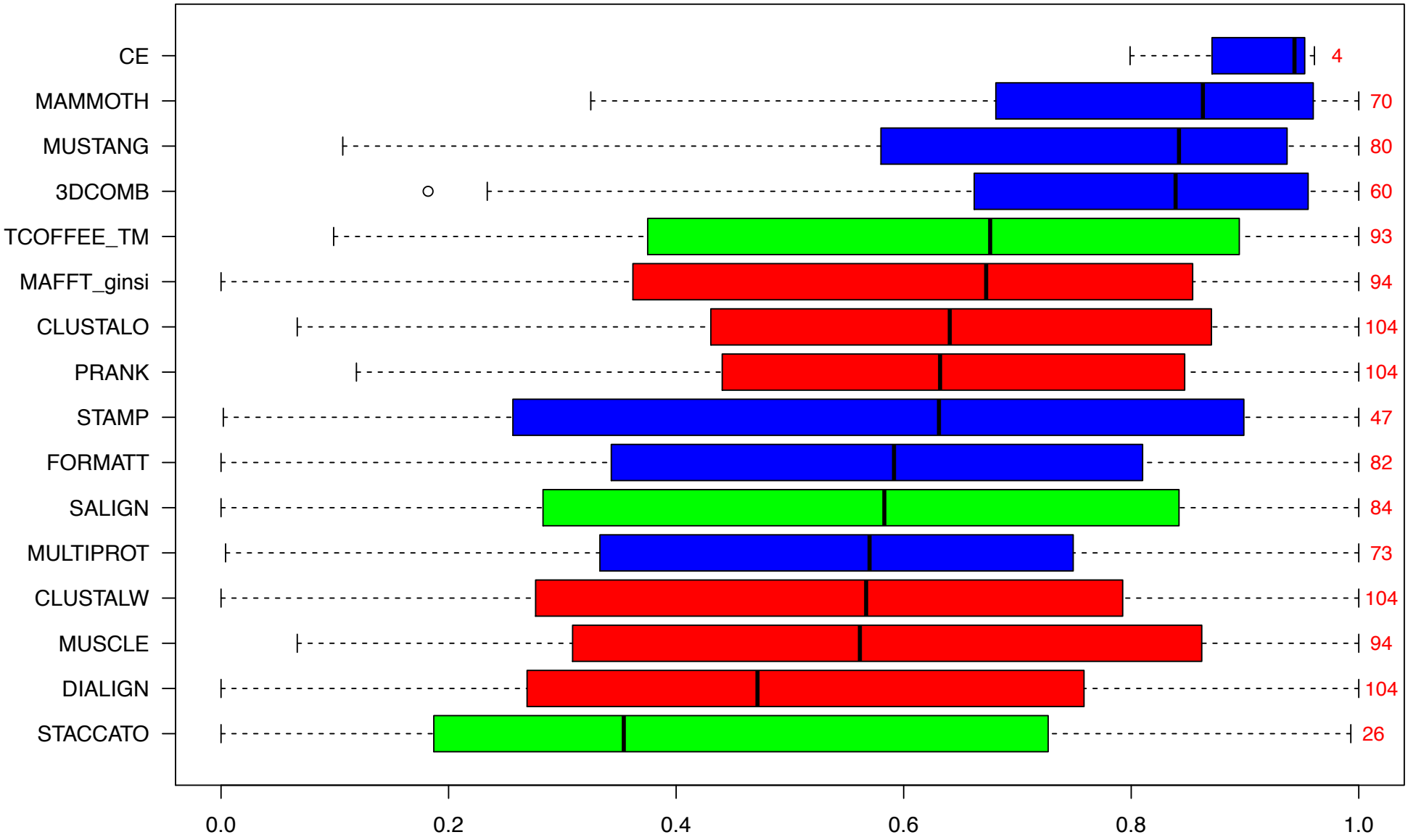
SP Other residues



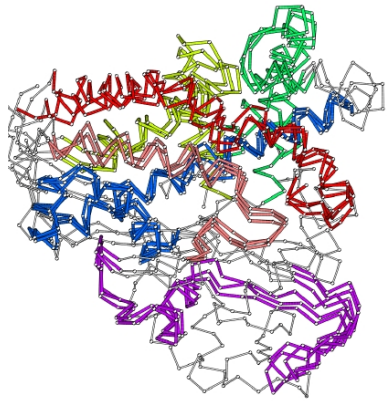
SP Buried residues



SP Exposed residues



Conclusion



Protein structure alignments

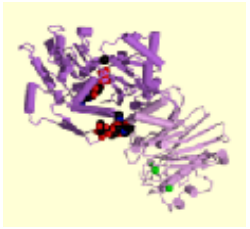
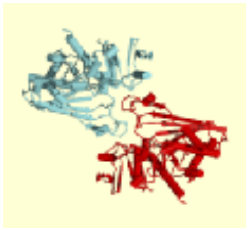
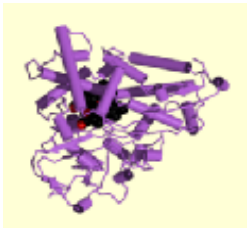
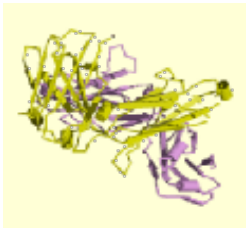
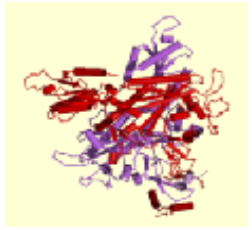
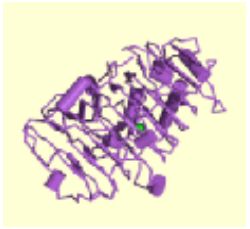
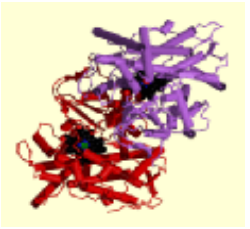
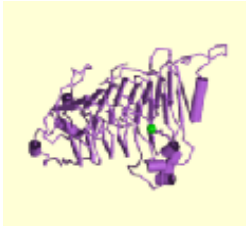
- Do structure aligments methods detect homology?
  => yes
- Are they better than sequence alignment methods?
  => yes
- Does structure really more conserved than sequence?
  => yes

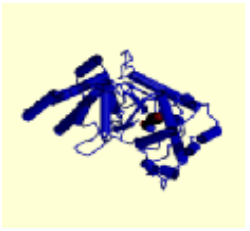
  How to combine structure and sequence information ?

Structure evolution

- How a structure is modified by a subsitution or an insersion/deletion?
- Can we define a « structural profile »?
- Is it possible to define an evolutionary model taking into account structure and sequence ?

Merci de votre attention !



Et Merci à :

ABI

Joël Pothier
Henry Soldano
Nadia Pisanti
Sophie Brouillet
Guillaume Achaz
Martine Boccara
Guillaume Santini

IMPMC

Jacques Chomilier

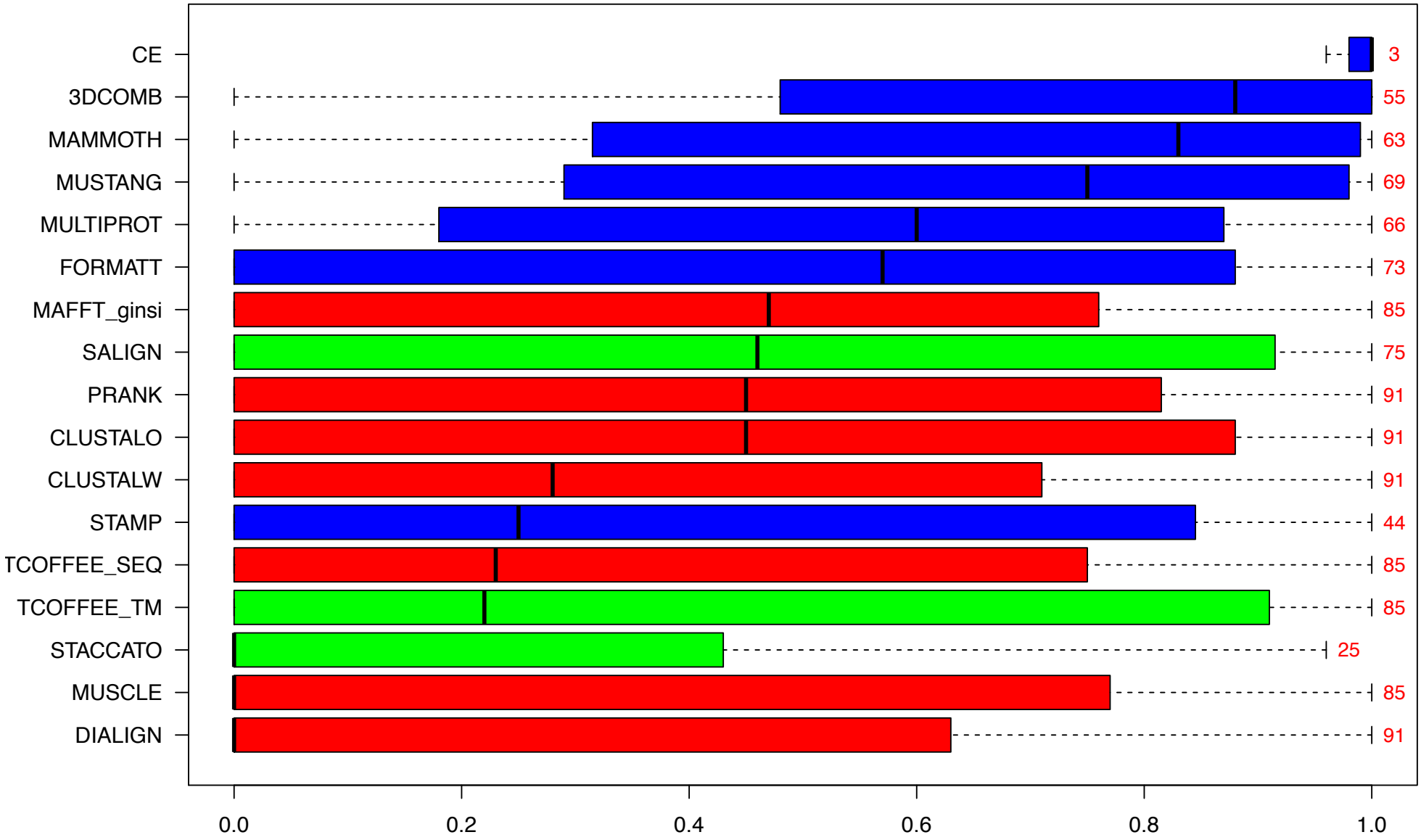
UFIP

Yves-Henri SANEJOUAND

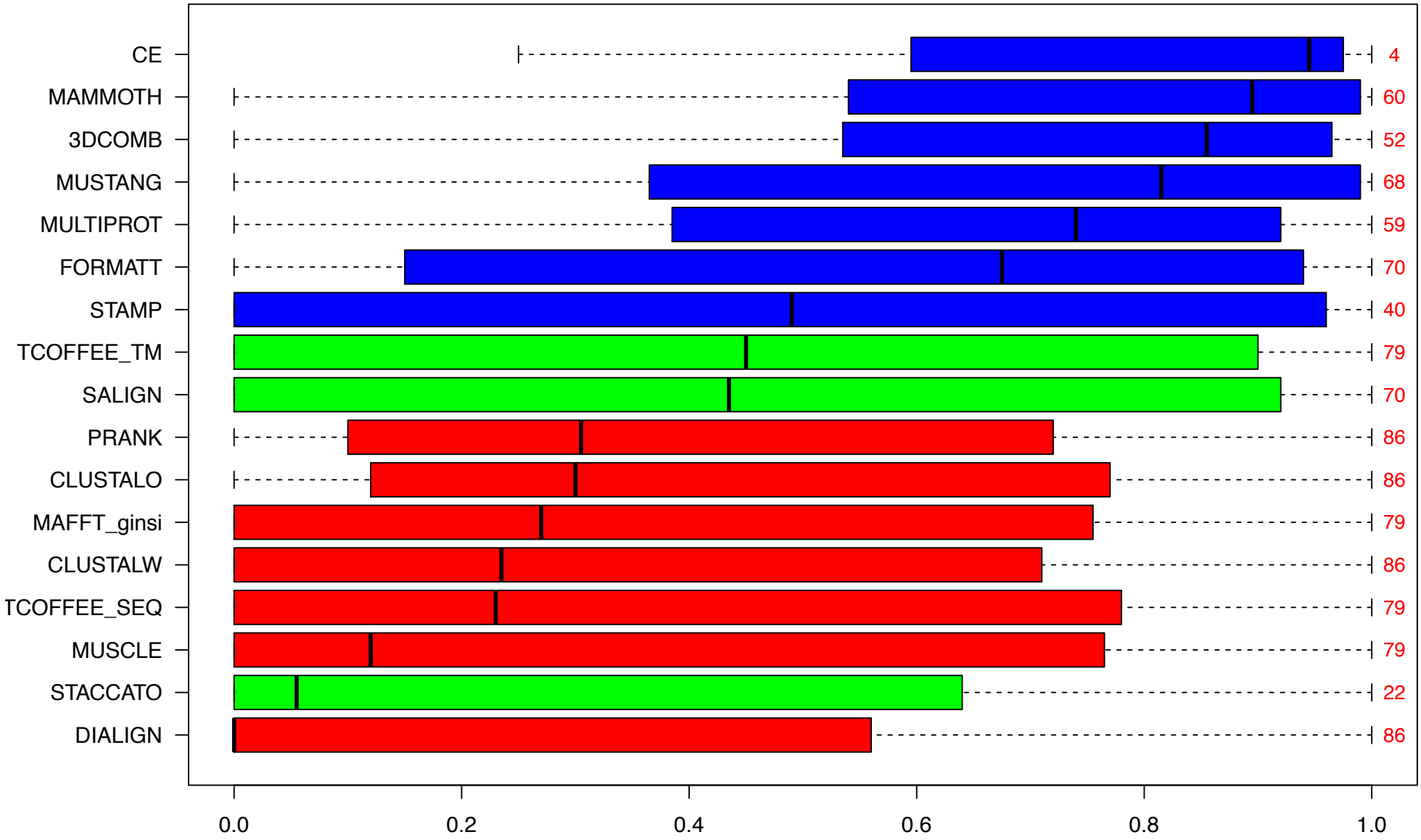
Et merci aux étudiants :

