



**HAL**  
open science

# Protein structure analysis, prediction and flexibility in the light of a structural alphabet

Alexandre de Brevern

► **To cite this version:**

Alexandre de Brevern. Protein structure analysis, prediction and flexibility in the light of a structural alphabet. 1ère réunion du GT MASIM du GDR BIM, GDR BIM (CNRS), Nov 2017, Paris, France. inserm-01636581

**HAL Id: inserm-01636581**

**<https://inserm.hal.science/inserm-01636581>**

Submitted on 16 Nov 2017

**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.

Copyright

GT MASIM

Méthodes Algorithmiques pour les  
Structures et Interactions Macromolé-  
culaires

GDR  
BiM

SIMB  
Bioinformatique & Interactions

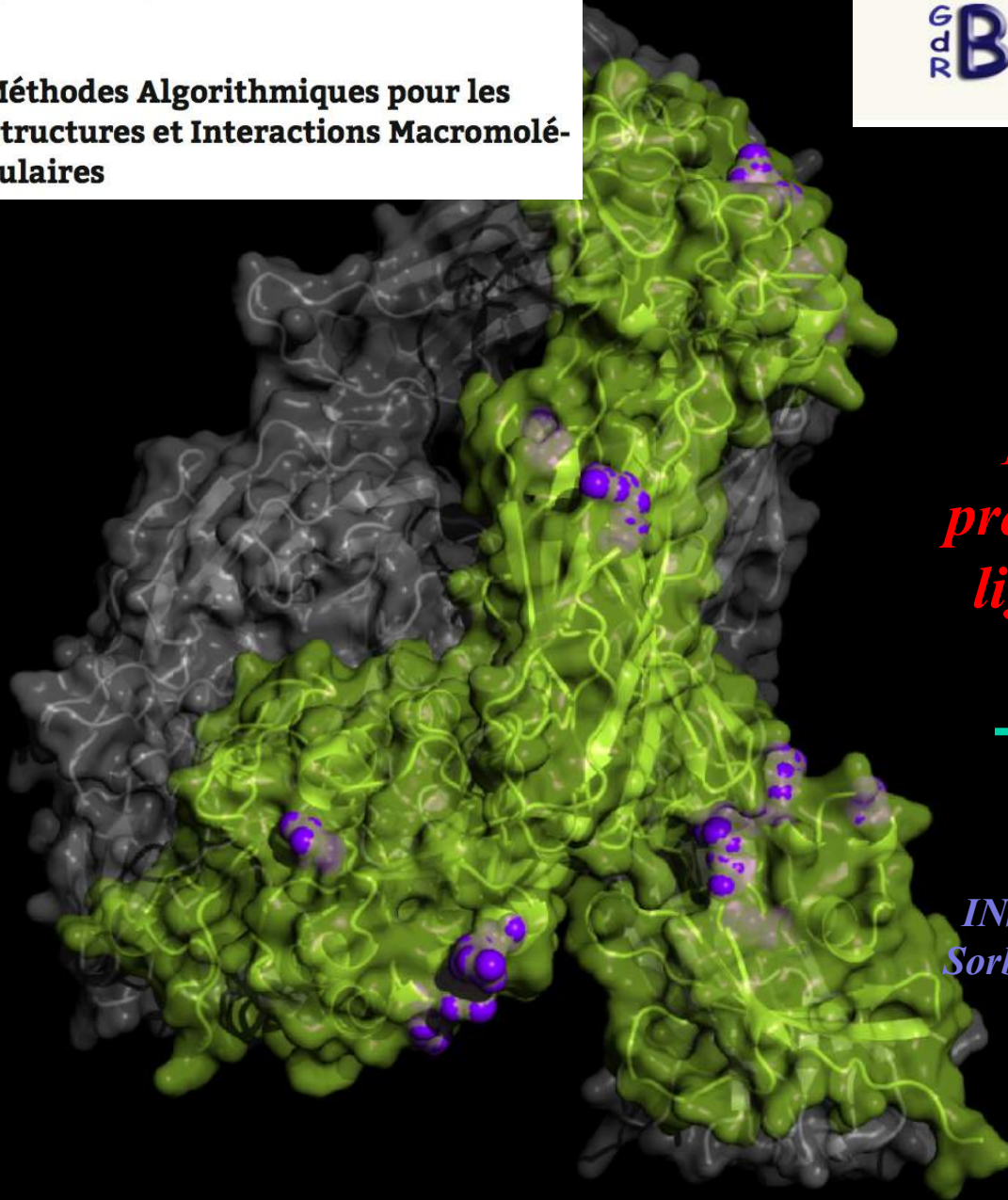
*GDR 3003 BIM – GT MASIM  
Thursday November 16th 2017  
Paris, France*

*Protein structure analysis,  
prediction and flexibility in the  
light of a structural alphabet.*

---

*Alexandre G. de Brevern*

*INSERM UMR\_S 1134, Univ. Paris Diderot,  
Sorbonne Paris Cité, Univ de la Réunion, Univ  
Antilles INTS, GR-Ex*



1. The team
2. Different web-db-tools of the team
3. Protein structure analysis, prediction and flexibility in the light of a structural alphabet

## 1. The team

Unit INSERM UMR\_1134 – Red Blood Cell Integrated Biology (BIGR)

<http://www.u665.inserm.fr/lang/en>



Team n°2 - Structural Dynamics and Interactions of  
Biological Macromolecules (DSIMB)

<http://www.dsimb.inserm.fr>



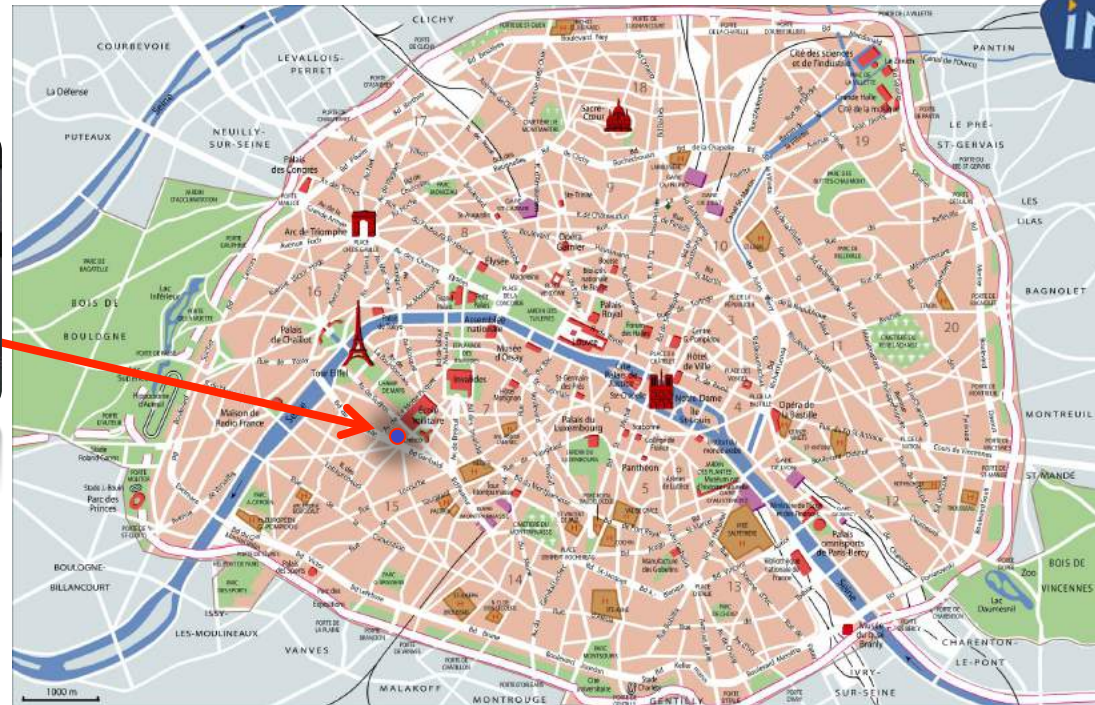
@DSIMB\_Lab



Other 3 teams are (wet) experimental ones.

## 1. The team

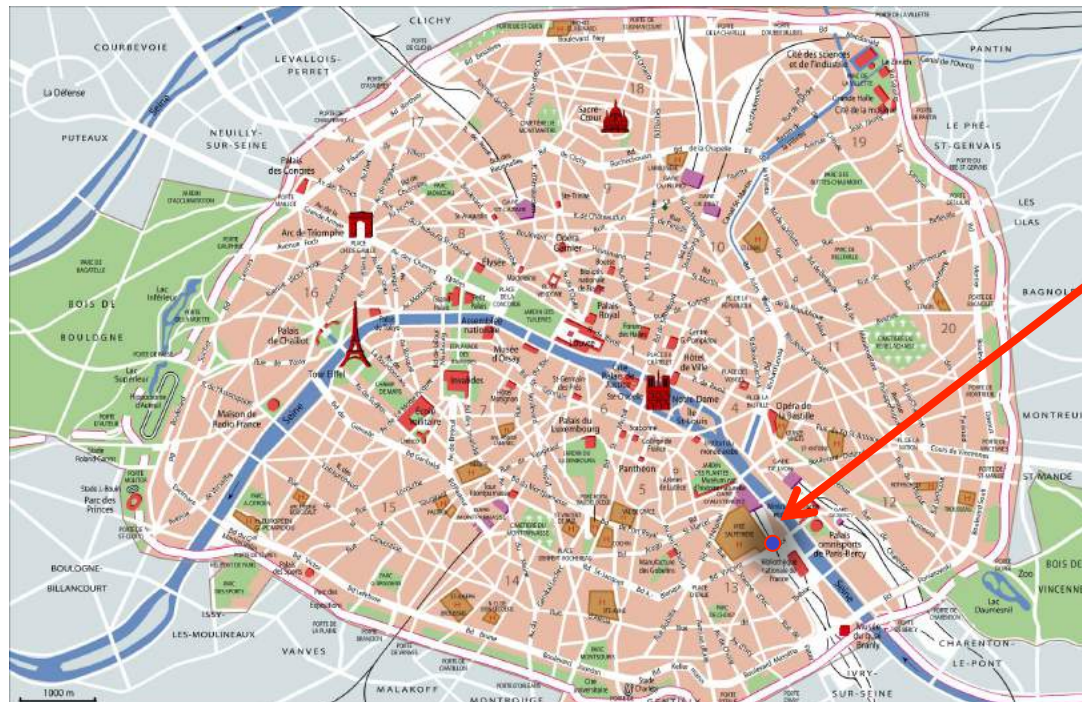
Localisation: National Institute for Blood Transfusion (PARIS)



## 1. The team



Associated to University Paris Diderot, Sorbonne Paris Cité



université  
**PARIS  
DIDEROT**  
PARIS 7



## 1. The team



Associated to University de la Réunion



## 1. The team



Team 1 is associated to University Antilles





## 1. The team



Pr. C. Etchebest (Pr. Univ Paris Diderot, HDR)

Pr. F. Cadet (Pr. Univ Réunion, HDR)

Dr. A.G. de Brevern (DR INSERM, HDR)

Dr. J.-C. Gelly (Ass. Pr. Univ Paris Diderot, HDR)

Dr. F. Gardebien (Ass. Pr. Univ Réunion, HDR)

Dr. Ph. Chartron (Ass. Pr. Univ Réunion)

1 Ing Univ Réunion, 1 Ing Univ Paris Diderot,  
1 Ing INSERM,  
1 ATER - 7 PhD students



---

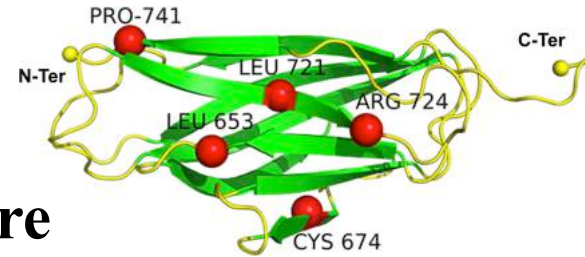
```
>ITB3_HUMAN Integrin beta-3
MDSSNVLQLIVDAYGKIRSKVELEVRDLPEELSLSFNATCLNN
EVIPGLKSCMGLKIGDTVFSIEAKVRGCPQEKEKSFTIKPVG
FKDSLIVQVTFDCDCACQAQAEPNSHRCNNGNGTFECGVCRCG
PGWLGSQCECSEEDYRPSQQDECSPREGQPVCSQRGECLCGQC
VCHSSDFGKITGKYCECDDFSCVRYKGE..
```

**sequence**

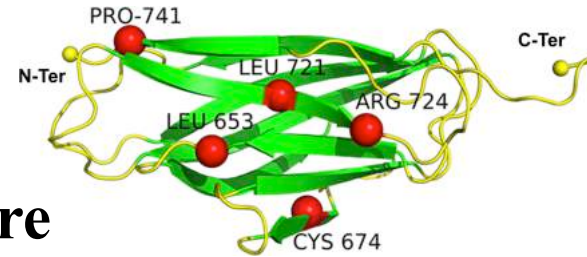
```
>ITB3_HUMAN Integrin beta-3  
MDSSNVLQLIVDAYGKIRSKVELEVRDLPEELSLSFNATCLNN  
EVIPGLKSCMGLKIGDTVVSFSIEAKVRGCPQEKEKSFTIKPVG  
FKDSLIVQVTFDCDCACQAQAEPNHRNNGNGTFECGVCRCG  
PGWLGSQCECSEEDYRPSQQDECSPREGQPVCSQRGECLCGQC  
VCHSSDFGKITGKYCECDDFSCVRYKGE..
```

**sequence**

**structure**



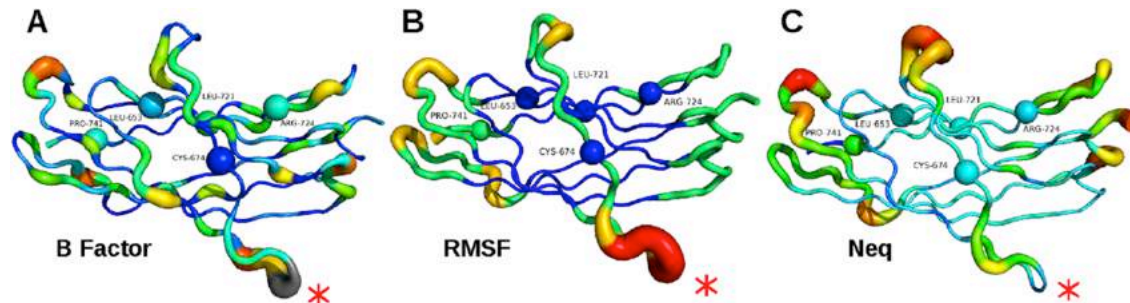
```
>ITB3_HUMAN Integrin beta-3
MDSSNVLQLIVDAYGKIRSKVELEVRLPEELSLSFNATCLNN
EVIPGLKSCMGLKIGDTVFSFSIEAKVRGCPQEKEKSFTIKPVG
FKDSLIVQVTFDCDCACQAQAEPNSHRCNNGNGTFECGVCRCG
PGWLGSQCECSEEDYRPSQQDECSPREGQPVCSQRGECLCGQC
VCHSSDFGKITGKYCEDDFSCVRYKGE..
```



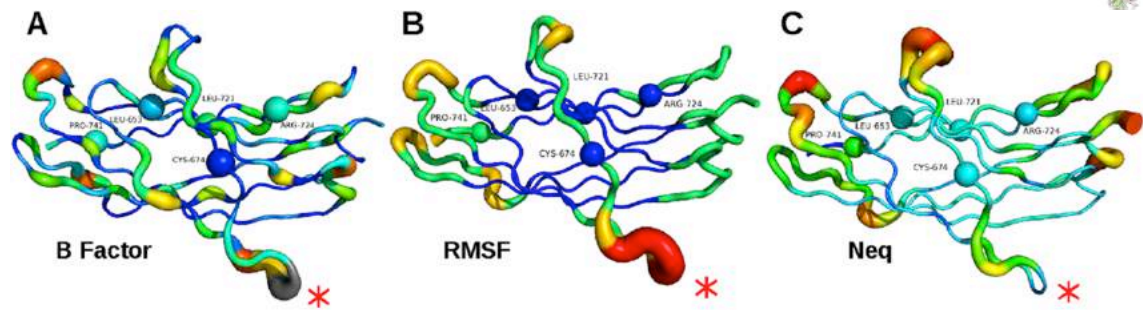
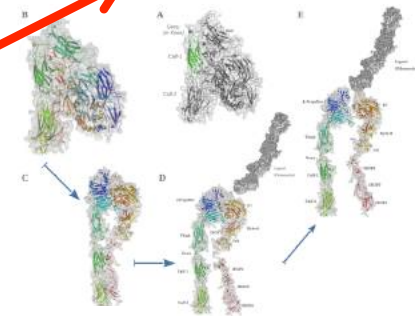
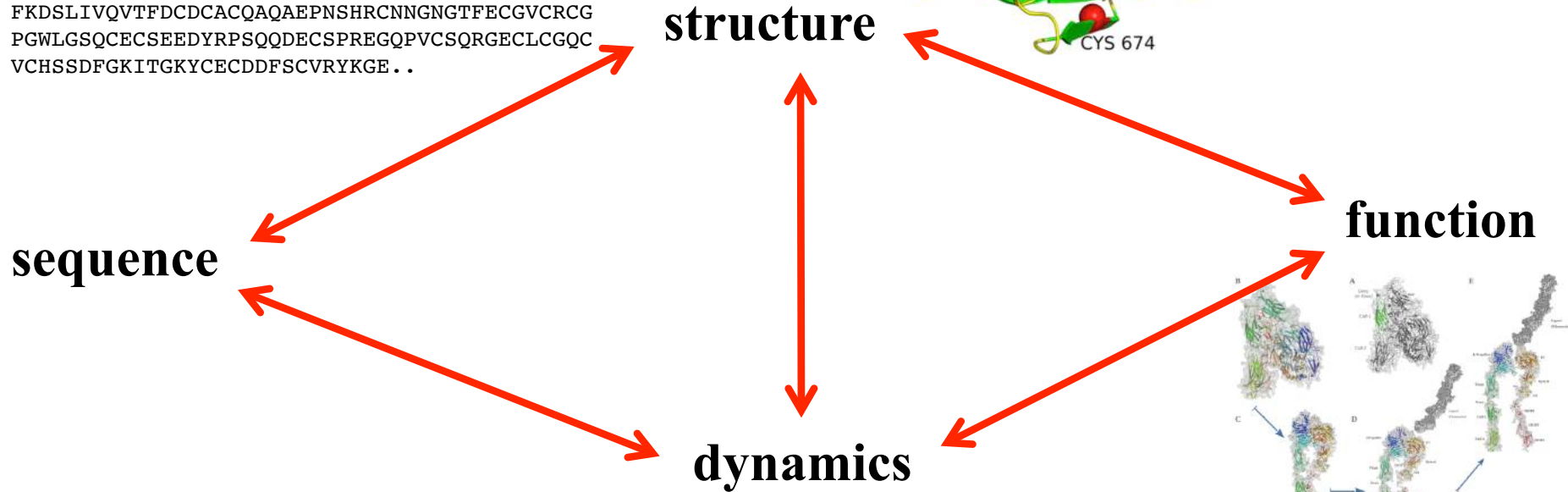
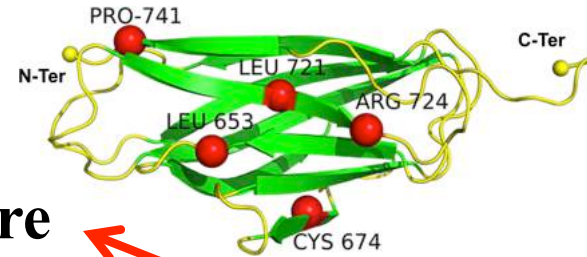
**sequence**

**structure**

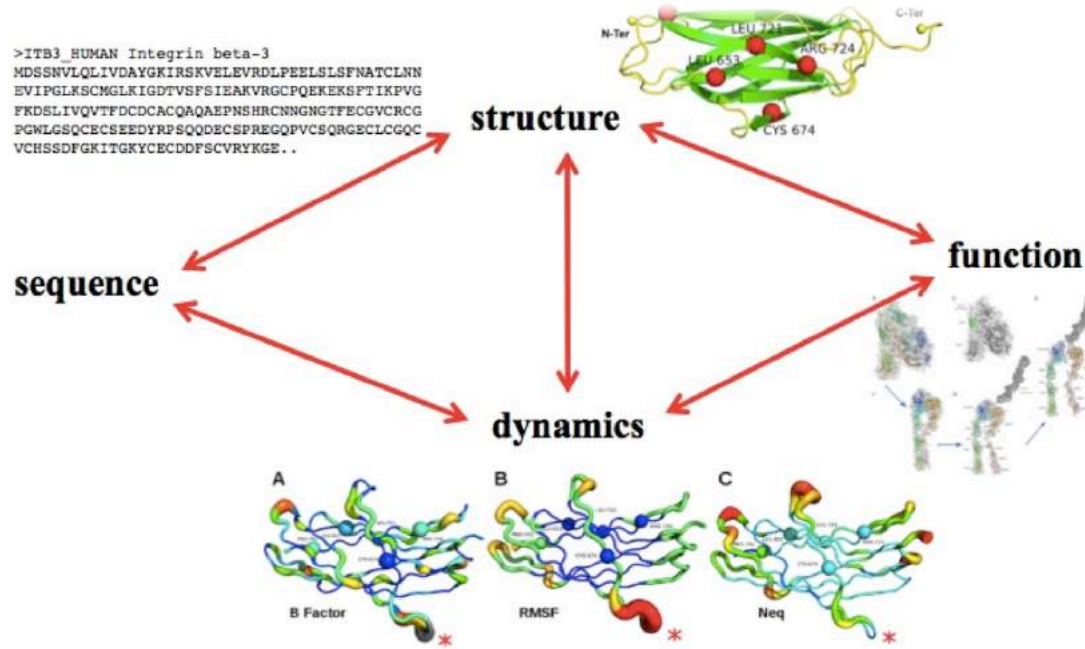
**dynamics**



```
>ITB3_HUMAN Integrin beta-3
MDSSNVLQLIVDAYGKIRSKVELEVRDLPEELSLSFNATCLNN
EVIPGLKSCMGLKIGDTVFSFSIEAKVRGCPQEKEKSFTIKPVG
FKDSLIVQVTFDCDCACQQAEPNSHRCNNGNGTFECGVCRCG
PGWLGSQCECSEEDYRPSQQDECSPREGQPVCSQRGECLCGQC
VCHSSDFGKITGKYCECDFSCVRYKGE..
```





