



**HAL**  
open science

# Analysis of allosteric effect of pathologic variants at the light of local protein conformations

Alexandre de Brevern

► **To cite this version:**

Alexandre de Brevern. Analysis of allosteric effect of pathologic variants at the light of local protein conformations. Belgrade BioInformatics Conference – BelBi 2018, Jun 2018, Belgrade, Serbia. inserm-01853665

**HAL Id: inserm-01853665**

**<https://inserm.hal.science/inserm-01853665v1>**

Submitted on 3 Aug 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**BELBI 2018**



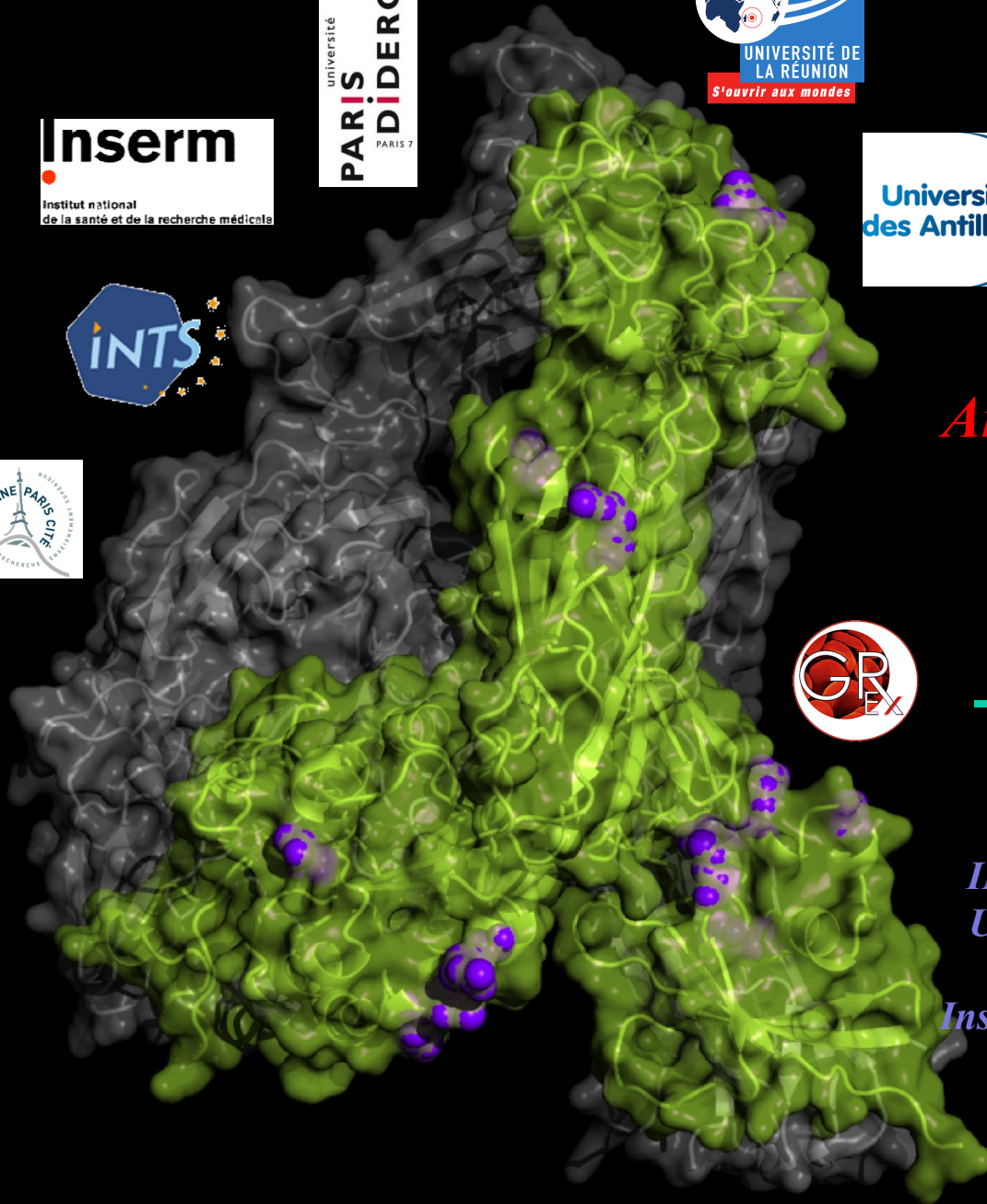


*BelBi 2018*

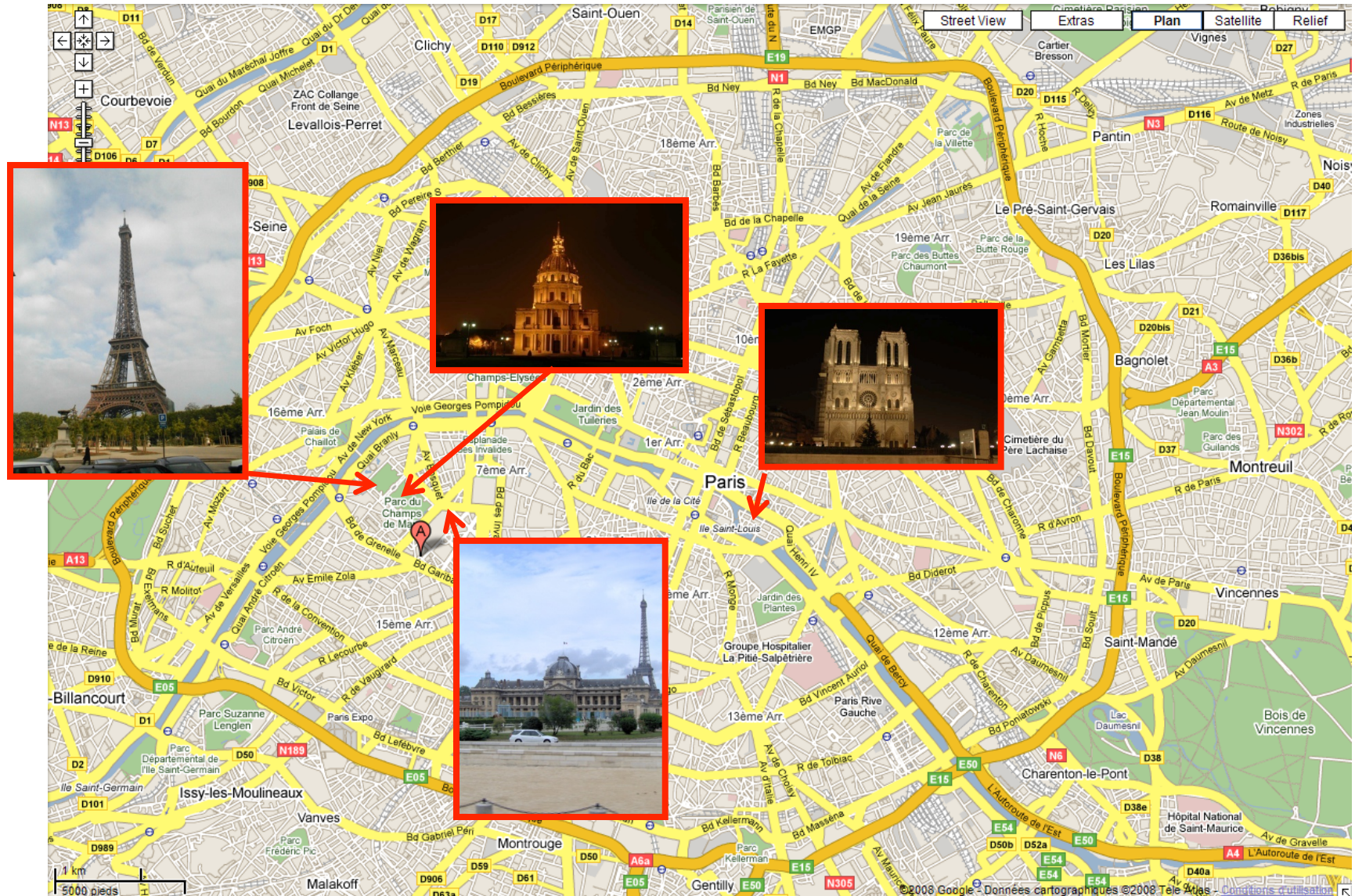
*Analysis of allosteric effect of  
pathologic variants at  
the light of local protein  
conformations*

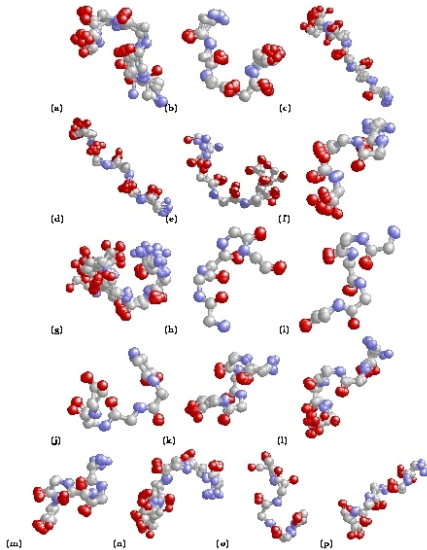
*Alexandre G. de Brevern*

*INSERM UMR\_S 1134, team 2 – DSIMB,  
Univ. Paris Diderot, Sorbonne Paris Cité,  
Univ de la Réunion, Univ des Antilles,  
Institut National de la Transfusion Sanguine  
(INTS), GR-Ex*



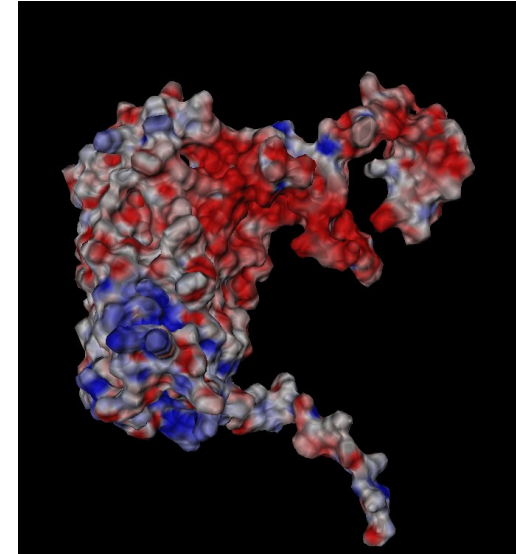
# DSIMB: a bioinformatics team

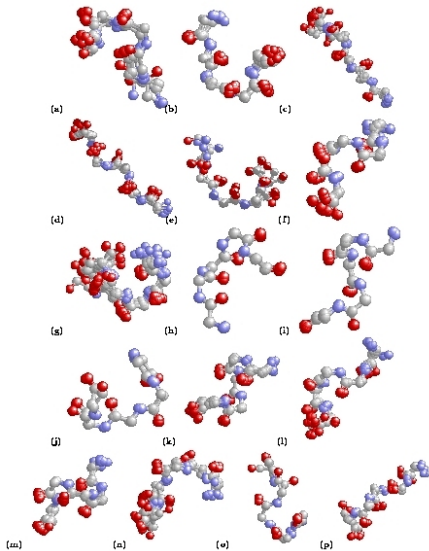




**Our main goal:**  
Sequence - Structure  
Function & Dynamics

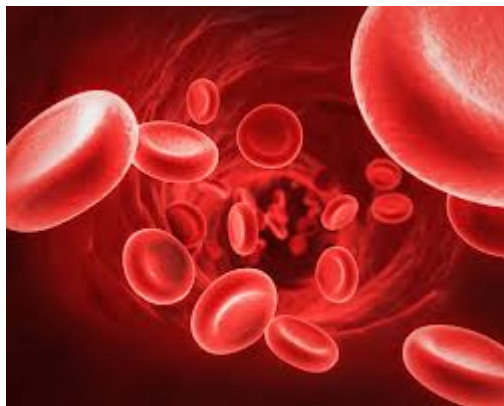
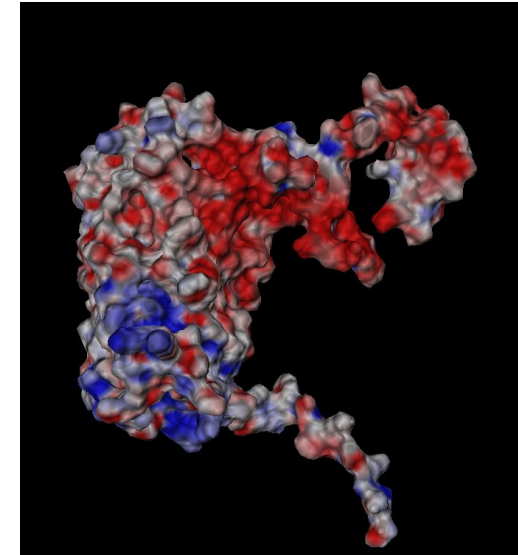
**Methodological  
developments  
and applications**



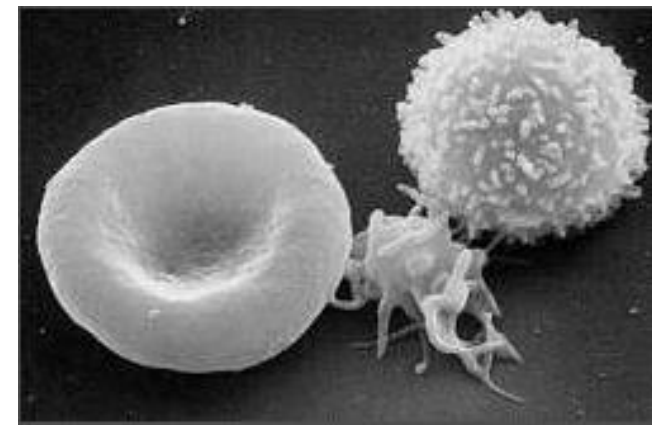


**Our main goal:**  
Sequence - Structure  
Function & Dynamics

**Methodological  
developments  
and applications**



**Red Blood Cells  
and Platelets**



- ❖ Why going beyond classical secondary structure ?
- ❖ How ?
- ❖ How is it useful for analysis of protein dynamics ?
- ❖ Can it be used also for protein disorder ?