Clément Joubert
Suvethigaa Shanthirabalan

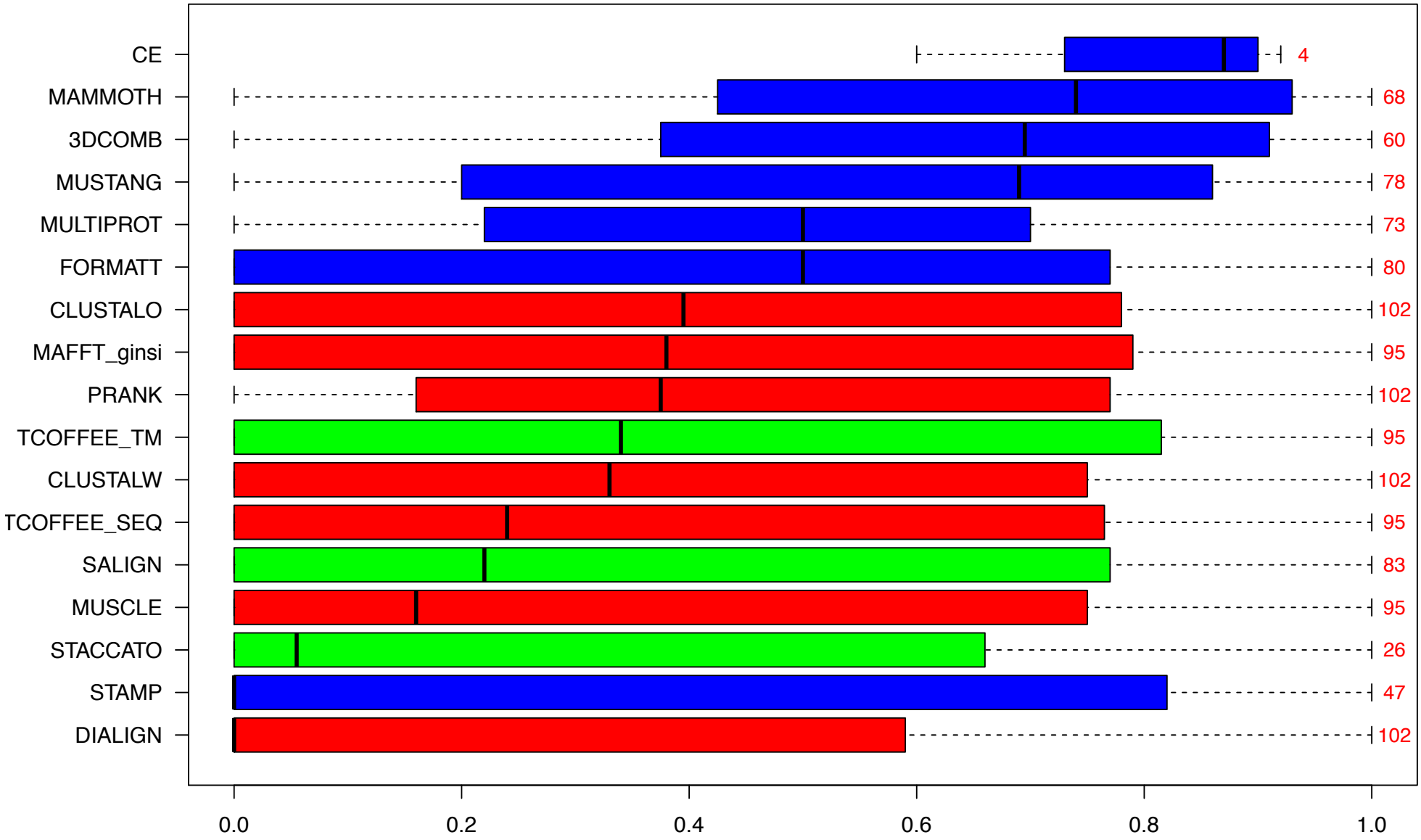
TC Residues in helices



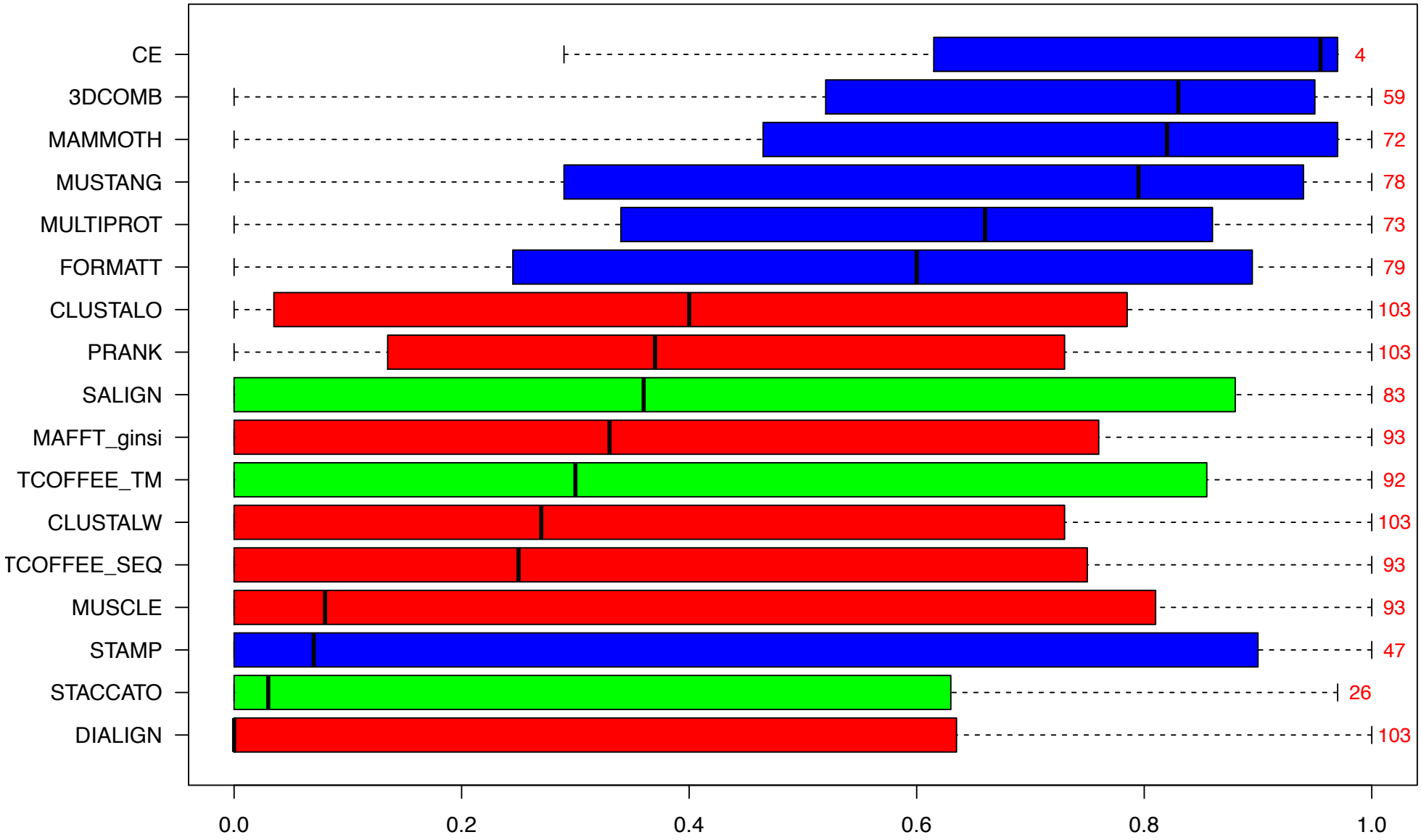
TC Residues in strands



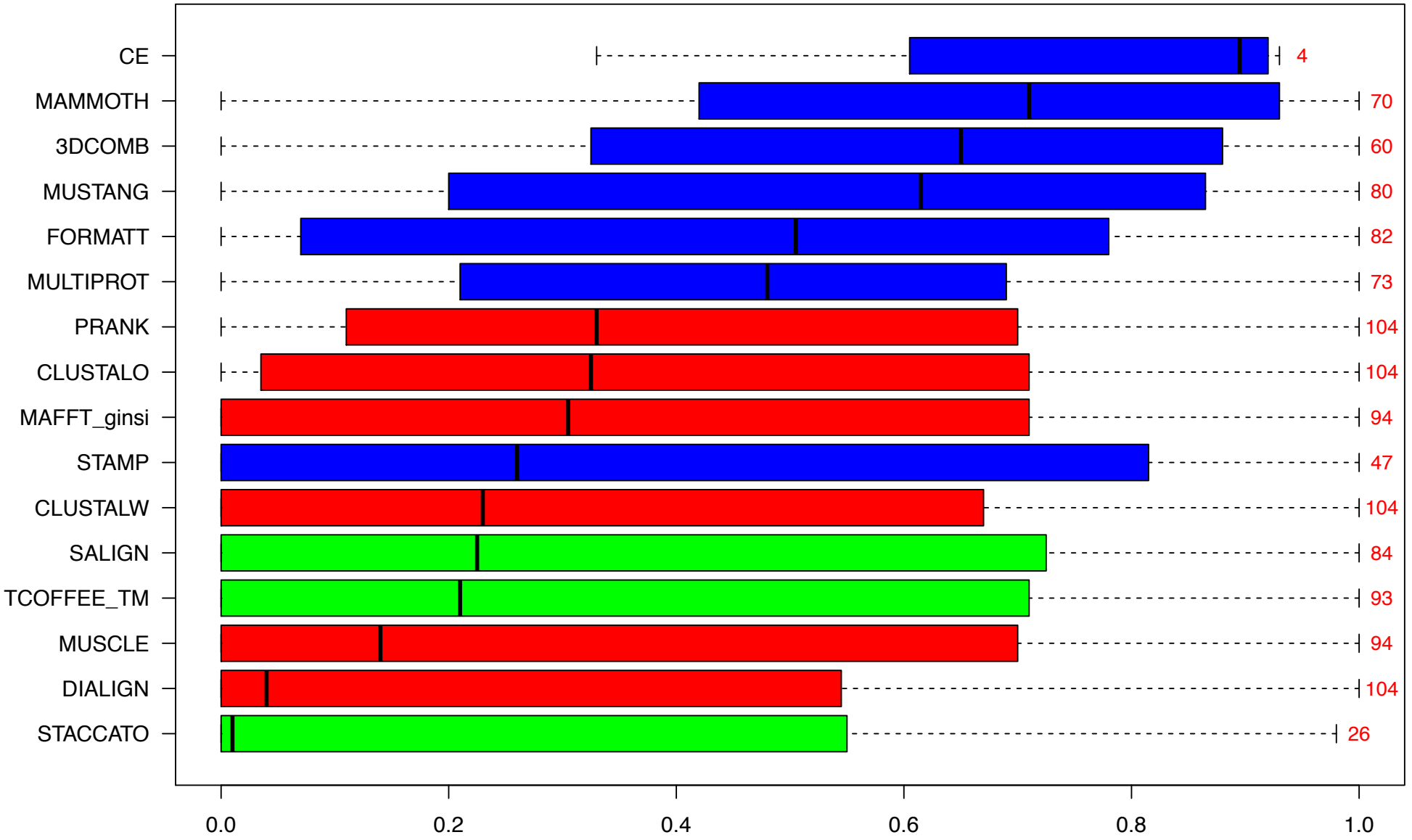
TC Other residues



TC Buried residues

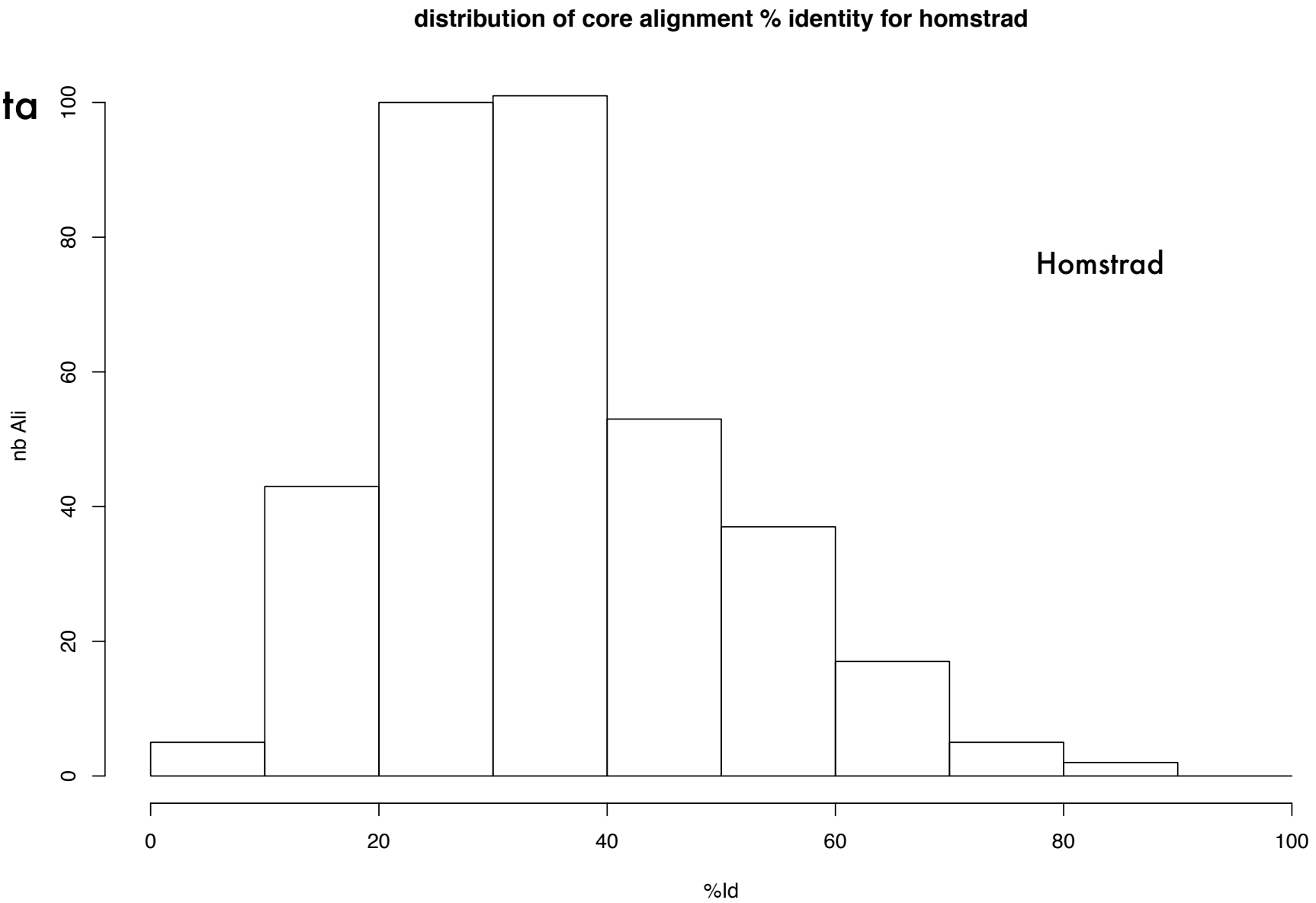


TC Exposed residues



Comparison of sequence and structure alignment methods

Data



Un exemple

- Etude de la résistance à la pénicilline : 5 mutations dans une β–lactamase augmente la résistance par ~100 000:

  g4205a, A42G, **E104K**, M182T, **G238S**

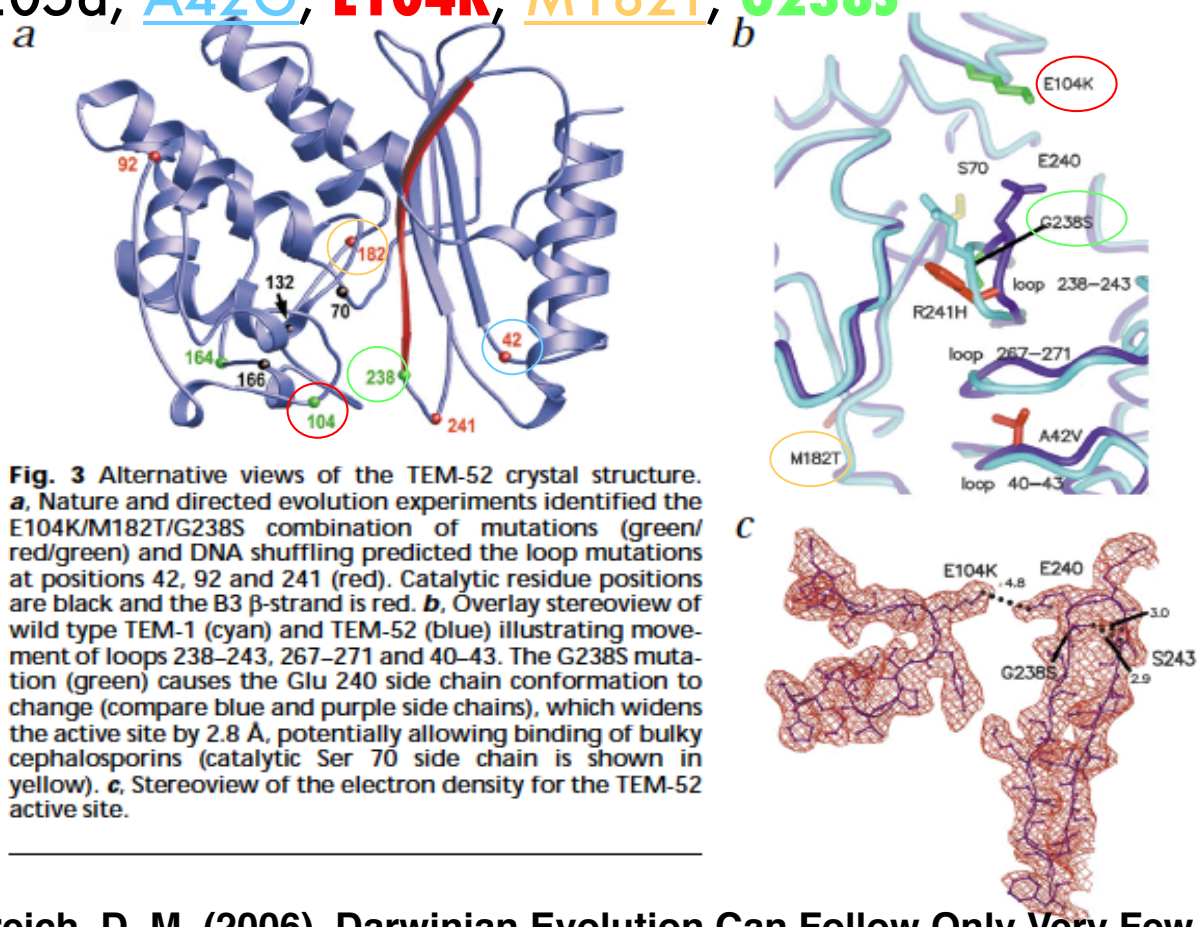


Fig. 3 Alternative views of the TEM-52 crystal structure. ***a***, Nature and directed evolution experiments identified the E104K/M182T/G238S combination of mutations (green/red/green) and DNA shuffling predicted the loop mutations at positions 42, 92 and 241 (red). Catalytic residue positions are black and the B3 β-strand is red. ***b***, Overlay stereoview of wild type TEM-1 (cyan) and TEM-52 (blue) illustrating movement of loops 238–243, 267–271 and 40–43. The G238S mutation (green) causes the Glu 240 side chain conformation to change (compare blue and purple side chains), which widens the active site by 2.8 Å, potentially allowing binding of bulky cephalosporins (catalytic Ser 70 side chain is shown in yellow). ***c***, Stereoview of the electron density for the TEM-52 active site.

• **Weinreich, D. M. (2006). Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science (New York, NY)*, *312*(5770).**

• **Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P., & Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis *Nature structural biology*, *8*(3).**

Mettre de la dynamique dans les structures protéiques pour comprendre leur évolution.

Les modes normaux semblent conservés dans les familles structurales[1]
Les structures se déformeraient dans le même sens que les modes normaux basses fréquences[2,3,4]. et aussi [5].

1. Maguid S, Fernandez-Alberti S, Echave J (2008) Evo- lutionary conservation of protein vibrational dynamics. Gene 422:7–13.
2. Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR (2005) An analysis of core deformations in protein superfamilies. Biophys J 88:1291–1299.
3. Friedland GD, Lakomek N-A, Griesinger C, Meiler J, Kortemme T (2009) A Correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. PLoS Comput Biol.
4. Velazquez-Muriel JA, Rueda M, Cuesta I, Pascual- Montano A, Orozco M, Carazo J-M (2009) Comparison of molecular dynamics and superfamily spaces of protein domain deformation. BMC Struct Biol 9:6.
5. Echave, J. & Fernández, F. M. A perturbative view of protein structural variation. Proteins 78, 173–180 (2010).

+ Aussi Liberles, D. A. *et al*. The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci* 21, 769–785 (2012).

Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR (2005) An analysis of core deformations in protein superfamilies. Biophys J 88:1291–1299.

Alignements multiples calculés avec MAMMOTH de 35 familles SCOP de 11 à 36 protéines

- Calculs des modes normaux (ANM) et PCA pour capturer les déformations principales

- Calcul du RMSi pour les comparer:

The overlap between both spaces is calculated from the root mean-square inner product (root mean-square inner product) (Amadei et al., 1999) of the PCA eigenvectors with the vibrational ones:

$$RMSIP = \left(\frac{1}{D} \sum_{i=1}^{D} \sum_{j=1}^{K} (\boldsymbol{\eta}_i \cdot \boldsymbol{v}_j)^2 \right)^{1/2}.$$

Here, ηi and νj are, respectively, the set of eigenvectors of the evolutionary and ANM spaces, with dimensionality equal to three times the number of residues defined. D is the dimensionality of the evolutionary space (five dimensions were used on average), and k is the dimensionality of the ANM space (the slowest 50 modes were employed).

+ calcul d'une distribution aléatoire et d'un z-score

Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR (2005) An analysis of core deformations in protein superfamilies. Biophys J 88:1291–1299.

⇒ « **70% of the total variance** in the core fluctuations can be explained with an average of **4.5 + ou -1.2 components**. »

⇒ « For most superfamilies there is a moderate degree of **correlation** between the **root mean-squared** fluctuations observed in the core, as computed from the alignments, and the fluctuations predicted by **ANM**, with correlations in the range of 0.3–0.8. »

⇒ « More interesting is the finding that the adaptive movements responsible for these fluctuations are highly cooperative, taking place in a space of **low dimensionality**, of only 4–5 dimensions, and similar in all superfamilies. Because side chain degrees of freedom in the protein core are basically dictated by the backbone conformation (Levitt et al., 1997), this finding suggests that in fact, and as far as the core region is concerned, the **conformational space to sample in model refinement is fairly small**. »

⇒ « We conclude that, to a significant extent, **the structural response of a protein topology to sequence changes takes place by means of collective deformations along combinations of a small number of low-frequency modes**. The findings have implications in structure prediction by homology modeling. »

Echave, J. & Fernández, F. M. A perturbative view of protein structural variation. *Proteins* 78, 173–180 (2010).

The ENM potential is of the form

$$V_{\mathrm{wt}} = \frac{1}{2}(\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}})^T \mathbf{K} (\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}}), \qquad (1)$$

where, for a protein of N sites, $\mathbf{r}$ is a column vector with $3N$ elements: the x, y, z coordinates of the N C_α, $\bar{\mathbf{r}}_{\mathrm{wt}}$ is the equilibrium structure, and $\mathbf{K}$ is the "stiffness" matrix, which represents the network's topology and spring force constants.