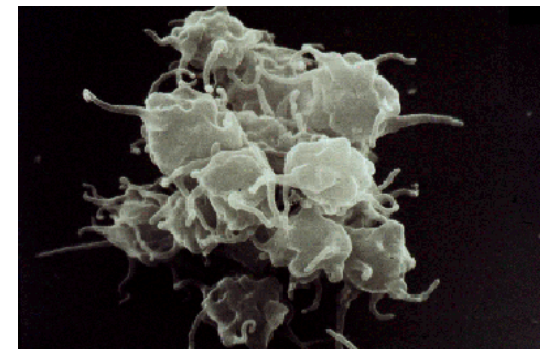


*Labex GREx*



**Methodological  
 developments  
 and applications**

**Red Blood Cells  
 (transfusion, platelets)**



## 2. The different webservers, databases & tools

<http://www.dsimb.inserm.fr/tools/>

**13 webservers**  
**9 databases**  
**7 tools**

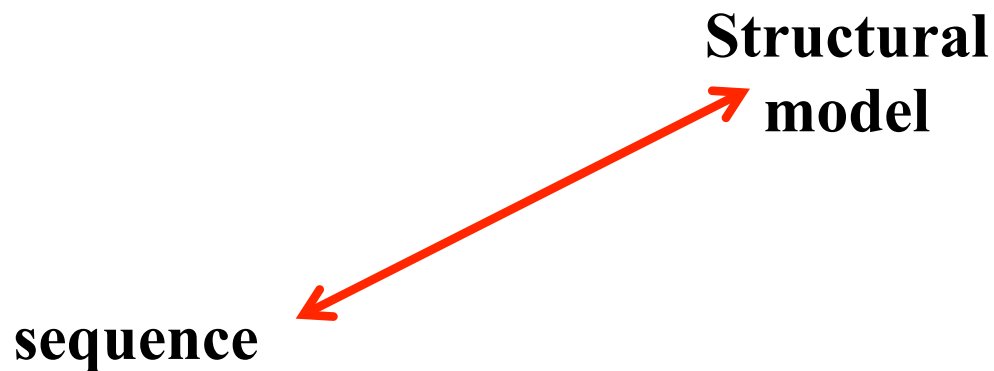
The screenshot shows the DSIMB website interface. At the top, there is a search bar and the DSIMB logo with the text "Dynamique des Systèmes et Interactions des Macromolécules Biologiques". Below the logo is a navigation menu with items: Presentation, Research, People, Publications, Web Tools (selected), Teaching, Symposium, News, and Contact. The main content area is titled "Webservice" and lists several tools with brief descriptions:

- OREMPRO**: A tool for the structural assessment of protein transmembrane domains.
- ORION**: ORION (Optimized protein fold Recognition) is a sensitive method for protein template detection.
- IPBA**: IPBA is a tool for protein structures mining and superimposition based on similarity in the local backbone conformation.
- MuIPBA**: muIPBA is an efficient tool for comparison of protein structures based on similarity in the local backbone conformation.
- Protein Block Expert (PBE)**: Web-based protein structure analysis server using a structural alphabet.
- PredyFlexy**: PredyFlexy predicts from the sequence: (i) the series of Long Structural Prototypes (LSPs), a sophisticated structural alphabet and (ii) the protein flexibility, this last being defined from data coming from X-ray structures and molecular dynamics simulation.
- VLDP**: VLDP webserver is designed to determine protein contacts, accessibility and residue volume using Laguerre diagram.
- SWORD**: A partitioning algorithm capable of producing multiple alternative domain assignments for a given protein structure.

On the right side, there is a sidebar with a language selector (Français, English) and a "News" section containing several short articles with dates and titles, such as "Recrutement d'un ingénieur 6 June 2017" and "International Research Network on Hematology meeting 2 March 2016".



- 
- I. The specific case of transmembrane protein structural model



**TOOL**

**MAIDEN** (Model quality Assessment for Intramembrane Domains using an ENergy criterion) is a statistical potential (or pseudo-energy function) optimized on native membrane protein structures. (<http://www.dsimb.inserm.fr/maiden/>)



Postic G, Ghouzam Y, Gelly JC (2015) An empirical energy function for structural assessment of protein transmembrane domains. *Biochimie*. **115**:155-61.



## Introduction

MAIDEN (Model quality Assessment for Intramembrane Domains using an ENergy criterion) is a statistical potential optimized on native alpha-helica and beta-sheet membrane protein structures. MAIDEN computes pseudo-energies for transmembrane domains of native or predicted protein structures, in order to evaluate their quality.

Reference:

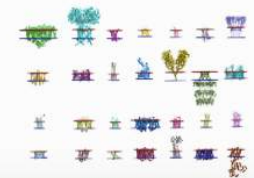
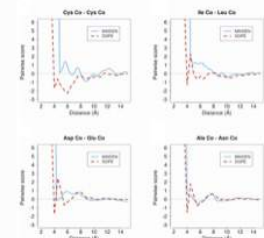
- Postic G, Ghouzam Y, Gelly JC. An empirical energy function for structural assessment of protein transmembrane domains. *Biochimie*. 2015 Aug;115:155-61.

[Download MAIDEN](#)

## Training set

The set of structures used for the calculation of MAIDEN contains 66 representative structures (41 alpha and 25 beta) from the PDBTM database (Tusnady, et al., 2005) determined by crystallography at <2.5 Å resolution and with an R-factor <0.3 (see below):

1JB0	2F93	2VDF	3EMN	3SZV	4AL0
1KQF	2FGQ	2WZE	3GIA	3TDS	4AMJ
10KC	2GR7	2WJR	3GP6	3TIJ	4DVE
1QD5	2HDI	2WSW	3HD6	3TX3	4DX5
1QJP	2J58	2X27	3KVN	3V5U	4E1S
1U19	2J7A	2X55	3M73	3VY8	4E1Y
1UUN	2MPR	2X9K	3N5K	3W54	4ENE
1XKW	204V	2XOV	3PCV	3WBN	4E2C
1XG0	2067	2XEV	3DSU	3TUX	4UET

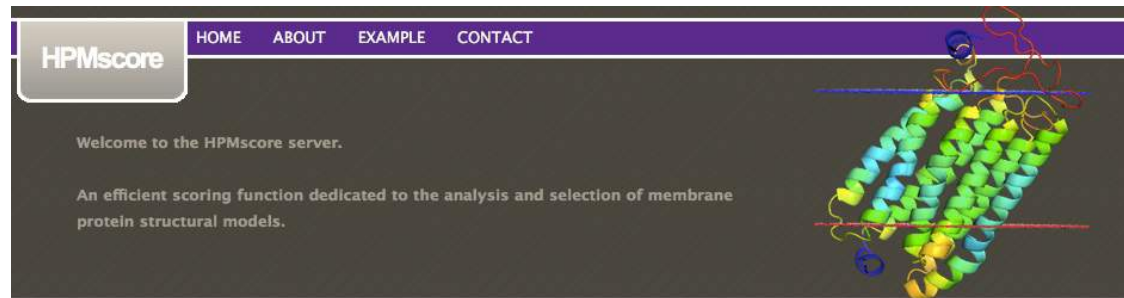


webservice

**HPMScore** a specific transmembrane protein learning to select best structural model. ([http://www.dsimb.inserm.fr/dsimb\\_tools/hpmscore/](http://www.dsimb.inserm.fr/dsimb_tools/hpmscore/))



Esque J, Urbain A, Etchebest C, de Brevern AG (2015) Sequence-Structure relationship study in all-alpha transmembrane proteins using an unsupervised learning approach. *Amino Acids*. 47(11):2303-22 for v.1



## Introduction

Transmembrane proteins are essential proteins implicated in many major biological and pathological processes. Nonetheless, due to experimental difficulties, the number of available membrane protein structures remains limited. Hence, building and assessment of protein structural models are complicated tasks. HPMScore is a dedicated tool designed to score membrane protein structural models.

HPMScore was extensively tested using the [HOMEPE2 database](#) as a reference; it proved its efficiency being more discriminative than the other available tools.

The principle of the algorithm is detailed in the [documentation page](#).

## HPMScore (list of models in an archive)

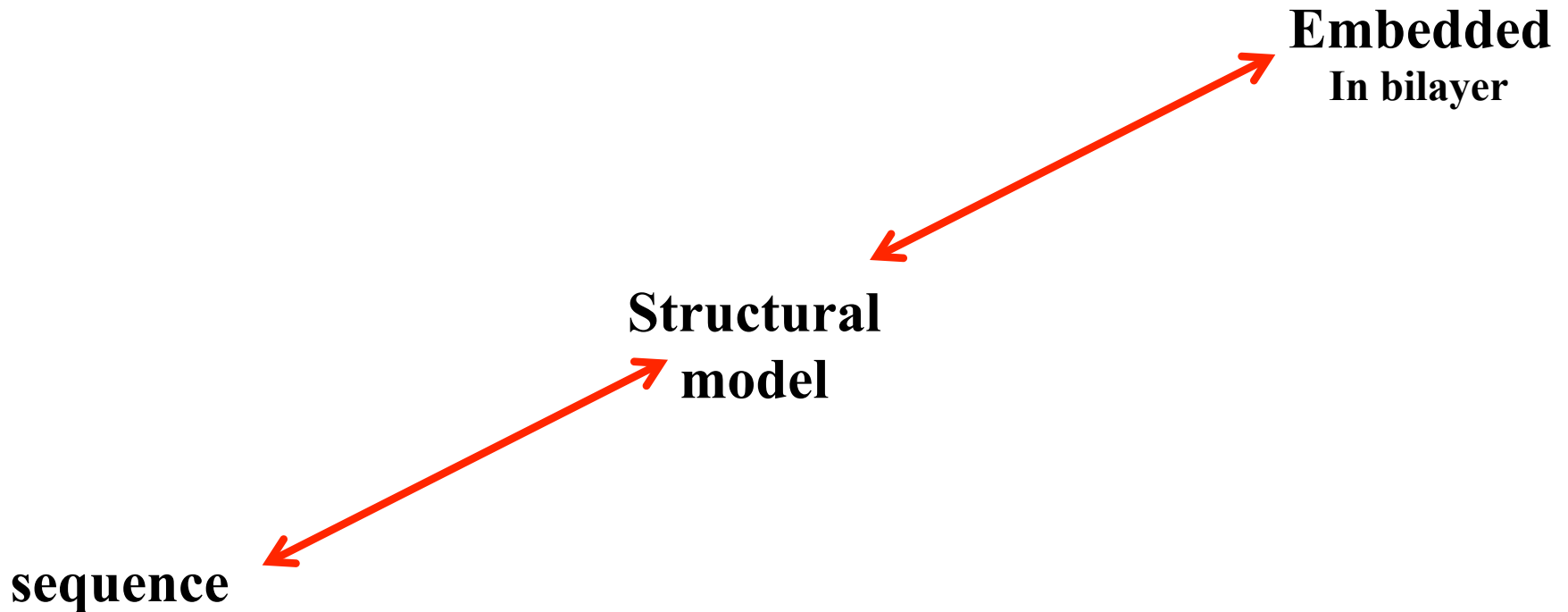
Upload one archive containing all your pdb files of containing the models of one protein. The archive must be in zip or tar.gz format.

Each PDB file will be processed to keep atom coordinates, **only the first chain will be analyzed**. Please be sure to have valid pdb files prior the upload.

Click on the DEMO button for a demonstration of the web service for each input.

Archive Demo Mode

- I. The specific case of transmembrane protein structural model



**TOOL**

**ANVIL** (Assignment aNd Visualization of the Lipid bilayer) is aimed at assigning membrane boundaries to a protein 3D structure. ([http://www.dsimb.inserm.fr/dsimb\\_tools/anvil/](http://www.dsimb.inserm.fr/dsimb_tools/anvil/))

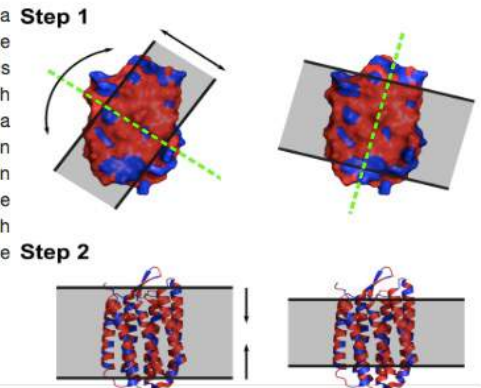


Postic G, Ghouzam Y, Guiraud V, Gelly J-C (2016) Membrane positioning for high- and low-resolution protein structures through a binary classification approach. *Protein Eng. Des. Select.* **29**: 87–92.

## ANVIL

### Assignment aNd Visualization of the Lipid bilayer

ANVIL is aimed at assigning membrane boundaries to a protein 3D structure, by using the spatial coordinates of the alpha carbons. It can also process coarse-grained models of protein structures. The algorithm follows an approach that treats the problem of membrane assignment as a binary classification. The software is implemented in Python and intended to be run with the PyPy interpreter. In output, ANVIL writes a PDB file that contains the coordinates of the membrane, which can be visualized with softwares, such as RasMol or PyMOL. The ANVIL source code is freely available under a CeCILL License.



[Download ANVIL](#)

© Copyright G. Postic & J.-C. Gelly, 2015

webservice

**OREMPRO** A tool for the structural assessment of protein transmembrane domains. ([http://www.dsimb.inserm.fr/dsimb\\_tools/OREMPRO/en/index.php](http://www.dsimb.inserm.fr/dsimb_tools/OREMPRO/en/index.php))



Postic G, Ghouzam Y, Gelly J-C  
(2016) OREMPRO web server:  
orientation and assessment of  
atomistic and coarse-grained  
structures of membrane proteins.  
*Bioinformatics*. **32**:2548–2550

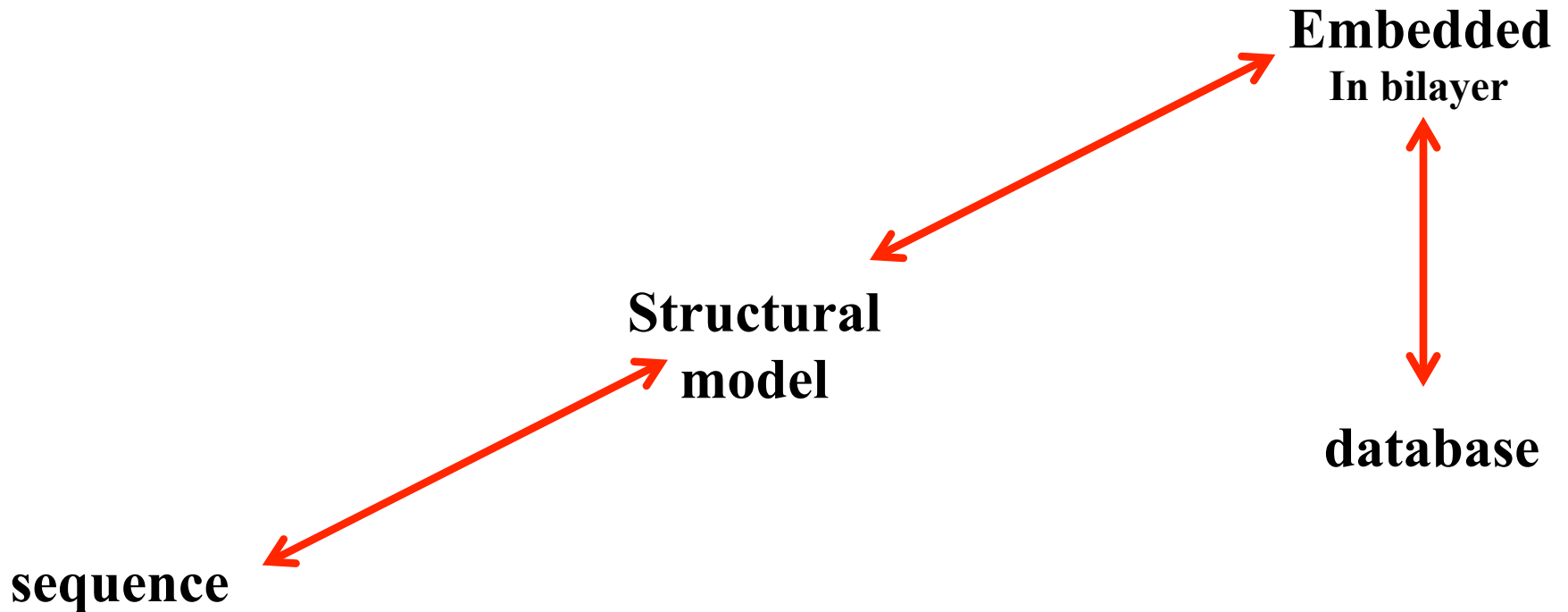
The screenshot shows the OREMPRO web server interface. At the top, there's a navigation bar with 'DSIMB Tools' and a logo. Below that, the main heading is 'OREMPRO Orientation and Evaluation of Membrane Proteins'. The interface includes a search bar for 'Protein structure' with a file upload button and a text input field. There are sections for 'Membrane assignment parameters' with radio buttons for 'Alpha-helical' and 'Beta-sheet', and a checkbox for 'Coarse-grained'. An 'Advanced options' section is partially visible. At the bottom, there are buttons for 'Proceed', 'Run example', and 'Atom'. A sidebar on the right contains a text box with the following content:

This web server<sup>1</sup> combines the **ANVIL** algorithm for positioning protein structures in the lipid bilayer<sup>2</sup> and the **MALDEN** statistical potential for the structural assessment of protein transmembrane domains<sup>3</sup>. Both methods are based on the sole coordinates of alpha carbons, which makes OREMPRO able to process both high- and low-resolution models.

<sup>1</sup>Postic G, Ghouzam Y, Gelly JC. *Bioinformatics* (2016) 31:3782-3789 #  
<sup>2</sup>Postic G, Ghouzam Y, Guireaud V, Gelly JC. *PEDS* (2016) 29:87-92 #  
<sup>3</sup>Postic G, Ghouzam Y, Gelly JC. *Biochimie* (2015) 115:155-161 #

© 2016 Guillaume Postic, Yassine Ghouzam and Jean-Christophe Gelly; Inserm UMR\_S 1134, Univ. Paris Diderot, Sorbonne Paris Cité

- I. The specific case of transmembrane protein structural model



database

**TMPL.** A database of transmembrane protein structures ( $\alpha$ -helical and  $\beta$ -sheet) positioned in the lipid bilayer. (<http://www.dsimb.inserm.fr/TMPL/>).



Postic G, Ghouzam Y, Etchebest C, Gelly J-C (2017) TMPL: a database of experimental and theoretical transmembrane protein models positioned in the lipid bilayer. *Database*, 2017, bax022.

**TMPL** Home Search Browse Deposit Download Help Contact

**Welcome to the TMPL database**  
Transmembrane protein Models Positioned in the Lipid bilayer

The TMPL database<sup>1</sup> contains 3D structures of transmembrane proteins with bilayer boundaries assigned by the ANVIL algorithm<sup>2</sup> and represented by two planes. The structural quality of the transmembrane domains delimited by ANVIL is evaluated with the MAIDEN statistical potential<sup>3</sup>; both methods are available through the [OREMPRO](#) web server<sup>4</sup>. The molecular structures in TMPL are i) either atomistic or coarse-grained representations of proteins, and ii) either theoretical or experimental models.

Users have the possibility of adding new entries to the TMPL database, by uploading protein structures orientated in the lipid bilayer. This can be achieved by using either available membrane protein structures (native or predicted), or new theoretical models, such as those that users can generate with the [ORION](#) web server<sup>5,6</sup>.