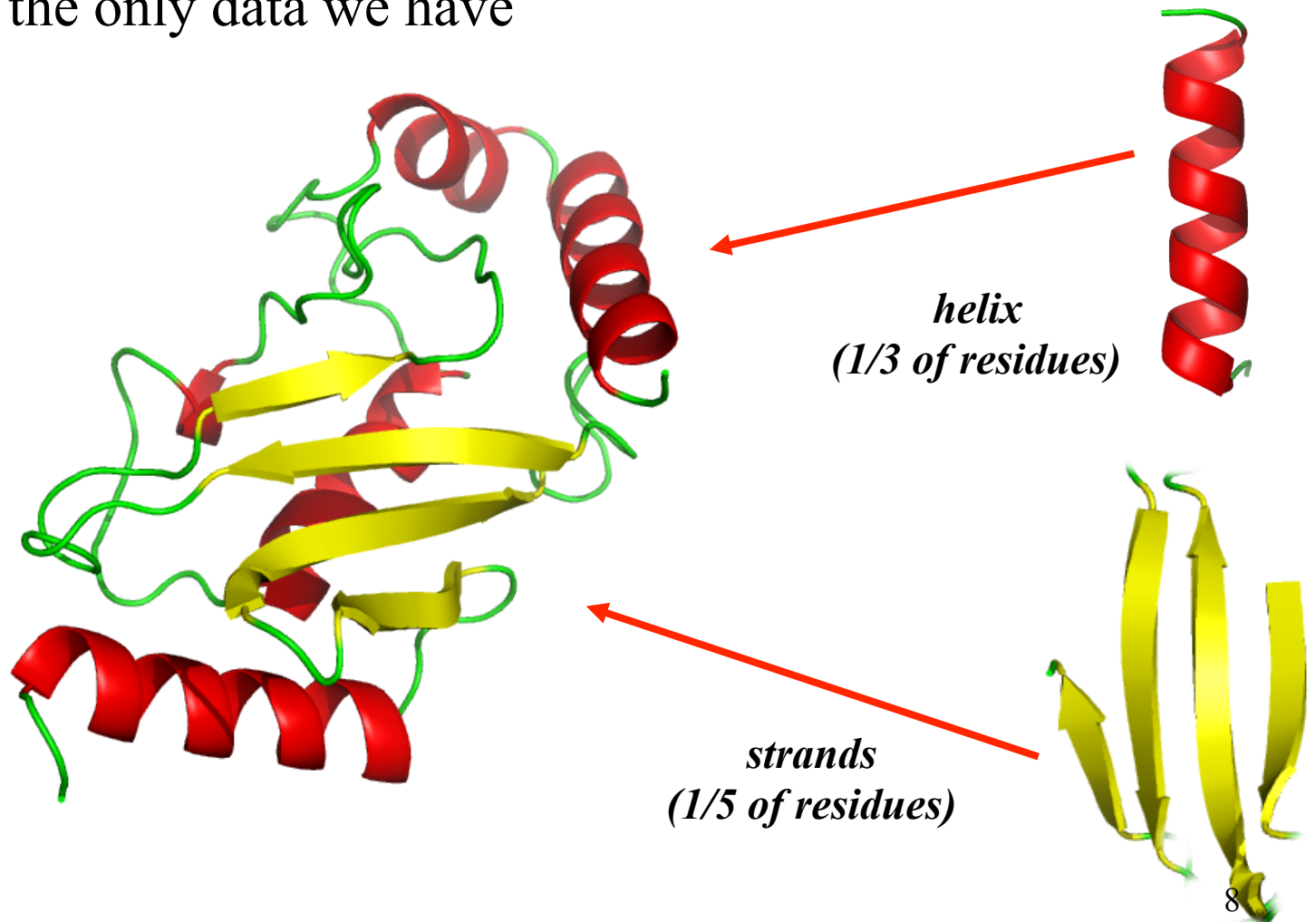
---

# SECONDARY STRUCTURE

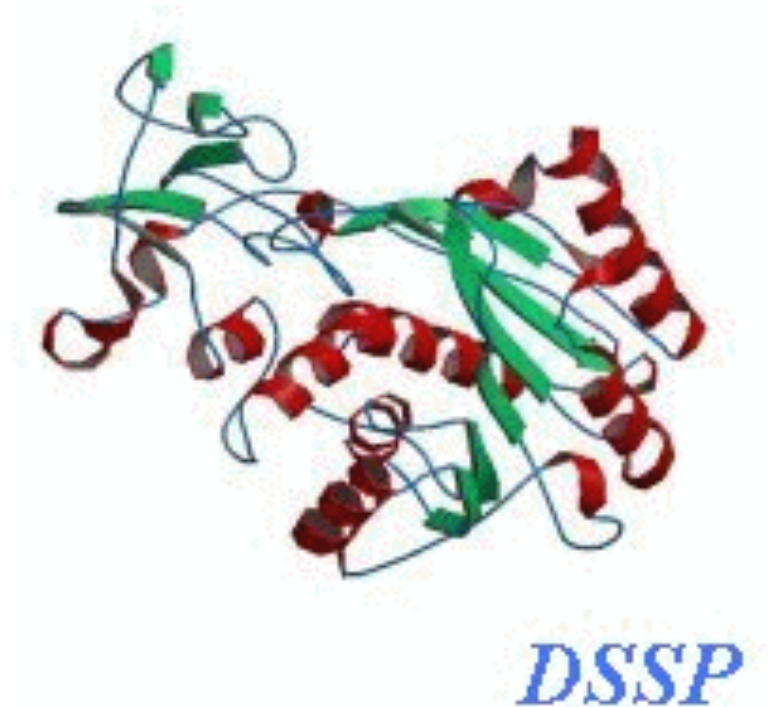


# The secondary structures

- That's the only data we have



# Secondary structures: some questions



No simple  
consensus  
between methods

Assignment example for the *Hhai Methyltransferase* protein (PDB code :10MH) by DSSP, STRIDE, PSEA, DEFINE, PCURVE, XTLSSTR and SECSTR.

L. Fourrier, C. Benros & A.G. de Brevern (*BMC Bioinfo*, 2004)

# Secondary structures: some questions

DSSP



KAKSI



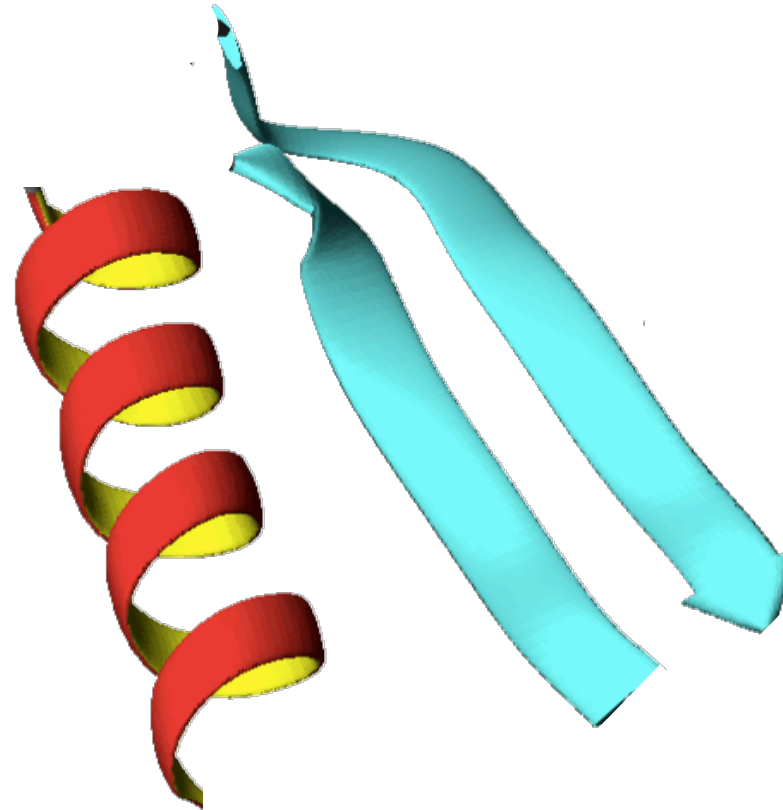
J. Martin, G. Letellier, J.F. Taly, A. Martin, A.G. de Brevern & J.F. Gibrat (*BMC Structural Biology*, 2005)

[http://migale.jouy.inra.fr/mig/mig\\_fr/servlog/kaksi/](http://migale.jouy.inra.fr/mig/mig_fr/servlog/kaksi/)

# Secondary structures: some questions

---

*Moreover .*



# Secondary structures: some questions

---

- It is possible to look at protein structures in a different ways
- At a local level.

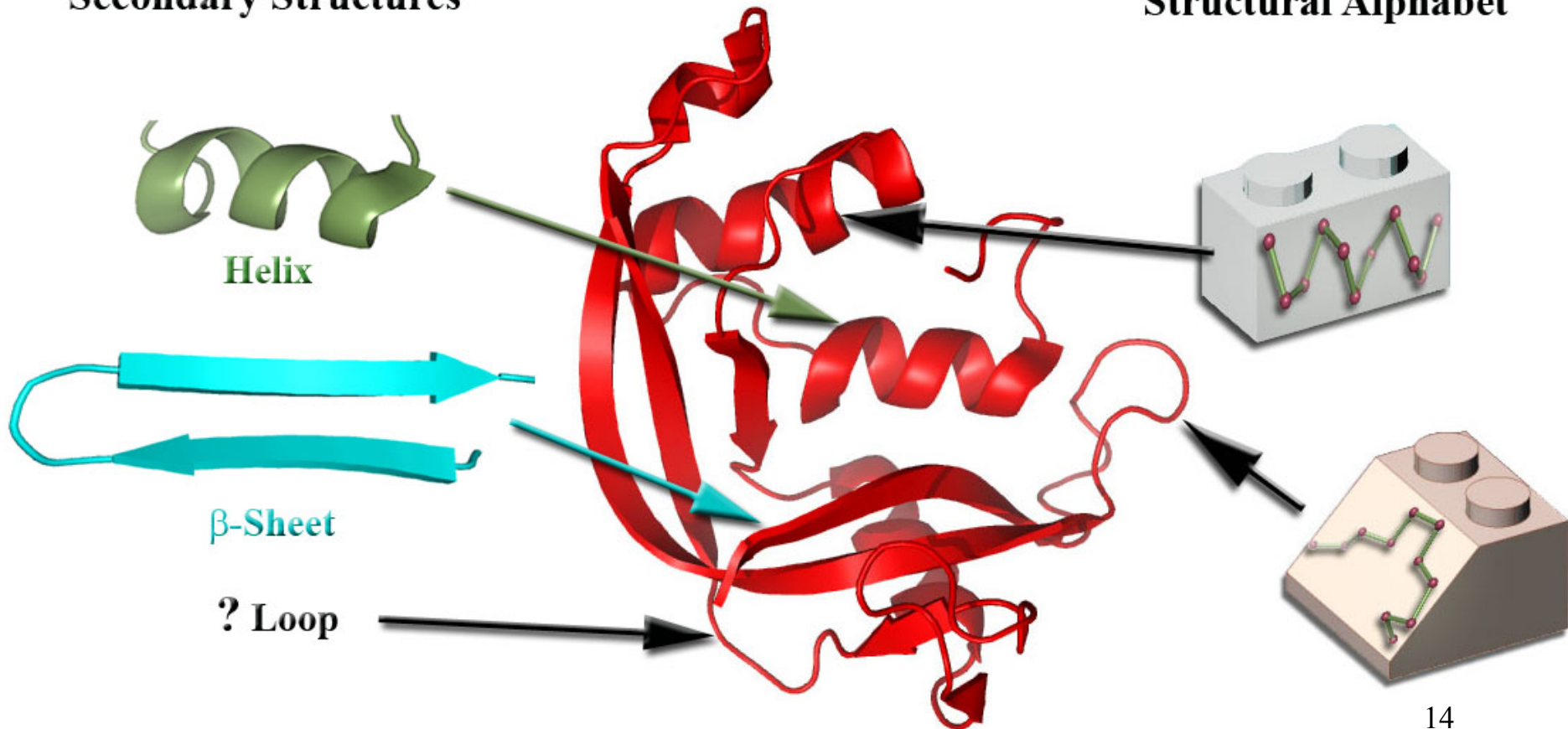
---

# STRUCTURAL ALPHABET

➤ Another vision

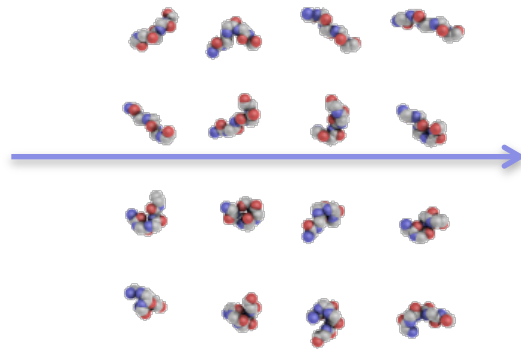
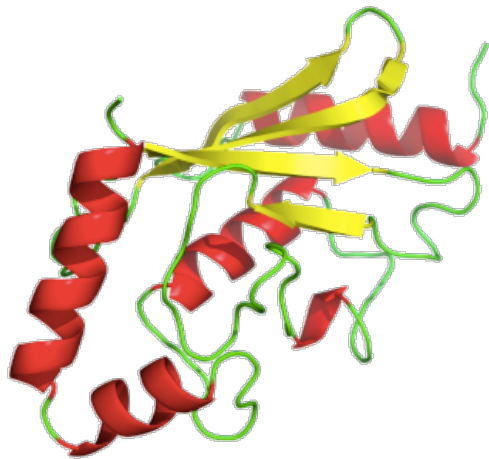
## Secondary Structures

## Structural Alphabet



➤ Structural Alphabet (SA):

Structure translation (3D -> 1D)



>2aak

ZZfk lmmmmmmmmmmmmmmcccehi  
 acdddfk bmkbcc dddddddehia  
 kgopac dddddd fkbckbcc dd  
 ddfbdc ddfkl cfkopac d fklm  
 mmgcfk bcfkl mmmmmmmmmmmmmc  
 c fkl c fbl m l mmmmmmmmmml mmm  
 mmmmmmmmmmmZZ

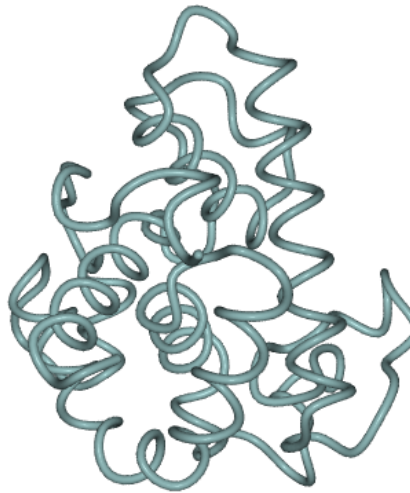
*3D protein structure*

*1D Sequence*

de Brevern A.G., Etchebest C. & Hazout, S. (*Proteins*, 2000).