We model a point mutation (amino acid replacement) by adding a linear perturbative term to the reference potential Eq. (1):

$$V_{\mathrm{mut}} = V_{\mathrm{wt}} - \mathbf{f}^T (\mathbf{r} - \bar{\mathbf{r}}_{\mathrm{wt}}), \qquad (2)$$

where $\mathbf{f}$ is a column vector with 3N elements that models the mutation. Since $\mathbf{f} = -\left(\frac{\partial V_{\mathrm{mut}}}{\partial \mathbf{r}}\right)_{\mathbf{r}=\bar{\mathbf{r}}_{\mathrm{wt}}}$, it can be interpreted as a force that drives the mutant's structure away from that of the wild type. Eq. (2) can be derived by expanding the ENM potential in Taylor series with respect to parameter variations and keeping only first and second order terms, as is shown in the Supporting Information Appendix.

The equilibrium structure of the mutant $\bar{\mathbf{r}}_{\mathrm{mut}}$ is the value of $\mathbf{r}$ that minimizes V_{mut}. Using Eqs. (1) and (2) we find the structural variation due to the mutation:

$$\delta\bar{\mathbf{r}} \equiv \bar{\mathbf{r}}_{\mathrm{mut}} - \bar{\mathbf{r}}_{\mathrm{wt}} = \mathbf{K}^{-1}\mathbf{f}. \qquad (3)$$

⇒ **Divergence structurale le long des modes basses fréquences, qu'il y ait sélection ou non**

Evolution des structures protéiques
Vers un modèle d'évolution structurale des protéines ?

Etude de l'effet :

- des mutations
- des insertions/délétions
- (de la co-évolution)

Effet des mutations : quel est l'impact d'une mutation sur une structure ?

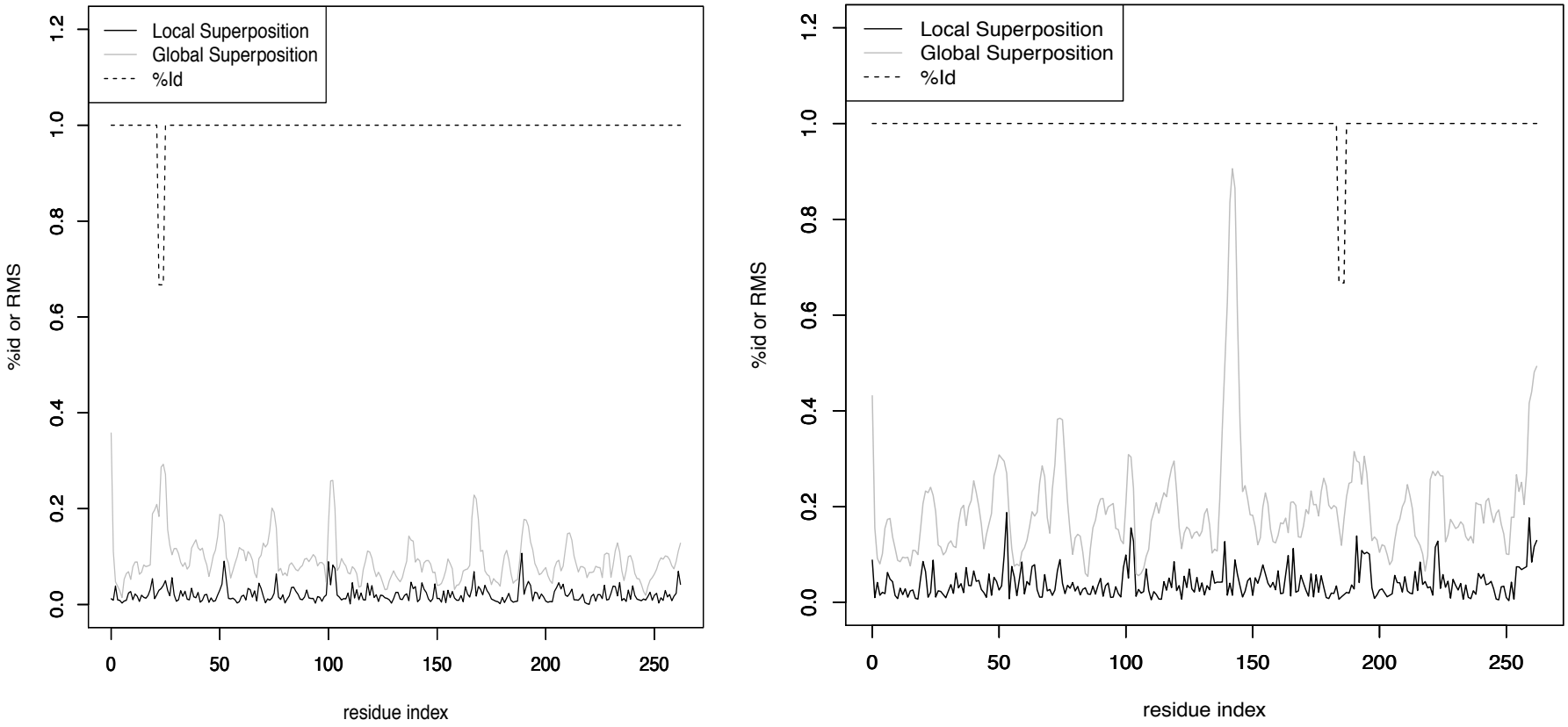
Banque de données:

Rank	Members	Reference	Length	Protein	Class	Cluster
1	147	1lw9	164	T4 lysozyme	Alpha	31255
2	124	2nwd	130	Human lysozyme	Alpha	37522
3	78	2dek	265	Transferase	Alpha beta	18272
4	69	2ili	255-260	Anhydrase II	Alpha beta	18267
5	31	1ey0	149	Staphylococcal nuclease	Beta	34381
6	29	4bfl	753	Catalase HPII	Multi-domain	796
7	25	2e3w	124	Ribonuclease A	Alpha beta	38031
8	24	2vb1	129	Hen lysozyme	Alpha	37731
9	22	4fi8	126-127	Transthyretin	Beta	37628
10	22	2j8c	302-314	Reaction centre	Alpha beta	13574
11	20	5dei	524-536	Benzoylformate decarboxylase	Alpha beta	2739
Total	591					

En collaboration avec J. Chomilier, S. Shanthirabalan

Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

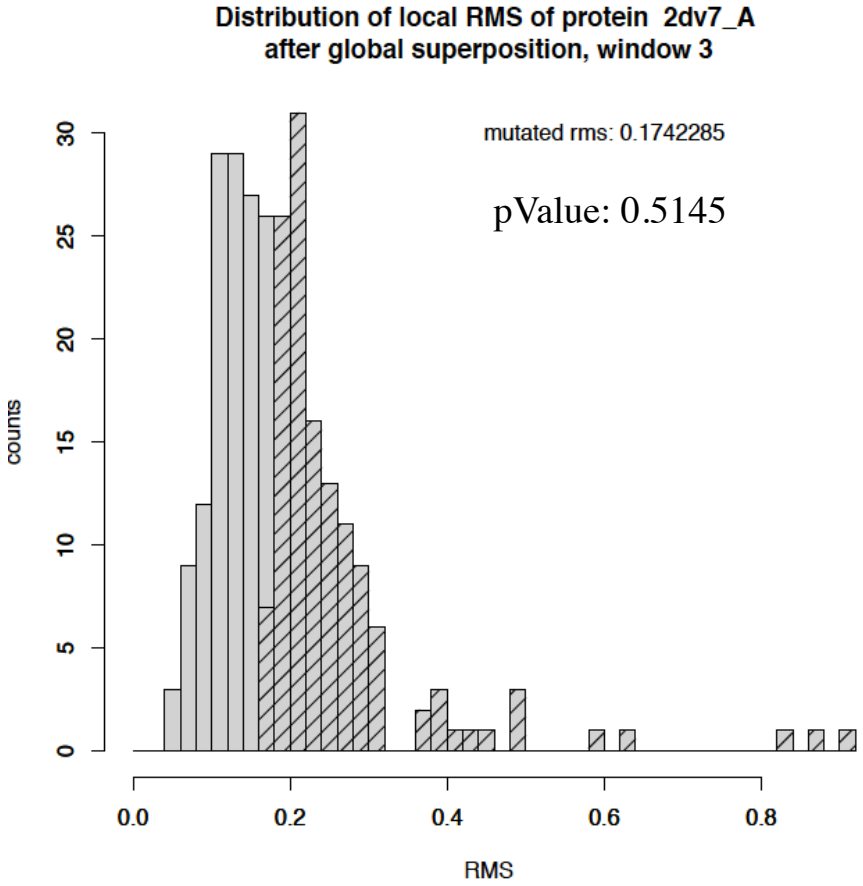
RMS calculés sur 3 Cα, à toutes les positions.

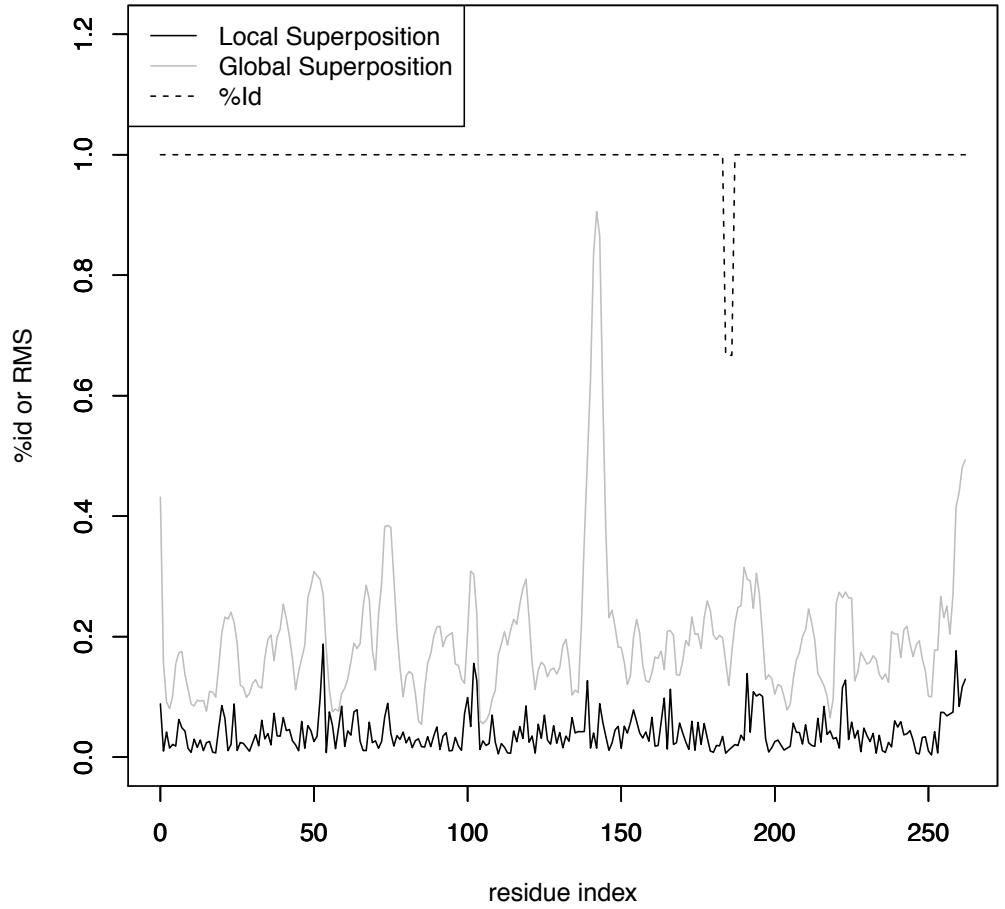


Distribution of the local rms with a 3 residues window for two cases taken from cluster with 2dek as a reference: 2e8r (left) and 2dv7 (right). Superimpositions are either local (dark line) or global (grey line). The dashed upper dotted line well indicates the location of the mutation.

Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

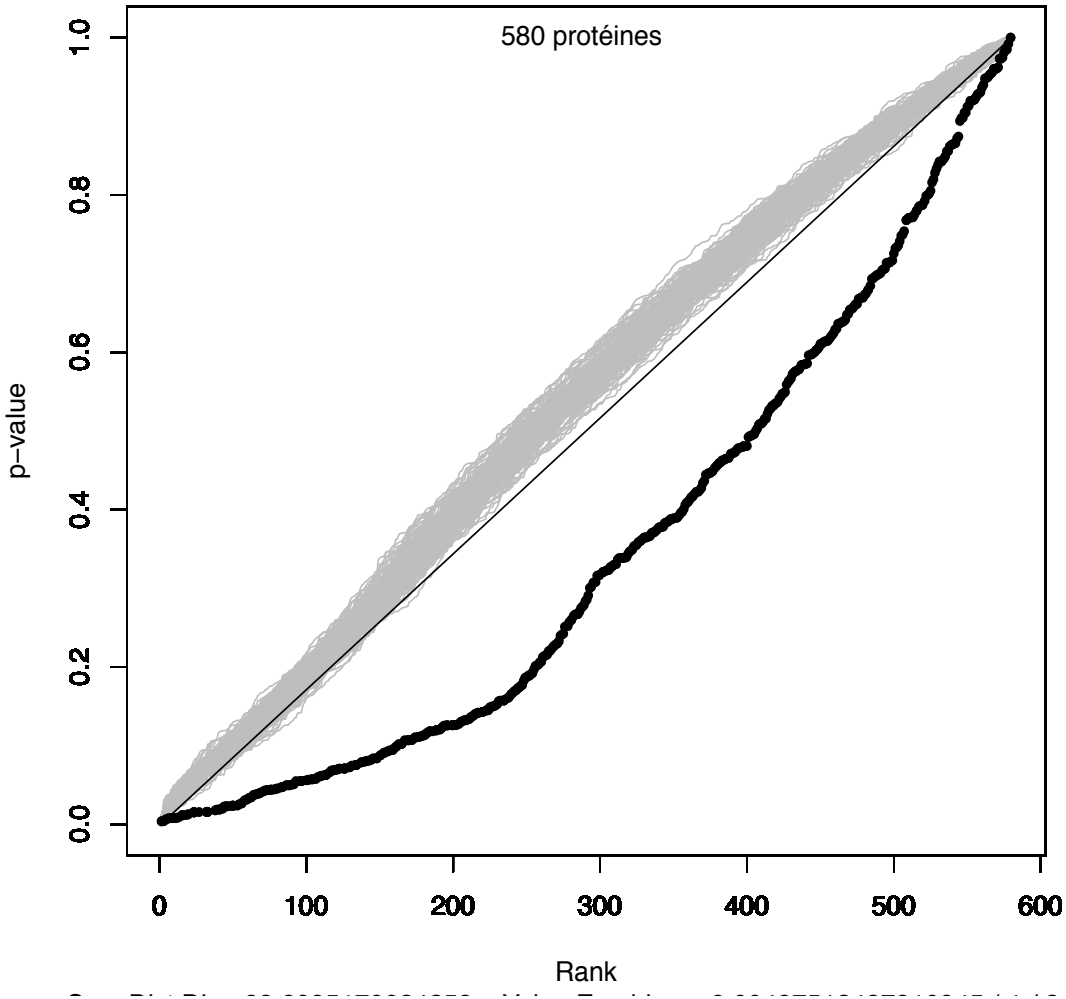
On calcule une p-value empirique pour chaque mutation





Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

On classe les p-value et on les affiche selon leur rang.



Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

On classe les p-value et on les affiche selon leur rang.

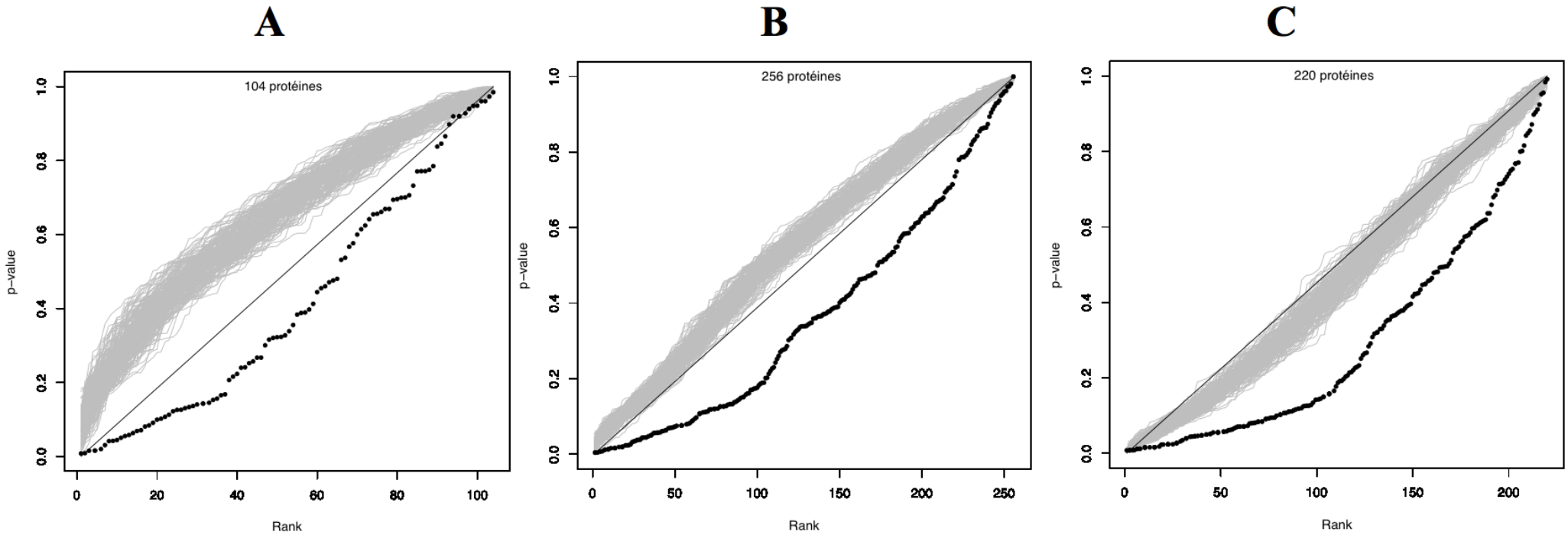


Figure 6. Curves of the p-value sorted as a function of their rank for mutations occurring only in: A) strands; B) helices; C) loops.

Effet des mutations : quel est l'impact d'une mutation sur une structure ?

On classe les p-value et on les affiche selon leur rang.

