Protein local conformations at the light of a structural alphabet

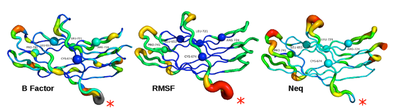
Alexandre G. de Brevern[1].

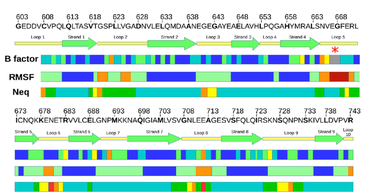
[1]INSERM, UMR_S 1134, DSIMB, Univ Paris Diderot, Sorbonne Paris Cité, University de la Reunion, University des Antilles, INTS, GR-Ex, Laboratoire d'Excellence, F-75739 Paris, France.

ABSTRACT

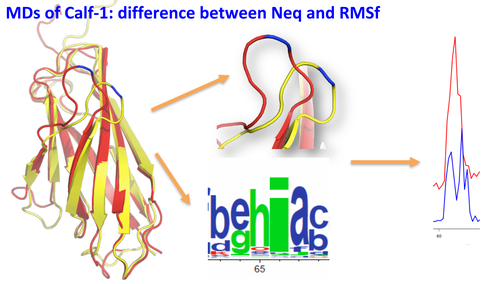
Protein structures are classically described in terms of secondary structures. Even if the regular secondary structures have relevant physical meaning, their recognition from atomic coordinates has some important limitations such as uncertainties in the assignment of boundaries of helical and β-strand regions. Further, on an average about 50% of all residues are assigned to an irregular state, i.e., the coil. Thus different research teams have focused on abstracting conformation of protein backbone in the localized short stretches. Using different geometric measures, local stretches in protein structures are clustered in a chosen number of states. A prototype representative of the local structures in each cluster is generally defined. These libraries of local structures prototypes are named as "structural alphabets". We have developed a structural alphabet, named Protein Blocks, not only to approximate the protein structure, but also to predict them from sequence [1,2]. Since its development, we and other teams have explored numerous new research fields using this structural alphabet. I will review here some of the most interesting applications: (i) the most efficient protein superimposition methods [3,4], (ii) new ways to analyze protein structures [2], and (iii) new tool for analysis of protein dynamics and allostery [5,6].

New tool for analysis of protein dynamics and allostery



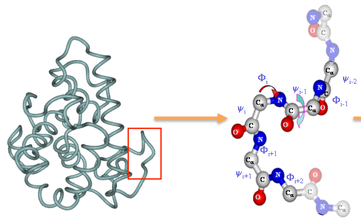


MDs of Calf-1: difference between Neq and RMSf



Flexibility: no ! rigid part surrounded by deformable regions

Encoding From 3D Protein structure to 1D sequence in terms of Protein Blocks

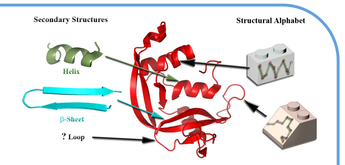


fragment *f*
(5 aa, 8 dihedral angles)

$\ldots \psi_{i-2}, \phi_{i-1}, \psi_{i-1}, \omega_{i-2}, \phi_i, \psi_i, \phi_{i+1}, \psi_{i+1}, \phi_{i+2} \ldots$

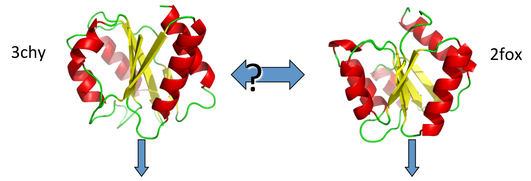
distance computation => 16 scores

The smallest




```
>153L
ZZmnopfklpccebjafklmmmnopabecjklmmmmmmmmm
mmmmmmmmmmmmmmmmmmnopafklmmmmmmmmmnooolapgeh
iafkopagcjkopafklmccehjfklmklmmmmmmmmmmmmm
mmmmmmmmmmbcfklmmmmmmmmmmmnomklmnbfklmmgoia
hilmmmmmmmmmmmmmmmnoZZ
```

iPBA: pairwise protein structure superimposition (http://www.dsimb.inserm.fr/dsimb_tools/ipba/)

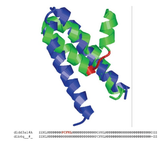



```
KBCCDDDDDFBFKLMMMMMMMMMNOPABDCDDFB
FKLMMMMMMNGOIABDCDDFBDGHILMLMMMMM
MMPMKLMMPCCDDDDDFBDCFKLMMMMMMNOPABD
CDDDDDFKLMMMMMMMMMMNO

DDDDFKOMMMMMMMMMMMMMMNOPACDDDDFKL
PCFKLMMMPFBDCDDDDDEHJLPACFKLMMMMMM
MMMMMMGHIACDDEHIAFKLNOMMMMMMMMMMMM
NOPACFBACDDDEHJACKLMMMMMMMMMMMMM
```