# Structural alphabet



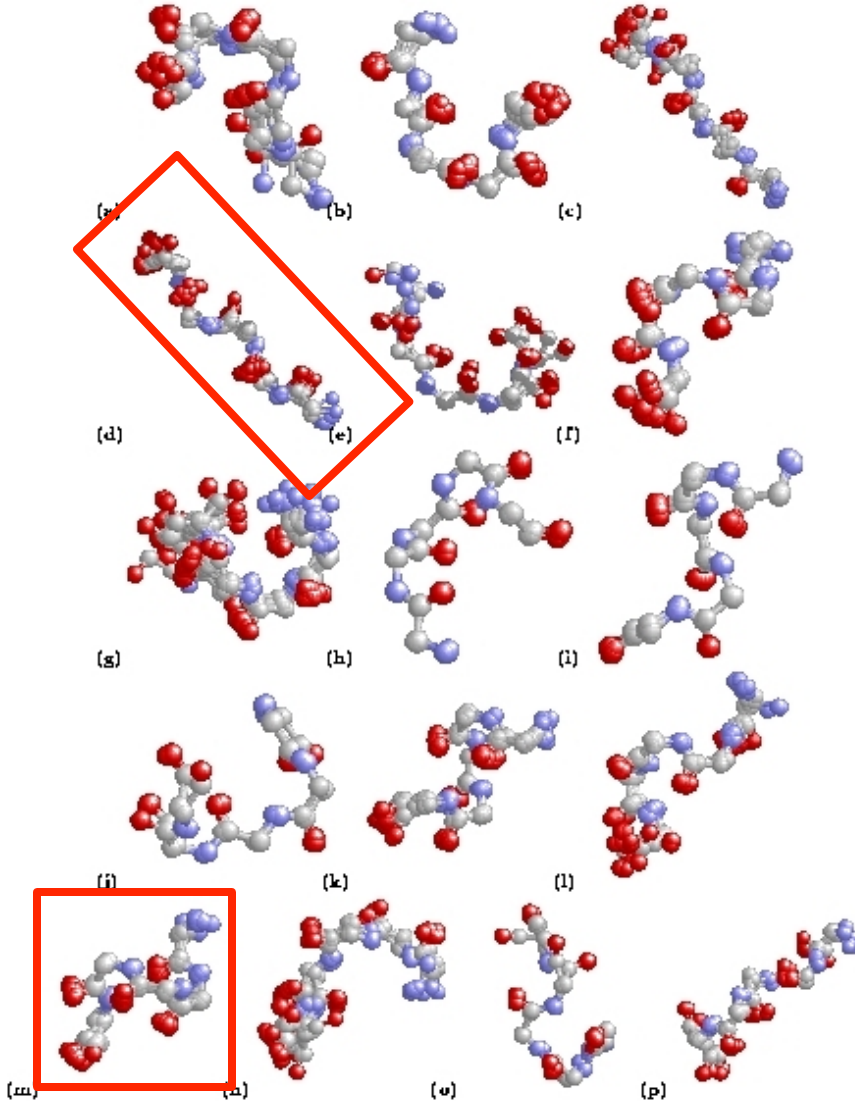
```
>153L [secondary structures]
```

```
HHHH              HHHHHH              HHHHHHHH  
HHHHHHHHHHHHHHHHHHHH              HHHHHHHHHH  
                                         HHHHHHHHHHHHHHHHHH  
HHHHHHHHHH      HHHHHHHHHHHHH      EEEEE  
      HHHHHHHHHHHHHHHHHHH
```

```
>153L [structural alphabet]
```

```
ZZmnopfklpccebjafklmmnopabecjklmmmmmmmmmm  
mmmmmmmmmmmmmmmmmmnopafklmmmmmmmmmmnooolapeh  
iafkopagcjkopafklmccehjfkklmklmmmmmmmmmmmm  
mmmmmmmmmmmbcfklmmmmmmmmmmmmnomklmnbfklmmgoia  
hilmmmmmmmmmmmmmmmmmmnoZZ
```

# Structural alphabet

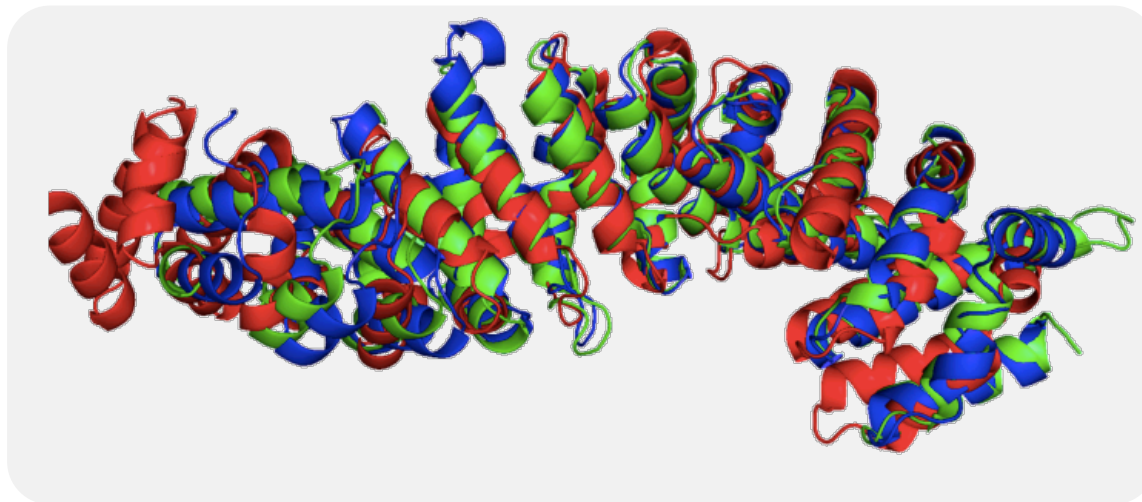


Pr. Serge Hazout

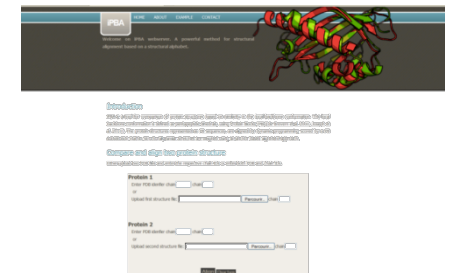
de Brevern A.G., Etchebest C. & Hazout, S. (*Proteins*, 2000).

16 Protein Blocks

## ➤ Superimposition of 3D protein structures

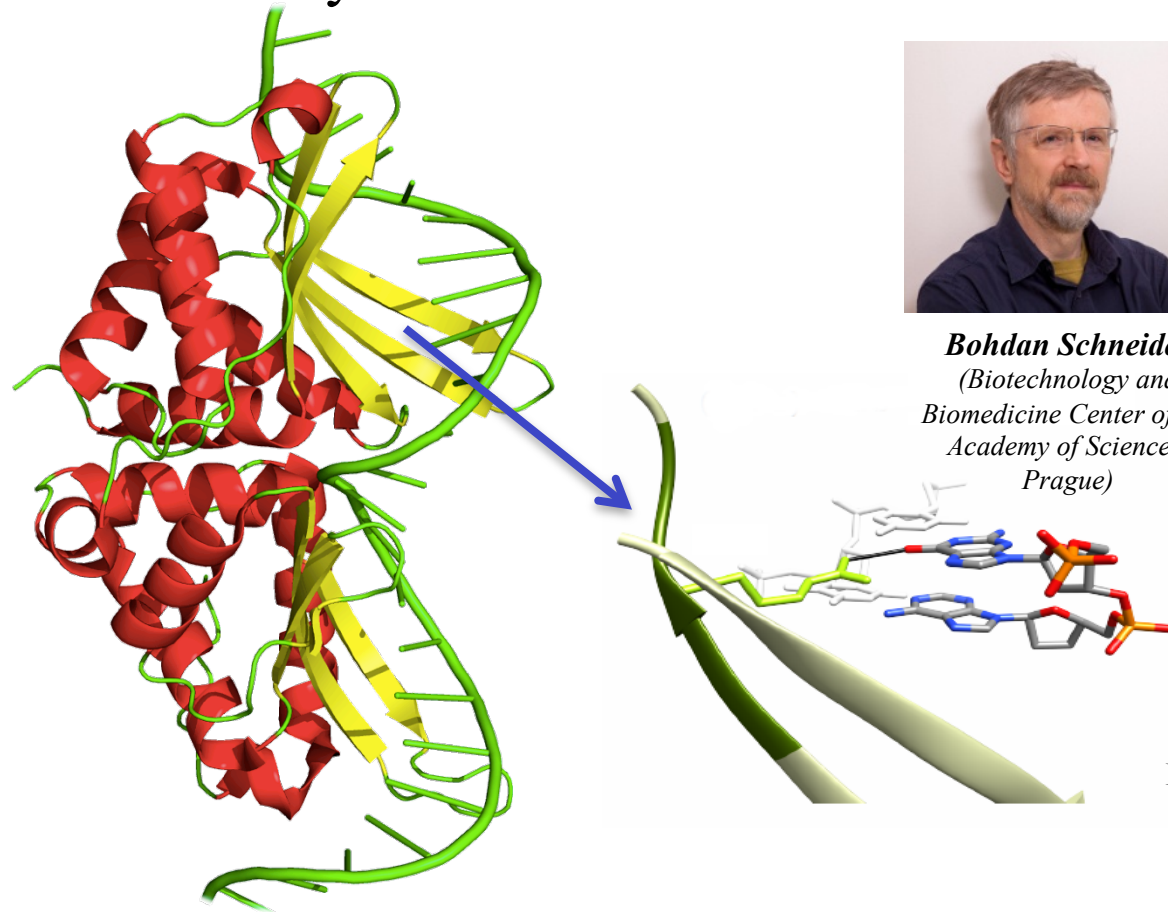


*Naraswami Srinivasan*  
(IISc Bangalore)



- Joseph A.P., Srinivasan N. & de Brevern A.G. (2011) *Biochimie*, **93**(9):1434-45.  
Agarwal G., Mahajan S., Srinivasan N. & de Brevern A.G. (2012) *PLoS ONE*, **6**(3):e17826.  
Gelly J.-C., Joseph A.P., Srinivasan N. & de Brevern A.G. (2011), *Nucleic Acid Res*, **39**:W18-23.  
Joseph A.P., Srinivasan N. & de Brevern A.G. (2012) *Biochimie*, **94**:2025-34.  
Léonard S., Joseph A.P., Srinivasan N., Gelly J.-C. & de Brevern A.G. (2014) *J Biol Struct Dyn*, **32**(4):661-8.

## ➤ Protein interface analyses



**Bohdan Schneider**  
(Biotechnology and  
Biomedicine Center of the  
Academy of Sciences,  
Prague)

PB *d* + Nt 19  
PDB code 3eh8

# Structural alphabet

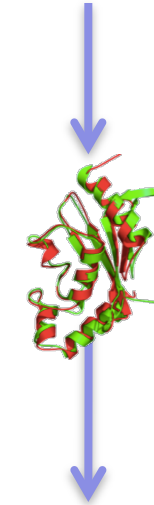
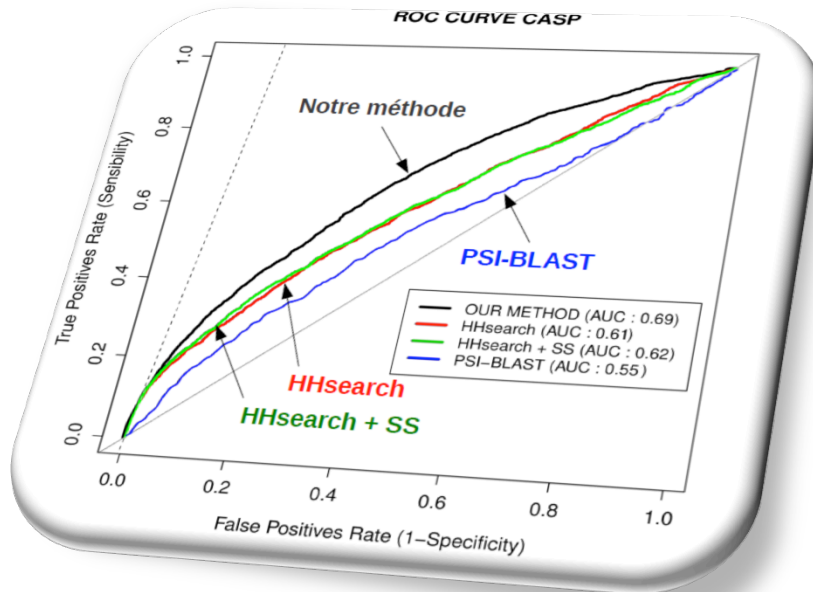
>153L  
 TDCYGNVNRIDTTGASCKTAKPEGL  
 SYCGVSASKKIAERDLQAMDRYKTI  
 IKKVGEKLCVEPAVIAGIISRESHA  
 GKVLKNGWDRGNGFGLMQVDKRSH  
 KPQGTWNGEVHITQGTTLINFIKT  
 IQKKFPSWTKDQQLKGGISAYNAGA  
 GNVRSYARMDIGTTHDDYANDVVAR  
 AQYYKQHGY



*Prediction  
 of PBs*

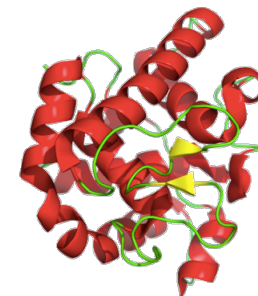


>153L  
 ZZmnopfkllpccebjafklmmnopa  
 becjklmmmmmmmmmmmmmmmmmm  
 mmmmmnopafklmmmmmmmmnoola  
 pgehiafkopagcjkopafklmccch  
 jfklmklmmmmmmmmmmmmmmmmmm  
 mmbcfklmmmmmmmmmmmmmmmmmm  
 klmngoiiahilmmmmmmmmmmmmmm  
 oZZ



*Structural  
 analogy*

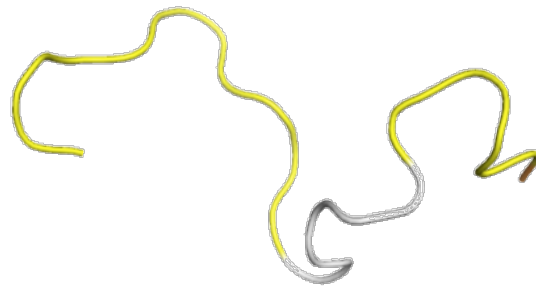
*Structural fold*



*Dr. J.-C. Gelly*

*(Bioinformatics, 2015; Sci Rep, 2016)*

- ❖ The most used structural alphabet at this day (Web Of Sciences: 156 citations, Google scholar: 288)
- ❖ More than 20 teams worldwide
- ❖ Used for local and global analysis, prediction (locale, specific, flexibility), le *protein design*, building of structural models, etc.