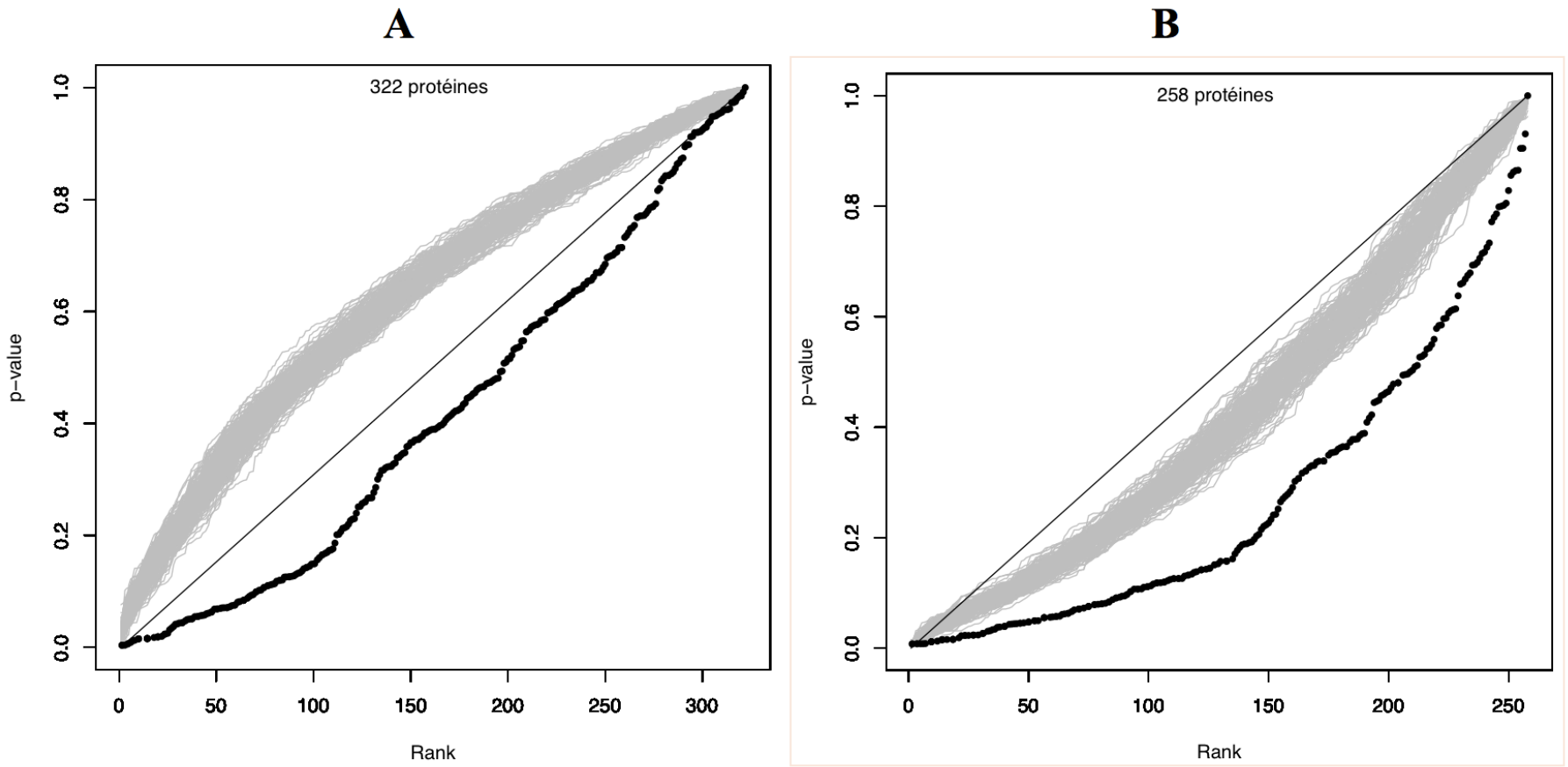
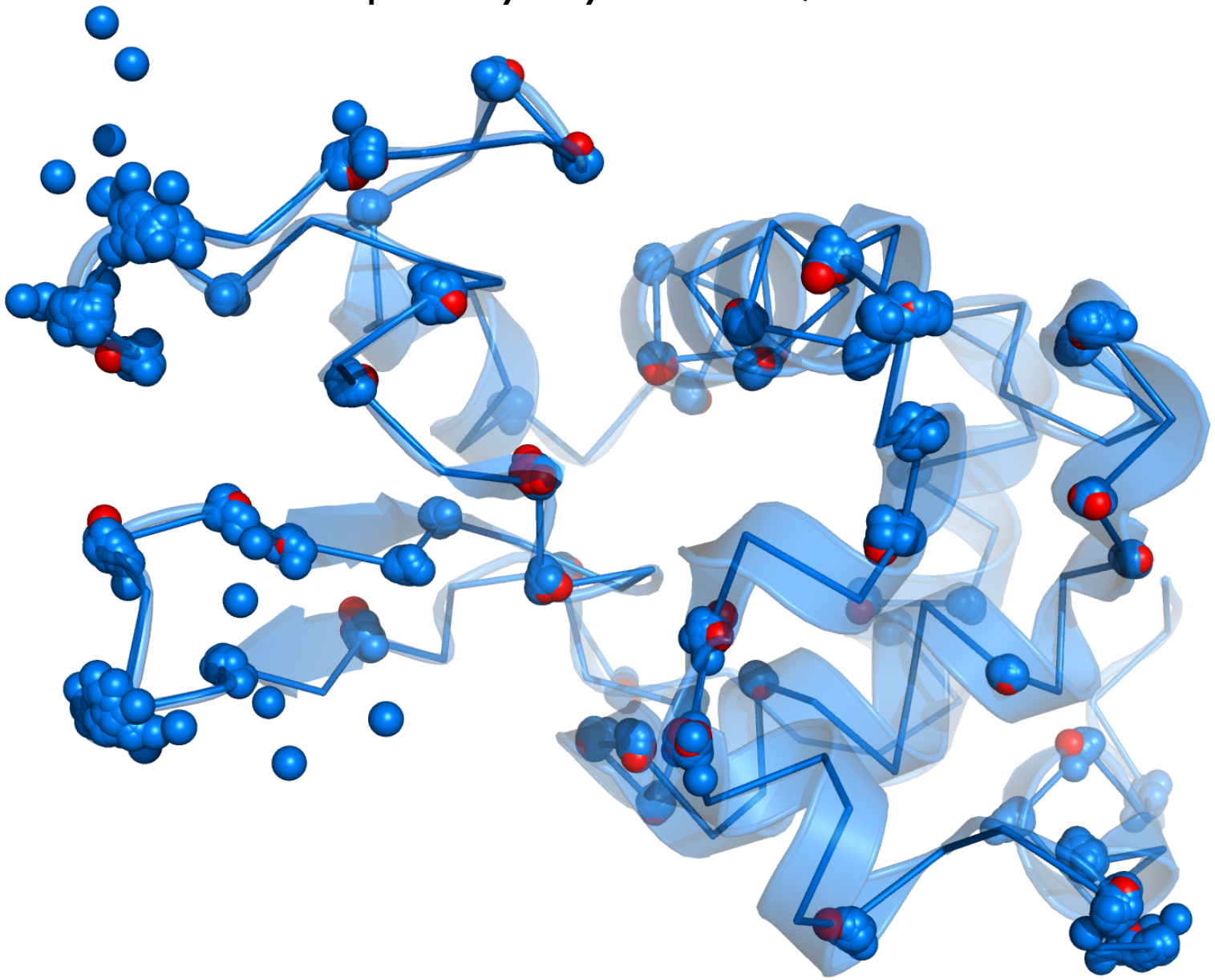
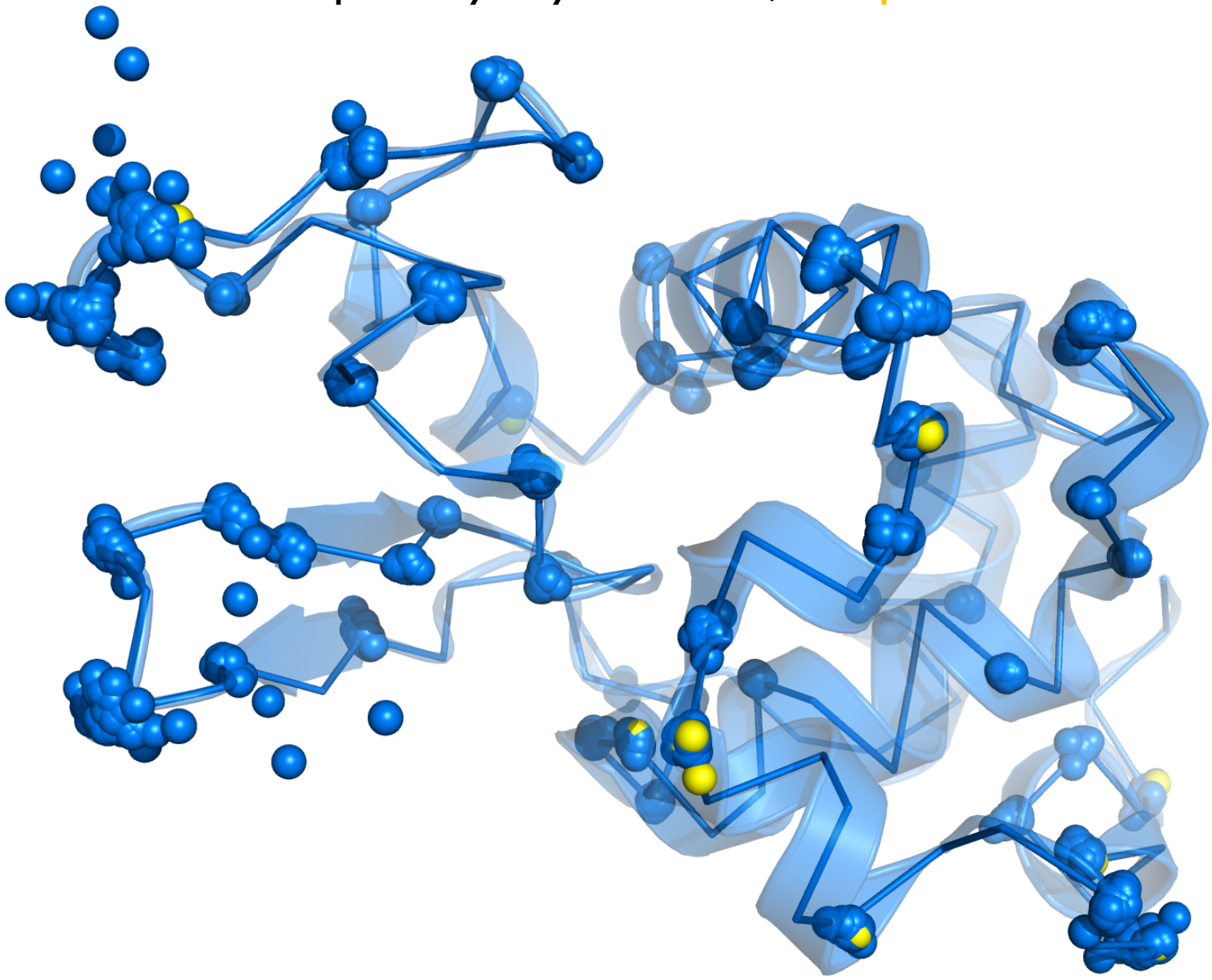


Figure 7. Distribution of the p-value as a function of their rank, among all members of the mutated dataset, according to their accessibility: a) buried; b) exposed.

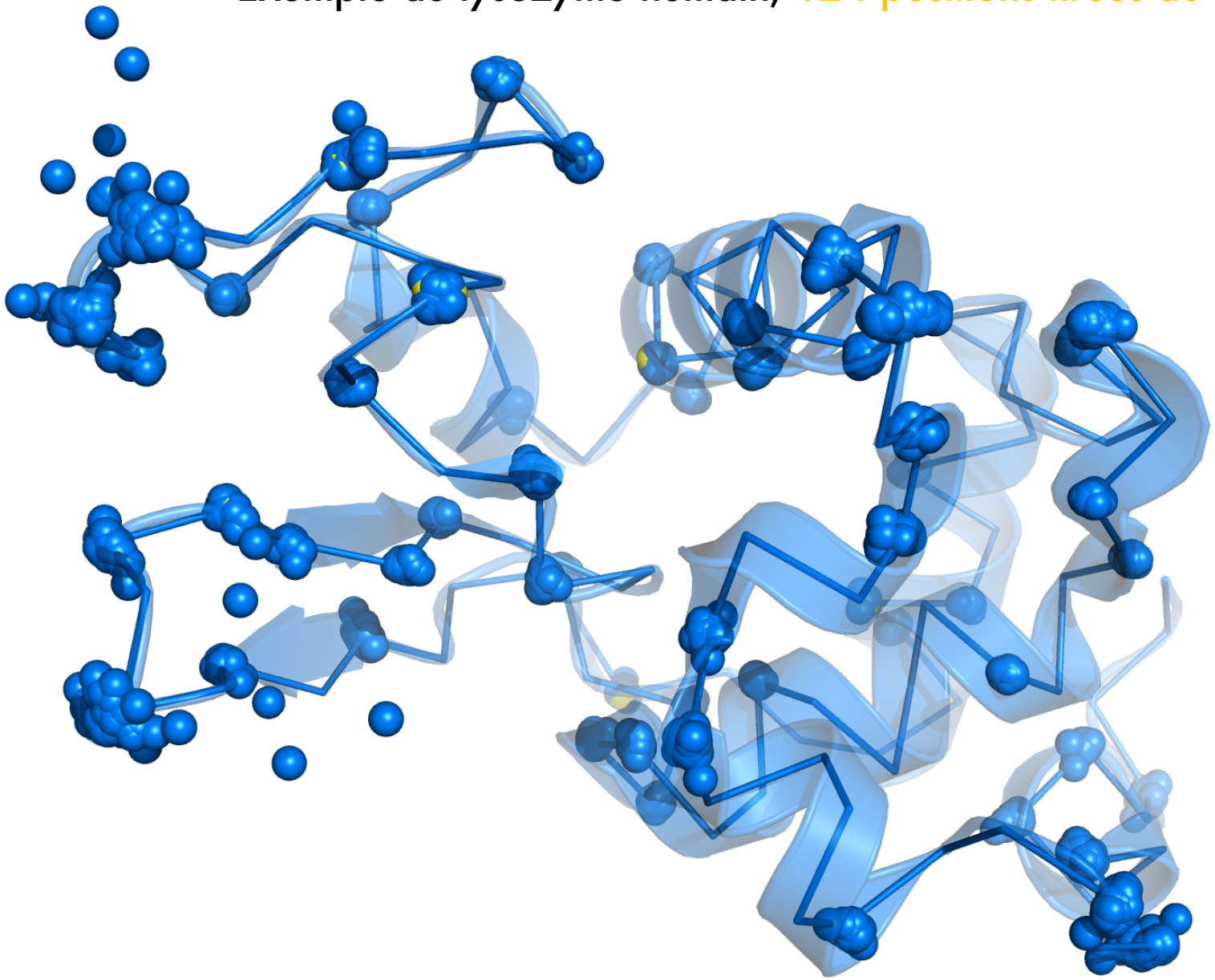
Exemple du lysozyme humain, 124 mutations



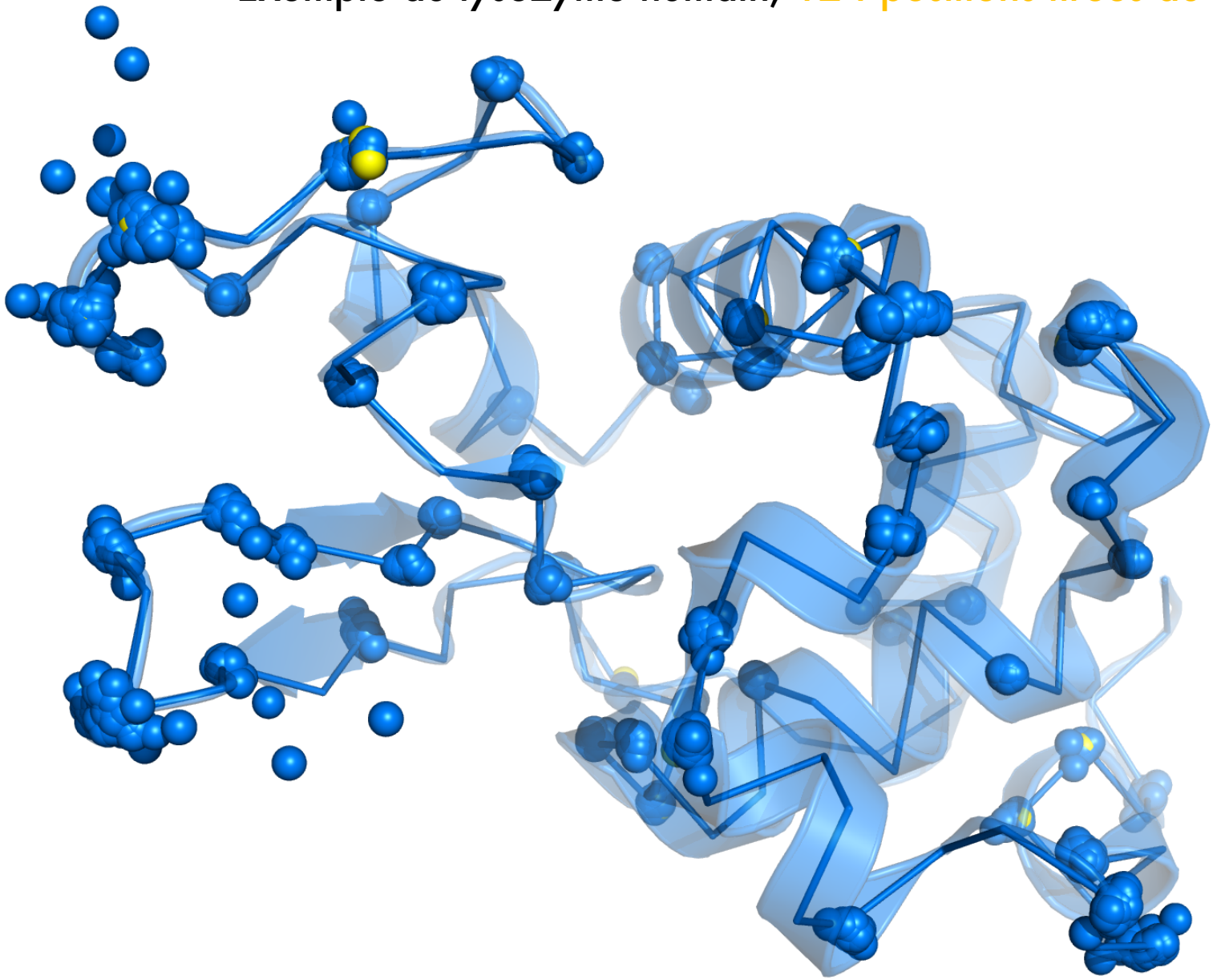
Exemple du lysozyme humain, 124 positions tirées au hasard