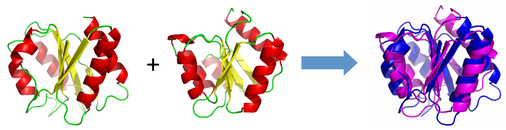
Substitution matrix based on superimposed proteins of PALI
+ Local / global alignment + anchor-based approach

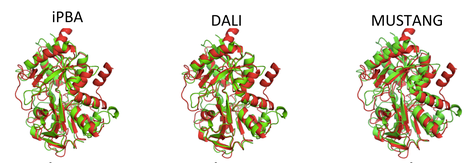
mulPBA: multiple protein structures superimposition (http://www.dsimb.inserm.fr/dsimb_tools/mulpba/)

```
3chy_: KBCCDDDDDFBF--KLMMMMMMMMMNOPABDCDDFBFKLMMMMMMNGOIABDCDDFBDGHI
2fox_: ---DDDDFKOMMMMMMMMMMMMMNOPACDDDD--FKLPCFKLMMM-PFBDCDDDDDEHJ

3chy_: LM--LMMMMMMMPMKLMMPCCDDDDDFBDCFKL----MM-MMM-M-NOPA--BDCDDDD-
2fox_: LPACFKLMMMMMMMMMMMMGHIACDDEHIAFKLNOMMMMMMMMMMMMNOPACFBACDDDEH

3chy_: -DFKLMMMMMMMMMMMNO
2fox_: JACKLMMMMMMMMMMMMM
```



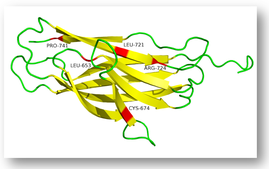
Performance of iPBA was better than methods like DALI and MUSTANG in more than 80% of the pairwise alignments.



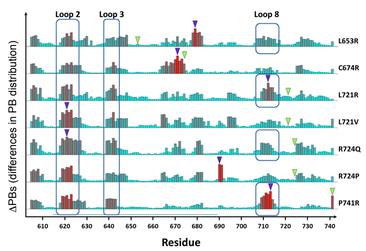


New tool for analysis of protein dynamics and allostery

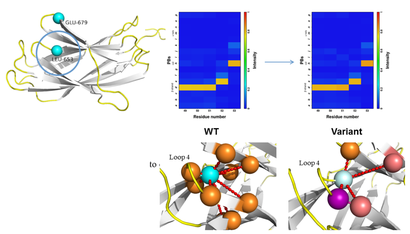
Calf-1: no impact at the variant sites !



Calf-1: Global variant mobility

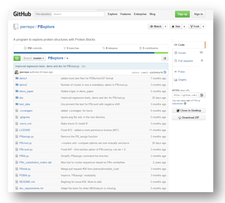


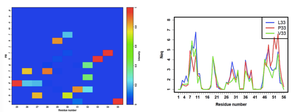
Variant L653R



Wild-type: 9 interactions
Variant: only one conserved but 3 new ones that compensate !
→ No structural changes

https://github.com/pierrepo/PBxplore





Long range impacts – observed with PBs

REFERENCES

[1] de Brevern A.G., Etchebest C. and Hazout S. (2000), Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks, *Proteins*, **41**:271-287.
[2] Joseph A.P., Agarwal G., Mahajan S., Gelly J.-C., Swapna L.S., Offmann B., Cadet F., Bornot A., Tyagi M., Valadié H., Schneider B., Etchebest C., Srinivasan N. and de Brevern A.G. (2010), A short survey on protein blocks. *Biophysical reviews*, **2**:137-147.
[3] Gelly J.-C. , Joseph A.P. , Srinivasan N. and de Brevern A.G. (2011), iPBA: A tool for protein structure comparison using sequence alignment strategies, *Nucleic Acid Res.*, **39** :W18-23,
[4] Joseph A.P., Srinivasan N. and de Brevern A.G. (2012), Progressive structure-based alignment of homologous proteins: Adopting sequence comparison strategies, *Biochimie*, **94**:2025-2034.
[5] Barnoud J., Santuz H., Craveur P., Joseph A.P., Jallu V., Poulain P., de Brevern A.G. (2017) PBxplore: a tool to analyze local protein structure and deformability with Protein Blocks. *PeerJ*, **5**:e4013.
[6] Goguet M., Narwani T.J., Petermann R., Jallu V., de Brevern A.G. (2017) In silico analysis of Glanzmann variants of Calf-1 domain of αIIbβ3 integrin revealed dynamic allosteric effect, *Sci Rep.*, **7**:8001.

Acknowledgments

This work is supported by grants from the Ministry of Research (France), University Paris Diderot, Sorbonne Paris Cité, the National Institute for Blood Transfusion (INTS, France), the Institute for Health and Medical Research (Inserm, France) and CEFIPRA (Indo-French collaborative grant numbers 3903-E & 5302-2).