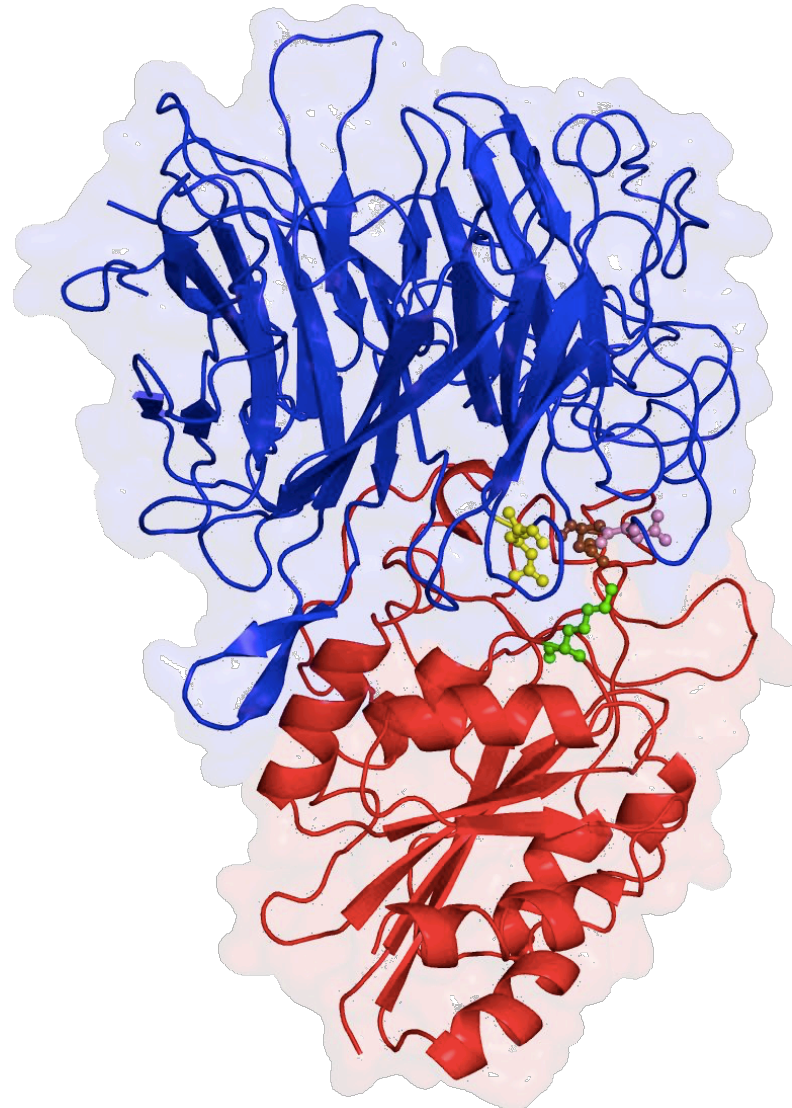
*Analysis of molecular  
dynamics of protein and  
peptide*

---

# STRUCTURAL ALPHABET & FLEXIBILITY: MD

# Integrins

## Integrins



**Dr. Vincent Jallu**  
**Platelet departement**



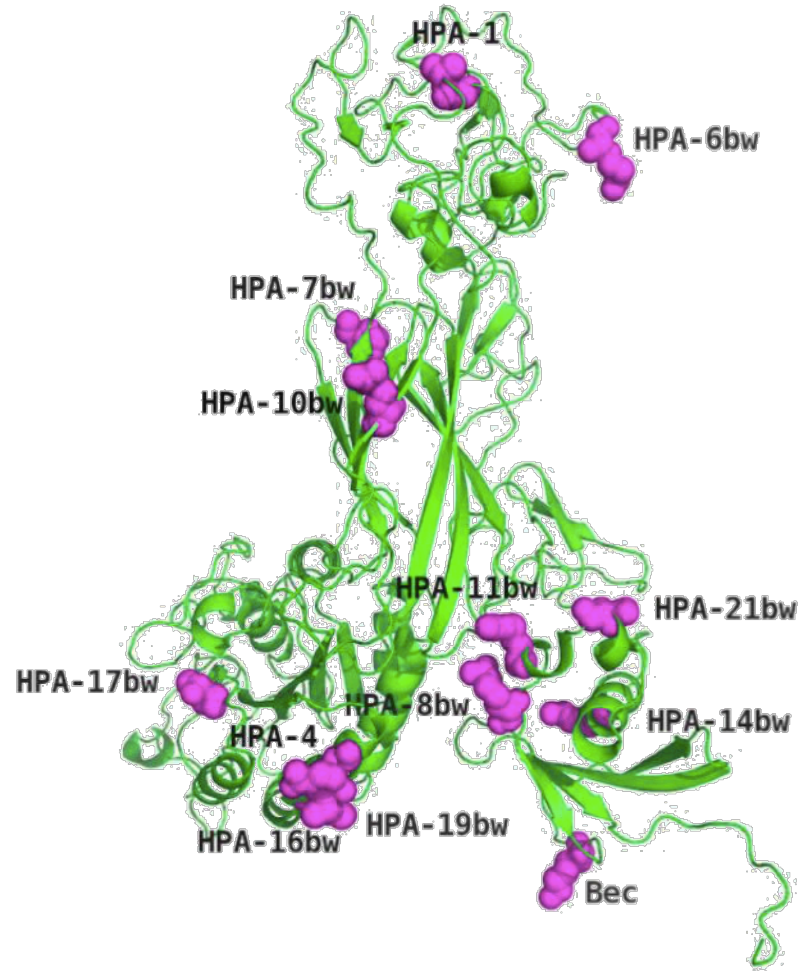


---

## One complex (Integrin):

- **Glanzmann's thrombasthenia** is an abnormality of the platelets, an extremely rare **coagulopathy**.
- **Fetal / Neonatal alloimmune thrombocytopenia (FNAIT)**, a severe bleeding syndrome in which fetal / neonatal platelet destruction is mediated by **maternal antibodies** directed to specific antigens (or alloantigens) inherited from the father  
Also the **only blood group not on Red Blood Cells**.

# Integrins



- This protein is *very* flexible



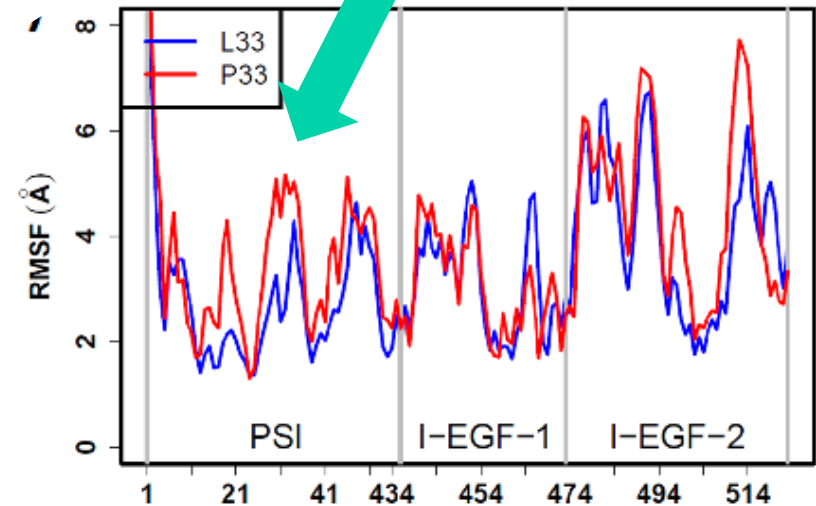
*We compute Root Mean Square  
Fluctuation (RMSF)*

$$RMSF_i = \sqrt{\frac{1}{nstep} \sum_{t=1}^{nstep} \|\vec{r}_i(t) - \langle \vec{r}_i \rangle\|^2}$$

- This protein is *very* flexible



*Residue 33 is associated to high RMSF*



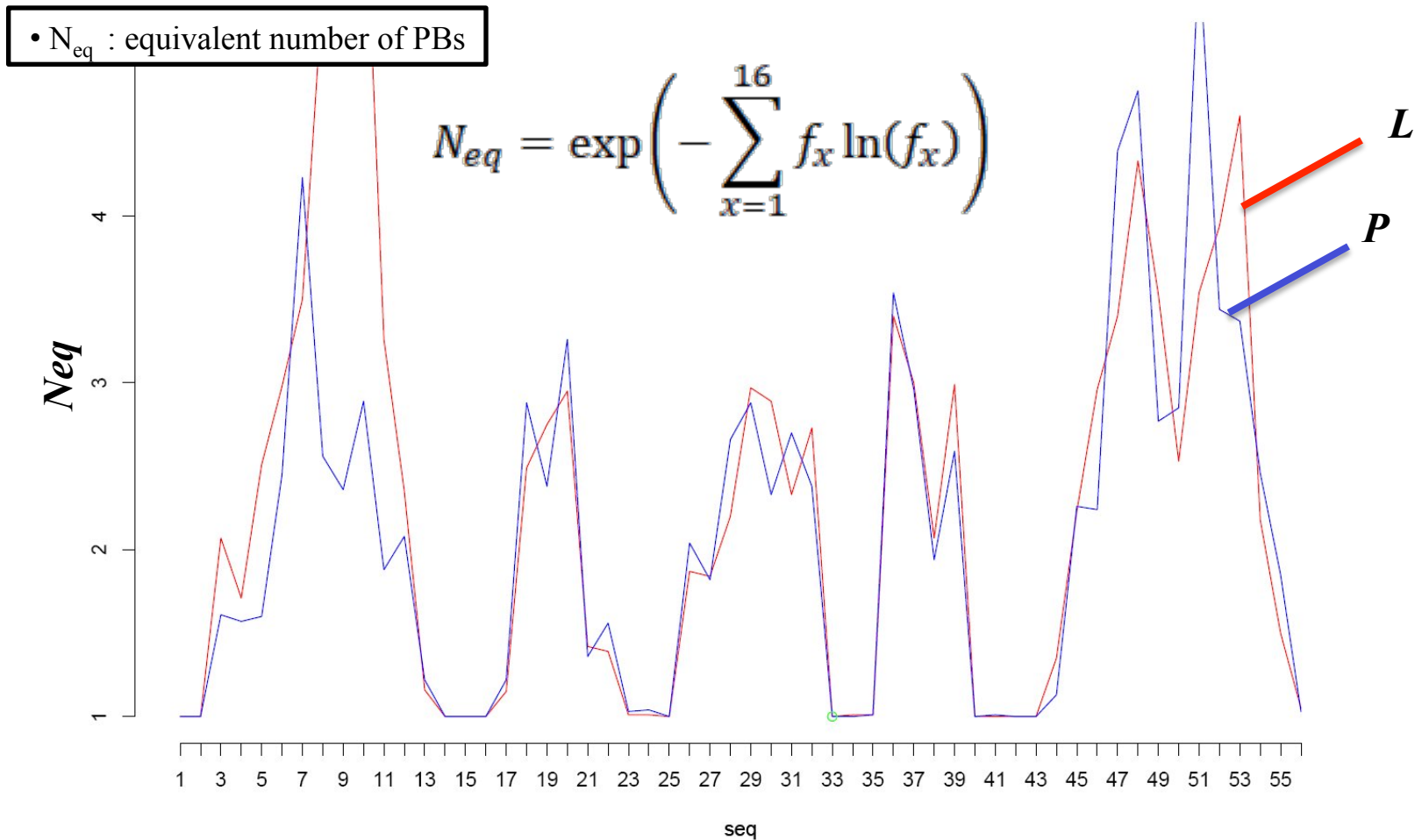
➤ How to use Protein Blocks here ?