**Statistics**

• Native structures: 997	• Predicted models: 10,170
• Alpha-helical: 830	• Pfam domains: 729
• Beta-sheet: 167	• Deposited models: 2
• Martinized: 1,039	

<sup>1</sup>Postic G, Ghouzam Y, Etchebest C, Gelly JC. *Database* (2017) doi:10.1093/db

<sup>2</sup>Postic G, Ghouzam Y, Guiraud V, Gelly JC. *Protein Eng Des Sel* (2016) 29:87-92

<sup>3</sup>Postic G, Ghouzam Y, Gelly JC. *Biochimie* (2015) 115:155-161

<sup>4</sup>Postic G, Ghouzam Y, Gelly JC. *Bioinformatics* (2016) 32:2548-2550

<sup>5</sup>Ghouzam Y, Postic G, de Brevern AG, Gelly JC. *Bioinformatics* (2015) 31:3782-3789

<sup>6</sup>Ghouzam Y, Postic G, Guerin PE, de Brevern AG, Gelly JC. *Sci Rep* (2016) 6:28268

© 2016 Postic G., Ghouzam Y., Etchebest C. & Gelly J.-C.; Inserm UMR\_S 1134, Univ. Paris Diderot, Sorbonne Paris Cité



---

## II. Analysis of protein structures.

webservice

**VLDP**. A webservice designed to determine protein contacts, accessibility and residue volume using Laguerre diagram. ([http://www.dsimb.inserm.fr/dsimb\\_tools/vldp/](http://www.dsimb.inserm.fr/dsimb_tools/vldp/))



### Introduction

VLDPws is a tool for analysing protein structures based on Laguerre diagram, a powerful mathematical-geometric method. The protein structure is represented as a graph (Esque et al. 2010, Esque et al. 2011). From this diagram, information can be extracted pertaining to the packing of proteins, residue volume, contacts, its accessibility and even topological genus calculations. VLDP software has been developed by Jeremy Esque and Christophe Oguey (IPTM, CNRS, Univ Cergy-Pontoise), while the web development has been done by Sylvain Leonard (DSIMB, Inserm, INTS, Univ Paris Diderot).

### PDB data

Use this form to upload a chain from the Protein Data Bank. Input can be below.

#### Protein Data Bank uploader

PDB Identifier & Chain : 1F8VA

Submit

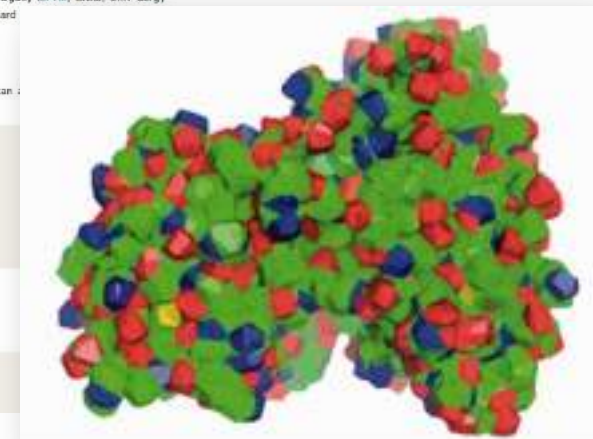
Stick the chain label (A, B, C, ...) to the end of the PDB identifier.

### Input File

Upload a PDB file and enter the chain IDs.

#### Personal PDB file

Filename: Choisir le fichier / aucun fichier sélectionné chain



Esque J, Leonard S, Oguey C, de Brevern AG (2013) VLDP webserver: a powerful geometric tool for analysing protein structures in their environment. *Nucl Acid Res.* 41:W373–378.

database

**PolyPrOnline.** Polypronline is a webserver dedicated to the assignment of PolyProline II helices ([http://www.dsimb.inserm.fr/dsimb\\_tools/polyproline/](http://www.dsimb.inserm.fr/dsimb_tools/polyproline/)).



Chebrek R, Leonard S, de Brevern AG, Gelly J-C (2014) PolyprOnline: polyproline helix II and secondary structure assignment database. *Database*. 2014:bau102.

The screenshot shows the PolyPrOnline website interface. At the top, there is a navigation bar with links for Home, Search, How to use, and About. Below this is an 'Introduction' section with a 3D molecular model of a protein structure. The text in the introduction discusses the importance of secondary structures and the specific focus on PolyProline II (PPII) helices. To the right of the introduction is a large 3D visualization of a protein structure, labeled 'E', showing a series of repeating helical segments. Below the introduction is a 'Database statistics' section, which states that the database contains 24761 protein structures. At the bottom of the page, there is a small footer with copyright information and a last update date.

database

**PTM-SD** provides an access to proteins for which Post Translational Modifications are both experimentally annotated and structurally resolved ([http://www.dsimb.inserm.fr/dsimb\\_tools/PTM-SD/](http://www.dsimb.inserm.fr/dsimb_tools/PTM-SD/)).



Craveur P., Rebehmed J., de Brevern A.G. (2014) PTM-SD: a database of structurally resolved and annotated post-translational modifications in proteins. *Database*. 2014: bau041.

PTM STRUCTURAL DATABASE

HOME BROWSE DATABASE TOOLS CREATE YOUR DATASET CONTACT LINKS

ABOUT PTM-SD

HOME

PTM STRUCTURAL DATABASE DATA SOURCE BROWSE CHART STATISTICS

Welcome to the Post Translational Modifications Structural Database (PTM-SD).

Proteins are composed of a succession of amino acids. Their 3D structures are the support of major crucial biological functions.

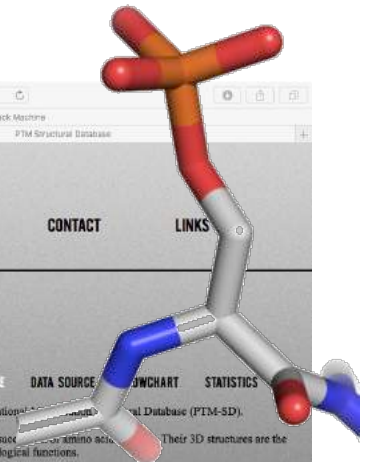
Post translational modifications (PTMs) are covalent chemical modifications of proteins that typically occur after the protein synthesis.

They allow a fine and sometimes obligatory modulation of biological functions; they can determine proteins activity state, localization, turnover, and interactions with other proteins.

PTM-SD provides an access to proteins for which PTMs are both experimentally annotated and structurally resolved.

[Read more...](#)

Oxidation on C-115 in P13448



webservice

**Protein Peeling 1, 2 & 3.** An approach for splitting a 3D protein structure into compact fragments ([http://www.dsimb.inserm.fr/dsimb\\_tools/peeling3/](http://www.dsimb.inserm.fr/dsimb_tools/peeling3/))



Gelly J-C, de Brevern AG (2011)  
Protein Peeling 3D: New tools  
for analyzing protein structures.  
*Bioinformatics*. 27(1):132-3.

**Protein Peeling 3 D**  
Splitting a protein structure into proteins units and domains

**Menu**  
Home  
About method  
Example of output  
Contacts

**Announcement**  
08/2010 - Peeling3 le lancement  
25/02/2009 - Peeling3 le nouveau  
20/02/2008 - Peeling3 le lancement  
18/01/2008 - Peeling le lancement

**Peeling a protein**  
This method take an input pdb file and peel it.

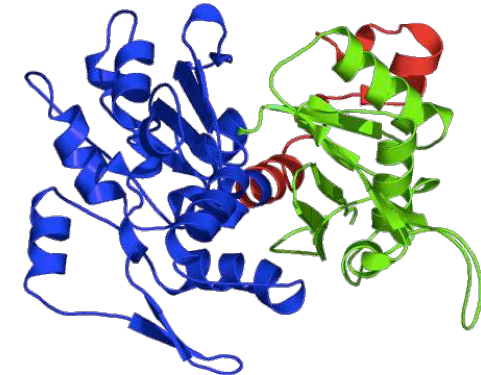
Browse for your 3D structure PDB (pdb file format):  
Chercher le fichier : aucun fichier sélectionné.  
[PEEL IT]

Note that data older than 7 days are deleted.

**Options**

Prune tree :	<input type="checkbox"/> Yes • <input checked="" type="checkbox"/> No	CF out-off: 0.1
Choose a specific E-value :	0.5	
Minimal size of protein unit :	10	
Minimal size of secondary structures :	8	
Take into account only regular secondary structure for Peeling :	<input type="checkbox"/> Yes • <input checked="" type="checkbox"/> No	
Cut-off distance between atom :	8 C	
Delta parameter in logistic function :	1.5	
Maximal size of protein unit :	0	

**References**  
Gelly JC, de Brevern AG  
Protein Peeling 3D: New tools for analyzing protein structures  
*Bioinformatics*. 2011 Jan 1;25(1):132-3.  
Faure C, Brevel A, de Brevern A G.  
Analysis of protein contacts into Protein Units  
*Biochimie* 2006 91(7):876-87.  
Gelly JC, de Brevern AG, Hatzut S.  
Protein Peeling: an approach for splitting a 3D protein structure into compact fragments.  
*Bioinformatics*. 2008 Jun 15;22(12):149-51. Epub 2008 Nov 14.



webservice

**SWORD.** A partitioning algorithm capable of producing multiple alternative domain assignments for a given protein structure (<http://www.dsimb.inserm.fr/sword/>)



Postic G, Ghouzam Y, Chebreck R, Gelly J-C (2017) An ambiguity principle for assigning protein structural domains, *Science Advances*. **3**(1): e1600552.

The screenshot displays the DSIMB SWORD webserver interface. At the top, three protein structure models are shown with different domain colorings. Below them, navigation links include [Tutorial], [Download], [Contact], and [DSIMB Tools]. The main form allows users to select a protein structure by PDB code (1x4) and chain (A), or to upload a coordinate file. A 'Proceed' button is visible. On the right, a text box explains the SWORD algorithm's goal of producing multiple alternative domain assignments for ambiguous protein structures. Below the interface, three alternative domain assignments for Lysozyme (1lys) are shown as ribbon diagrams:

- Lysozyme (1lys) Three domains (optimal)**: Residues 1-39, 40-98, and 99-129.
- Lysozyme (1lys) Two domains (1st alternative)**: Residues 1-39; 99-129 and 40-98.
- Lysozyme (1lys) Two domains (3rd alternative)**: Residues 1-98 and 99-129.

---

## III. Protein families and functions.

database

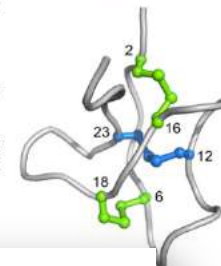
**KNOTTIN.** The reference database on protein knottin structural properties.  
(<http://www.dsimb.inserm.fr/KNOTTIN/>).



Welcome to the KNOTTIN database

### What are knottins?

- Knottins are small disulfide-rich proteins characterized by a very special "disulfide through disulfide knot"
- This knot is achieved when one disulfide bridge crosses the macrocycle formed by the two other disulfides and the interconnecting backbone.
- The knot implies that knottins contain at least 3 disulfide bridges.
- The structural family of knottins have the disulfide between cysteines III and VI (blue) going through disulfides I-IV and II-V (green).
- The growth factor cystine knots also contain a knot but the connectivity is different and they cannot be superimposed onto knottins. These proteins belong to a distinct structural family not described in this site.
- Knottins are sometime referred to as "Inhibitor Cystine Knots".



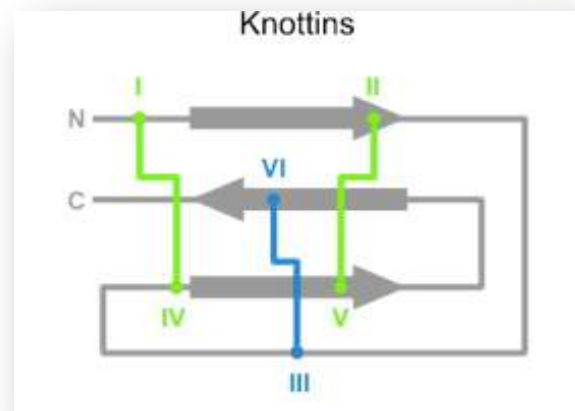
### Updates

• June 2017  
Major update of the KNOTTIN database  
+ Addition of theoretical 3D models  
+ Visualization with JSmol  
+ New design of the web interface

• February 2014  
The KNOTTIN database is available  
We had to stop the server for technical reasons. It is now available again.  
Please let us know about any trouble you could encounter with this server.  
Also note that we are currently working on the next update of the database.  
Thank you for using KNOTTIN!



Postic G, Gracy J, Périn C, Chiche L, Gelly JC. (2018) KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res.* gkx1084





database

**RESPIRE.** The database aims at providing a comprehensive structural view of the Red Blood Cell protein content. (<http://www.dsimb.inserm.fr/respire/>).



## Welcome to RESPIRE

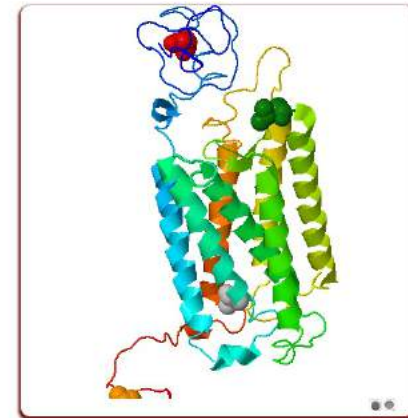
Welcome to the **Repository of Enhanced Structures of Proteins Involved in the Red Blood Cell Environment**. This database aims at providing a reference, ordered and protein-driven information on proteins available in the erythrocyte.

You will find on this database annotations for proteins detected in the latest differentiation stages of erythropoiesis. To help in finding useful content, reference external databases are queried regularly and then RESPIRE methods allow the integration of both this data and of additional processing into dedicated pages, one per protein ([See Documentation for details](#)).

With a combination of structural information, clinical information and links to reference external databases, it simplifies the access to RBC components for biologists, clinicians and structural biologists.

The protein structures presented in the carousel display models produced by the DSIMB laboratory for some protein of interest.

Start browsing the database by clicking an image in the carousel, by using pre-set searches in the **Queries** dropdown menu, or by searching by protein name, gene name or uniprot identifier in the top search box. Many links are also provided in protein entries to propose alternative browsing facilities.



## RESPIRE.

*database*

RESPIRE Search - Documentation Statistics Contact History

### Atypical chemokine receptor 1

The protein contains 336 amino acids for an estimated molecular weight of 35553 Da.

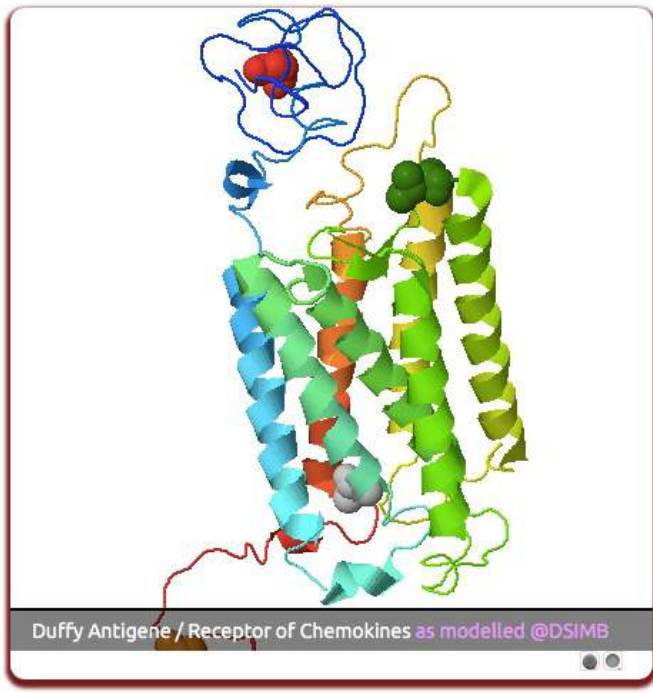
[Previous protein](#) [Next protein](#)

Domains Structure Interactions Gene Ontologies Genetics (OMIM) History Links

```

...0...5...0...5...0...5...0...5...0...5...0...5...0...5...100
NSSYGVNDSFPDGDYGANLEAAAPCHSCNLLDSDALPFFILTSVLGILAVSTVLFMLFRPLFRWQLCPGIPVLA
...0...5...0...5...0...5...0...5...0...5...0...5...0...5...200
ALCSLGVCVWYGSFAFAQALLGCHASLGHR LGAGQVPGLTLGLTVGIWGVALLTLPVTLASGASGGLCTLIYS
...0...5...0...5...0...5...0...5...0...5...0...5...0...5...300
GAKGLKALGMIPGPWMNILWAWIFWPHGVVGLDFLVRSKLLLLSTCLAQQA DLLLNLAEAAILHCVAT
...0...5336
SHLDTLGSKS
    
```

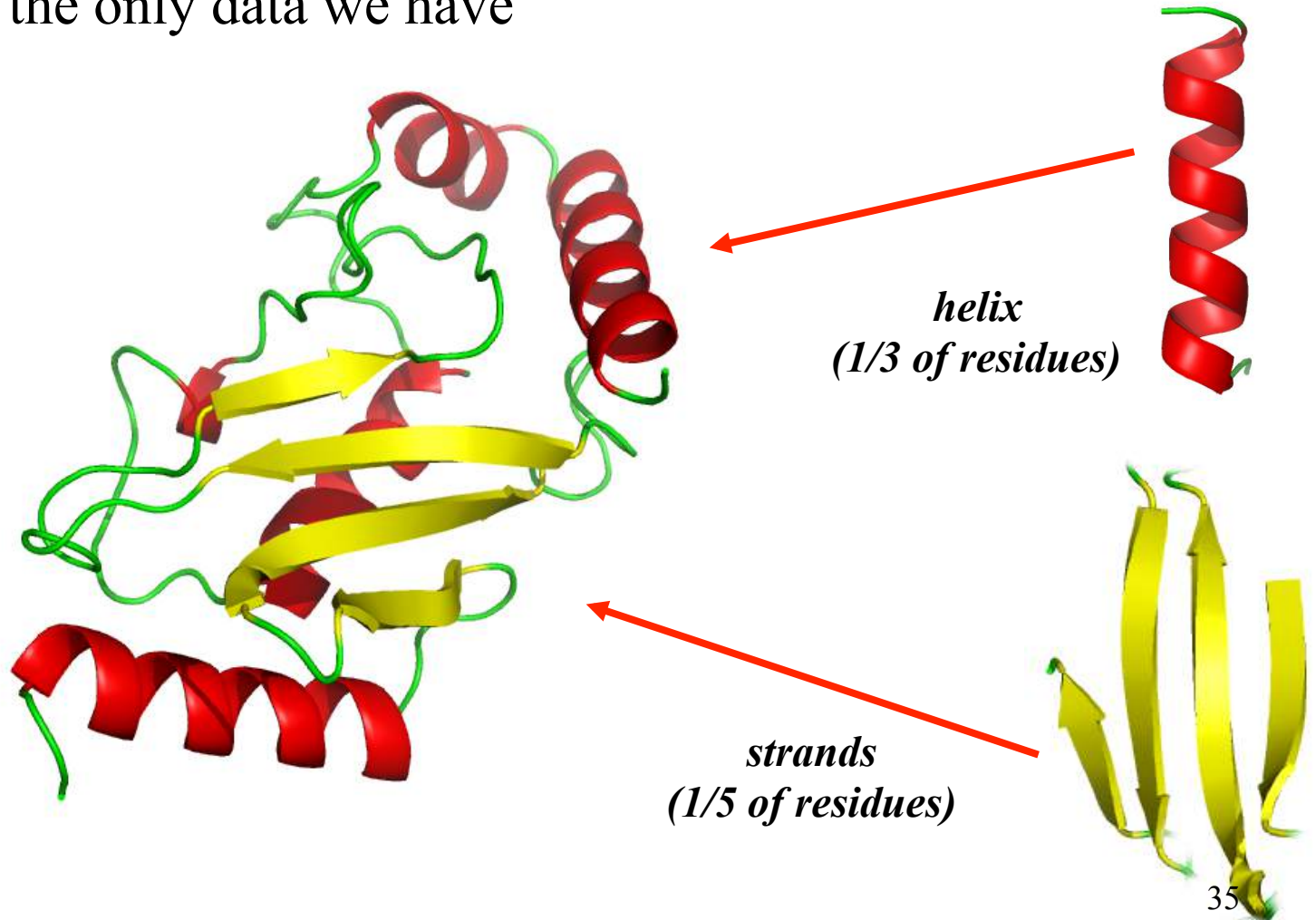
[Protein coevolution profile \(FreeContact\)](#) [Multiple Sequence Alignment \(Muscle\)](#)



- 
3. Protein structure analysis, prediction and flexibility in the light of a structural alphabet