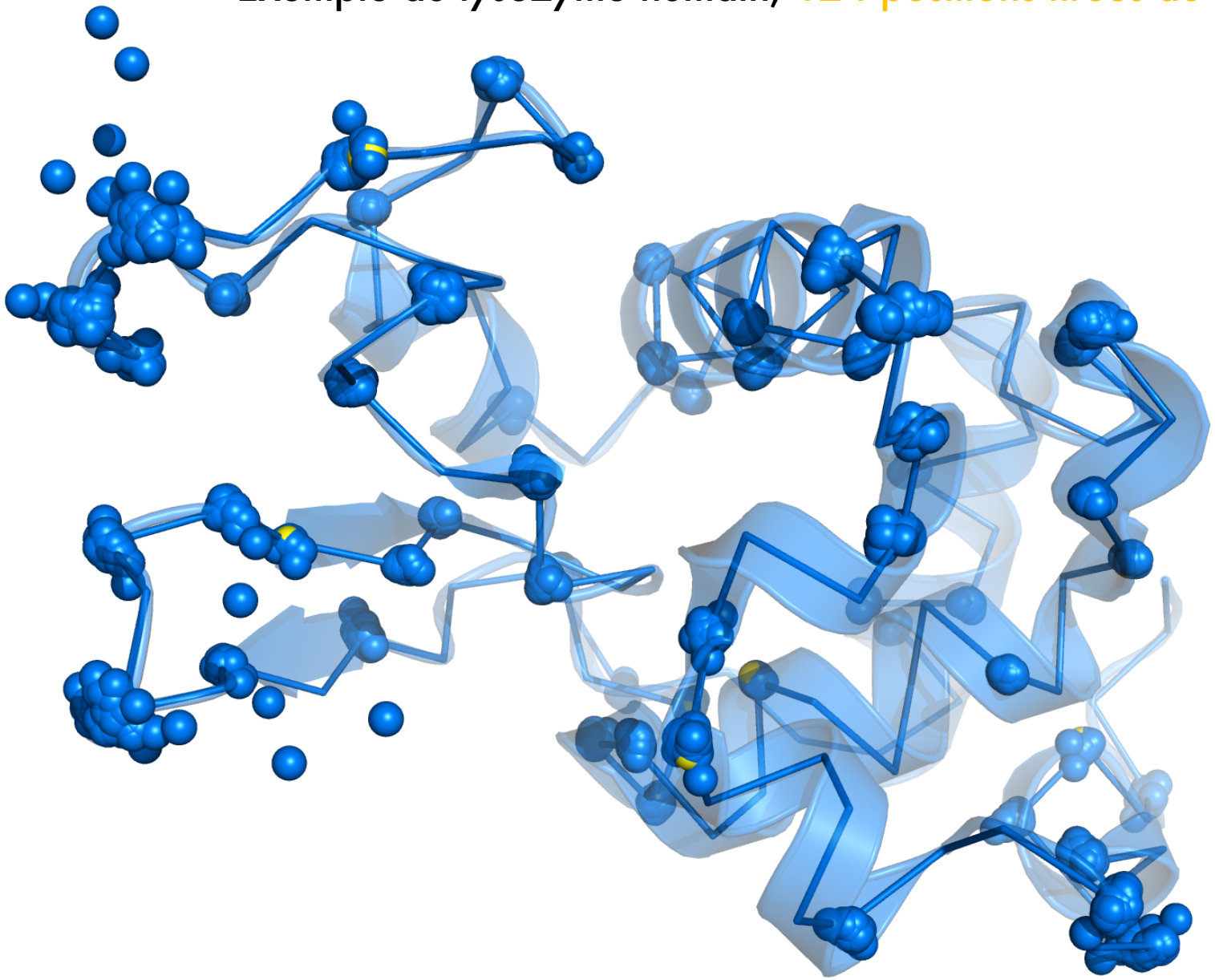
Exemple du lysozyme humain, 124 positions tirées au hasard



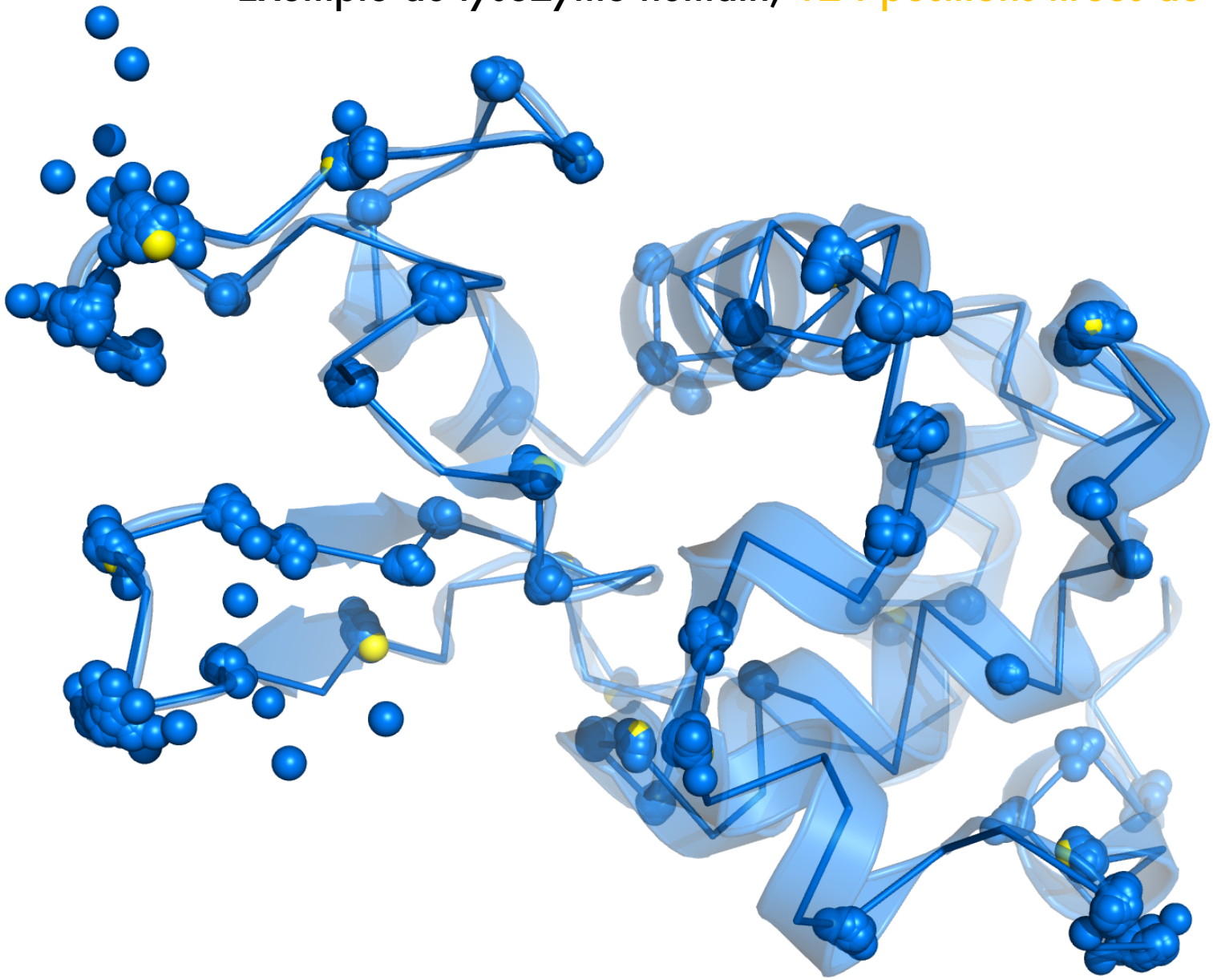
Exemple du lysozyme humain, 124 positions tirées au hasard



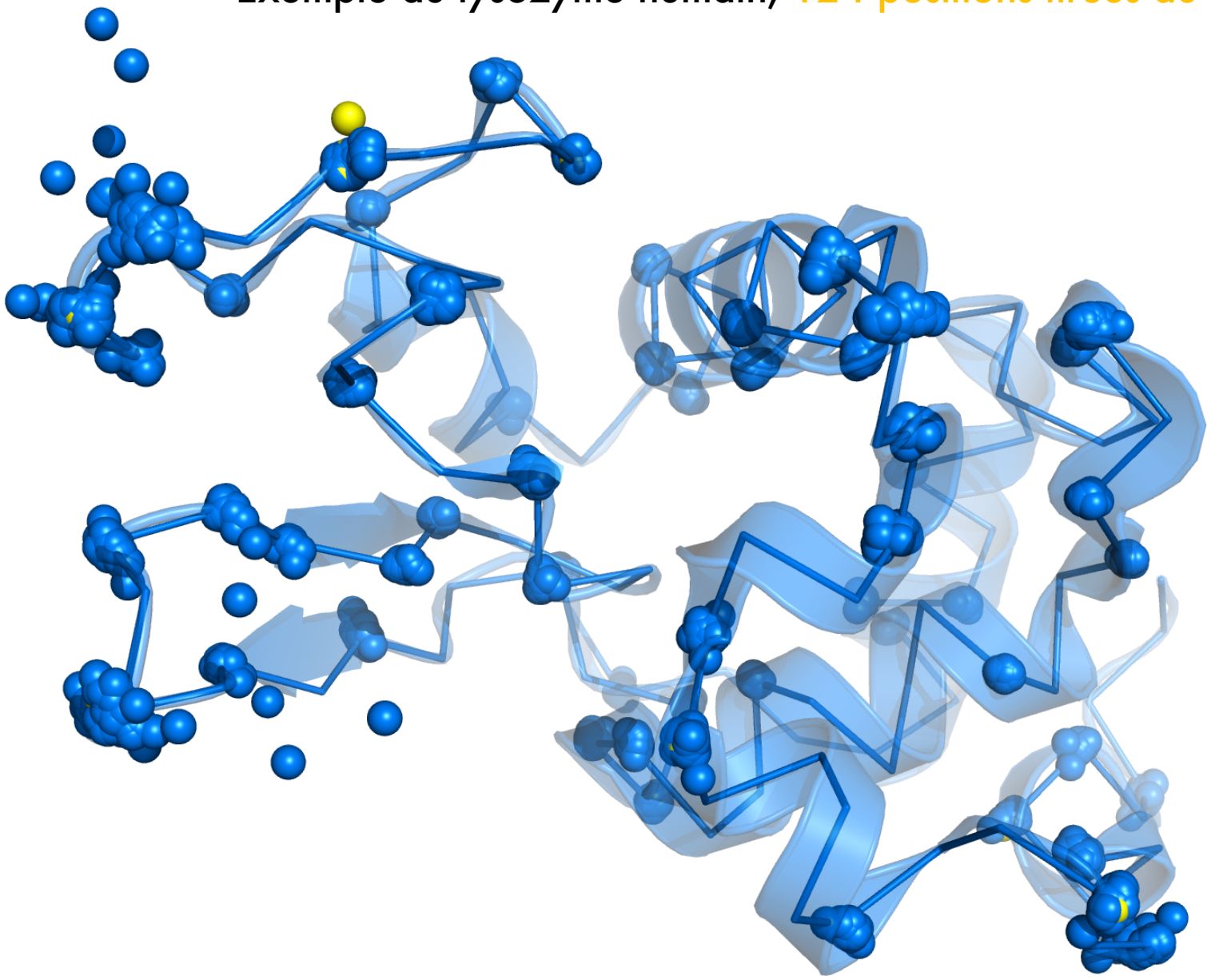
Exemple du lysozyme humain, 124 positions tirées au hasard



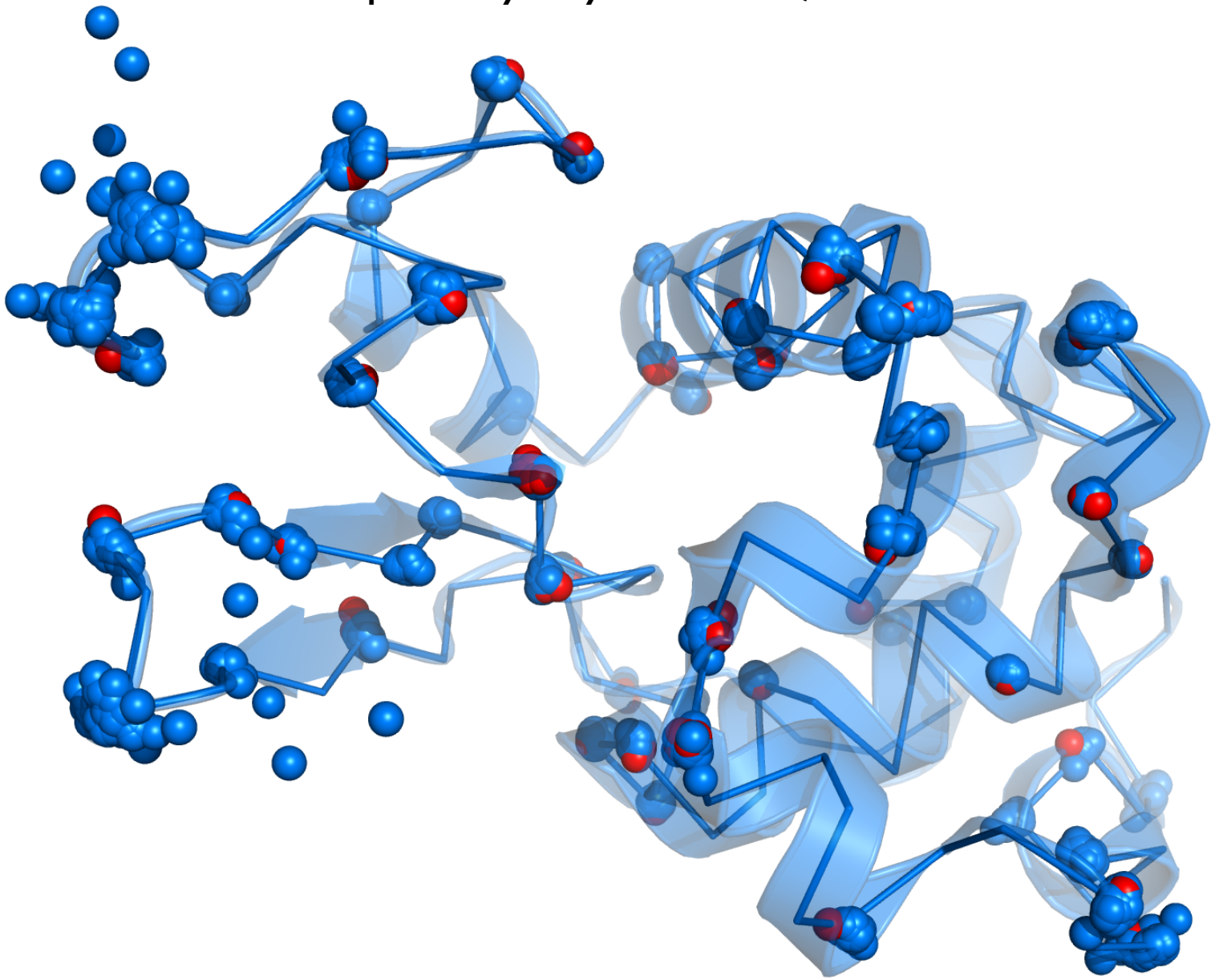
Exemple du lysozyme humain, 124 positions tirées au hasard



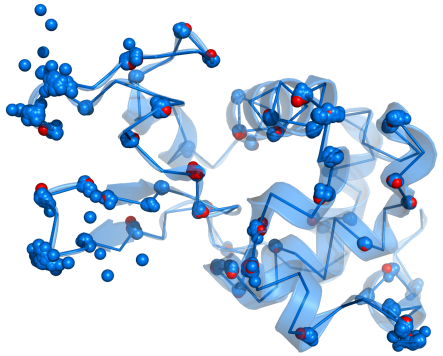
Exemple du lysozyme humain, 124 positions tirées au hasard



Exemple du lysozyme humain, 124 mutations



Effet des mutations :
quel est l'impact d'une mutation sur une structure ?



Conclusion:

=> Il y a un effet visible, quelque soit la localisation de la mutation

=> L'effet se propage mais il reste local (2/3 résidus en séquences, résidus en contact)

=> Il n'y a par contre pas de lien (évident) entre les variations de stabilité (DDG) et la déformation (RMS)

=> Il existe une faible corrélation positive entre les RMS et les erreurs de prédiction de DDG (FoldX[1]).

Et maintenant :

Comment se font se déformations ?

1 Schymkowitz, Joost et al. "The FoldX Web Server: An Online Force Field." *Nucleic Acids Research* 33.Web Server issue (2005): W382–W388. *PMC*. Web. 1 May 2017.

Effet des mutations :
Quel est l'impact d'une insertion ou d'une délétion ?
La structure se déforme-t-elle dans le sens des modes normaux basses fréquences ?

Données :

-> Paires de protéines ayant plus 90% d'identité

-> 134 délétions artificielles

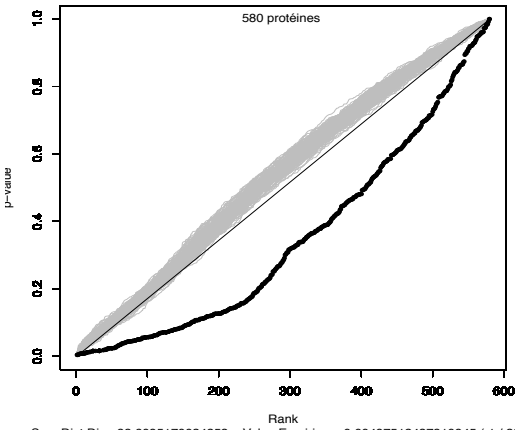
-> 17 insertions artificielles

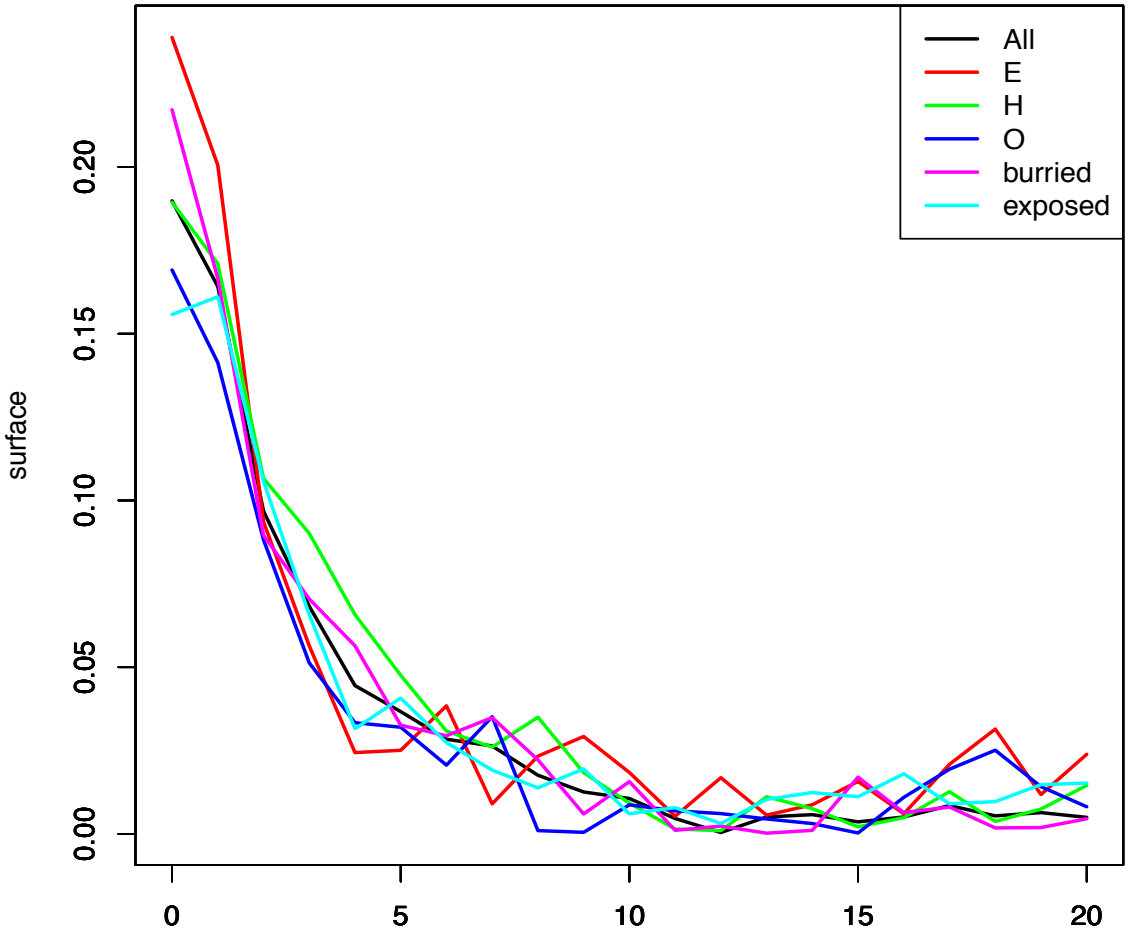
-> 505 indels autres (naturels ?)