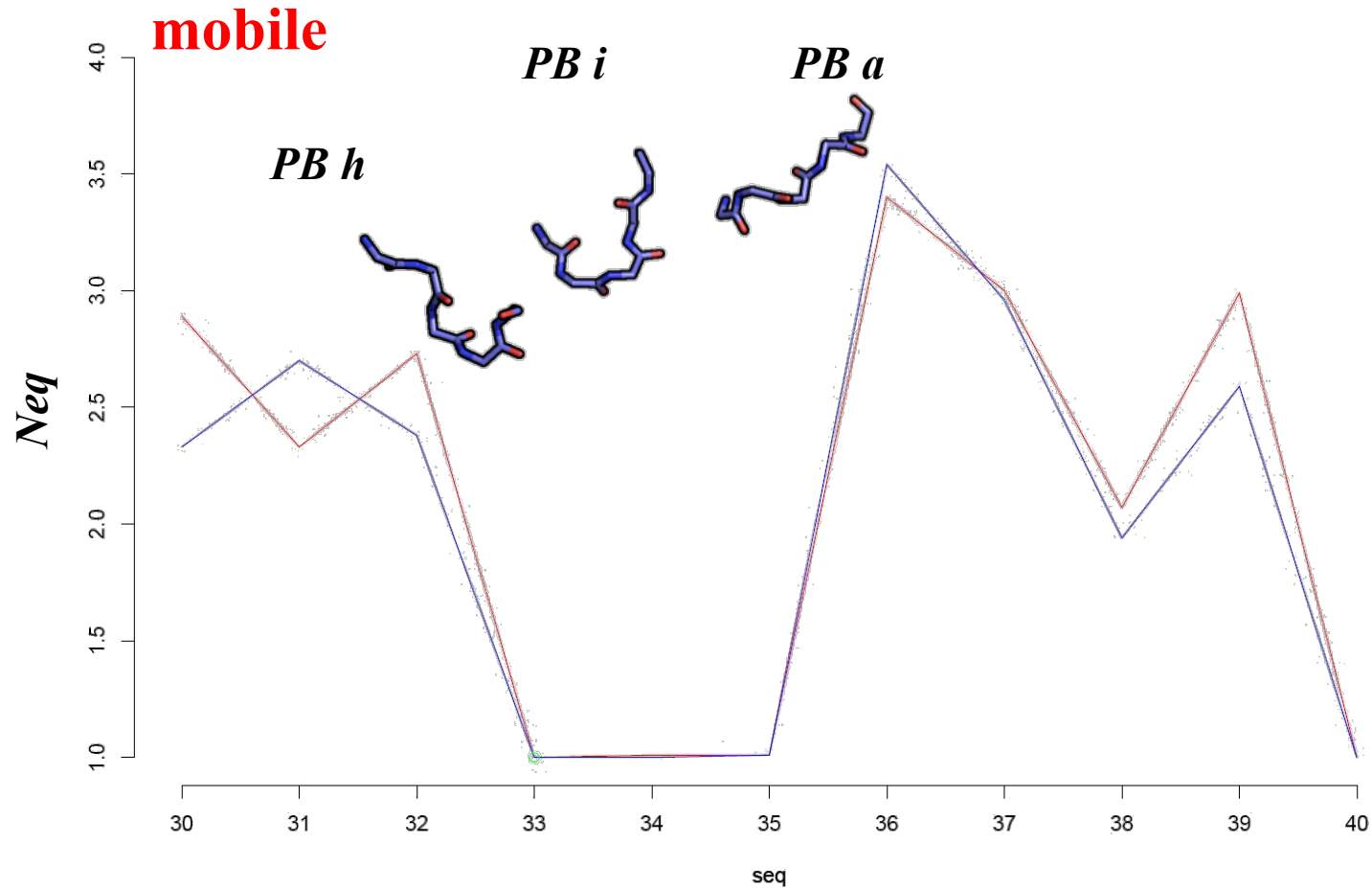
ZZfklmnopabfklmmnnpfkbccdfblkbcchiaeh  
iafklmmnopafklpccddddddfbdcdddfblmb  
ckbccdfbdcfbdcddddehiacddddddddehiac  
dddddddfbfklmmmmmmmmgojlmmmmmmmnpcb  
ccccccccccccdfklpcfbdflmnomclmmbdcdd  
debiadfbdcfbfklmmmmmmmmccdddfblbdcfk  
lmmmmmmommmnopafkbccdddddfbdcdehkl  
mmnopabccehpacbdfkopafblmpccddfkllmm  
mmmmnopacdddbdfklmmmmmmmmmmghiacddd  
fklmmmmmmmmmmmmmmnccccdddehiafkbccdd  
ddfahjbccehiabdehiaghiacddddddddehjd  
fddddddddehiabdcddddddfbdfklmpcfbg  
cknojhiaccehhiacdehiacjkopafgehjlmng  
jiafbfblkbcehiopaccehhiacdfbdchiaaedj  
kopaccblbccfblpehhlpaccehhiacdehiacj  
kopacfbfklmmgfkopghhlpaccehhiacdfbghi  
edjkopaccbccfklmmmmmmmmnopajklmmk  
lmmmomcdcdddfbdcdebiacdcdddddfkopacdd  
dddddfkopacdddfbdcdddZZ



- Both sets of simulations are quite similar



- But position 33 is not 'flexible', it stays highly constraint →

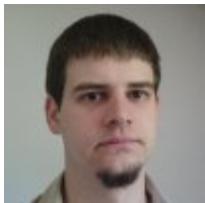




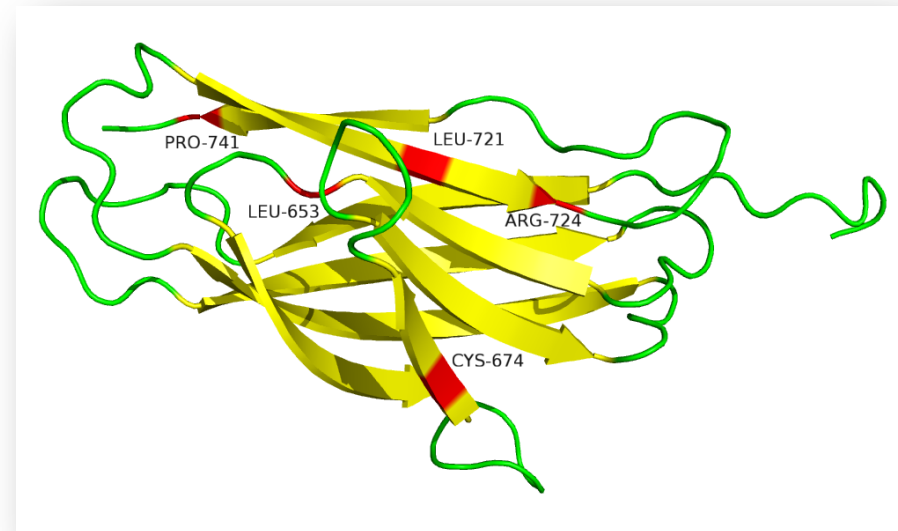
## Conclusion

- Not real change of the epitope, always highly accessible
- But not flexible, it is **mobile** within **deformable regions**
- Proof of interest of **Protein Blocks** for the analysis => does not correlate with RMSf

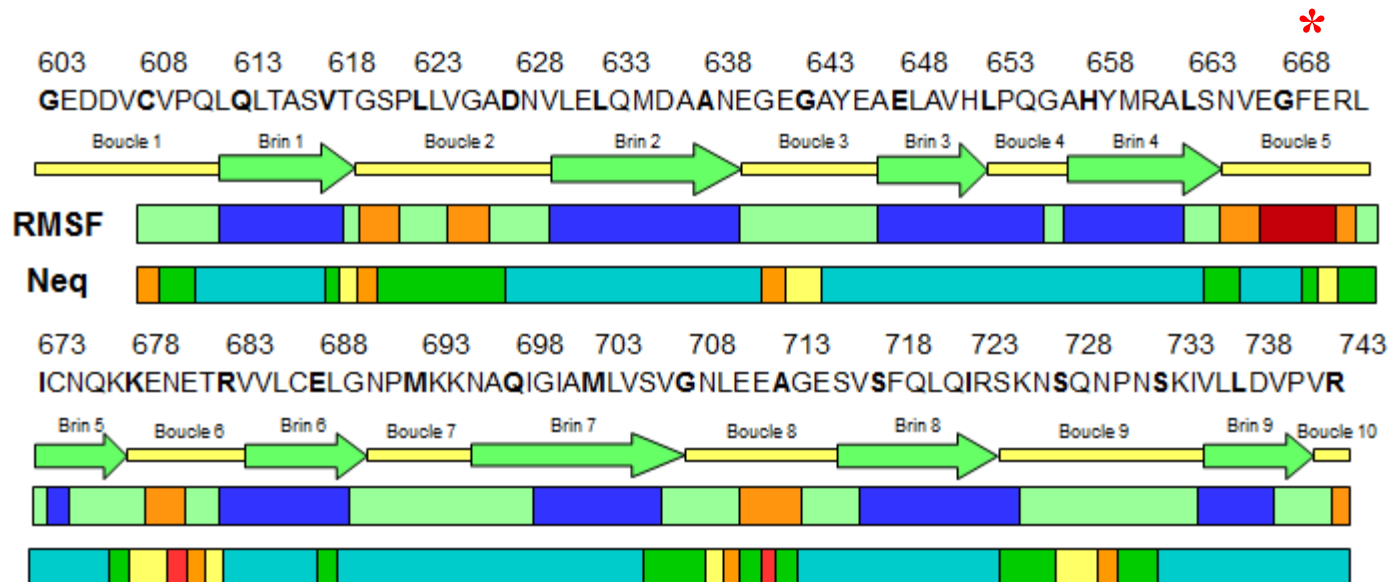
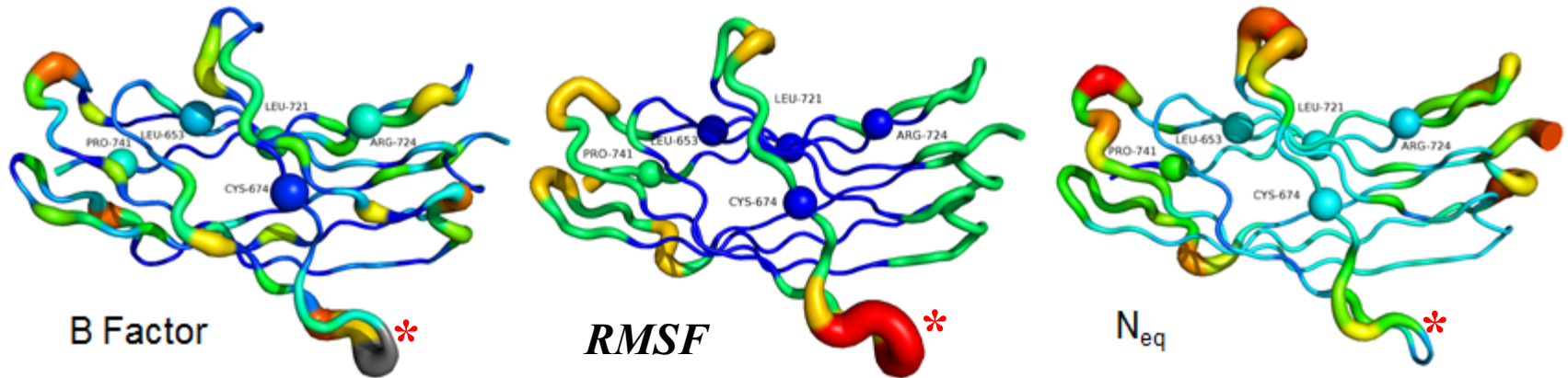
- A more systematic approach: an example of Calf-1 domain
- Calf-1 domain : 141 residus (603-743) from  $\alpha$  chain
- 7 mutations involved in GT:



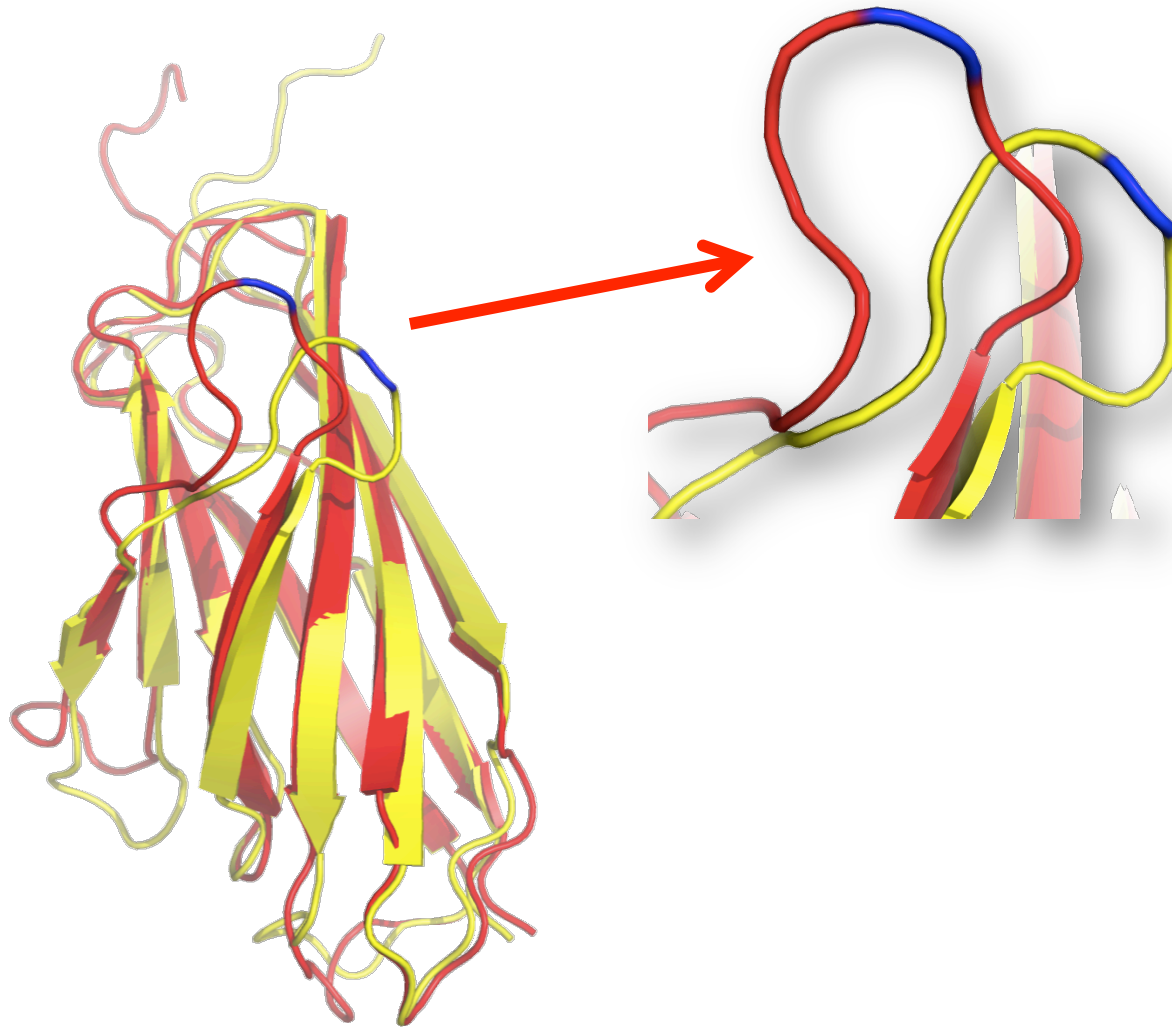
- L653R  
- C674R  
- L721V / L721R  
*Matthieu Goguet* R724P / R724Q  
(master 2 student) P741R