# Structural alphabet

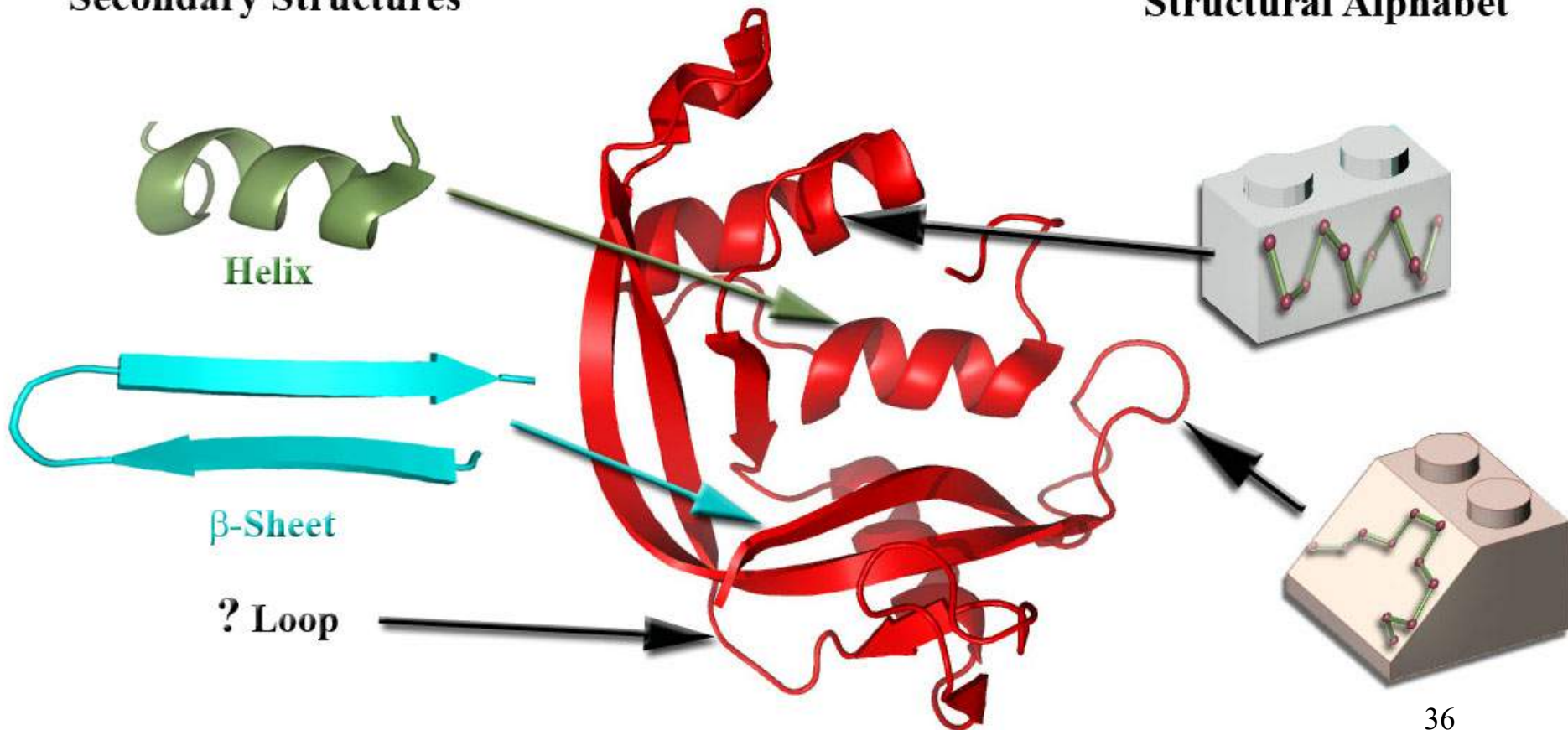
- That's the only data we have



➤ Another vision

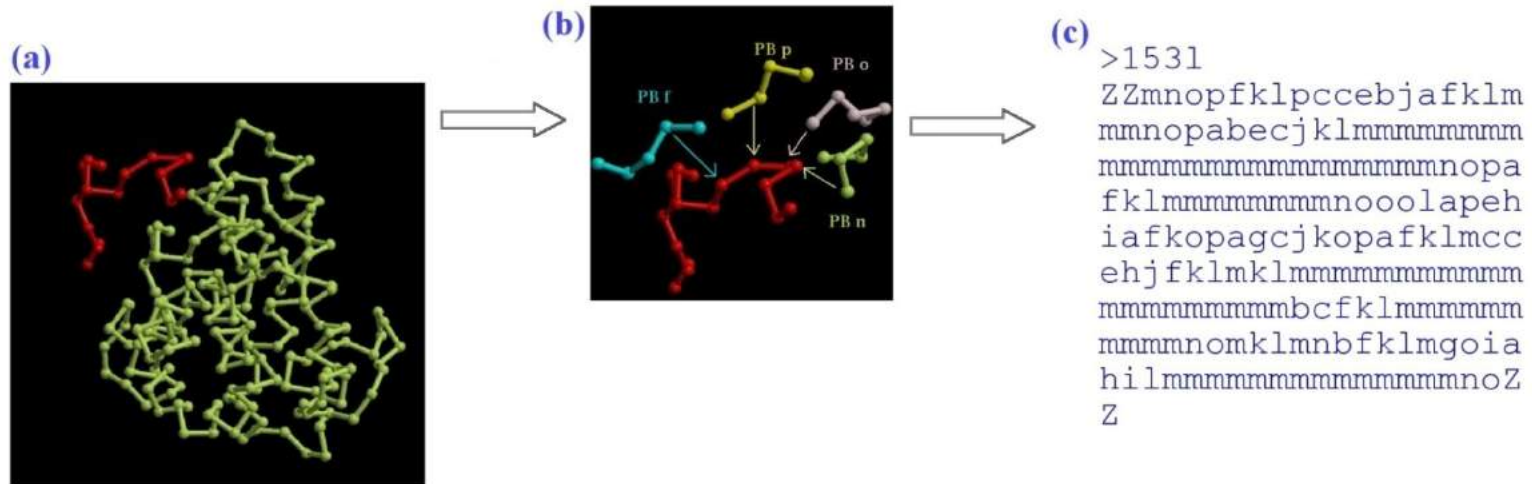
## Secondary Structures

## Structural Alphabet



➤ Structural Alphabet (SA):

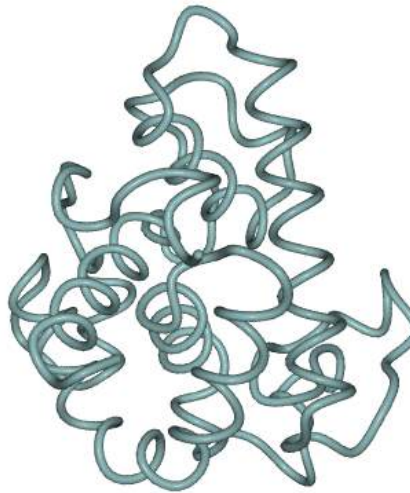
Structure approximation (3D -> 1D)



## *PB Sequence*

de Brevern AG, Etchebest C, Hazout S (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins*. **41**:271-287

# Structural alphabet: principle



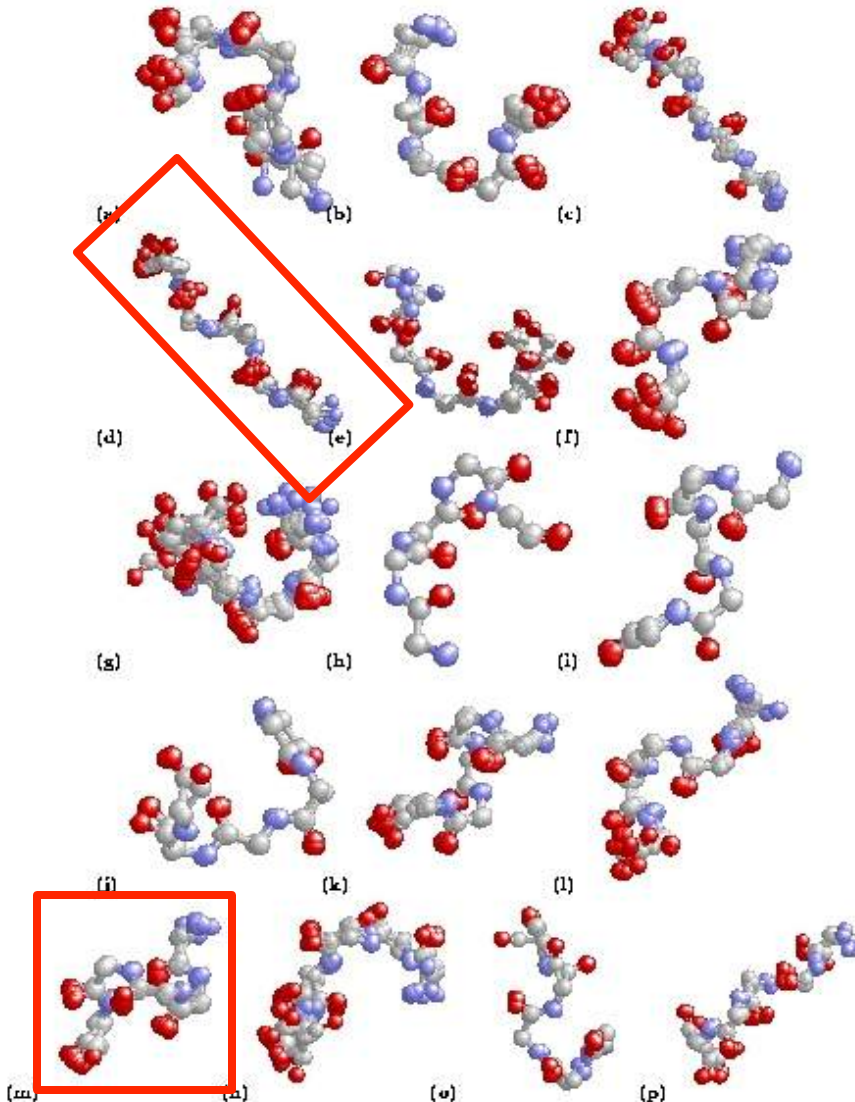
>153L [*secondary structures*]

```
HHHHH          HHHHHH          HHHHHHHH
HHHHHHHHHHHHHHHHHHHH          HHHHHHHHHH
          HHHHHHHHHHHHHHHHHHH
HHHHHHHHHHH          HHHHHHHHHHHH          EEEEE
          HHHHHHHHHHHHHHHHHHH
```

>153L [*structural alphabet*]

```
ZZmnopfklpccebjafklmmnopabecjklmmmmmmmmmm
mmmmmmmmmmmmmmmmmmnopafklmmmmmmmmmmnooolapgeh
iafkopagcjkopafklmccehjfkmlklmmmmmmmmmmmmmmmm
mmmmmmmmmmmbcfklmmmmmmmmmmmmmmmmnomklmnbfgklmmgoia
hilmmmmmmmmmmmmmmmmmmnoZZ
```

# Structural alphabet: principle



Pr. Serge Hazout

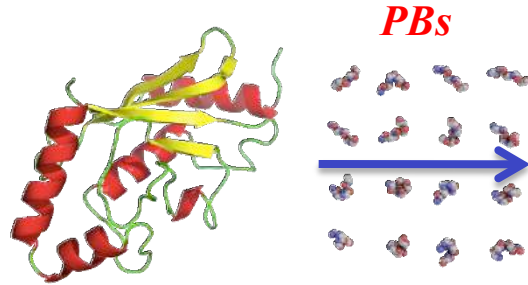
de Brevern AG, Etchebest C, Hazout S (2000)  
Bayesian probabilistic approach for predicting  
backbone structures in terms of protein blocks.  
*Proteins*. 41:271-287

16 Protein Blocks

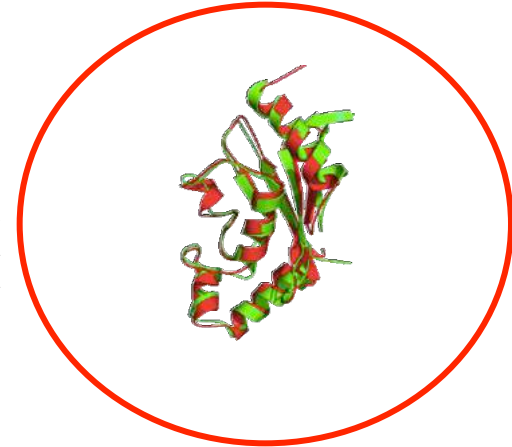


# Structural alphabet : superimposition

*Similarity detection?*



```
>2aak
zzfklmmmmmmmmmmmmccch
iacdddfkmbkbcddddddeh
iakgopacdddddddfkbcabc
cdddfbdcddfkclcfkopacd
fklmmmgcfkbcfklmmmmmm
mmmmccfklcfblmlmmmmmm
mmlmmmmmmmmmmmmmmzz
```



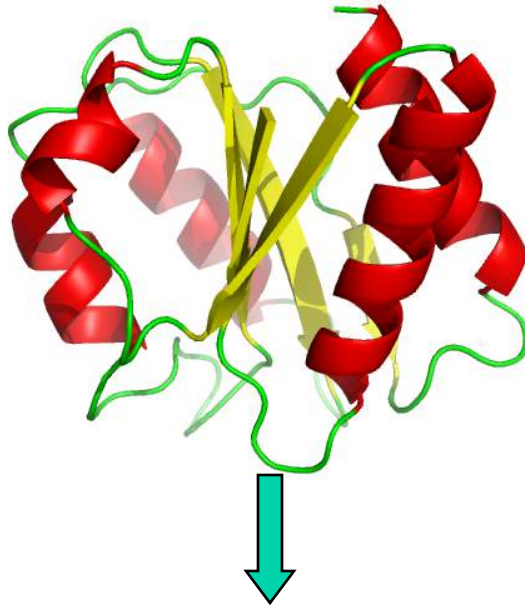
*N. Srinivasan*



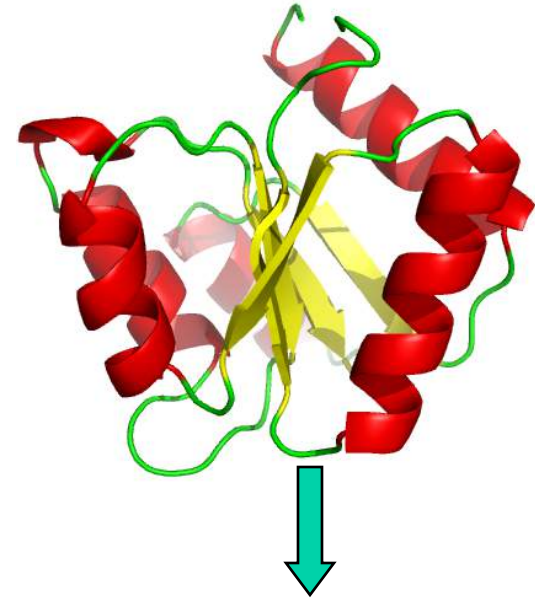
*Manoj Tyagi*

# Structural alphabet : superimposition

3chy



2fox

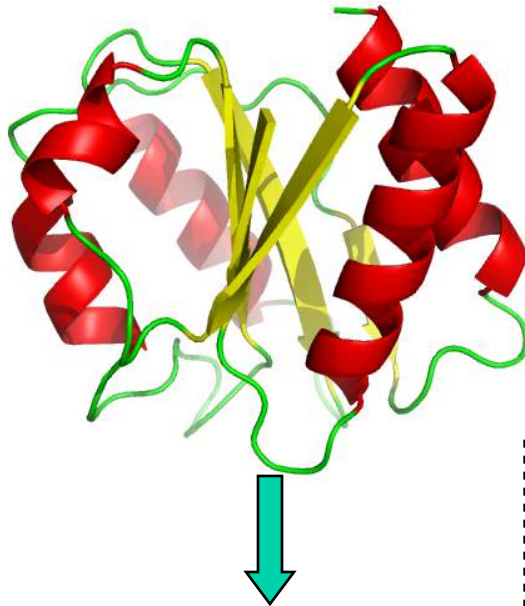


KBCCDDDDFBFKLMMMMMMMMNNOABDCDDFB  
 FKLMMMMMNGOIA BDCDDFBDGHILMLMMMMM  
 MMPMKLMMPCCDDDDFBDCFKLMMMMMNOABD  
 CDDDDFBKLMMMMMMMMMMNO

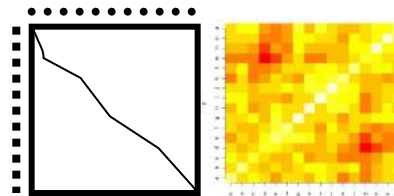
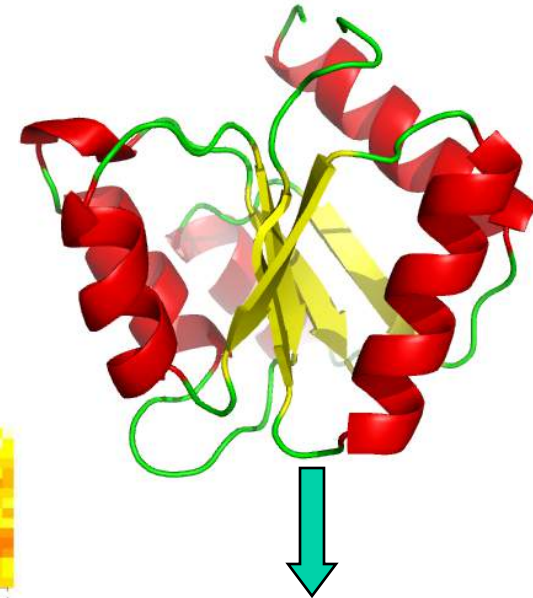
DDDDFKOMMMMMMMMMMMMMMMMMNNOACDDDDFKL  
 PCFKLMMMPFBDCDDDDDEHJLPACFKLMMMMM  
 MMMMMGHIA CDDDEHIAFKLNOMMMMMMMMMMMM  
 NOPACFBACDDDEHJACKLMMMMMMMMMMMMMM

# Structural alphabet : superimposition

3chy



2fox



KBCCDDDDFBFKLMMMMMMMMNOPABDCDDFB  
 FKLMMMMMNGOIA BDCDDFB DGHILMLMMMMM  
 MMPMKLMPCCDDDDFBDCFKLMMMMMNOPABD  
 CDDDDFBKLMMMMMMMMMMNO

DDDDFKOMMMMMMMMMMMMMMMMMNOPACDDDDFKL  
 PCFKLMMMPFBDCDDDDDEHJLPACFKLMMMMM  
 MMMMMGHIA CDDEHIAFKLNOMMMMMMMMMMMM  
 NOPACFBACDDDEHJACKLMMMMMMMMMMMMM



local / global alignment

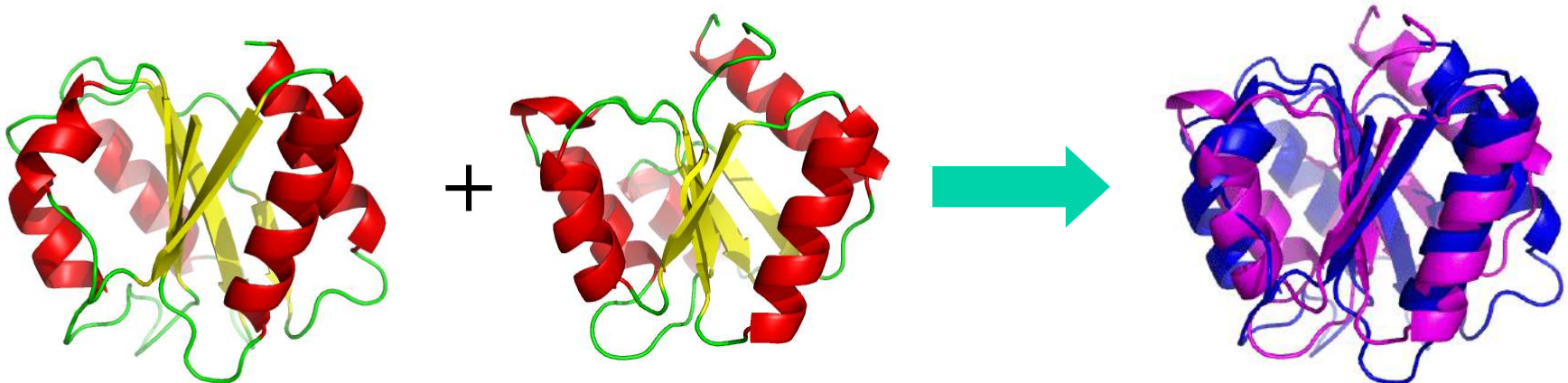
# Structural alphabet : superimposition

```

3chy_ : KBCCDDDDFBF--KLMMMMMMMMNOPABDCDDFBFKLMMMMMNGOIABCDDFBGHI
2fox_ : ---DDDDFKOMMMMMMMMMMMMMMMMNOPACDDD--FKLPCKFLMMM-PFBDCDDDDDEHJ

3chy_ : LM--LMMMMMMMPKLMPCDDDDFBDCFKL---MM-MMM-M-NOPA--BDCDDDD-
2fox_ : LPACFKLMMMMMMMMMMMMMNGHIACDDEHJAFKLNOMMMMMMMMMMMMNOPACFBACDDDEH

3chy_ : -DFKLMMMMMMMMMMMMMNO
2fox_ : JACKLMMMMMMMMMMMMMM
  
```



Tyagi M, Sharma P, Swamy CS, Cadet F, Srinivasan N, de Brevern AG, Offmann B (2006) Protein Block Expert (PBE): a web-based protein structure analysis server using a structural alphabet. *Nucleic Acids Res.* **34**(Web Server issue) W119-23.

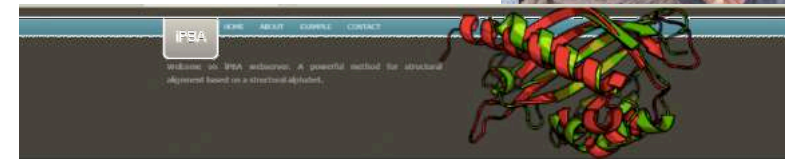
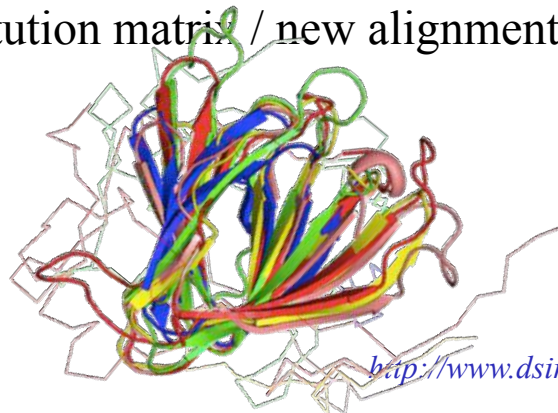
# Structural alphabet : superimposition

- French-Indian collaboration: CEFIPRA grant (2008-2012):  
with Pr. N. Srinivasan (Bangalore)

*Joseph Agnel Praveen*



pairwise superimposition and mining of similar folds  
(new substitution matrix / new alignment)



*iPBA:*

[http://www.dsimb.inserm.fr/dsimb\\_tools/ipba/](http://www.dsimb.inserm.fr/dsimb_tools/ipba/)

**Introduction**  
iPBA is a tool for comparison of protein structures based on similarity in the local backbone conformation. The local backbone conformation is defined as post-proline-peptides, using Protein Ducts (PD) (de Brevern et al. 2007, Joseph et al. 2010), the protein structure represented as PD sequences, are aligned by dynamic programming using a PD substitution matrix. Structurally similar structures are assigned using an anchor based alignment approach.