Effet des mutations :
quel est l'impact d'une insertion ou d'une délétion ?

- Construction de la banque de données ✓
- Calculs des modes normaux ✓
- Comparaison des modes normaux basses fréquences avec les déformations structurales …
- Analyse en fonction de la localisation structurale des indels
- Identification des insertions et délétions dans les événements naturels
- Comparaison entres les insertions et délétions naturels et artificiels
- Recherche de co-évolution entre ces évènements et les mutations.

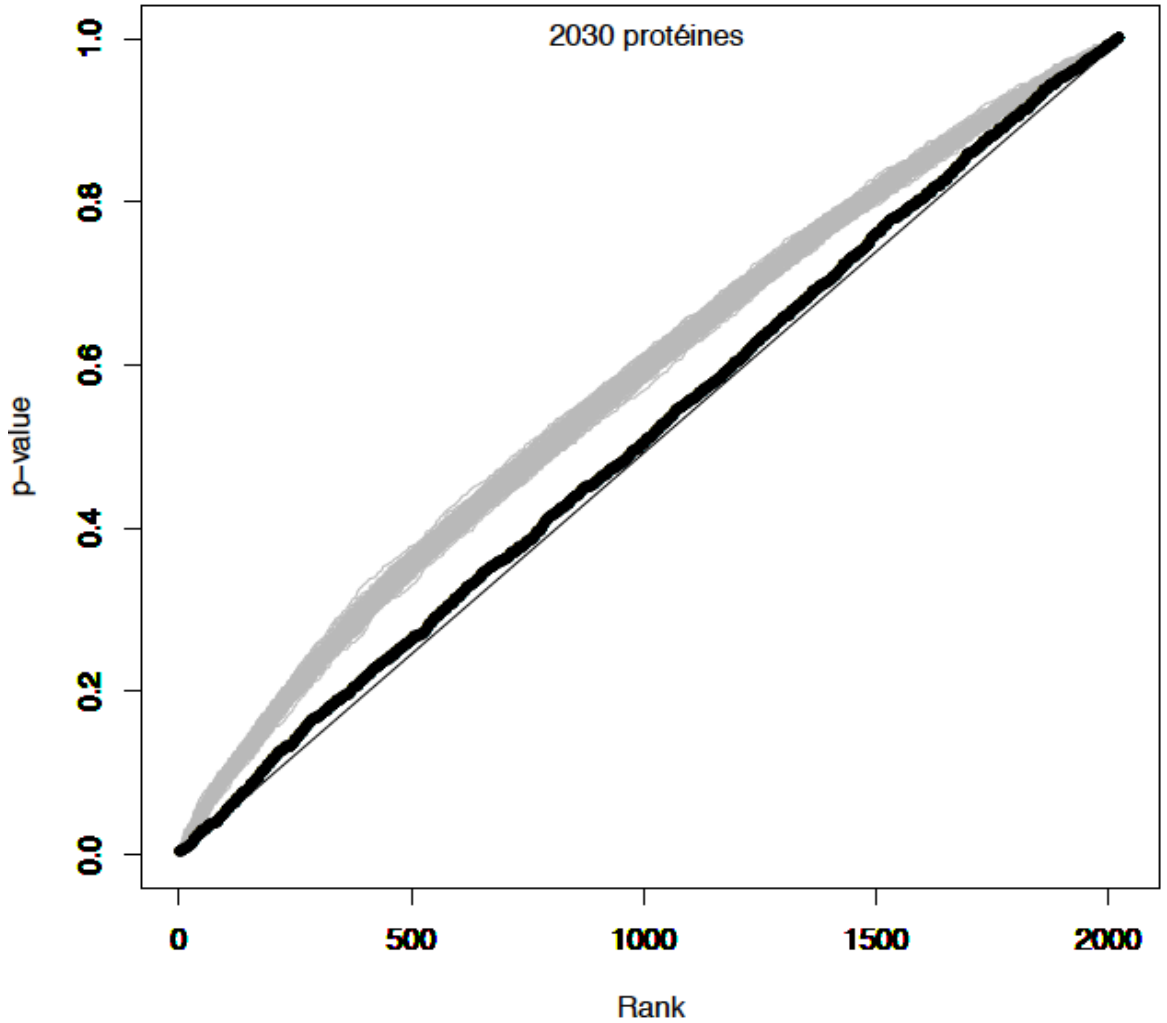
Effet des mutations :
quel est l'impact d'une mutation sur une structure ?





Surfaces between the average random and the mutated p-values for various windows along which the rms is calculated: centred on the mutated residue, and shifted on both sides by 1 to 20 residues. The surfaces have been calculated for several subsets of residues: mutated residue in helices, strands or loop (green, red and blue lines) and buried or exposed residues (magenta and cyan lines).

Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

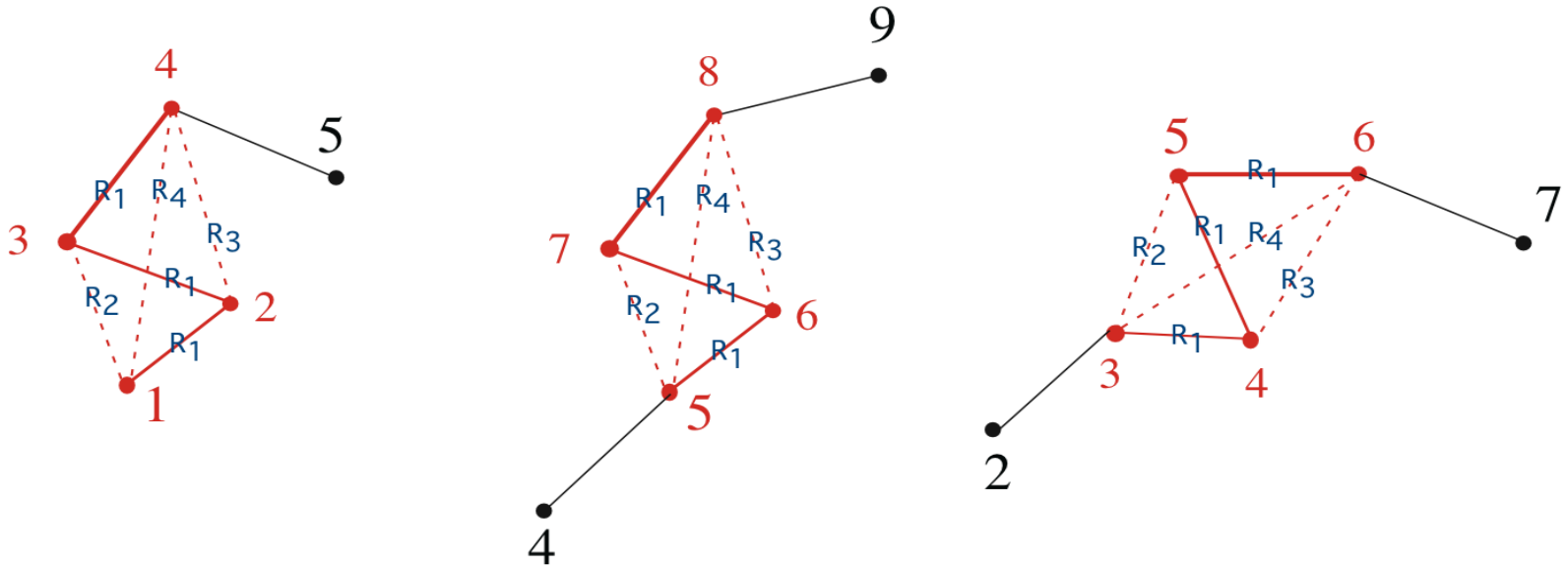


Rank vs p-values for residues that are geometrically close but not sequentially adjacent to the mutated position, for the whole mutated dataset.

Alignement multiple de structures

Recherche de motifs répétés : Triades

Distances internes entre Cα



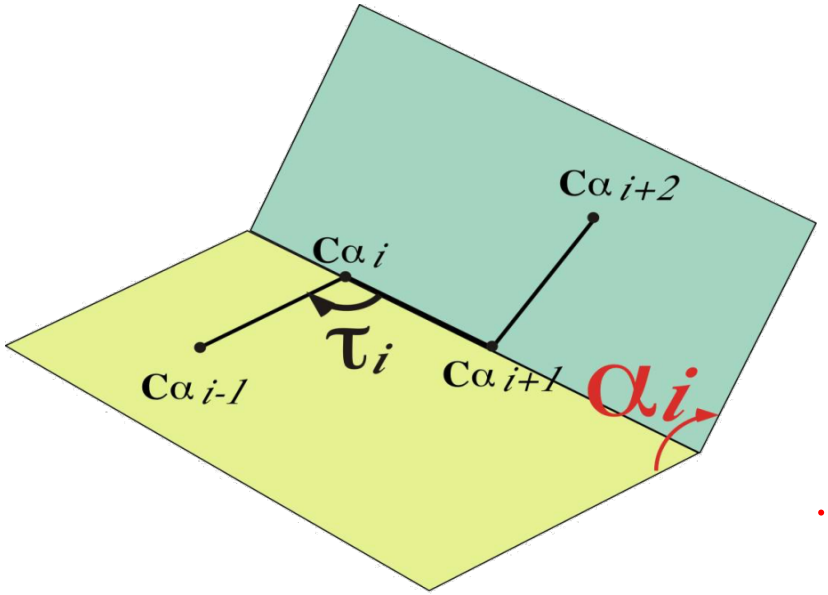
Construction des motits de taille k à partir des motits chevauchants de taille intérieure.

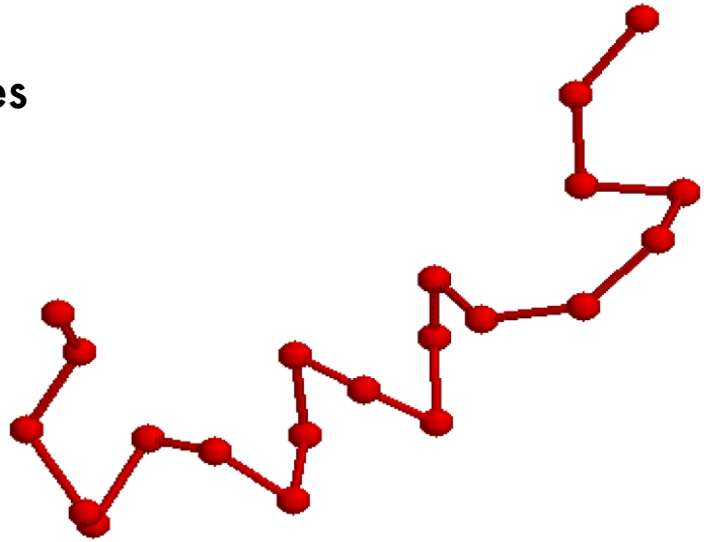
N. Pisanti, H. Soldano, M. Carpentier, Incremental Inference of Relational Motifs with a Degenerate Alphabet , Lecture Note in Computer Science (proceedings CPM, Combinatorial Pattern Matching, Volume 3537, May 2005, Pages 229 - 240).

N. Pisanti, H. Soldano, M. Carpentier, J. Pothier, Implicit and Explicit Representation of Approximated Motifs KCL series book, edited by C. Iliopoulos, K. Park and K. Steinhfel (à paraître en 2005)

Les angles α

Les structures comme des séries de symboles



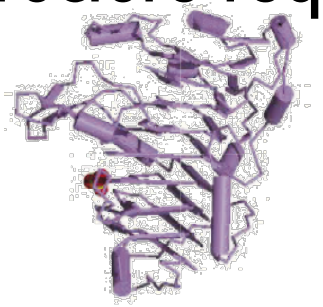


...206–55–52–63–…–79–46–150–250–…

...α_{21}–α_{05}–α_{05}–α_{06}–…–α_{07}–α_{04}–α_{15}–α_{25}–…

Recherche de similarités structurales dans une banque : YAKUSA

Structure requête

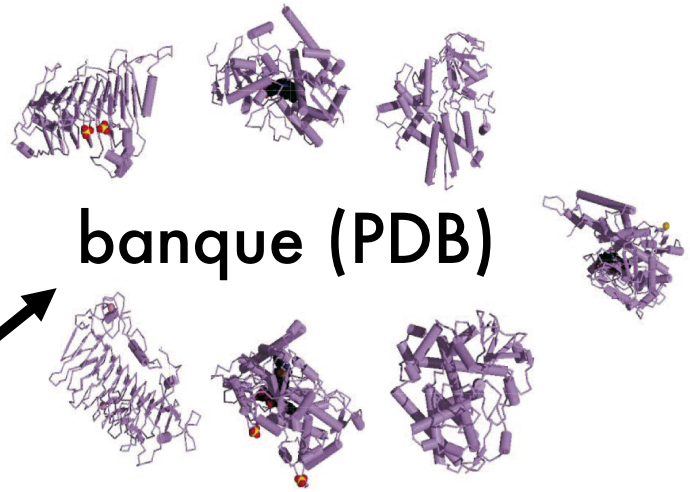


... 55–72–…–150–250–…

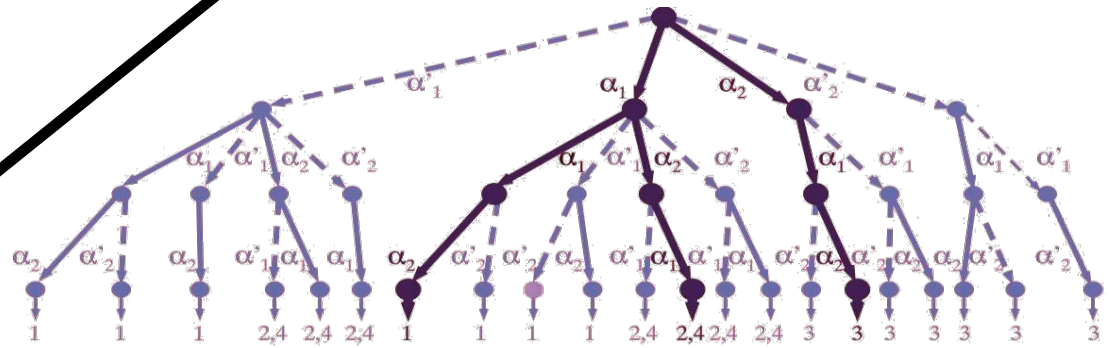
... α_{05}–α_{07}–…–α_{15}–α_{25}–…

Recherche sur banque

1 min 30 pour 15000 structures



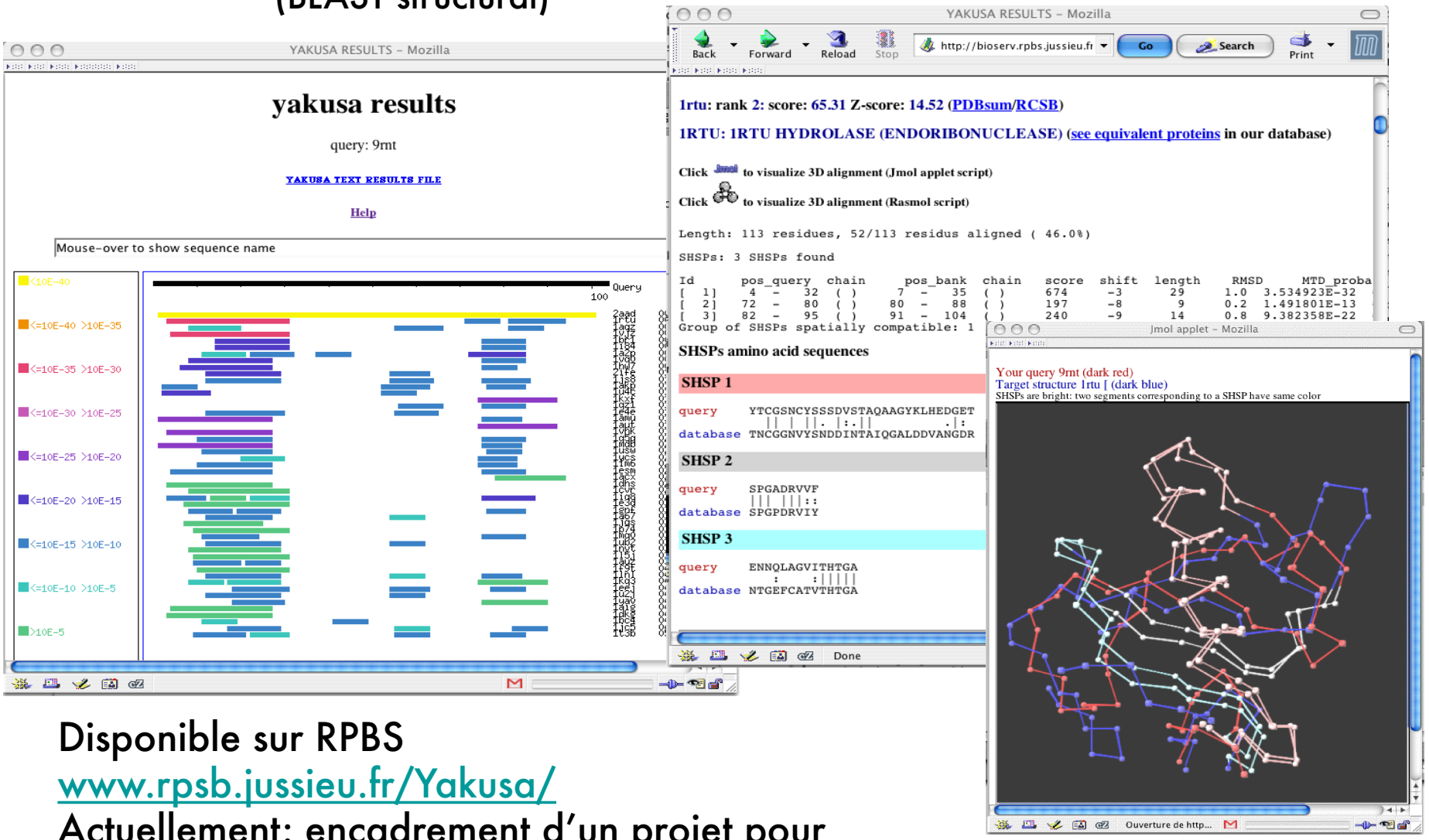
banque (PDB)



M. Carpentier, S. Brouillet, J. Pothier, YAKUSA: a fast structural databases scanning method, Proteins: Structure, Function, and Bioinformatics, volume 61, issue 1, pages 137-51.