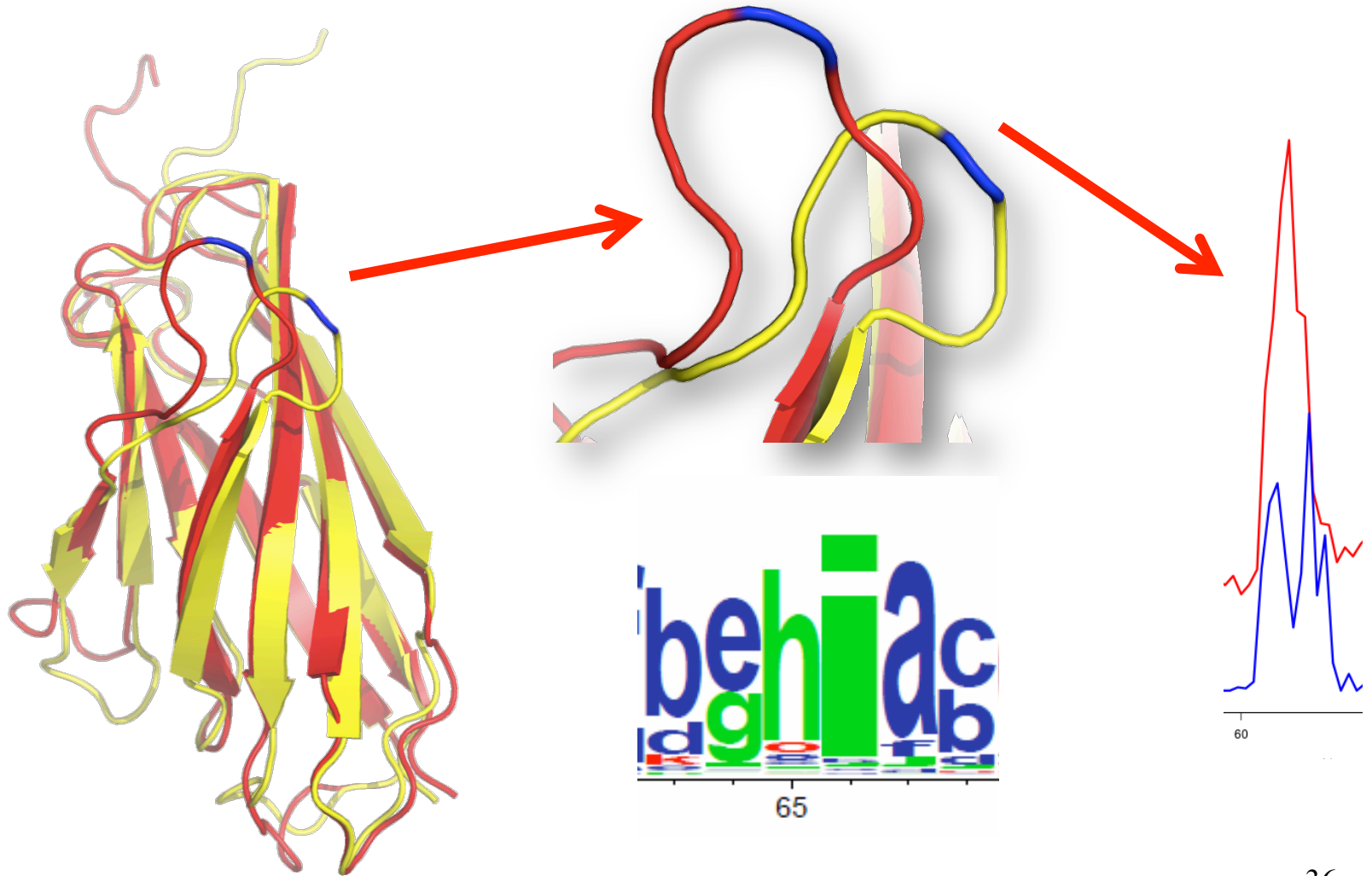
➤ Analysis of WT Calf-1



# Allostery



# Allostery



➤ More details in

Craveur *et al.*, (*Frontiers in Mol Sci*, 2015)



## Protein flexibility in the light of structural alphabets

Pierrick Craveur<sup>1,2,3,4</sup>, Agnel P. Joseph<sup>5</sup>, Jeremy Esque<sup>6</sup>, Tarun J. Narwani<sup>1,2,3,4</sup>, Floriane Noël<sup>1,2,3,4</sup>, Nicolas Shinada<sup>1,2,3,4</sup>, Matthieu Goguet<sup>1,2,3,4</sup>, Sylvain Leonard<sup>1,2,3,4</sup>, Pierre Poulain<sup>1,2,3,4,7</sup>, Olivier Bertrand<sup>1,3,4</sup>, Guilhem Faure<sup>8</sup>, Joseph Rebehmed<sup>9</sup>, Amine Ghozlane<sup>10</sup>, Lakshmpuram S. Swapna<sup>11,12</sup>, Ramachandra M. Bhaskara<sup>11,13</sup>, Jonathan Barnoud<sup>1,2,3,4,14</sup>, Stéphane Teletchéa<sup>1,2,3,4,15</sup>, Vincent Jallu<sup>16</sup>, Jiri Cerny<sup>17</sup>, Bohdan Schneider<sup>17</sup>, Catherine Etchebest<sup>1,2,3,4</sup>, Narayanaswamy Srinivasan<sup>11</sup>, Jean-Christophe Gelly<sup>1,2,3,4</sup> and Alexandre G. de Brevem<sup>1,2,3,4\*</sup>

<sup>1</sup> Institut National de la Santé et de la Recherche Médicale U 1134, Paris, France, <sup>2</sup> UMR\_S 1134, DSIME, Université Paris Diderot, Sorbonne Paris Cité, Paris, France, <sup>3</sup> Institut National de la Transfusion Sanguine, DSIME, Paris, France, <sup>4</sup> UMR\_S 1134, DSIME, Laboratory of Excellence GR-Ex, Paris, France, <sup>5</sup> Rutherford Appleton Laboratory, Science and Technology Facilities Council, Didcot, UK, <sup>6</sup> Institut National de la Santé et de la Recherche Médicale U964,7 UMR Centre National de la Recherche Scientifique 7104, IGBMC, Université de Strasbourg, Illkirch, France, <sup>7</sup> Ets Poulain, Pointe-Noire, Congo, <sup>8</sup> National Library of Medicine, National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA, <sup>9</sup> Centre National de la Recherche Scientifique UMR7590, Sorbonne Universités, Université Pierre et Marie Curie – MIA-N – IAD – IJC, Paris, France, <sup>10</sup> Metagenopols, INRA, Jouy-en-Josas, France, <sup>11</sup> Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Bangalore, India, <sup>12</sup> Hospital for Sick Children, and Departments of Biochemistry and Molecular Genetics, University of Toronto, Toronto, ON, Canada, <sup>13</sup> Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt, Germany, <sup>14</sup> Laboratoire de Physique, École Normale Supérieure de Lyon, Université de Lyon, Centre National de la Recherche Scientifique UMR 5672, Lyon, France, <sup>15</sup> Faculté des Sciences et Techniques, Université de Nantes, Unité Fonctionnelle et Ingénierie des Protéines, Centre National de la Recherche Scientifique UMR 6286, Université Nantes, Nantes, France, <sup>16</sup> Platelet Unit, Institut National de la Transfusion Sanguine, Paris, France, <sup>17</sup> Institute of Biotechnology, The Czech Academy of Sciences, Prague, Czech Republic

### OPEN ACCESS

#### Edited by:

Kris Pauwels,  
Vrije Universiteit Brussel, Belgium

#### Reviewed by:

Mark PIMH,  
King's College London, UK  
Elena Papaleo,  
University of Copenhagen, Denmark

#### \*Correspondence:

Alexandre G. de Brevem,  
Institut National de la Santé et de la  
Recherche Médicale U 1134, 6 Rue  
Alexandre Cabanel, 75015 Paris,  
France  
alexandre.dibrevem@  
univ-paris-diderot.fr

#### Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
*Frontiers in Molecular Biosciences*

Received: 28 February 2015

Accepted: 30 April 2015

Published: 27 May 2015

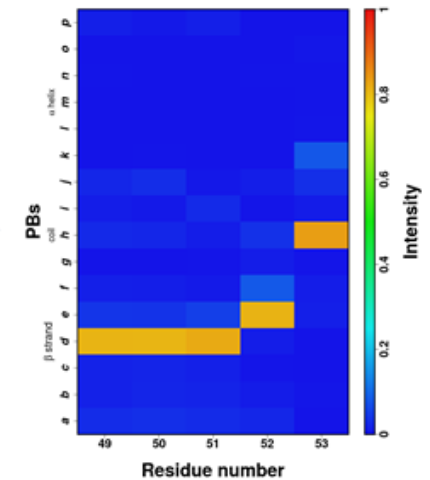
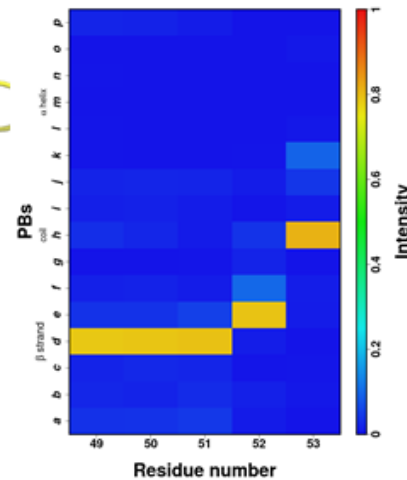
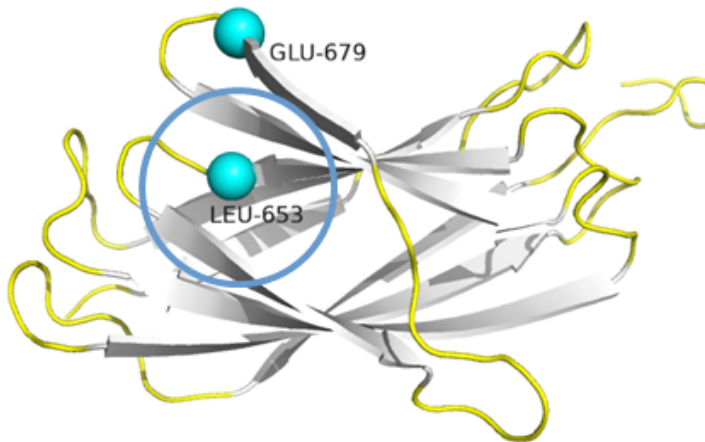
#### Citation:

Craveur P, Joseph AP, Esque J,  
Narwani TJ, Noël F, Shinada N,  
Goguet M, Leonard S, Poulain P,  
Bertrand O, Faure G, Rebehmed J,  
Ghozlane A, Swapna LS, Bhaskara  
RM, Barnoud J, Teletchéa S, Jallu V,  
Cerny J, Schneider B, Etchebest C,  
Srinivasan N, Gelly J-C and de  
Brevem AG (2015) Protein flexibility in  
the light of structural alphabets.  
*Front. Mol. Biosci.* 2:20.  
doi: 10.3389/fmolb.2015.00020

Protein structures are valuable tools to understand protein function. Nonetheless, proteins are often considered as rigid macromolecules while their structures exhibit specific flexibility, which is essential to complete their functions. Analyses of protein structures and dynamics are often performed with a simplified three-state description, i.e., the classical secondary structures. More precise and complete description of protein backbone conformation can be obtained using libraries of small protein fragments that are able to approximate every part of protein structures. These libraries, called structural alphabets (SAs), have been widely used in structure analysis field, from definition of ligand binding sites to superimposition of protein structures. SAs are also well suited to analyze the dynamics of protein structures. Here, we review innovative approaches that investigate protein flexibility based on SAs description. Coupled to various sources of experimental data (e.g., B-factor) and computational methodology (e.g., Molecular Dynamic simulation), SAs turn out to be powerful tools to analyze protein dynamics, e.g., to examine allosteric mechanisms in large set of structures in complexes, to identify order/disorder transition. SAs were also shown to be quite efficient to predict protein flexibility from amino-acid sequence. Finally, in this review, we exemplify the interest of SAs for studying flexibility with different cases of proteins implicated in pathologies and diseases.

**Keywords: protein structures, disorder, secondary structure, structural alphabet, protein folding, allostery, protein complexes, protein–DNA interactions**

## Variant L653R : Mutation site



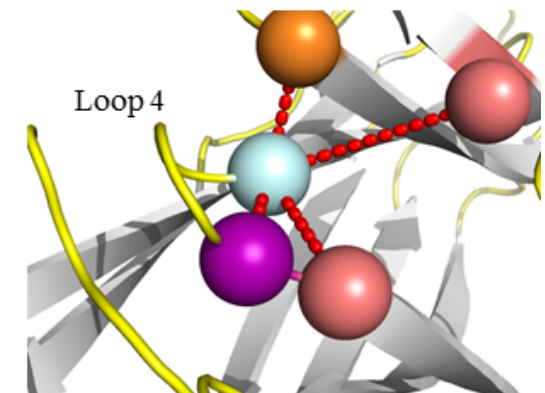
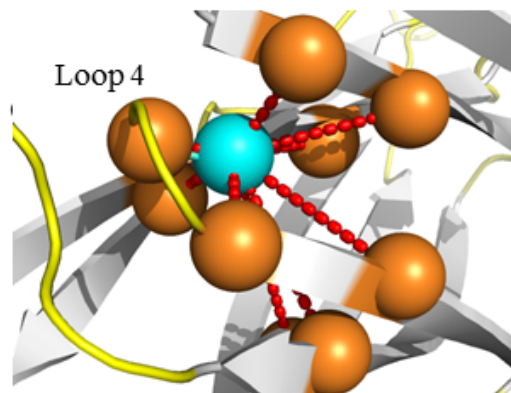
**WT**