**Compare and align two protein structure**  
Enter the PD sequences of the two structures to be compared and aligned. You can either enter the PD sequences or upload the PDB files.

**Protein 1**  
Enter PDB identifier chain:  chain:   
or  
upload file structure file:

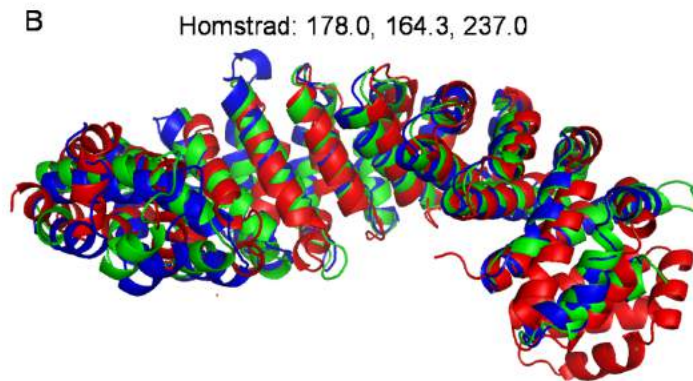
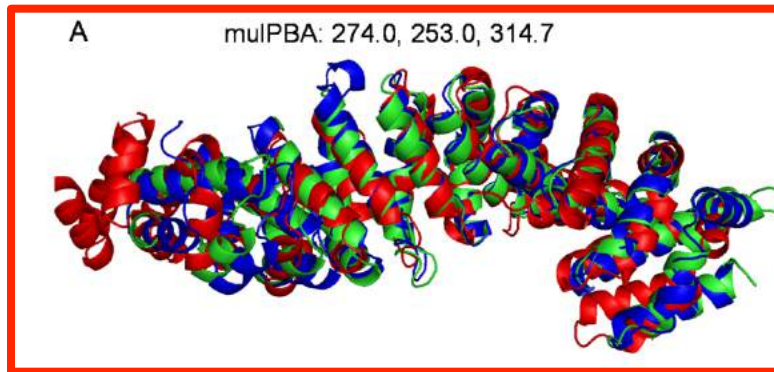
**Protein 2**  
Enter PDB identifier chain:  chain:   
or  
upload second structure file:

Joseph AP, Srinivasan N, de Brevern AG (2011) Improvement of protein structure comparison using a structural alphabet. *Biochimie* **93**(9):1434-45.

J-C Gelly, Joseph AP., Srinivasan N, de Brevern AG (2011) iPBA : A tool for protein structure comparison using sequence alignment strategies. *Nucleic Acid Res* **39**:W18-23.

# Structural alphabet : superimposition

- improvement of multiple structure alignments: improvement > 85% of cases



*mulPBA:*  
[http://www.dsimb.inserm.fr/dsimb\\_tools/mulpba](http://www.dsimb.inserm.fr/dsimb_tools/mulpba)

Introduction

mulPBA is a tool for comparison of protein structures based on similarity in the local backbone conformation. The local backbone conformation is defined as pentapeptide dihedrals, using Protein Blocks (PB) [de Brevern et al. 2000, Joseph et al. 2010]. The protein structures represented as PB sequences, are aligned by dynamic programming scored by a PB substitution matrix [Joseph et al. 2011]. A progressive alignment strategy similar to CLUSTALW was adopted for multiple PB sequence alignment. Highly similar stretches identified by the pairwise alignments are given higher weights during the progressive alignment. Encoding the information on backbone conformation as PB sequences, enables 'sequence-like' structural alignment, which is minimally influenced by structural flexibility (see example). The residue equivalences from PB based alignments are identified to obtain a three dimensional fit of the structures, followed by an iterative refinement of the structural superposition.

Compare and align multiple protein structure - PDB data

Use this form to analyse chain from Protein Data Bank files. In order to make multiple protein analysis a minimum of 3 protein are required. The files can also be uploaded see below.

Protein Data Bank uploader

1FBNA  
1GBAA  
1GBBA  
1NT2A  
1DVA

Align Clear Form

Data have to be formatted as below: PDBCODE:IAN.  
Separate identifiers by a newline or a space

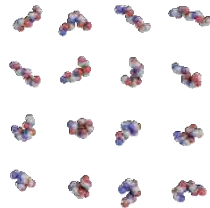
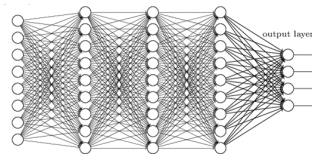
Joseph AP, Srinivasan N, de Brevern AG (2012) Progressive structure-based alignment of homologous proteins: Adopting sequence comparison strategies. *Biochimie*. **94**:2025-34.

Joseph AP, Leonard S, Srinivasan N, Gelly J-C, de Brevern A.G (2014) mulPBA : an efficient multiple protein structure alignment method based on a structural alphabet. *J Biol Struct Dyn*. **32**(4):661-8.



## Prediction of PBs

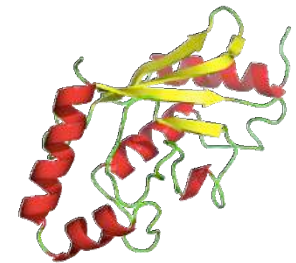
```
>protseq
MSSADKERLARENHSEIERRRR
NKMTAYITELSDMVPTCSALAR
KPKDLTILRMAVSHMKSLRGTG
NTSTDGSYKPSFLTDQELKHLI
LEAADGFLFIVSCETGRVVYVS
DSVTPVLNQPQSEWFGSTLYDQ
VHPDDVDKLRQLSTSE
```



```
>predPB
zzfklmmmmmmmmmmmmccch
iacdddfkmbkbccddddddeh
iakgopacdddddddfkbcabc
cdddfbdcdddfklcfkopacd
fklmmmgcfkbcfklmmmmmm
mmmmccfklcfblmlmmmmmm
mmlmmmmmmmmmmmmzz
```



## Protein structural model



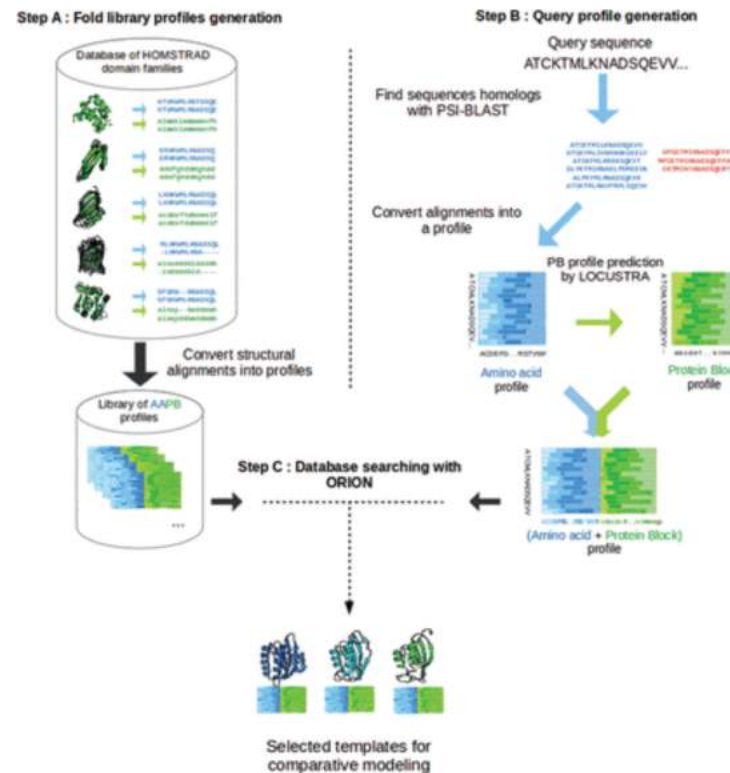


webservice

**ORION** (Optimized protein fold RecognitION) is a sensitive method for protein template detection (<http://www.dsimb.inserm.fr/ORION/>)

*Excellent CASP results*

Ghouzam Y, Postic G, de Brevern AG, Gelly JC (2015)  
Improving protein fold recognition with hybrid profiles combining sequence and structure evolution.  
*Bioinformatics*. **31**:3782-9.



webservice

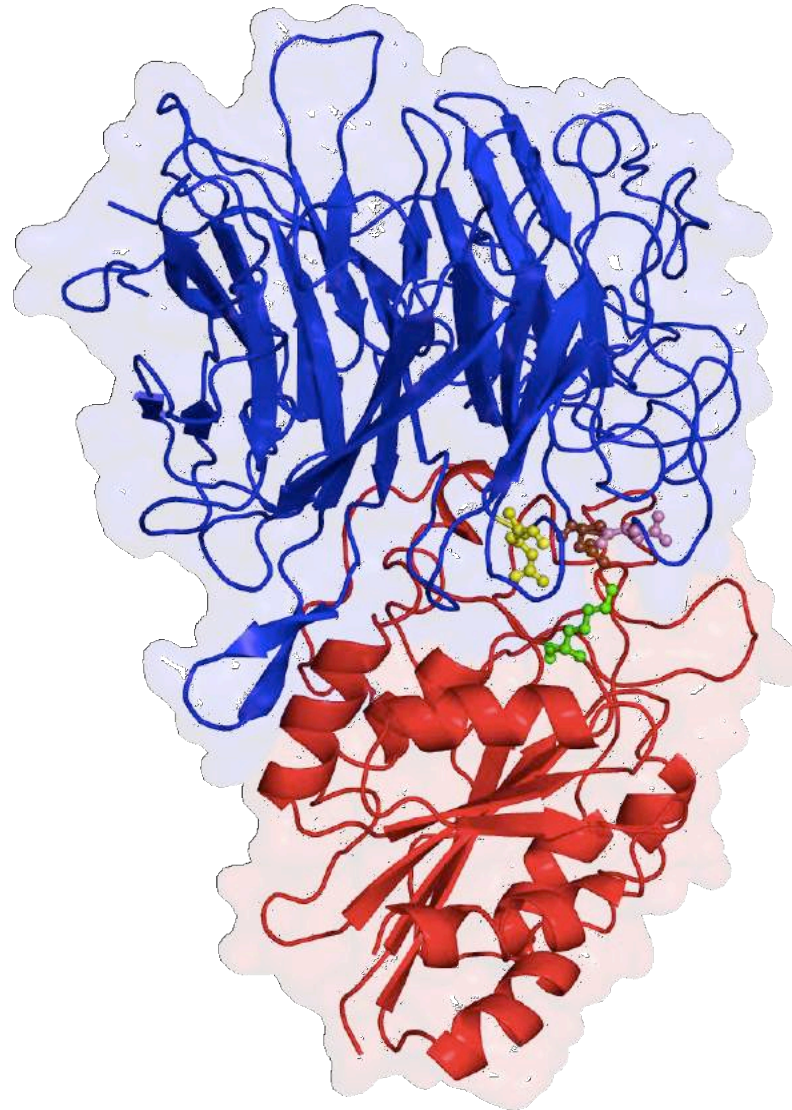
**ORION** (Optimized protein fold RecognitION) is a sensitive method for protein template detection (<http://www.dsimb.inserm.fr/ORION/>)



Ghouzam Y, Postic G, Guerin PE, de Brevern AG, Gelly JC (2016) ORION: a web server for protein fold recognition and structure prediction using evolutionary hybrid profiles. *Sci Rep.* 6:28268.

The screenshot shows the ORION web server interface. At the top, there is a navigation bar with links for Home, About, Example, Contact, and Help. Below the navigation bar, a brief description of ORION is provided. The main section is titled "Submit prediction" and contains a form for submitting a protein sequence. The form includes a text area for pasting the sequence, a "Sequence file" selection button, and search options for database selection, alignment mode, and the number of hits displayed. The "Database selection" dropdown is open, showing options like Scope95\_v2.06, Scope70\_v2.06, pcb95\_10March16, and pcb70\_05April6. The "Alignment mode selection" has radio buttons for glo/loc, global, and local. The "Maximum number of hits displayed" is set to 100. At the bottom, there is a "Job options" section with a checkbox for "Send results to email (Optional)" and a dropdown menu set to "Recommended".

## Integrins

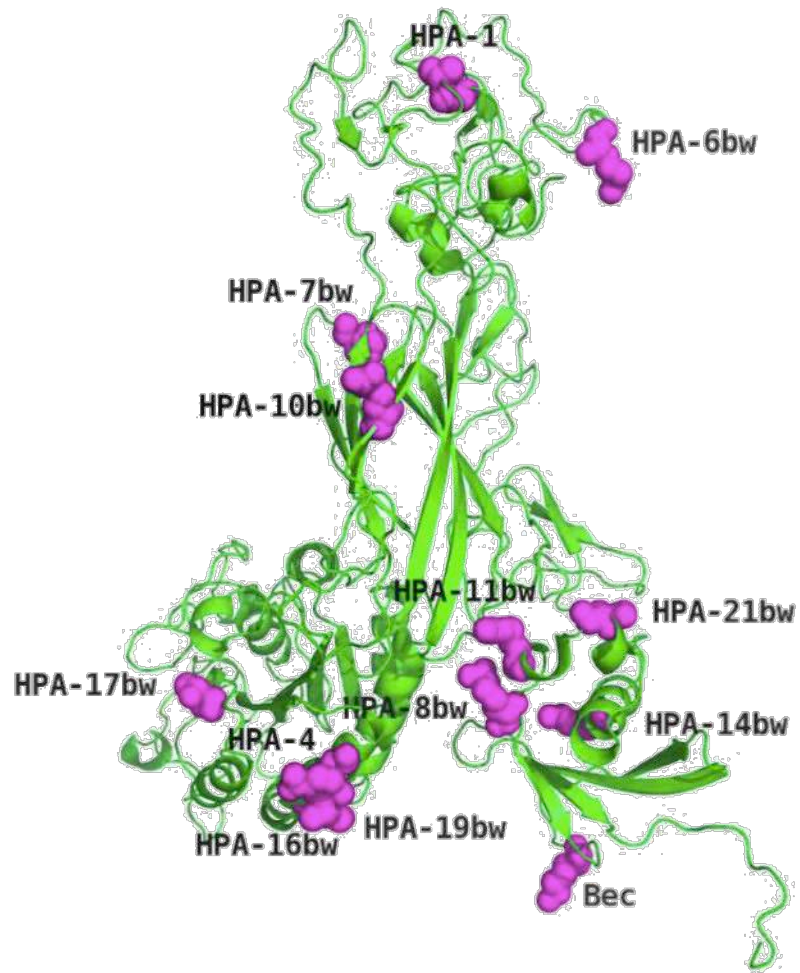


**Dr. Vincent Jallu**  
**Platelet departement**



## One complex (Integrins) / 2 pathologies

- **Bleeding:** Glanzmann's thrombasthenia is an abnormality of the platelets, an extremely rare coagulopathy.
- **Alloimmunization mother/fetus:** 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



- This protein is *very* flexible

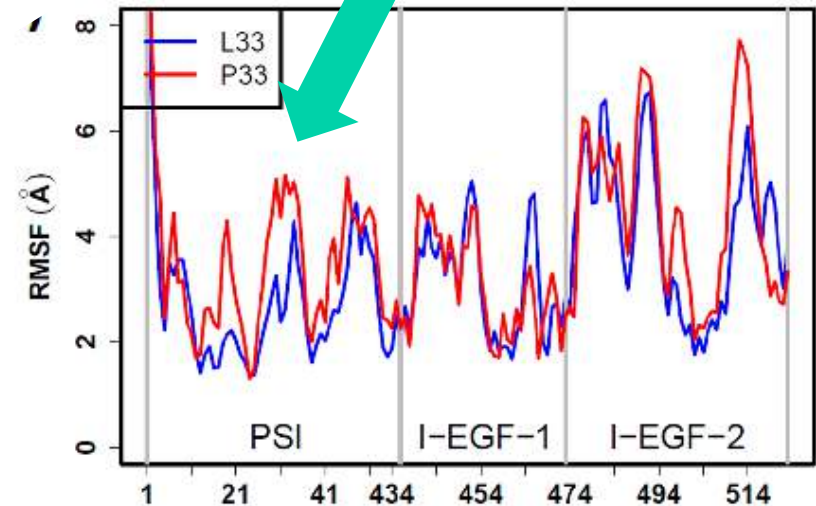


*We compute Root Mean Square  
Fluctuation (RMSF)*

- This protein is *very* flexible



*Residue 33 is associated to high RMSF*



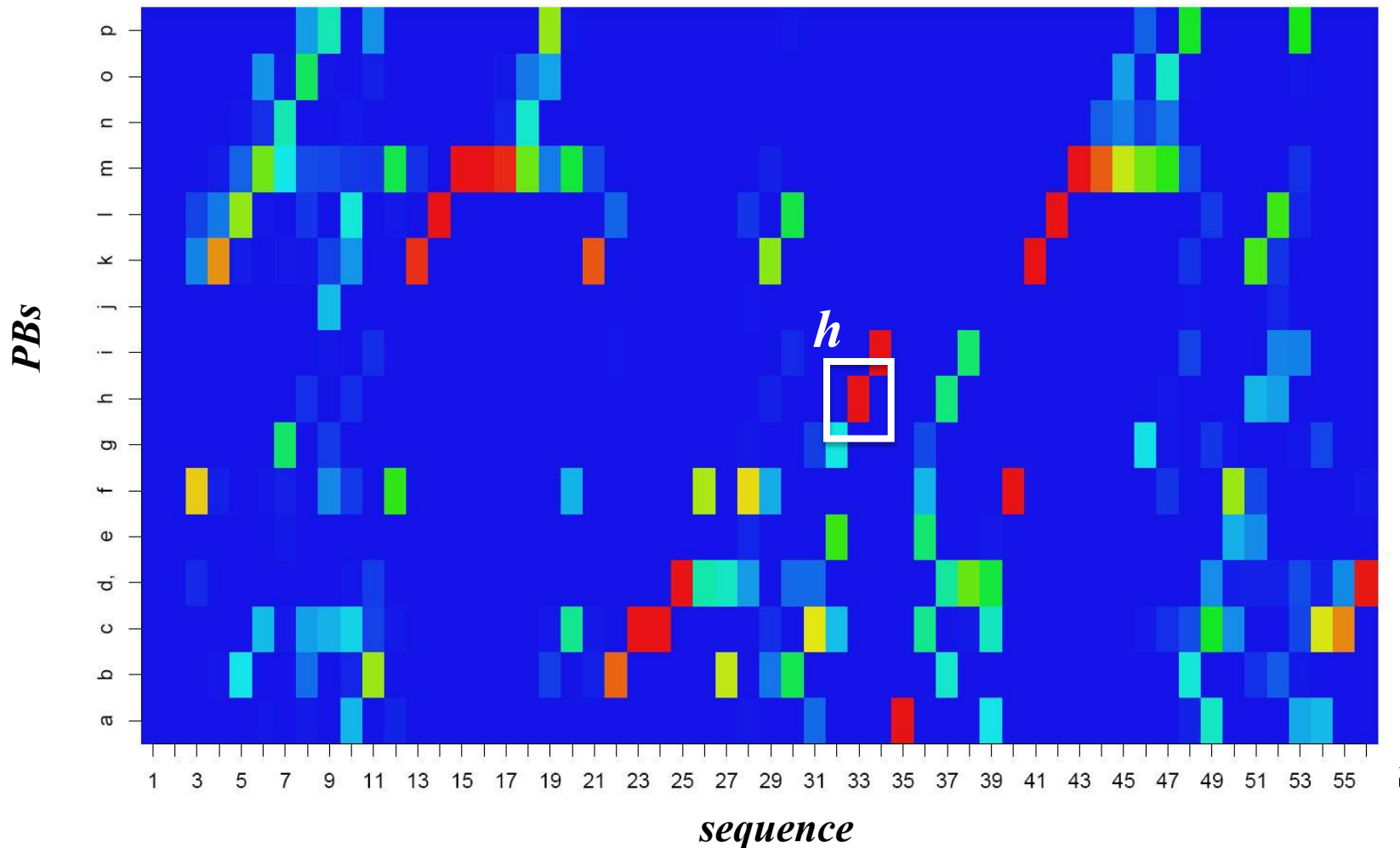
➤ How to use Protein Blocks here ?