Recherche de similarités structurales dans un banque : YAKUSA

(BLAST structural)



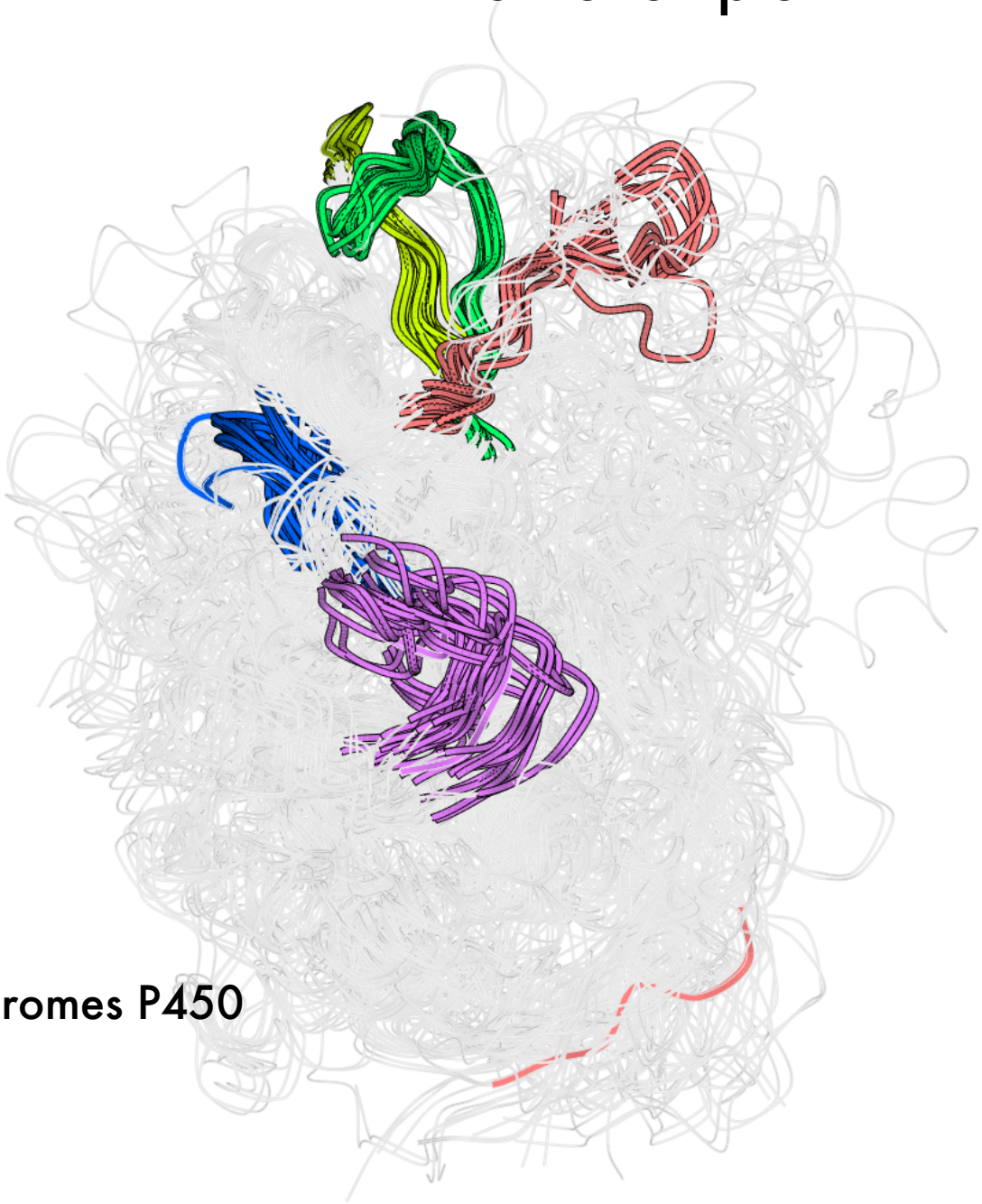
Disponible sur RPBS

www.rpsb.jussieu.fr/Yakusa/

Actuellement: encadrement d'un projet pour l'améliorer

C. Alland, F. Moreews, D. Boens, M. Carpentier, S. Chiusa, M. Lonquety, N. Renault, Y. Wong, H. Cantalloube, J. Chomilier, J. Hochez, J. Pothier, B.O. Villoutreix, J.-F. Zagury, P. Tuffery, ;
RPBS: a web resource for structural bioinformatics, Nucleic Acid Research, 2005, 33: W44–W49

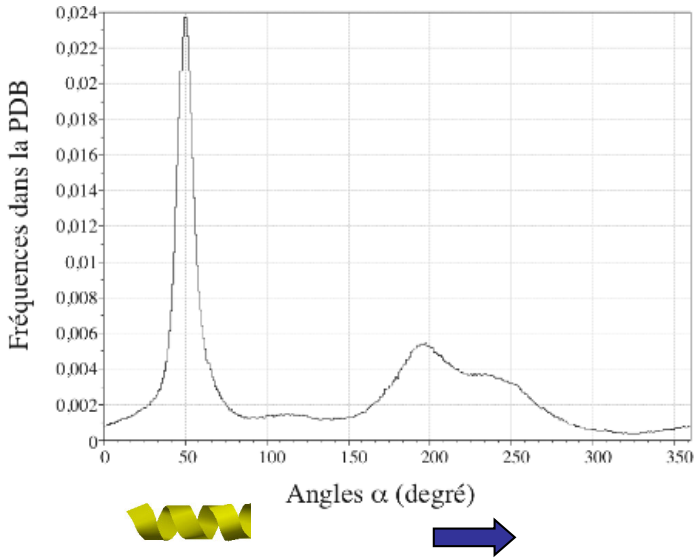
Un exemple



35 cytochromes P450

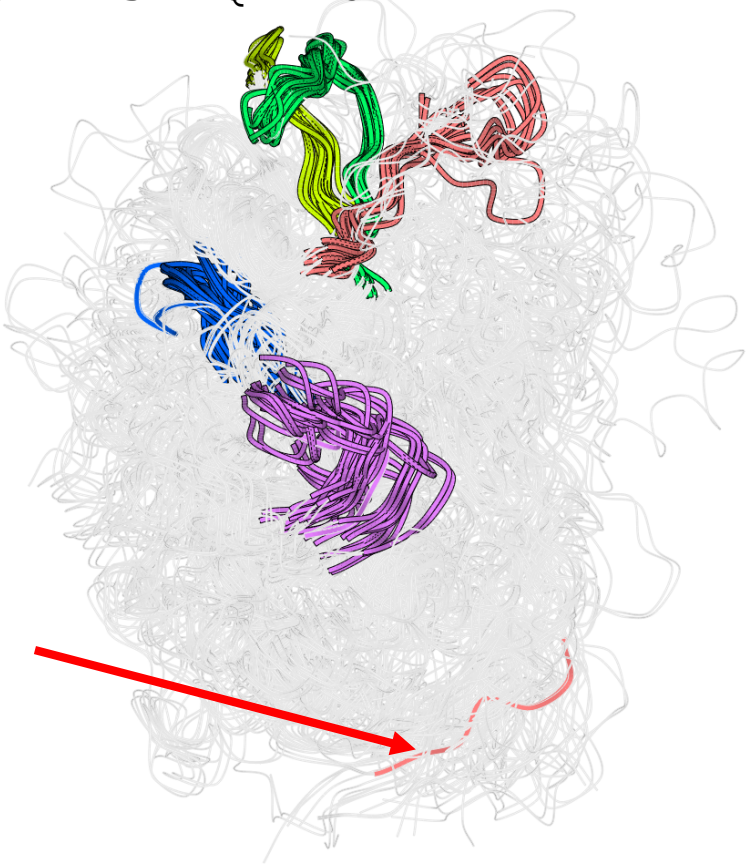
Exemple

```
1 -> 1CPT OXIDOREDUCTASE(OXYGENASE                DGECDFMTDCALY    148
2 -> 1DT6 OXIDOREDUCTASE                          KVSKGLGIAFSNA    105
3 -> 1E9X OXIDOREDUCTASE                          YEFEMAQPPESYR    415
4 -> 1IZO OXIDOREDUCTASE                          ADEVVLFEEAKEI    129
5 -> 1N40 OXIDOREDUCTASE                          GAPADLRNDFADP    128
6 -> 1N97 ELECTRON TRANSPORT                      GKPLSPSLAEHAL    146
7 -> 1OXA OXIDOREDUCTASE (OXYGENASE)              SGVVDIVDRFAHP    135
8 -> 1PHD OXIDOREDUCTASE(OXYGENASE)               QGQCNFTEDYAEP    145
9 -> 2HPD OXIDOREDUCTASE(OXYGENASE)               GFNYRFNSFYRDQ    157

1 -> |12|25|29| 9| 3|22|21|22|17|23|24|21|25|
2 -> |12|32| 1| 8|29|24|11|23|19|24|14| 5|26|
3 -> | 4| 3|35| 7| 1|21|12| 5|23|25|27|32| 7|
4 -> | 4|18| 1| 4| 2|23|22|22|23|23|22|22|22|
5 -> |16|23|35| 4| 1|22|20|21|19|21|22|22|23|
6 -> |23| 9| 7| 9| 3|23|23|22|23|23|22|22|22|
7 -> |19|13|34| 7| 1|23|21|22|17|23|22|22|24|
8 -> | 3|24|33| 6| 1|23|20|23|16|24|21|21|24|
9 -> |16| 7|31| 4| 0| 3|34|24|23|23| 2|22| 7|
```



Exemple

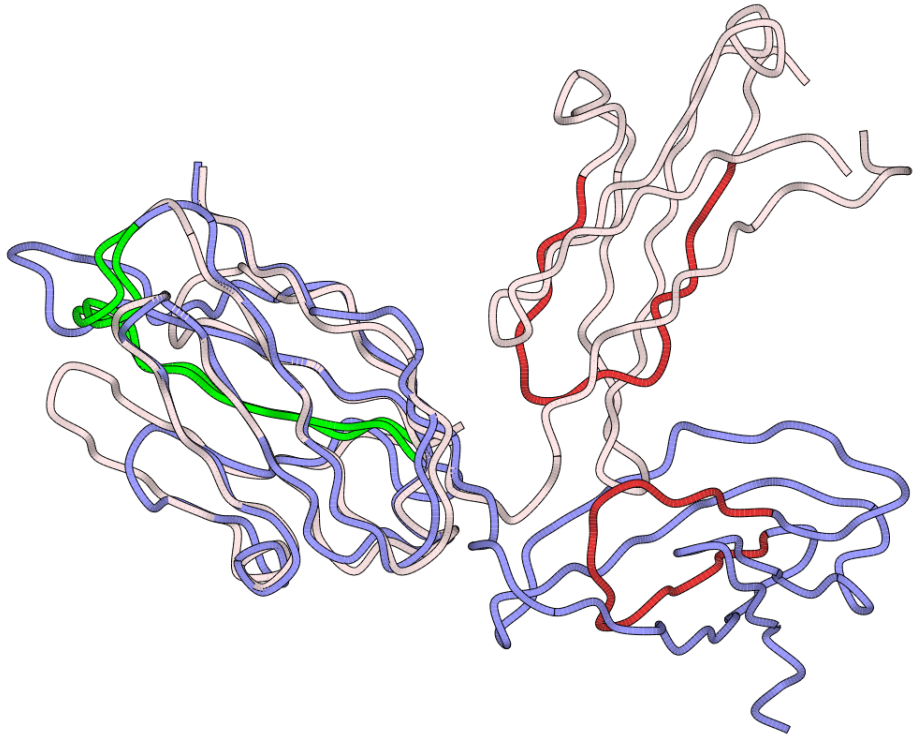
```
1 -> 1CPT OXIDOREDUCTASE(OXYGENASE                DGECDFMTDCALY    148
2 -> 1DT6 OXIDOREDUCTASE                          KVSKGLGIAFSNA    105
3 -> 1E9X OXIDOREDUCTASE                          YEFEMAQPPESYR    415
4 -> 1IZO OXIDOREDUCTASE                          ADEVVLFEEAKEI    129
5 -> 1N40 OXIDOREDUCTASE                          GAPADLRNDFADP    128
6 -> 1N97 ELECTRON TRANSPORT                      GKPLSPSLAEHAL    146
7 -> 1OXA OXIDOREDUCTASE (OXYGENASE)              SGVVDIVDRFAHP    135
8 -> 1PHD OXIDOREDUCTASE(OXYGENASE)               QGQCNFTEDYAEP    145
9 -> 2HPD OXIDOREDUCTASE(OXYGENASE)               GFNYRFNSFYRDQ    157

1 -> |12|25|29| 9| 3|22|21|22|17|23|24|21|25|
2 -> |12|32| 1| 8|29|24|11|23|19|24|14| 5|26|
3 -> | 4| 3|35| 7| 1|21|12| 5|23|25|27|32| 7|
4 -> | 4|18| 1| 4| 2|23|22|22|23|23|22|22|22|
5 -> |16|23|35| 4| 1|22|20|21|19|21|22|22|23|
6 -> |23| 9| 7| 9| 3|23|23|22|23|23|22|22|22|
7 -> |19|13|34| 7| 1|23|21|22|17|23|22|22|24|
8 -> | 3|24|33| 6| 1|23|20|23|16|24|21|21|24|
9 -> |16| 7|31| 4| 0| 3|34|24|23|23| 2|22| 7|
```



Modifications

Solutions testées :

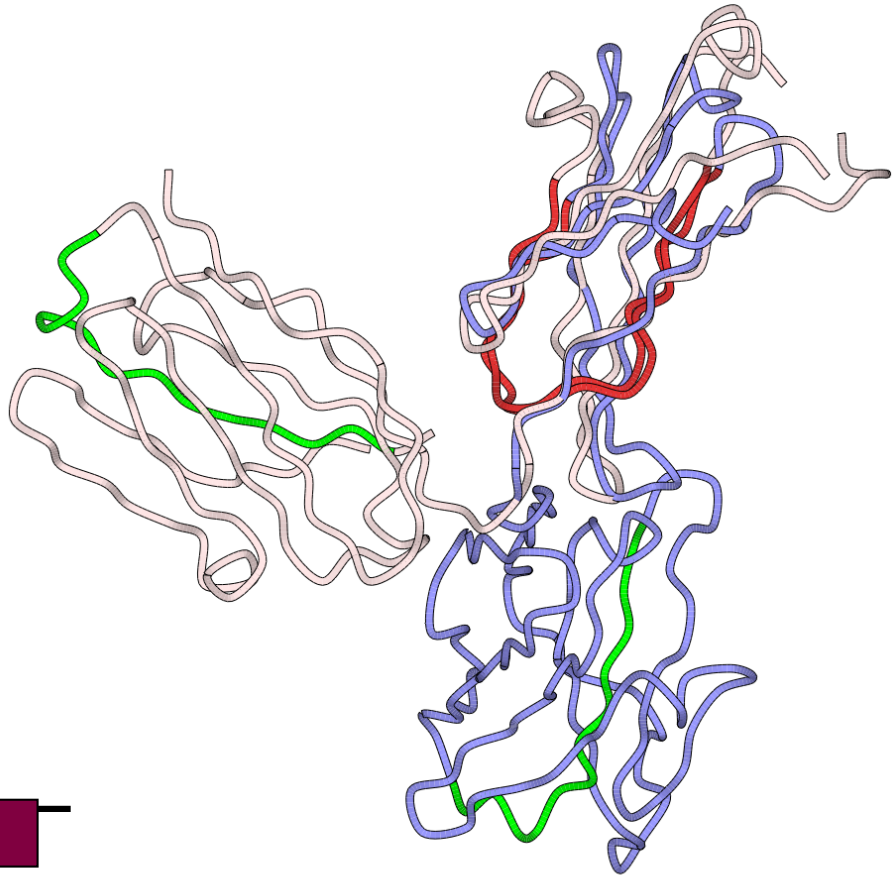
-> prise en compte du RMS



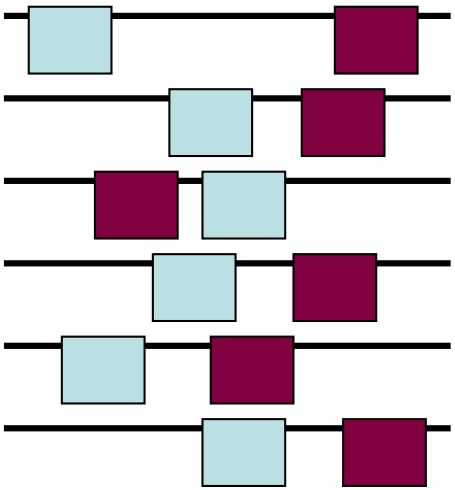
Modifications

Solutions testées :