**Variant**

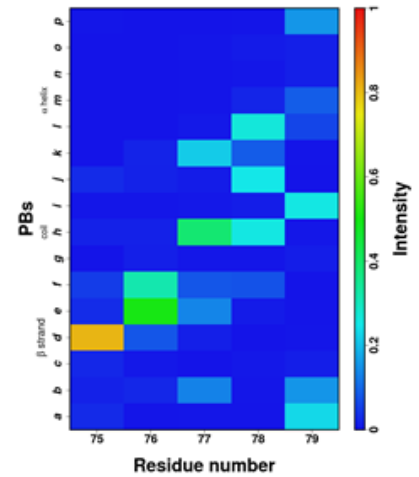
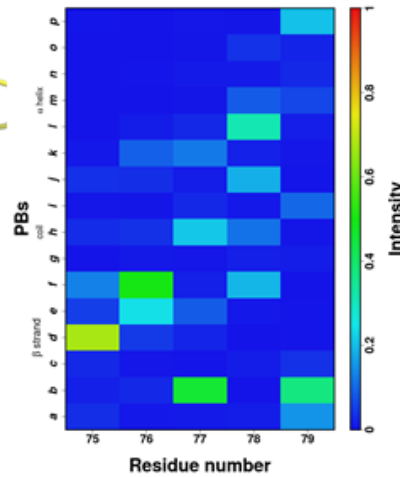
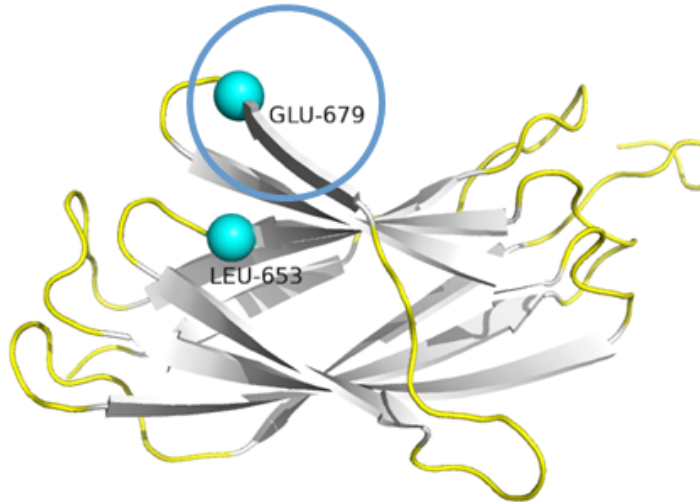
WT : 9 interactions

Variant : only 1 conserved but 3 news to

→ No structural changes



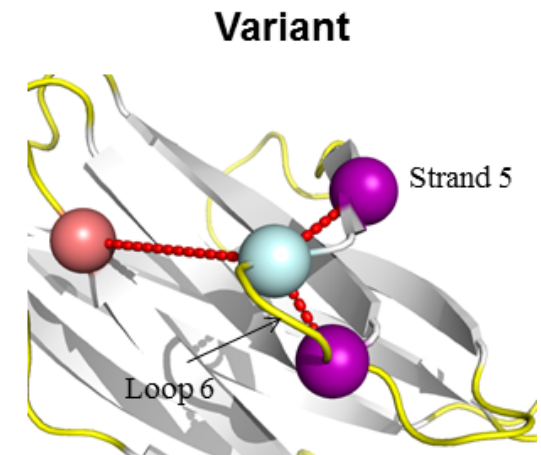
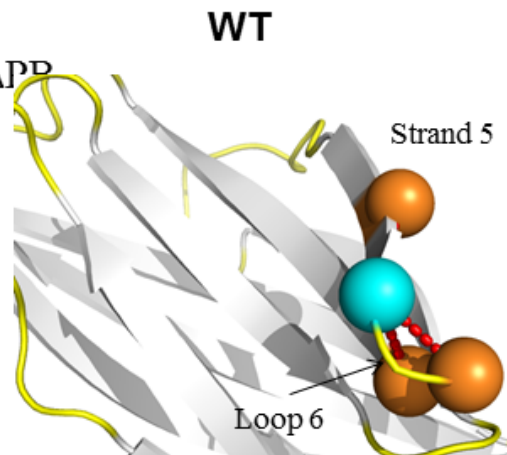
## Variant L653R : position E679



$N_{eq}$  similar in WT and variant but high  $\Delta PP$

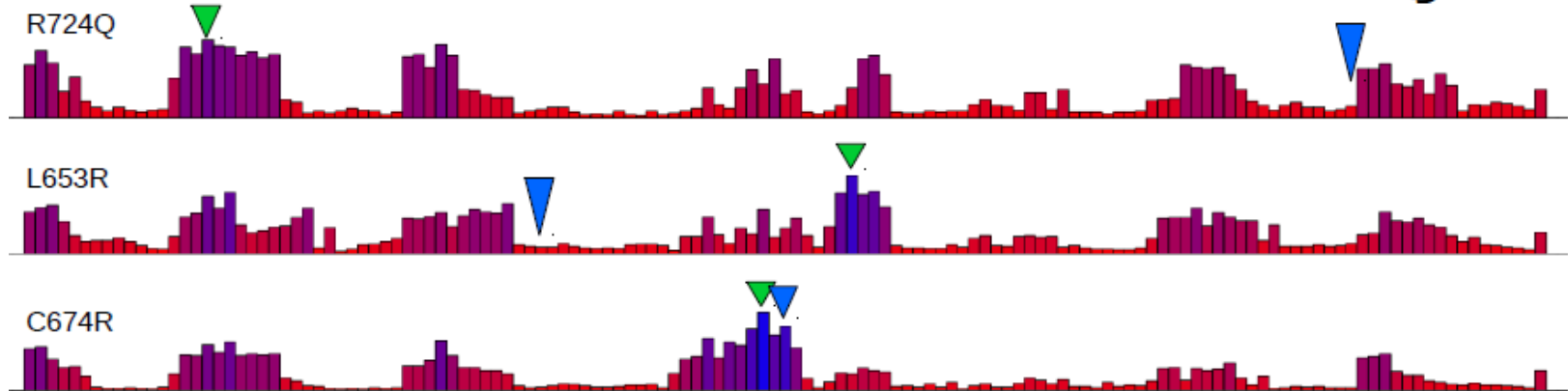
Interaction reshuffle (no conservation)

→ Structural desorganization

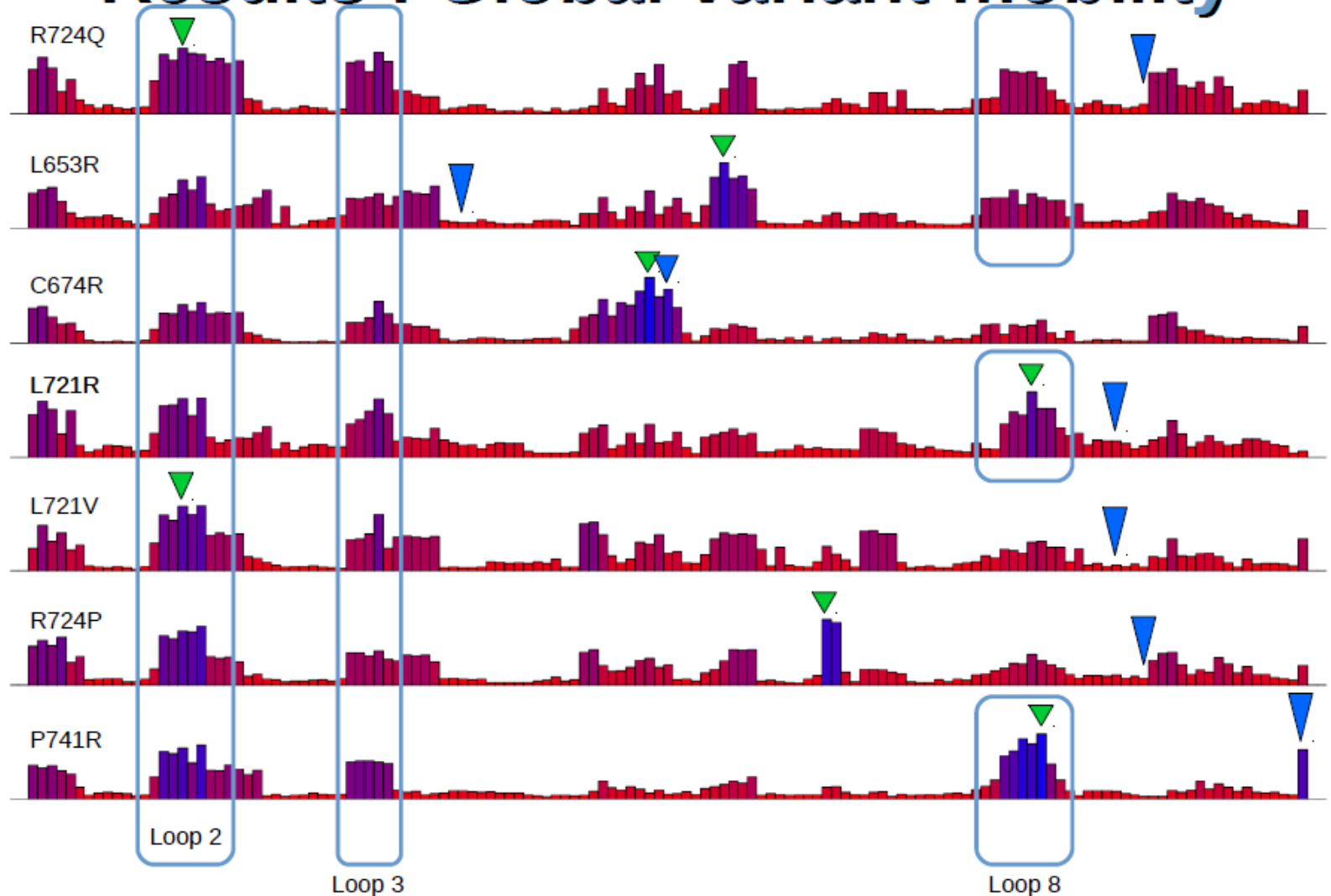




## Results : Global variant mobility



## Results : Global variant mobility



# Conclusion (1)

- 
- Reliable, automated and efficient molecular dynamics protocol with interesting use of Protein Blocks
  - Results on Calf-1 : no changes on mutation spots but structural local impact observed at another place.
  - Does these allosteric effects are common to all of these variants ?
  - Future: Study of this and others integrins

- Last developments: deposit on ArXiv, now published in PeerJ

<https://github.com/pierrepo/PBxplore>

GitHub This repository Search Explore Features Enterprise Blog Sign up Sign in

pierrepo / PBxplore Watch 5 Star 6 Fork 6

A program to explore protein structures with Protein Blocks

312 commits 3 branches 0 releases 3 contributors

Branch: master PBxplore / +

Improved regression tests, demo and doc for PBclust.py

demo1 added more test files for PDBx/mmCIF format 7 months ago

demo2 Number of cluster is now a mandatory option in PBclust.py. 3 months ago

demo\_paper Added images in demo\_paper 5 months ago

doc Improved regression tests, demo and doc for PBclust.py 26 days ago

test\_data Uncomment the test for PBcount with negative shift 3 months ago

.coveragerc added .coveragerc for travis 6 months ago

.gitignore Ignore png file only in the root directory 5 months ago

.travis.yml Make travis-CI install R 3 months ago

LICENSE Fixed #12 - added a more permissive license (MIT) 11 months ago

PBassign.py Remove the PB\_assign function 2 months ago

PBclust.py -clusters and -compare options are now mutually exclusive 26 days ago

PBcount.py Fixed #47 -first-residue option of PBcount.py can be < 0. 3 months ago

PBlib.py Simplify PBassign command line function 2 months ago

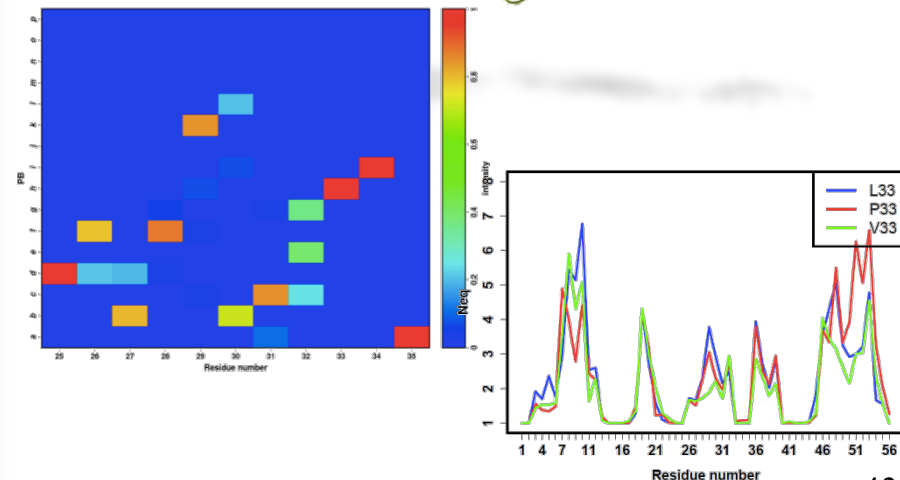
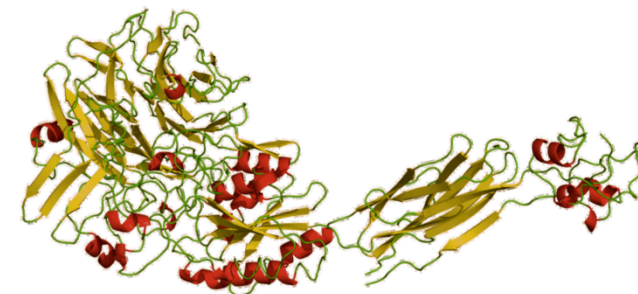
PBs\_substitution\_matrix.dat New tool to cluster sequences based on PBs similarities 2 years ago

PBstat.py Merge pull request #56 from jbamoud/module\_clust 3 months ago

PDBlib.py Improve 'PBassign' modularity 3 months ago

README.md Beginning for issue #30. More to add. 5 months ago

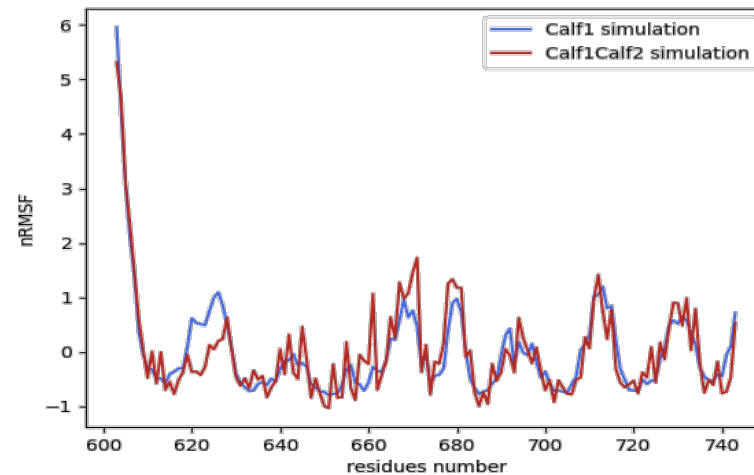
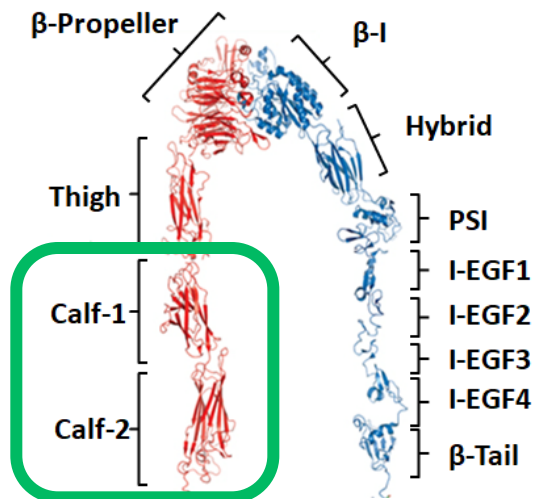
dev\_requirements.txt Adapt the tests for when MDAnalysis is missing 5 months ago