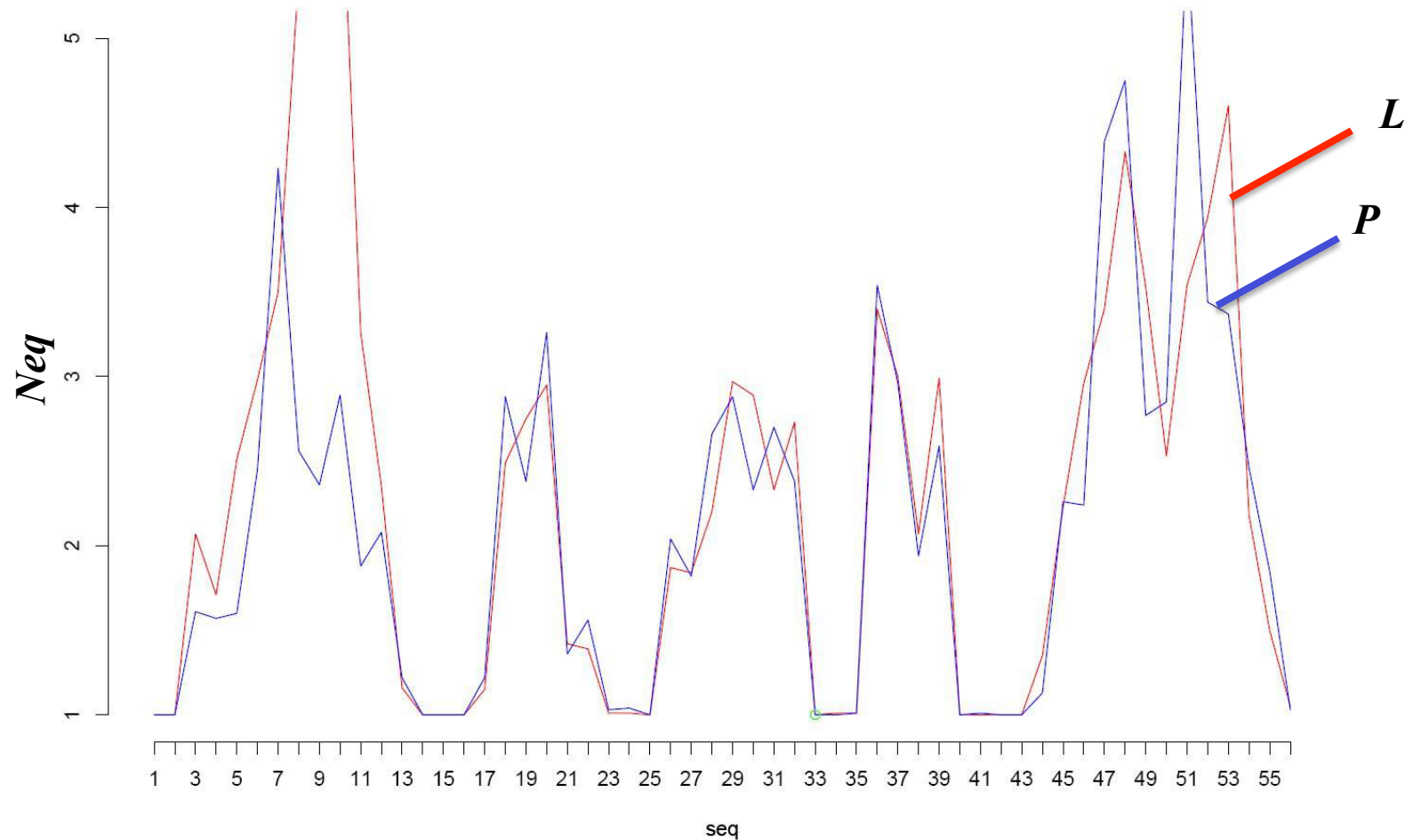
ZZfklmnopabfklmmnpfkbccdfblkbcchiaeh  
 ia fklmmnopafklpccddddddfbdcdddfblmb  
 ckbccdfbdcfbdcddddehiacddddddddehiac  
 ddddddfbfklmmmmmmmmgoj lmmmmmmnpcb  
 cdddddddddffklpcfbd fklmnomclmmmbdcdd  
 debiadfbdc bfbklmmmmmmmmccddd fblbdcfk  
 lmmmmmmommmnopafkbccddddfbdcdehkl  
 mmnopabccehpac bdfkopa fblmmpccdfklmmm  
 mmmnopacdddbdfklmmmmmmmmmmghiacddd  
 fklmmmmmmmmmmmmmmnccdddehiafbcccdd  
 ddfehjbccehiabdehiaghiacddddddddehjd  
 fddddddddehiabdcddddddfbdfklmmpcfbg  
 cknojhiaccehhiacdehiacjkopafgehjlmng  
 jiafbfblkbcehiopaccehhiacdfbdchiaaedj  
 kopaccblbccfb lpehhlpaccehhiacdehiacj  
 kopacfbfklmmgfkopghhlpaccehhiacdfbghi  
 edjkopaccbccfklmmmmmmmmnopajklmmmk  
 lmmmomcdcddfbdcdebiacdcdddddfkopacdd  
 ddddfkopacddd fbdccddZZ



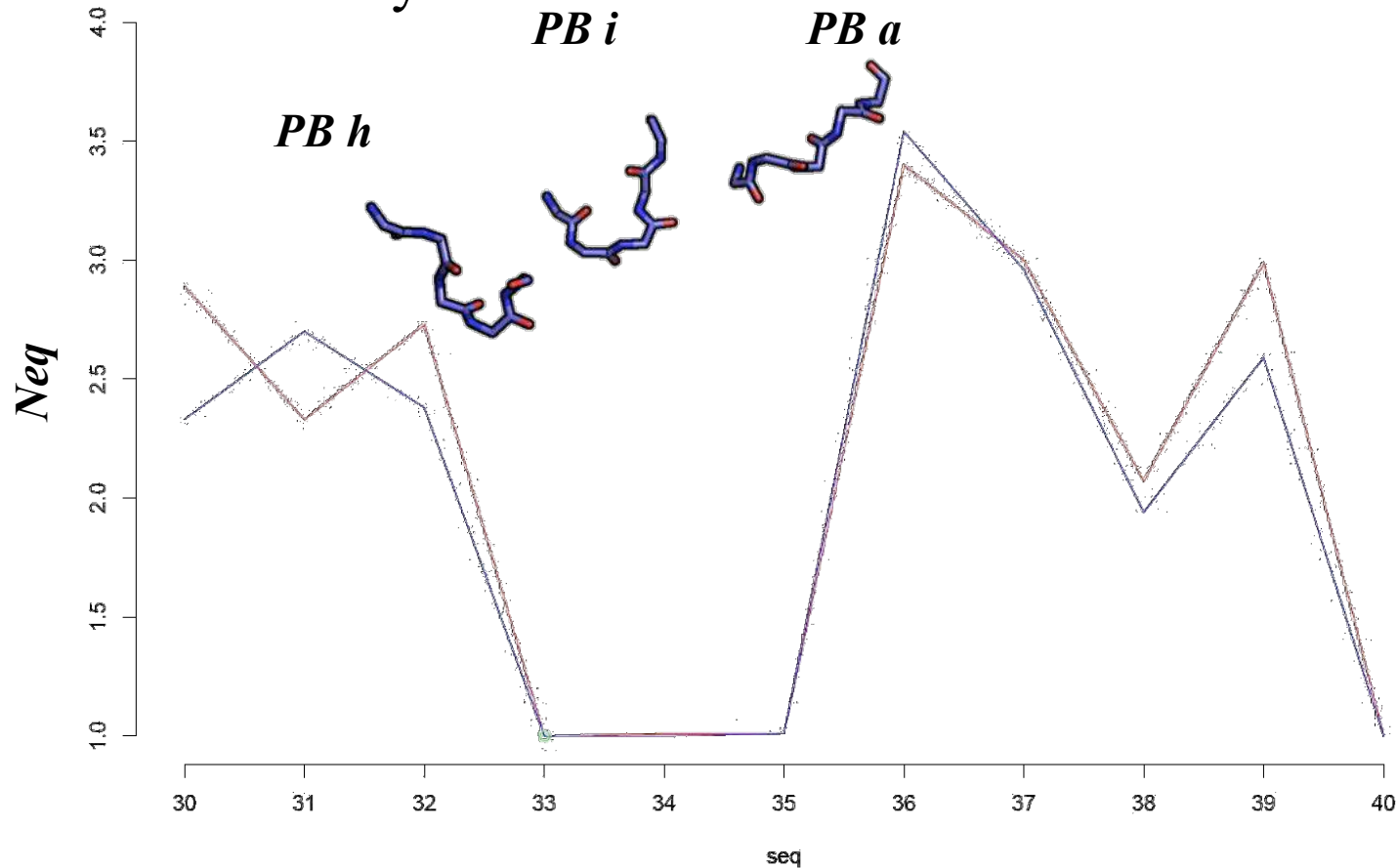
- Then count for every snapshots, the frequency of PBs



- Both sets of simulations are quite similar



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



## First Conclusion

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

Jallu V, Poulain P, Fuchs PF, Kaplan C, de Brevern AG (2012) Modeling and Molecular Dynamics of HPA-1a and -1b Polymorphisms: Effects on the Structure of the  $\beta 3$  Subunit of the  $\alpha \text{IIb}/\beta 3$  Integrin. *PLoS ONE*. 7(11):e47304.

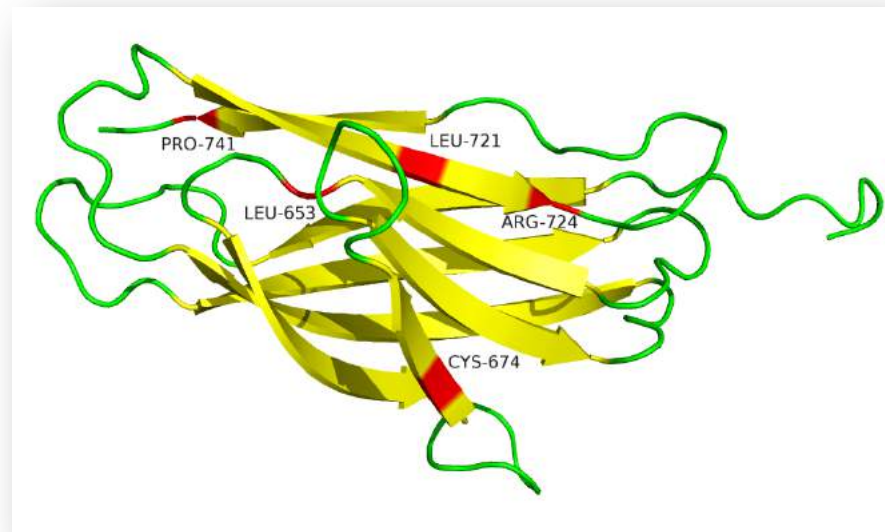
Jallu V, Poulain P, Fuchs PF, Kaplan C, de Brevern AG (2014) Modeling and molecular dynamics simulations of the V33 variant of the integrin subunit  $\beta 3$ : structural comparison with the L33 (HPA-1a) and P33 (HPA-1b) variants. *Biochimie*. 105:84-90.

- 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:



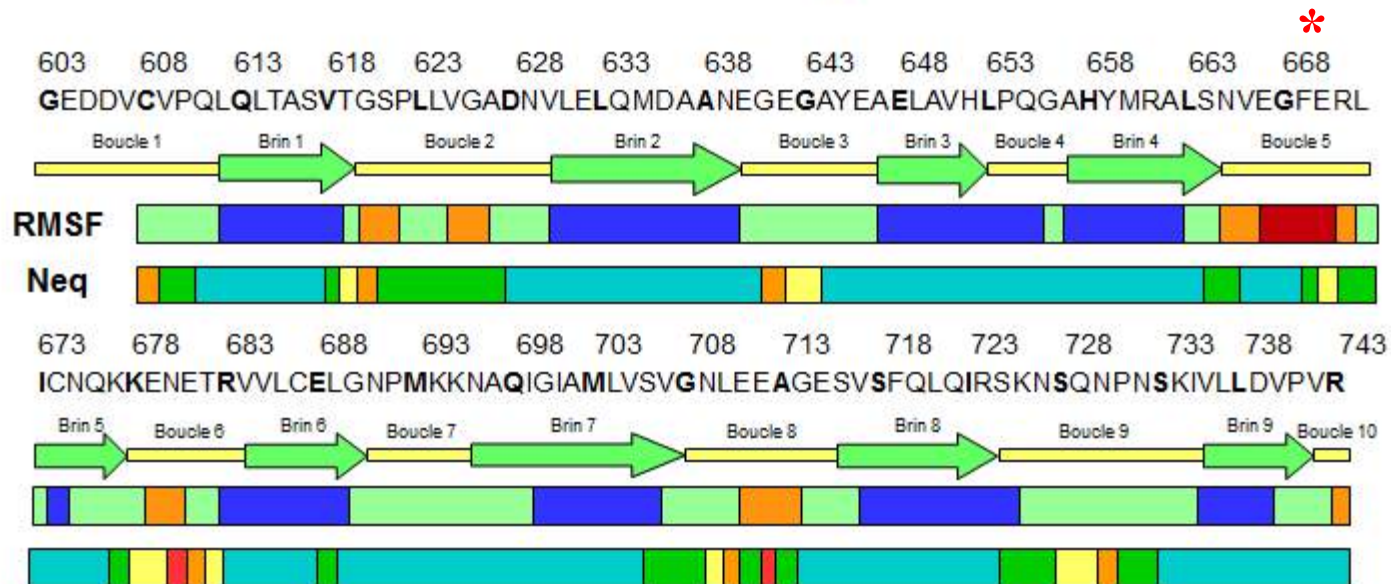
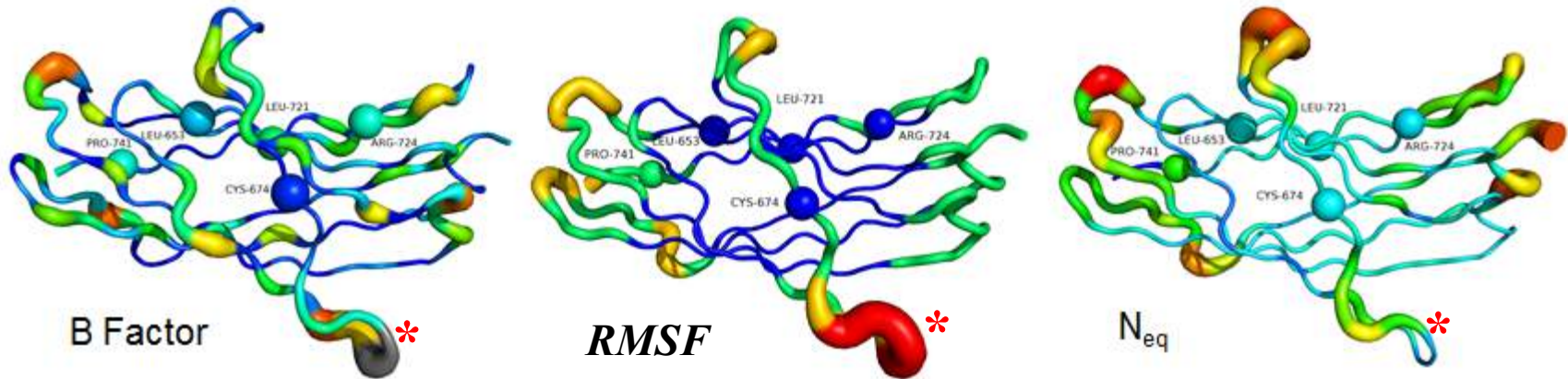
Matthieu Goguet  
(master 2 student)

- L653R
- C674R
- L721V / L721R
- R724P / R724Q
- P741R

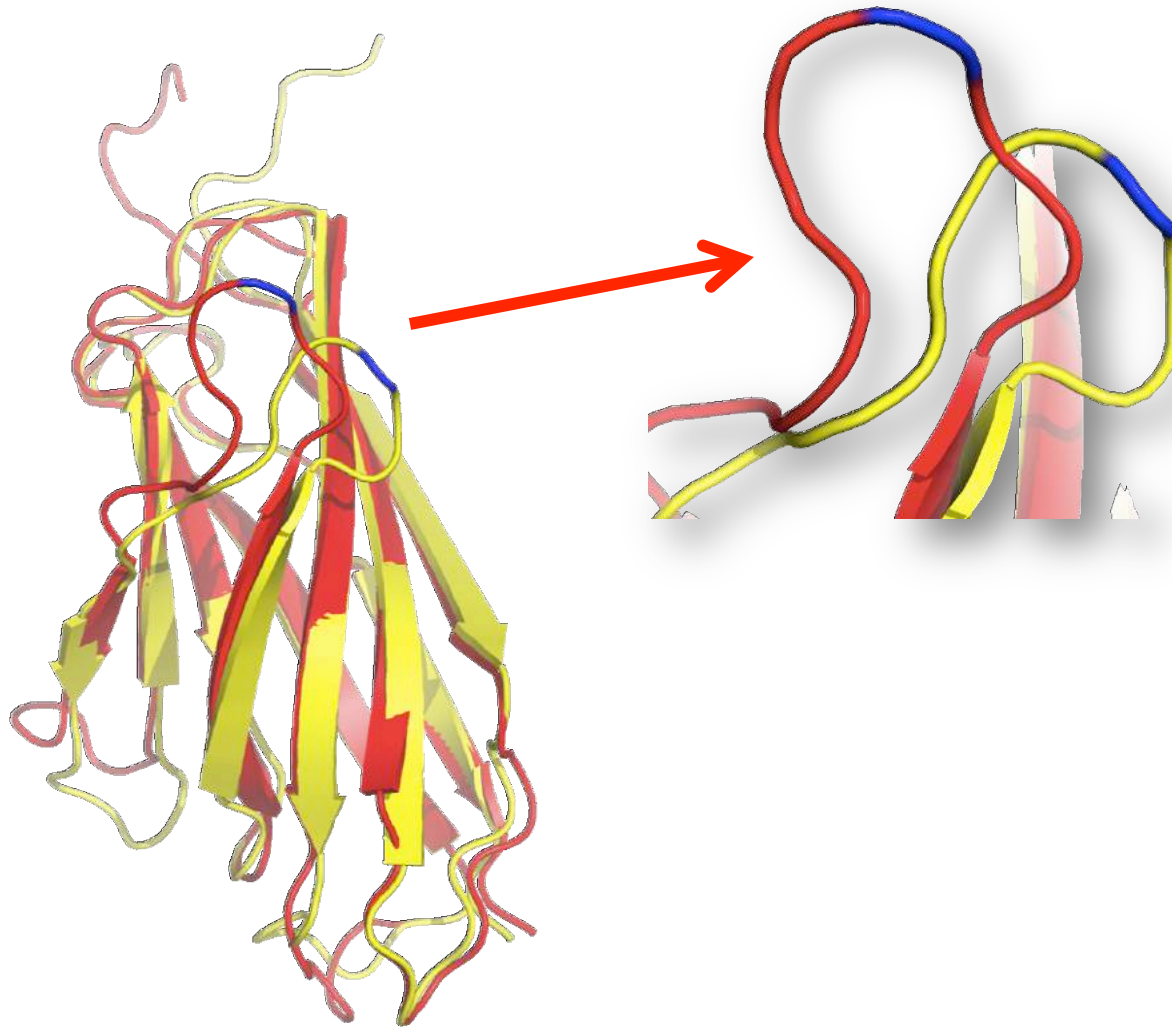


- MD : simulation *in silico* of a molecule's behavior.
- Gromacs 4.5.7 / Force field OPLS-AA / all atom
- 8 systems : 1 WT + 7 variants
- For each system : 11 independent MDs, 850ns simulation time