-> prise en compte du RMS

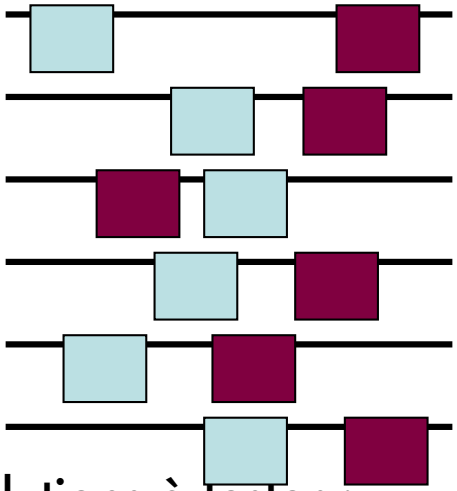


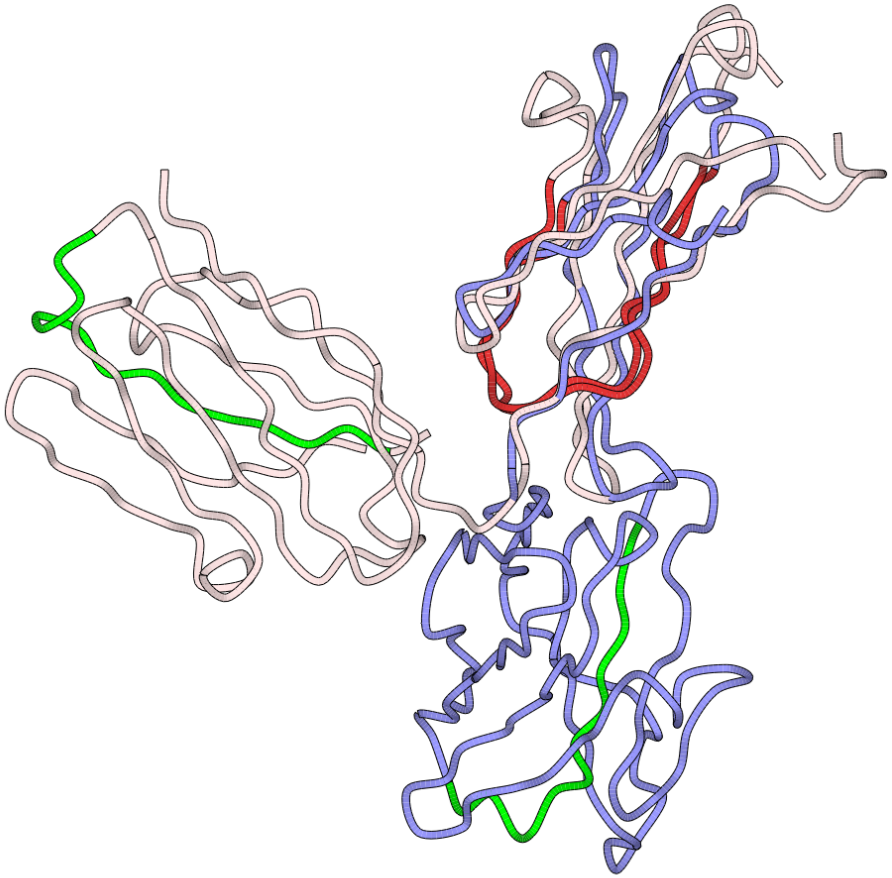
-> contrainte d'ordre



Perspectives

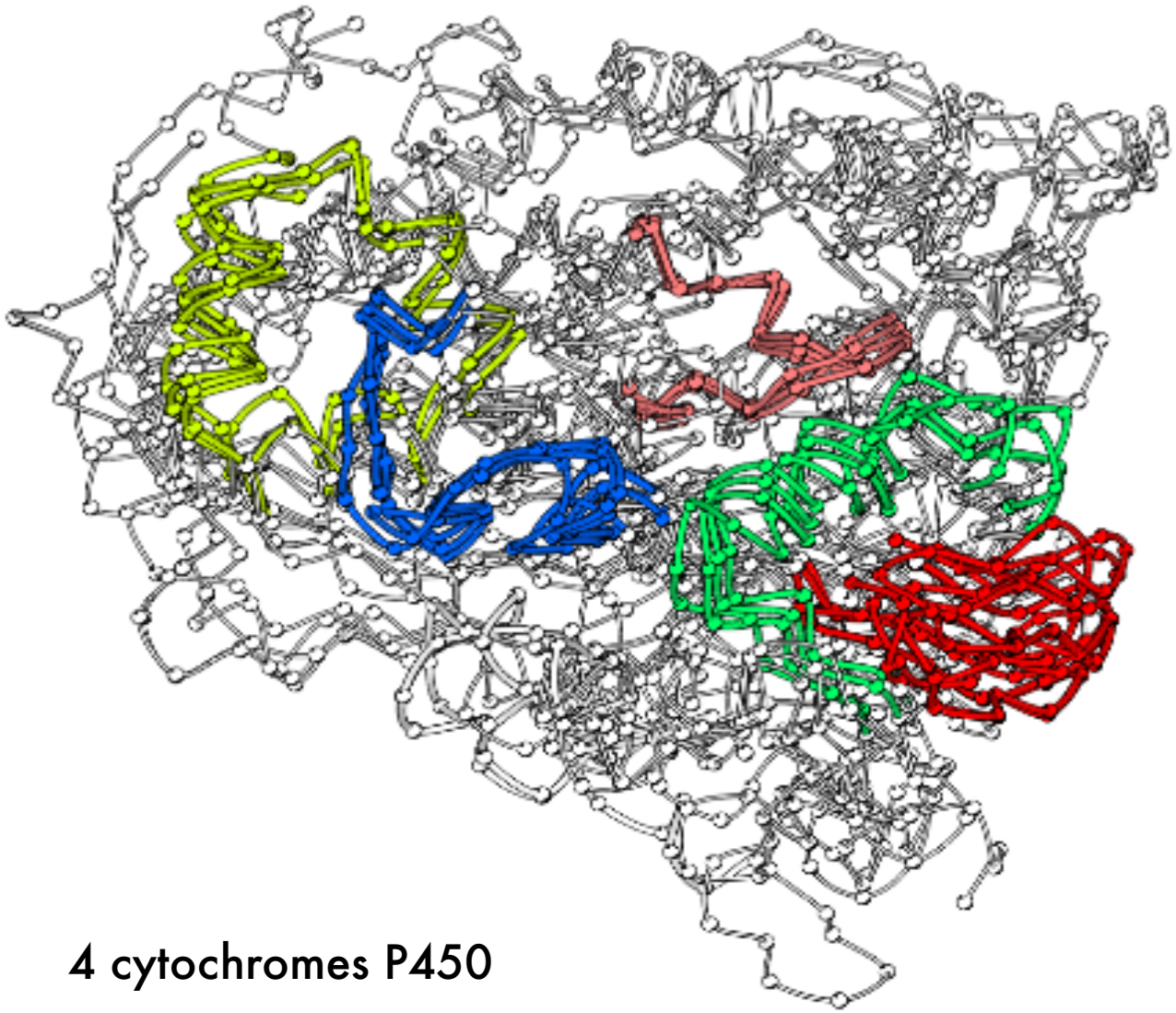
Solutions testées :
-> prise en compte du RMS
-> contrainte d'ordre





Solutions à tester :
-> ignorer les hélices
-> ne garder que les meilleurs blocs puis affiner l'alignement après superposition des structures selon ces blocs

Recherche de motifs répétés : Triades



4 cytochromes P450

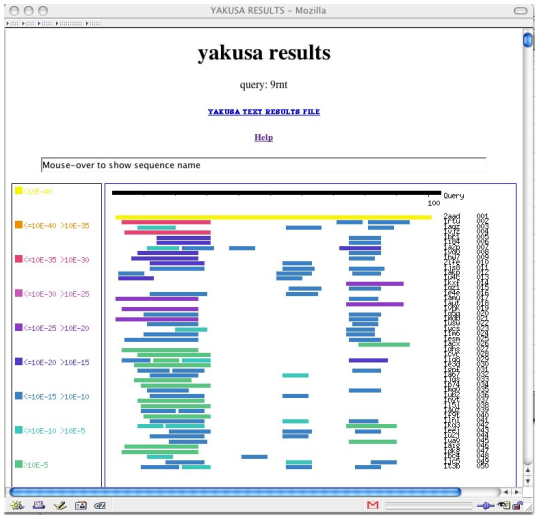
Comparaison des méthodes

- Méthode des « m-diagonales » :
  - quorum réglable
  - alignement des paires de structures
- méthode du Gibbs sampling :
  - pas d'alignement des paires, comparaison d'un très grand nombre de structures
  - quorum fixe
- Méthode de recherche de motifs répétés (Triades) :
  - pas d'alignement des paires, quorum réglable
  - exhaustivité
  - générique

Méthodes de comparaison de structures protéiques

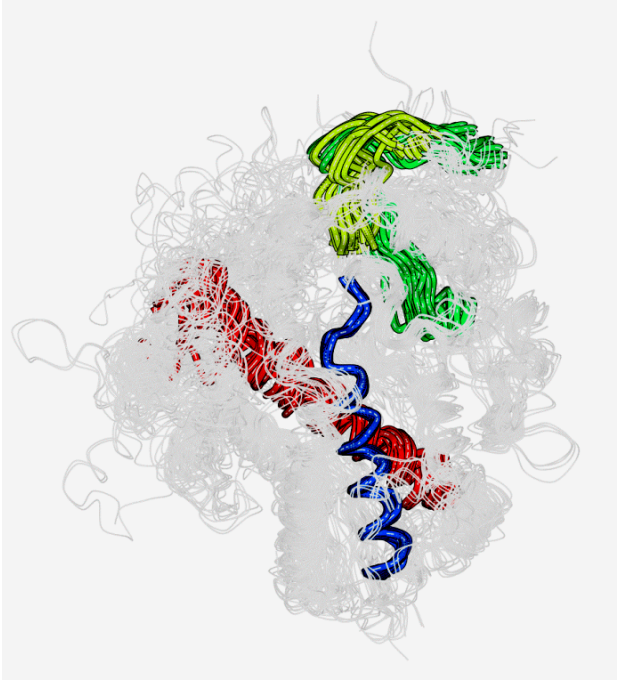
-Comparaison d'une structure avec toutes les structures d'une banque.

-> Yakusa[1]

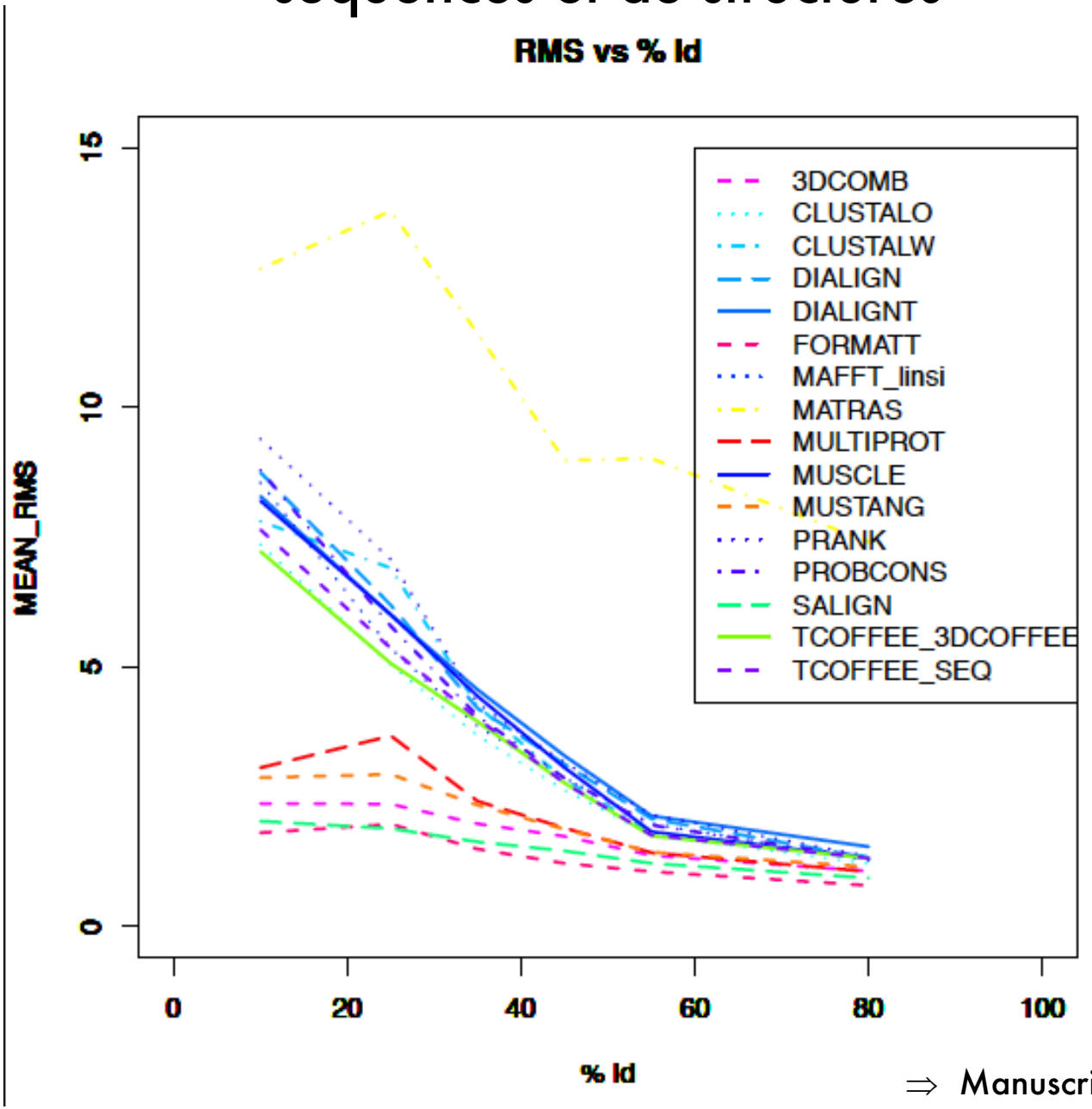


- Comparaison de plusieurs structures
3 méthodes:

* Méthode des « m-diagonales »
* Méthode du Gibbs sampling
* Méthode de recherche de motifs relationnels (Triades)

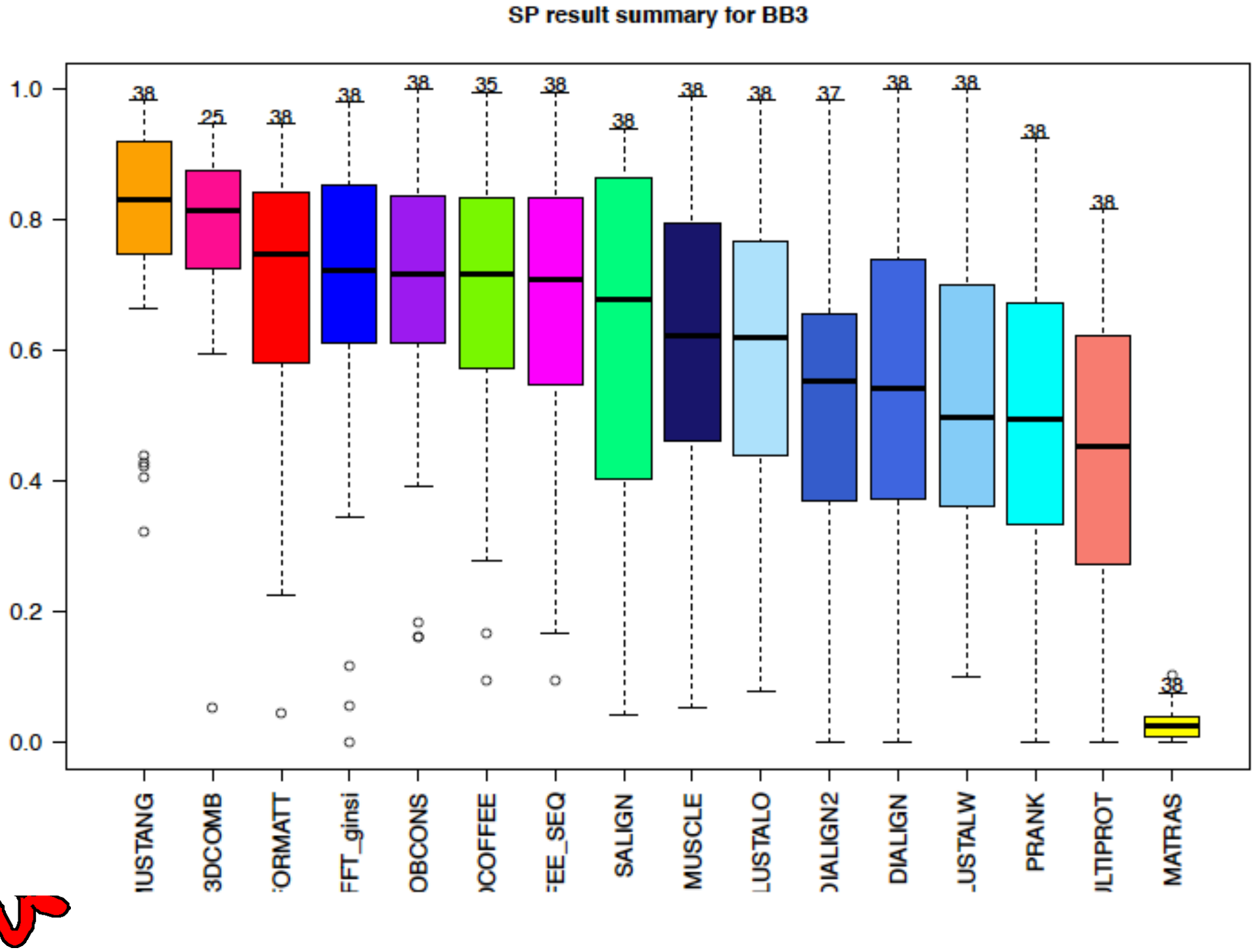


Comparaison des méthodes d'alignement multiple de séquences et de structures



⇒ Manuscrit en cours d'écriture

Alig
stru


SP result summary for BB3
1.0
0.8
0.6
0.4
0.2
0.0
38
25
38
38
38
35
38
38
38
37
38
38
38
38
IUSTANG
3DCOMB
ORMATT
FFT_ginsi
OBCONS
COFFEE
EE_SEQ
SALIGN
MUSCLE
LUSTALO
DIALIGN2
DIALIGN
USTALW
PRANK
LTIPROT
MATRAS

Echave, J. & Fernández, F. M. A perturbative view of protein structural variation. *Proteins* 78, 173–180 (2010).

Datasets

- LFENM: SIMULATED réf = 1a6m sperm whale oxy-myoglobin crystallized at pH 7, 151 residus. Start ing from 1a6m, we simulated 151x100=15,100 single-point mutants using random forces as described in the previous section. To take into account the heme into the ENM we placed five extra nodes at the posi- tions of the heme's Fe and the four CH porphyrin atoms.
- Random As a reference model, a null model to compare the LFENM with, we used a dataset of 1510 simulated structures obtained by adding to the wild-type (reference) structure a vector of dimension 3N with random elements picked from a uniform distribution with values in (2n, n) with n 5 0.1.
- Globin-like This dataset includes 1a6m and 21 members of the superfamily of globin-like homologous proteins, as classified in the SCOP database.9
- Mutants This dataset contains 119 sperm whale myoglobin mutant structures: 22 single mutants, 77 double mutants, 7 triple mutants, and 13 quadruple mutants. Members of this dataset may also differ from 1a6m in the ligand bound to Fe and/or pH, heme state. For most (108) cases the aspartic acid in position 122 is replaced by asparagine. Most of the rest of the mutations are at sites 29, 64, 68, and 67
- Wild-type variants This dataset includes 1a6m and 48 structures that have the same (wild-type) sequence. Different members of this set have different ligands, pH, and/or Fe oxidation state. There are also 3 members with Co(II) replacing Fe(II).

Echave, J. & Fernández, F. M. A perturbative view of protein structural variation. *Proteins* 78, 173–180 (2010).

To better understand the observed connection between evolutionary deformations and dynamical deformations, a model was proposed in which pertur- bation of Elastic Network Models accounts for the effect of mutations on equilibrium conformation.140

This model predicts that the equilibrium conforma- tion will diverge along the low-energy normal modes even under random unselected mutations, which casts doubt on the functional interpretation. If the perturbed ENM is correct, dynamical deformations (normal modes) should govern not only evolutionary divergence, but also the structural change due to perturbation. Further support to the idea of functional signal in ENM perturbation comes from the observation that the same pattern variation along normal modes is found for unselected engineered mutants and for structures of the same protein determined in different experimental conditions.141
To say that even under random mutations a pro- tein would diverge along the lowest normal modes is not to say that such modes are nonfunctional or that selection plays no role in molding structural diver- gence. It is possible that natural selection increases or decreases the contribution of a certain normal mode to structural variation. However, a careful assessment demands the use of a null model that takes into account the dominant effect of the lowest normal modes even in the absence of selection. There is some work that suggests that this could be the case for proteins that experience large functional conformational transitions.145 Disentangling the effects of natural selection from those of drift on the patterns of structural divergence is a subject on which further research is needed.

Effet des mutations :
quel est l'impact d'une mutation sur une structure ?

RMS