## Calf-1 + Calf- 2

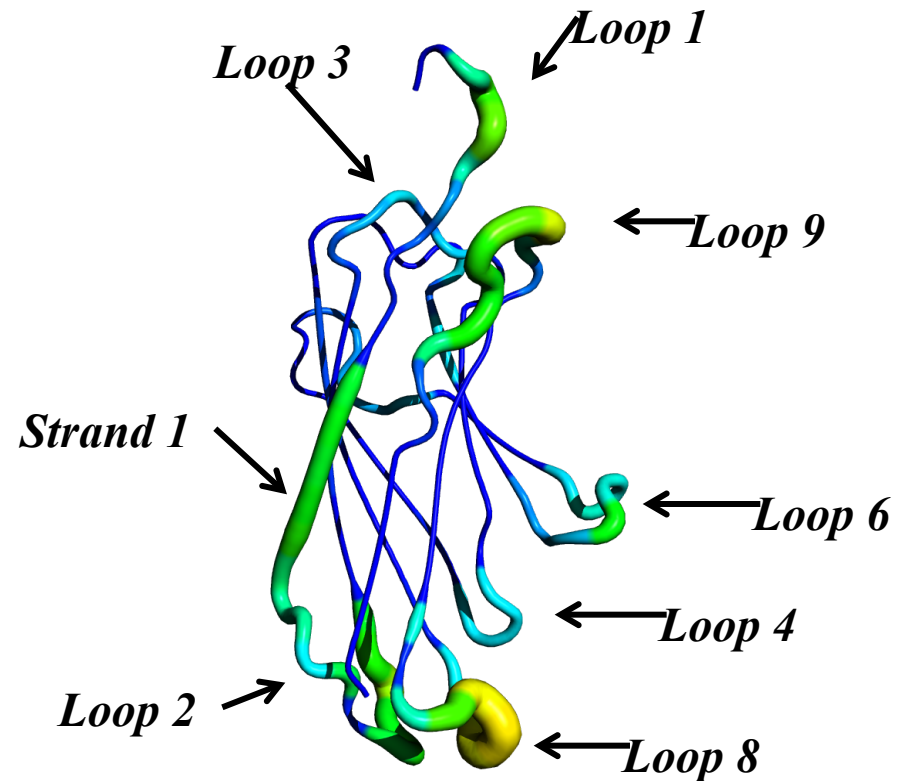
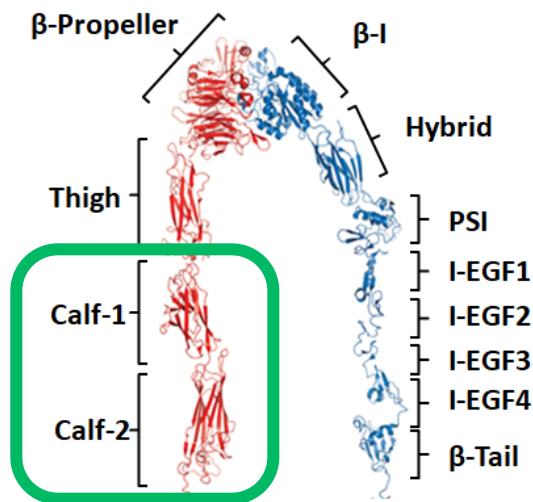
- No impact of the forcefield used
- No impact on the domains cutting

*Soubika Bisoo*  
(master 2 student)



## Calf-1 + Calf-2

- BUT allosteric effect of Calf-2 variant on Calf-1 domain !



*Calf-1*: ΔPB

- 
- Secondary structures are useful tools, but they have their own limitations
  - Other complementary approaches are interesting, i.e., a simple structural alphabet as the Protein Blocks
  - They can help to analyze molecular dynamics, i.e., you can see what you can see without.

- Thanks to last Belbi 2016, I am now really more interested into this question

The screenshot shows the top navigation bar with links for PED, Home, Browse, Help, and Contact. Below the navigation bar, the main heading reads "PED<sup>3</sup>: Protein Ensemble Database" with the subtitle "The database of conformational ensembles describing flexible proteins". A search bar is present with the text "Select experimental data type: NMR SAXS BOTH ANY" and a "Search" button.

## Welcome to the Protein Ensemble Database!

### What is an IDP?

Structural heterogeneity is intimate to the functioning of many proteins and thus describing a protein with a single native structure is often insufficient to elucidate its function. In particular, intrinsically disordered proteins (IDPs) can only be approached by solution techniques and described as structural ensembles. The same is true for multidomain proteins that have disordered linkers. Apparently, the ensemble representations of these proteins carry essential function-related information, yet they have not been available until now.

[Read more about IDPs...](#)

### What is an ensemble?

The goal of PED is to serve as an openly accessible database for the deposition of structural information on IDP- and denatured protein ensembles based on Nuclear Magnetic Resonance (NMR) and Small-angle X-ray Scattering (SAXS) data. We are also hosting purely computational models, typically from Molecular Dynamics (MD) simulations. The deposition of structural coordinates as well as primary data can be used for evaluating and re-calculating the ensembles, thus supporting the evolution of new modeling methods leading to much improved skills of connecting "unstructure" Å with function.

[Read more about ensembles...](#)

### Latest news

#### Tweets by @proteinEnsemble



**PED**  
@proteinEnsemble

The Protein Ensemble Database (PED) is going under a

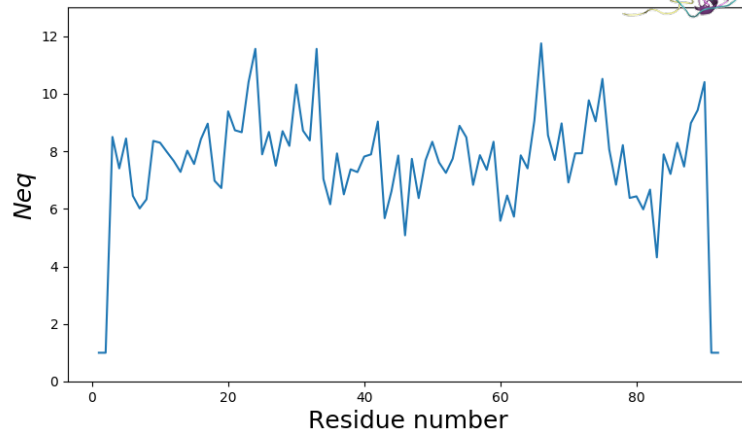
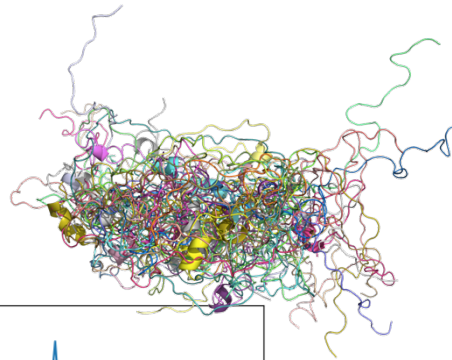
Data submission is temporarily suspended until further notice

PED stores 25473 protein structures of 60 ensembles in 24 entries as of 06/17/2018 04:43:16 pm

For citing PED, please refer to: Mihaly Varadi, Simone Kosol, Pierre Lebrun, Erica Valentini, Martin Blackledge, A. Keith Dunker, Isabella C. Felli, Julie D. Forman-Kay, Richard W. Kriwacki, Roberta Pierattelli, Joel Sussman, Dmitri I. Svergun, Vladimir N. Uversky, Michele Vendruscolo, David Wishart, Peter E. Wright and Peter Tompa "pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins" *Nucleic Acids Res.* 2014, Jan (Database issue)

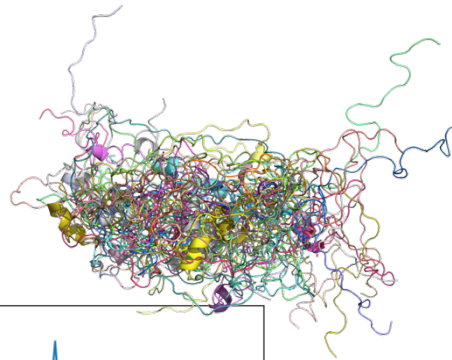


- Thanks to last Belbi 2016, I am now really more interested into this question

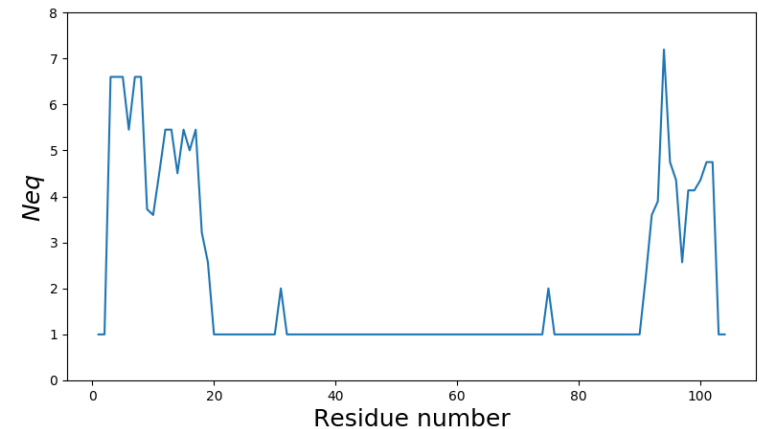
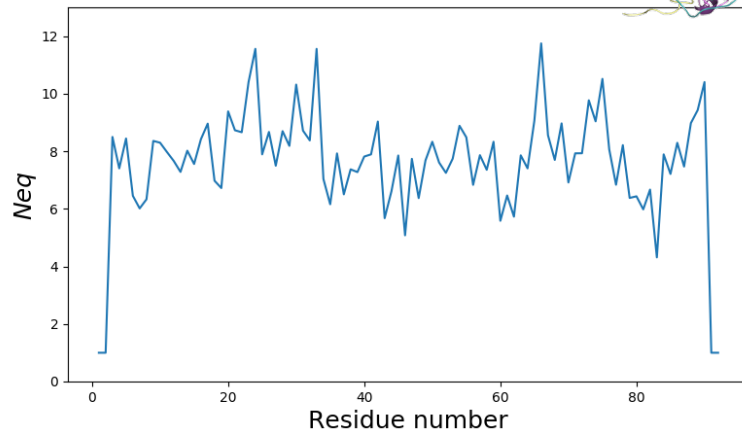


*Yes, real disorder everywhere*

- Thanks to last Belbi 2016, I am now really more interested into this question



*Not really*



- PBs can be a good tool for that... needs more works

*Rigid* → *flexible* → *disorder*

*Rigid / mobile* → *flexible/deformable* → *disorder*

# Acknowledgments

---

**S. Hazout**

**T.J. Narwani**

**M. Goguet**

**C. Etchebest**

S. Sali

S. Kaddhar

P. Poulain

S. Bisoo

P. Fuchs

N. Shinada

**V. Jallu**

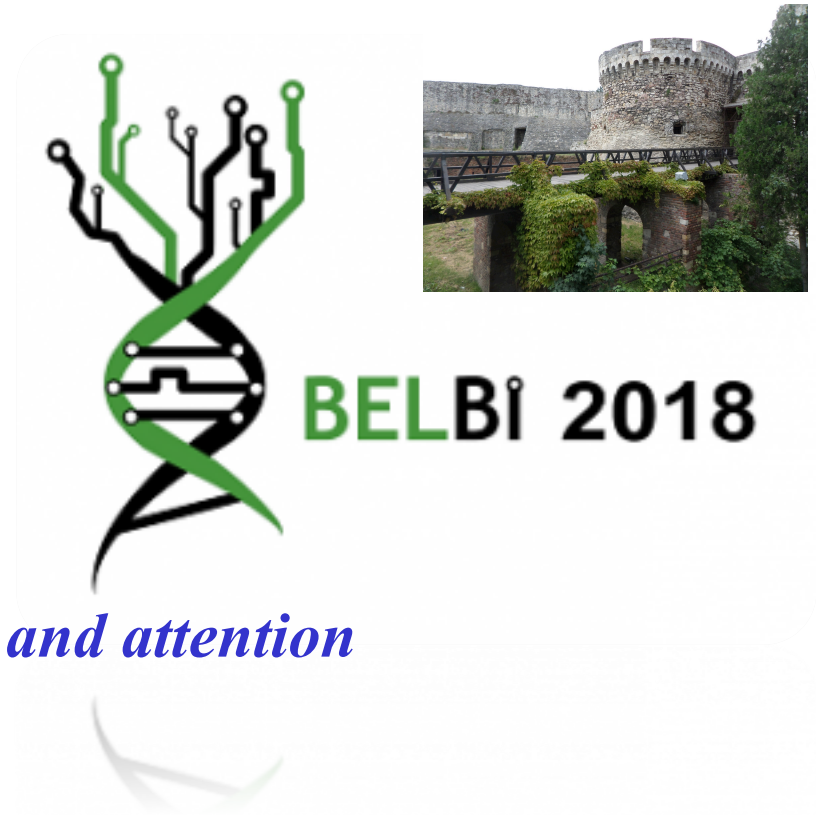
**N. Srinivasan**

C. Kaplan

**B. Schneider**

R. Petermann

**N. Mitic**



*Thank you for your time and attention*