➤ Analysis of WT Calf-1

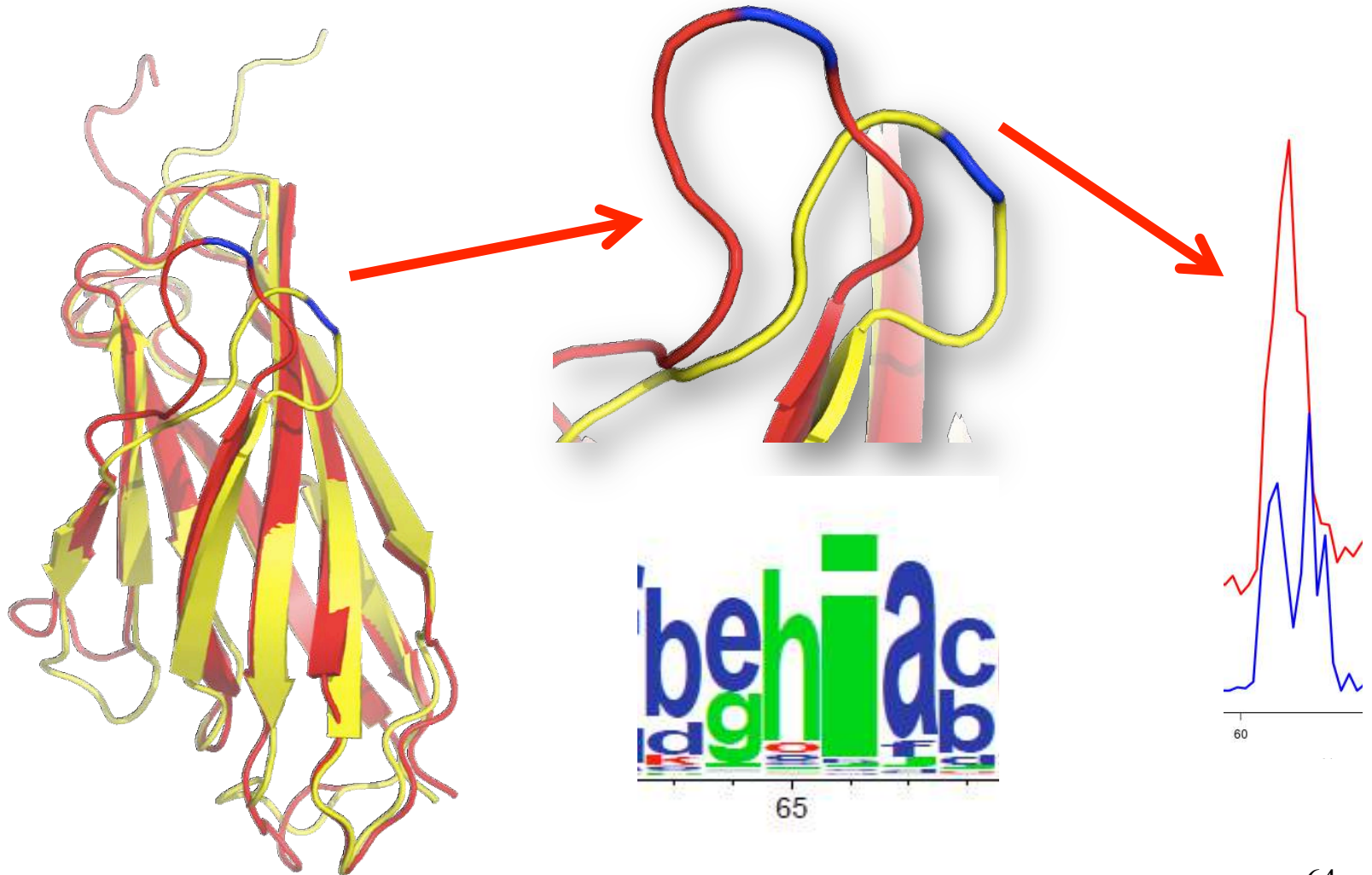


# Structural alphabet : analyses of flexibility

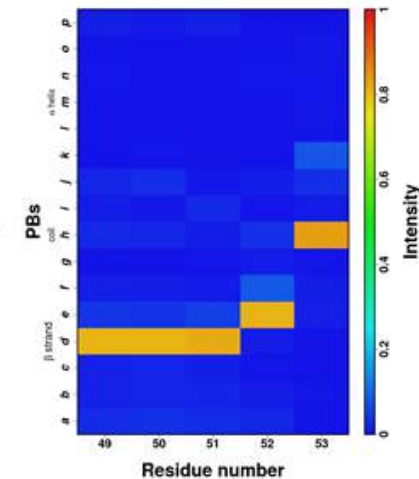
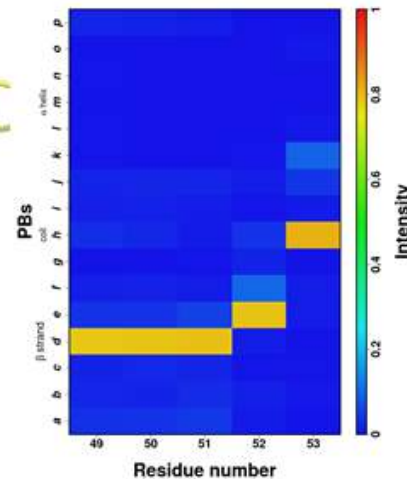
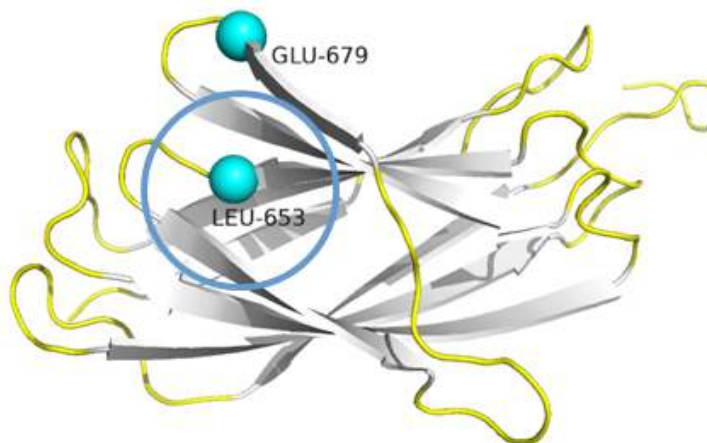




# Structural alphabet : analyses of flexibility



## Variant L653R : Mutation site



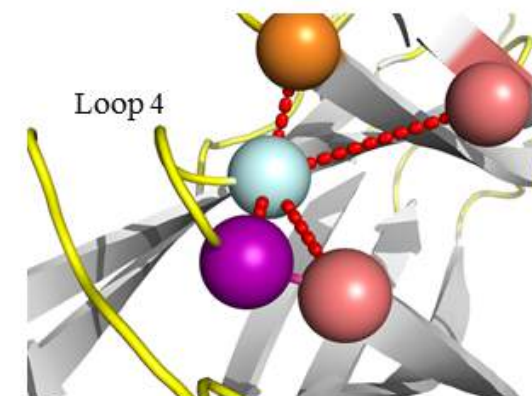
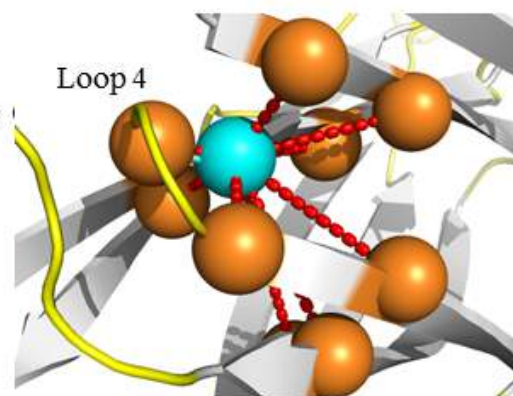
WT

Variant

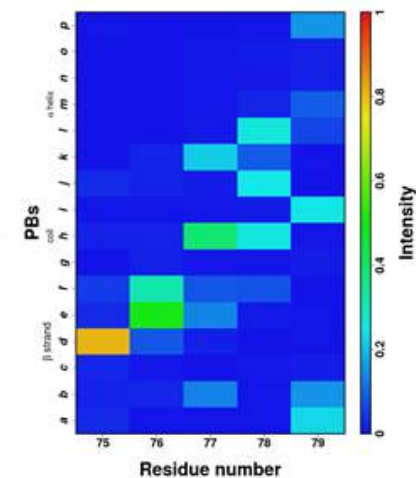
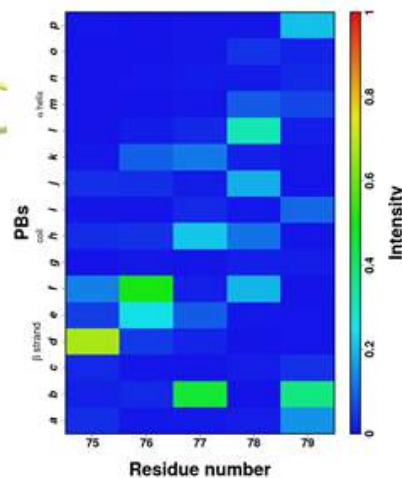
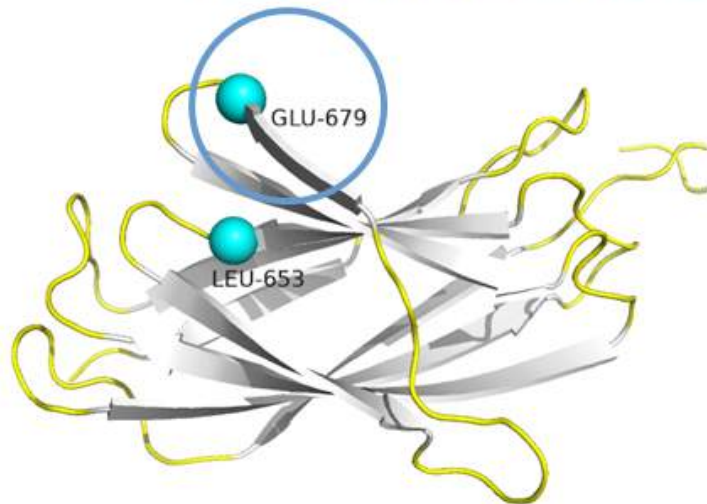
WT : 9 interactions

Variant : only 1 conserved but 3 news to

→ No structural changes



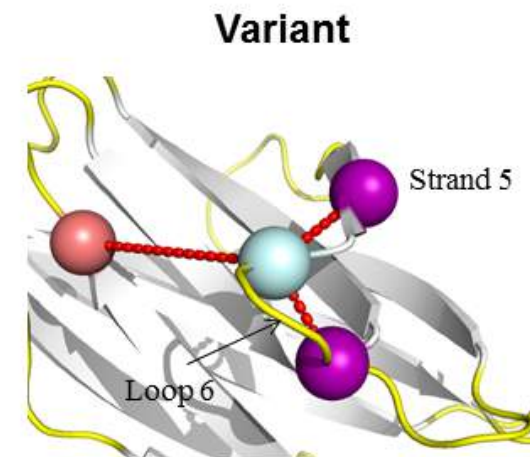
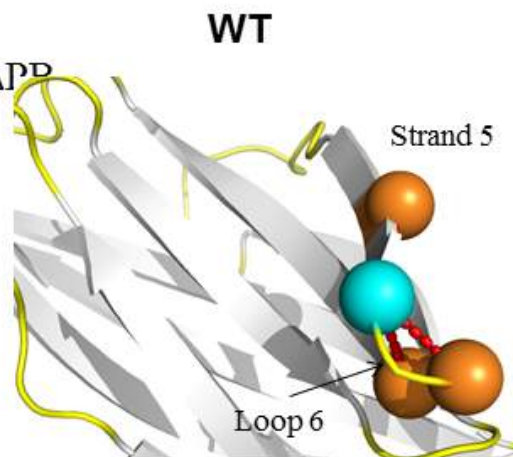
## Variant L653R : position E679



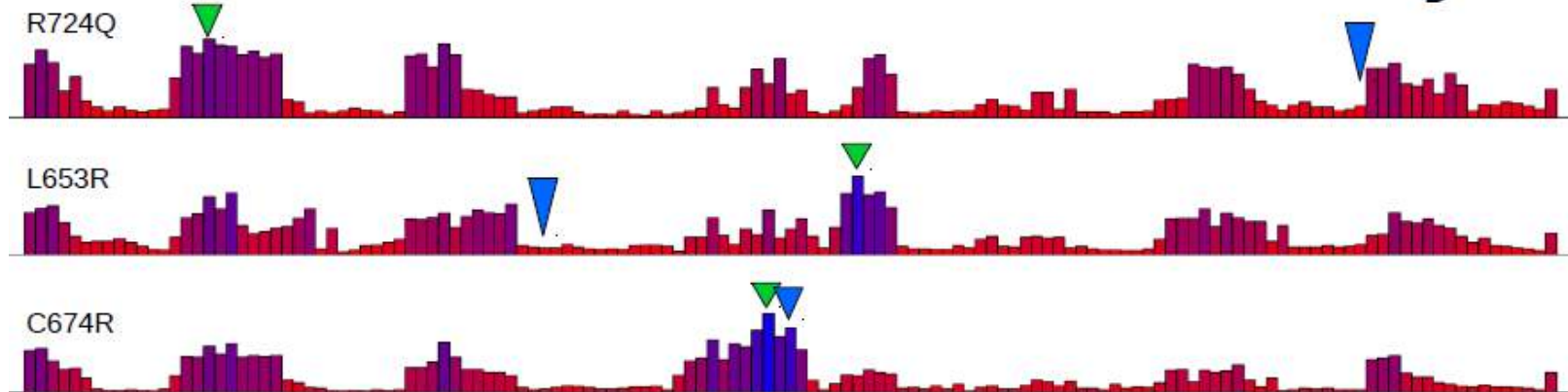
$N_{eq}$  similar in WT and variant but high  $\Delta PP$

Interaction reshuffle (no conservation)

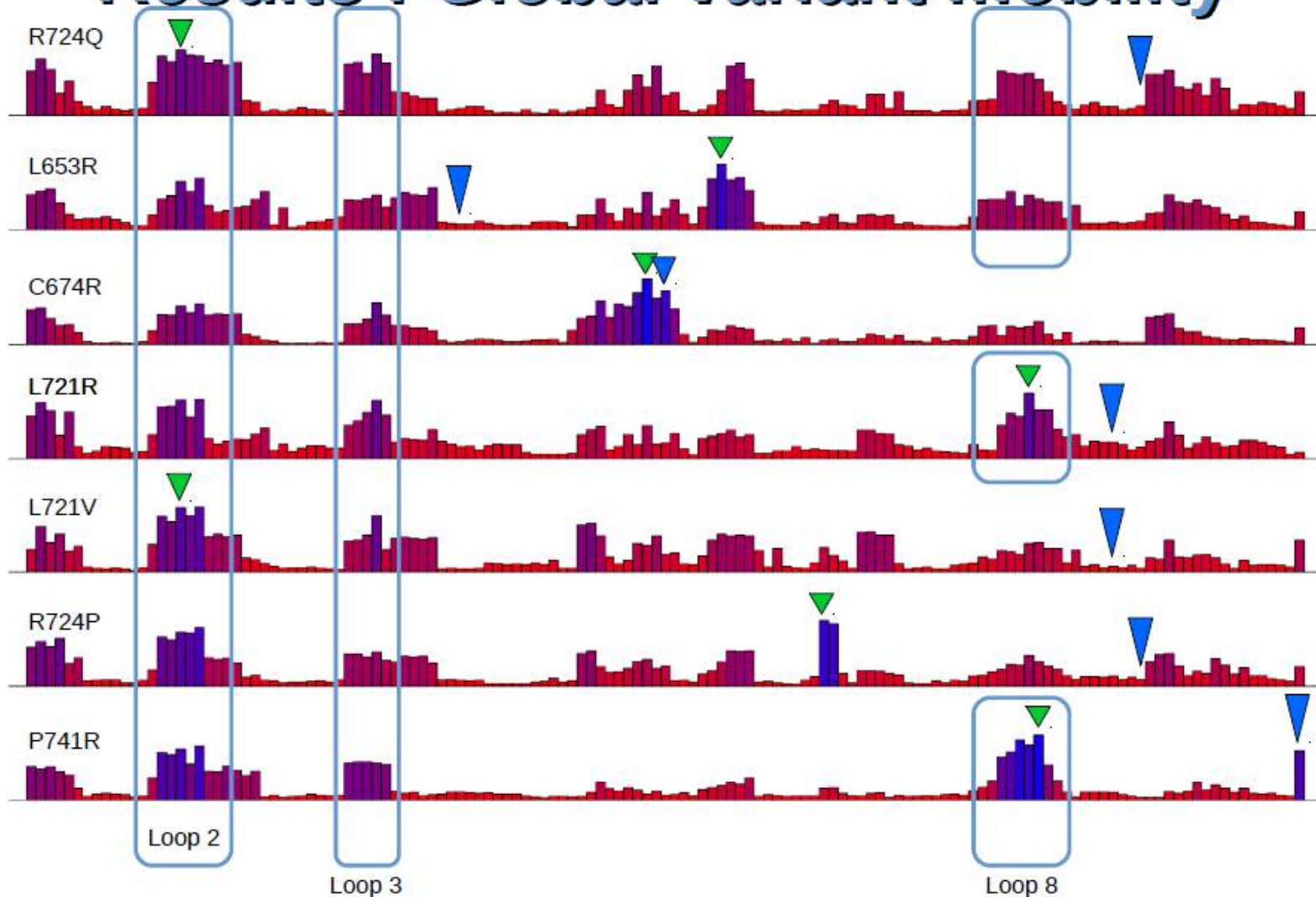
→ Structural desorganization



## Results : Global variant mobility



## Results : Global variant mobility



## Second Conclusion

- 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 ?

Goguet M, Narwani TJ, Petermann R, Jallu V, de Brevern AG (2017) In silico analysis of Glanzmann variants of Calf-1 domain of  $\alpha$ IIb $\beta$ 3 integrin revealed dynamic allosteric. *Sci Rep.* 7(1):8001.

➤ Last development: ArXiv & 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 / +

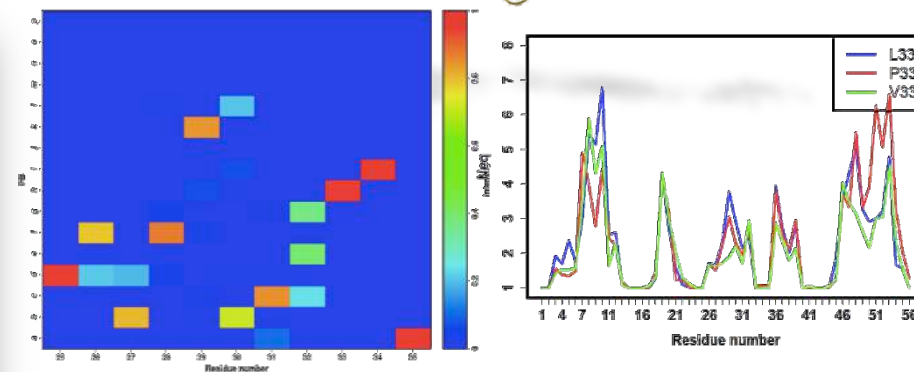
Improved regression tests, demo and doc for PBclust.py

pierrapo authored 26 days ago latest commit b2659d5e70

- 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 .coveragers 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
- PBsubstitution\_matrix.dat New tool to cluster sequences based on PBs similarities 2 years ago
- PBstat.py Merge pull request #56 from barnoud/module\_clust 3 months ago
- PBlib.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

Code Issues 11 Pull requests 0 Pulse Graphs

HTTPS clone URL https://github.com/ Clone in Desktop Download ZIP



Barnoud J, Santuz H, Craveur P, Joseph AP, Jallu V, de Brevern AG, Poulain P (2017) PBxplore: A Tool To Analyze Local Protein Structure And Deformability With Protein Blocks, *PeerJ. in press.*

# Inserm Structural alphabet : prediction of flexibility

webservice

**PredyFlexy** predicts from the sequence the protein flexibility, this last being defined from data coming from X-ray structures and molecular dynamics simulation ([http://www.dsimb.inserm.fr/dsimb\\_tools/predyflexy/](http://www.dsimb.inserm.fr/dsimb_tools/predyflexy/))



Bornot A, Etchebest C, de Brevern AG (2011) Predicting Protein Flexibility through the Prediction of Local Structures. *Proteins*. **79**(3): 839-52.

de Brevern AG, Bornot A, Craveur P, Etchebest C, Gelly J-C (2012) PredyFlexy: Flexibility and Local Structure prediction from sequence. *Nucleic Acid Res.* **40**:W317-22.

**PredyFlexy : Flexibility and Local Structure prediction from sequence**

Home  
Contacts  
About Method  
Example  
Download  
DSIMB

### Introduction

This server is designed to predict local protein structures and protein flexibility from its sequence. Results can be visualised at the amino acid level through a table and graphics.

### Protein Local Structure Prediction

It is now admitted that the folded state of proteins, that is, the native 3D structure, can be described by a limited set of recurring local structures (Flexee et al., Trends Biochem. Sci., 2005). This observation led to the development of fragment libraries designed to characterize in the most suitable way, the local structures of all proteins with known 3D structures. These libraries consist in a finite set of representative structural fragments. Nowadays, when no homologue protein is available, the most successful methods for predicting global 3D protein structures use fragment assembly techniques.

A library of 120 3D structural prototypes encompassing all known local protein structures has been developed (Barnos et al., Proteins, 2006). These Local Structure Prototypes (LSPs) were mean representative fragments of 120 overlapping structural classes of 11-residues fragments. They ensured a good quality of approximation. An associated local structure prediction method from sequence was also created. Its principal interest was to propose a limited number of relevant structural candidates for a given target sequence.

Recently, we achieved a balanced improvement of the prediction rate by coupling evolutionary information with support vector machines (SVMs). A very satisfying correct prediction rate of 63.1% was obtained for 5 proposed candidates. (Bornot et al., Proteins, 2009). This prediction method is implemented in this web service.

### Protein Flexibility Prediction

In the same way, protein structures are not rigid macromolecules. We analysed local structure flexibility features in proteins by relying on: (i) B-factors from X-ray experiments and (ii) backbone fluctuations in solution observed in molecular dynamics simulations. Finally, an original flexibility prediction method from sequence was developed (Bornot et al., Proteins, 2011). Three classes of flexibility are considered. Very low confusion between rigid and flexible classes was observed. Only 13.5% of rigid residues were predicted as flexible and reciprocally, only 5.8% of flexible ones were predicted as rigid. This method is implemented for this web service.

### Launch a prediction

Paste your sequence file (fasta file format):

```
PKKMGID  
MELKIDTPNNARHILVYVSTFPMIICDAKDDVYDAGDGLLNDFDQMGVHLS  
DQPIELVYVGVAVLQGLSLSHLSAPVBLA75LANITFPGIHALLSFGVSRHS  
ALINMLVYQFTLIVFADQWHTKQNNMAYTQSDIRBQFAPVBLA75LANITF  
PFTKNAITVLAELHSDIITQDQKQANITMSTLSSQIITLQFPAKALIK  
CEAHNMLIIDEAGTQVQVQVDEKADQDQVFDLISLRELGALPLAA75YDALS  
ENANMLQFTTETVSDHSDANMLDLPQDQDQVANKNSQMLHLLQDSEH  
DCDQVGRGLLQGLVYVHSDIKFAAGLGAITDSEWMLSNLIVLPHQGVTR  
IAPPVYVDEHIEGLSGLQATKRAL
```

**PREDICTION**



# Conclusion

- 
- DSIMB is working on many interesting topics 😊 !
  - I hope you have a nice (partial) view of our methodological developments.
  - Essential to have a good knowledge of biology and methodology to do good applications (not developed here)

- Three reviews to have a good views of our interest
1. SOA of secondary structures and structural alphabet

Offmann B., Tyagi M., de Brevern A.G. (2007)  
Local Protein Structures. *Current  
Bioinformatics* 2: 165-202.

## Local Protein Structures

Bernard Offmann<sup>1</sup>, Manoj Tyagi<sup>1,4</sup> and Akexandre G. de Brevern<sup>4,2</sup>

<sup>1</sup>Laboratoire de Biochimie et Génétique Moléculaire, Université de La Réunion, 15, avenue René Cassin, BP7151, 97715 Saint Denis Messag Cedex 09, La Réunion, France

<sup>2</sup>Équipe de Bioinformatique Génomique et Moléculaire (EBGM), INSERM UMR-S 726, Université Paris Diderot, case 7112, 2, place Jussieu, 75251 Paris, France

<sup>3</sup>Present Address: Computational Biology Branch, National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), 8600 Rockville Pike, Bethesda, MD 20894, USA

**Abstract:** Protein structures are classically described as composed of two regular states, the  $\alpha$  helices and the  $\beta$ -strands and one non-regular and variable state, the coil. Nonetheless, this simple definition of secondary structures hides numerous limitations. In fact, the rules for secondary structure assignment are complex. Thus, numerous assignment methods based on different criteria have emerged leading to heterogeneous and diverging results. In the same way, 3 states may over-simplify the description of protein structure, 50% of all residues, i.e. the coil, are not genuinely described even when it encompasses precise local protein structures. Description of local protein structures have hence focused on the elaboration of complete sets of small prototypes or "structural alphabets", able to analyze local protein structures and to approximate every part of the protein backbone. They have also been used to predict the protein backbone conformation and in *ab initio* *de novo* methods. In this paper, we review different approaches towards the description of local structures, mainly through their description in terms of secondary structures and in terms of structural alphabets. We provide some insights on their potential applications.

**Keywords:** Secondary structure, protein folds, structure-sequence relationship, structural alphabet, protein blocks, molecular modeling, *ab initio*.

### INTRODUCTION

Protein folds are often described as a succession of secondary structures. Their repetitive parts ( $\alpha$ -helices and  $\beta$ -strands) have been intensively analyzed since their initial description by Pauling and Corey [1]. Nonetheless, this description of the 3D structures in terms of secondary structures is not simple and different major drawbacks must be carefully addressed. Indeed, the rules for secondary structure assignments are not trivial, and so numerous assignment methods based on different criteria have emerged. The greatest discrepancies are found mainly at the caps of the repetitive structures. These small differences can result in different lengths for the repetitive structures, depending on the algorithm used. In addition, a classification of the backbone conformation limited to 3 states (the classical repetitive secondary structures and coils) does not precisely describe the protein structures, because it fails to describe the relative orientation of connecting regions. Besides, the coil state covering almost 50% of all residues corresponds to a large set of distinct local protein structures.

Thus, to circumvent these difficulties, other approaches were developed. They led to a new view of 3D protein structures which are now thought to be composed of a combination of small local structures or fragments, also called

prototypes. A given complete set of prototypes defines a "structural alphabet". Different groups described these local protein structures according to different criteria. Structural alphabets have been used to approximate and analyze local protein structures and to predict backbone conformation.

This paper is divided in two parts. First, we focus on the detailed analysis of known secondary structures with respect to the different secondary structure assignment methods. Second, we present the complete panoramas of known structural alphabets, i.e., libraries of protein local structures used in *ab initio* methods.

Thus, we present in the *Secondary structure section* the classical and less represented repetitive structures, the irregularities within these structures, the different kinds of turns, the Polyproline II and the loops. We focus on the problematic issue of secondary structure assignment and conclude on the step from secondary structure to 3D. The *Structural alphabets section* shows firstly the structural libraries dedicated to the structure approximation, secondly the different developed prediction methods based on structural alphabet description and finally some applications.

### SECONDARY STRUCTURES

**Introduction.** The description of protein structures in terms of secondary structures is widely used for analysis or prediction purposes (see the example of Human Liver Glycogen Phosphorylase A [2] taken from the Protein DataBank [3] in Fig. 1). The secondary structures are composed of well-known  $\alpha$ -helix [4] and  $\beta$ -sheet [5]. Secondary structure assignment is directly implemented in all 3D structure visu-

\*Address correspondence to this author at the Equipe de Bioinformatique Génomique et Moléculaire (EBGM), INSERM UMR-S 726, Université Paris Diderot, case 7112, 2, place Jussieu, 75251 Paris, France. Tel: (33) 1 44 27 77 31; Fax: (33) 1 43 26 38 30; E-mail: adbrevern@ebgm.jussieu.fr

- Three reviews to have a good views of our interest
- 1. SOA of secondary structures and structural alphabet
- 2. SOA on Protein Blocks

Joseph AP, Agarwal G, Mahajan S, Gelly J-C, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadié H., Schneider B, Etchebest C, Srinivasan N, de Brevern AG (2010) A short survey on Protein Blocks. *Biophysical Reviews* 2(3):137-145.

Biophys Rev (2010) 2:137–145  
DOI 10.1007/s12551-010-0036-1

REVIEW

### A short survey on protein blocks

Agnel Praveen Joseph · Garima Agarwal · Swapnil Mahajan · Jean-Christophe Gelly · Lakshmpuram S. Swapna · Bernard Offmann · Frédéric Cadet · Aurélie Bornot · Manoj Tyagi · Hélène Valadié · Bohdan Schneider · Catherine Etchebest · Narayanaswamy Srinivasan · Alexandre G. de Brevern

Received: 20 April 2010 / Accepted: 2 July 2010 / Published online: 5 August 2010  
© International Union for Pure and Applied Biophysics (IUPAB) and Springer 2010

**Abstract** Protein structures are classically described in terms of secondary structures. However, even if the regular secondary structures have relevant physical meaning, their recognition based on atomic coordinates has a number of important limitations, such as uncertainties in the assignment of the boundaries of the helical and  $\beta$ -strand regions. In addition, an average of about 50% of all residues are assigned to an irregular state, i.e., the coil. These limitations have led different research teams to focus on abstracting the conformation of the protein backbone in the localized short stretches. To this end, different geometric measures are being used to cluster local stretches in protein structures in a chosen number of states. A prototype representative of the local structures in each cluster is then generally defined.

Agnel Praveen Joseph and Garima Agarwal contributed equally to this article.

A. P. Joseph · J.-C. Gelly · A. Bornot · C. Etchebest · A. G. de Brevern (✉)  
Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), Université Paris Diderot Paris 7, 6, rue Alexandre Cabanel, Paris Cedex 15 75739, France  
e-mail: alexandre.debrevern@univ-paris-diderot.fr

A. P. Joseph · J.-C. Gelly · A. Bornot · C. Etchebest · A. G. de Brevern  
Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), INSERM, UMR-S 665, 6, rue Alexandre Cabanel, Paris Cedex 15 75739, France

A. P. Joseph · J.-C. Gelly · A. Bornot · C. Etchebest · A. G. de Brevern  
Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), Institut National de la Transfusion Sanguine (DVS), 6, rue Alexandre Cabanel, Paris Cedex 15 75739, France

G. Agarwal · S. Mahajan · L. S. Swapna · N. Srinivasan  
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

S. Mahajan  
National Centre for Biological Sciences, Tata Institute of Fundamental Research, UAS-GKVR, Campus, Bellary Road, Bangalore 560 065, India

B. Offmann · F. Cadet  
INSERM, UMR-S 665, Dynamique des Structures et Interactions des Macromolécules Biologiques (DSIMB), 15 Avenue René Cassin, BP 7151, 97715 Saint Denis Messag Cedex 09 La Réunion, France

B. Offmann · F. Cadet  
Faculté des Sciences et Technologies, Université de La Réunion, 15 Avenue René Cassin, BP 7151, 97715 Saint Denis Messag Cedex 09 La Réunion, France

M. Tyagi  
Computational Biology Branch, National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), 8600 Rockville Pike, Bethesda, MD 20894, USA

H. Valadié  
UMR 5168 CNRS-CEA-INRA-Université Joseph Fourier, Institut de Recherches en Technologies et Sciences pour le Vivant, 17 avenue des Martyrs, 38054 Grenoble Cedex 9, France

B. Schneider  
Institute of Biotechnology AS CR, Videnka 1083, 142 20 Prague, Czech Republic

Springer

# Conclusion

- Three reviews to have a good views of our interest

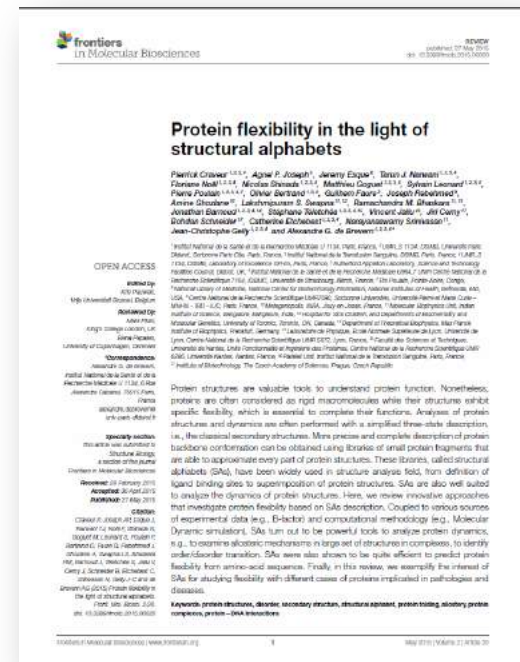
  1. SOA of secondary structures and structural alphabet
  2. SOA on Protein Blocks
  3. SOA on structural alphabet & flexibility



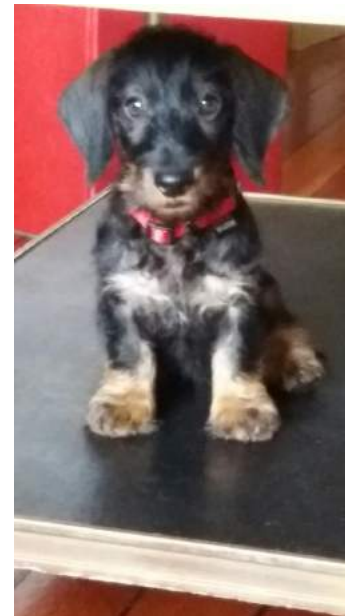
Craveur P, Joseph AP, Esque J, Narwani TJ, Noel F, Shinada N, Goguet M, Leonard S, Poulain P, Bertrand O, Faure G, Rebehmed J, Ghozlane A, Swapna LS, Bhaskara RM, Barnoud J, Teletchea S, Jallu V, Cerny J, Schneider B, Etchebest C, Srinivasan N, Gelly J-C, de Brevern AG (2015) Protein flexibility in the light of structural alphabets.

*Frontiers in Molecular Biosciences - Structural Biology.*

2:20.



➤ Thank you



---

# SUP SUP

- Various functions, various binding sites ...

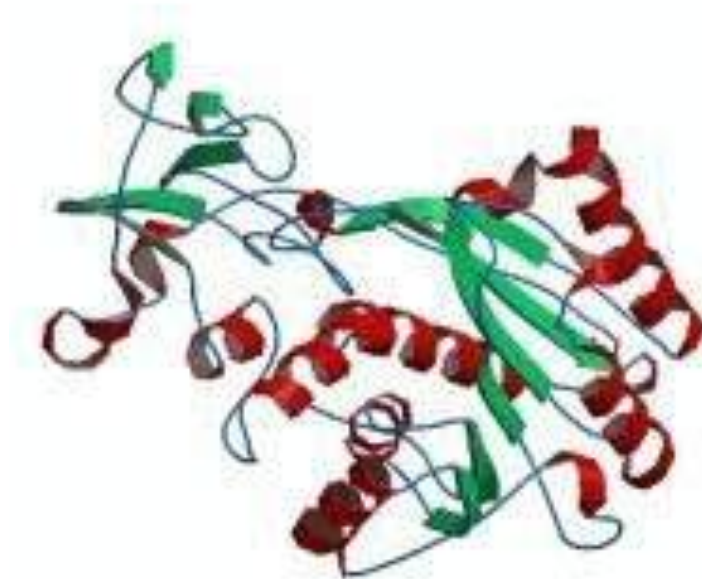


**René Roubíček** 1922 86

PROSTOROVÁ KOMPOZICE PRO  
ČESKOSLOVENSKÝ PAVILON NA SVĚTOVÉ  
VÝSTAVĚ EXPO 58 - II  
sklo, ocelová konstrukce  
realizace Studio Šimánek, Praha a sklárna  
Ajeto Lindava  
dar autora 2007

THREE-DIMENSIONAL COMPOSITION FOR THE  
CZECHOSLOVAK PAVILION AT THE WORLD  
EXHIBITION EXPO '58 - II  
Glass, steel frame  
Realized by Studio Šimánek, Prague, and Ajeto  
Lindava Glassworks  
Donated by the artist in 2007

# Secondary structures: some questions



No simple  
consensus  
between methods

**DSSP**

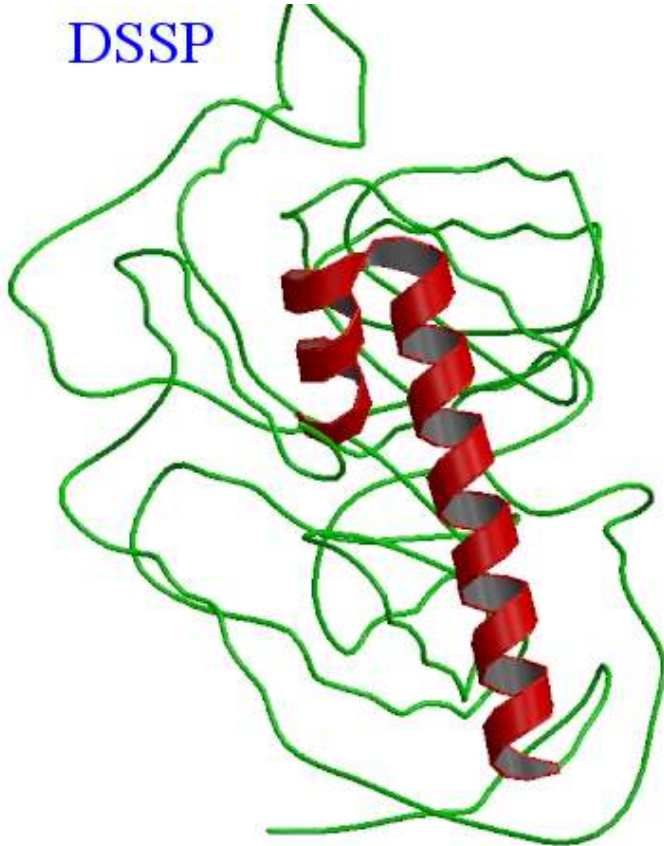
Assignment example for the *Hhai Methyltransferase* protein (PDB code :10MH) by DSSP, STRIDE, PSEA, DEFINE, PCURVE, XTLSSTR and SECSTR.

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

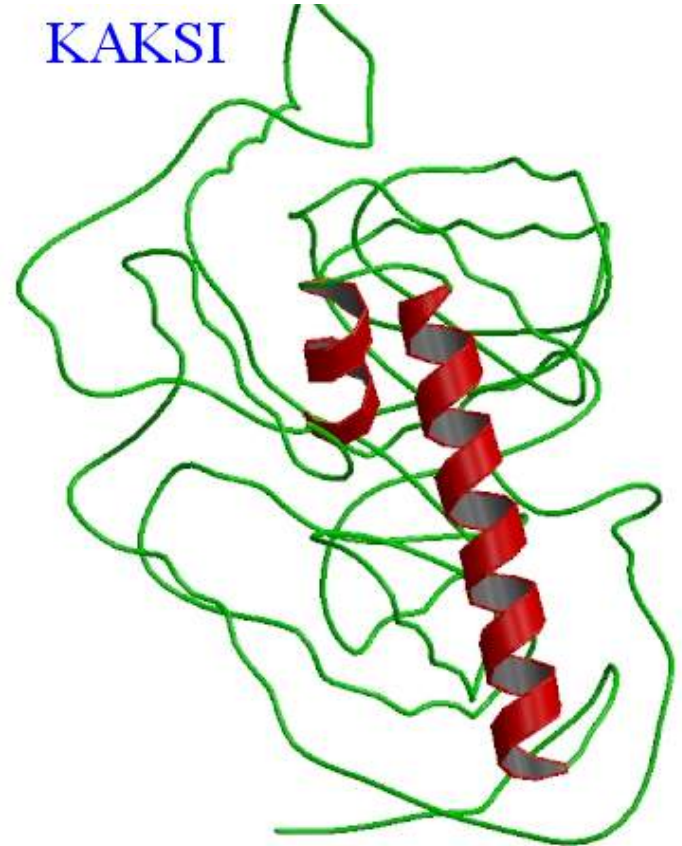


# Secondary structures: some questions

DSSP



KAKSI



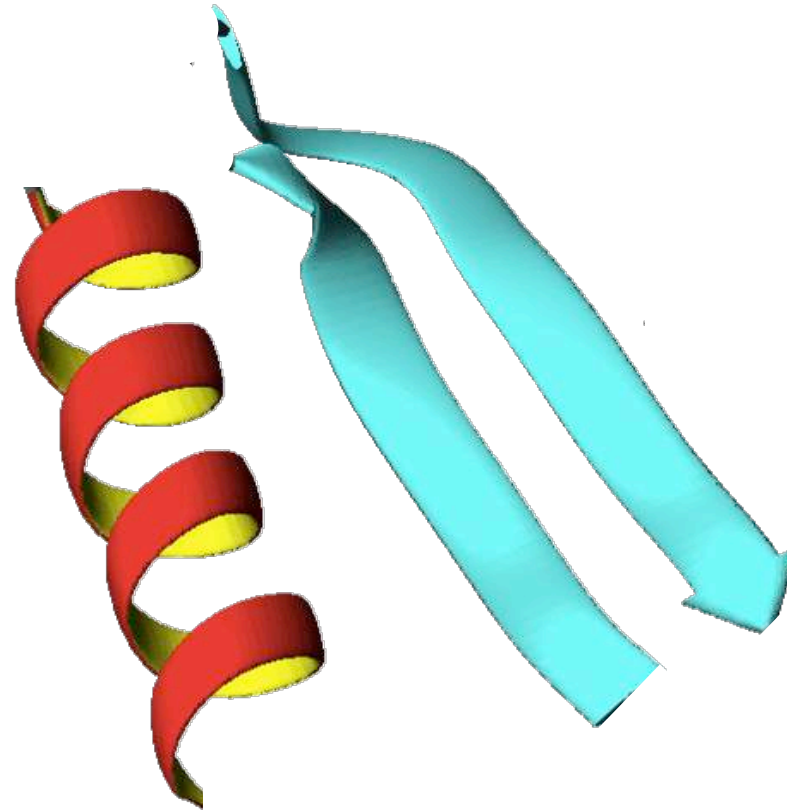
J. Martin, G. Letellier, J.F. Taly, A. Martin, A.G. de Brevern & J.F. Gibrat (*BMC Structural Biology*, 2005)

[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

---

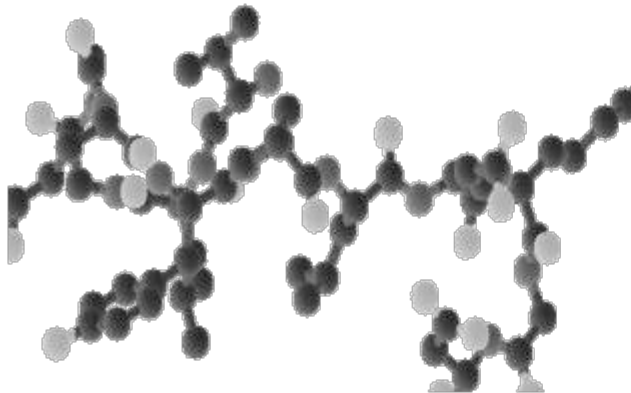
- It is possible to look at protein structures in a different ways
- At a local level.

---

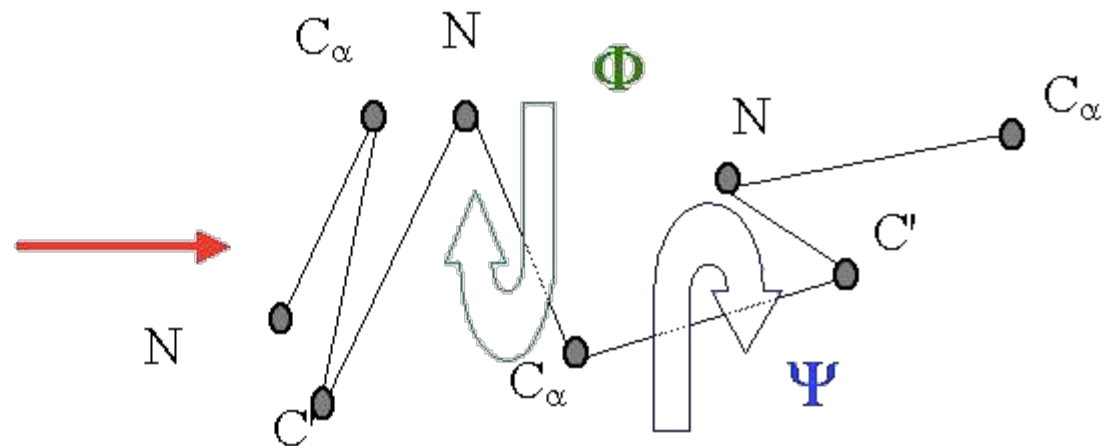
# STRUCTURAL ALPHABET

➤ Coding the structure

3D

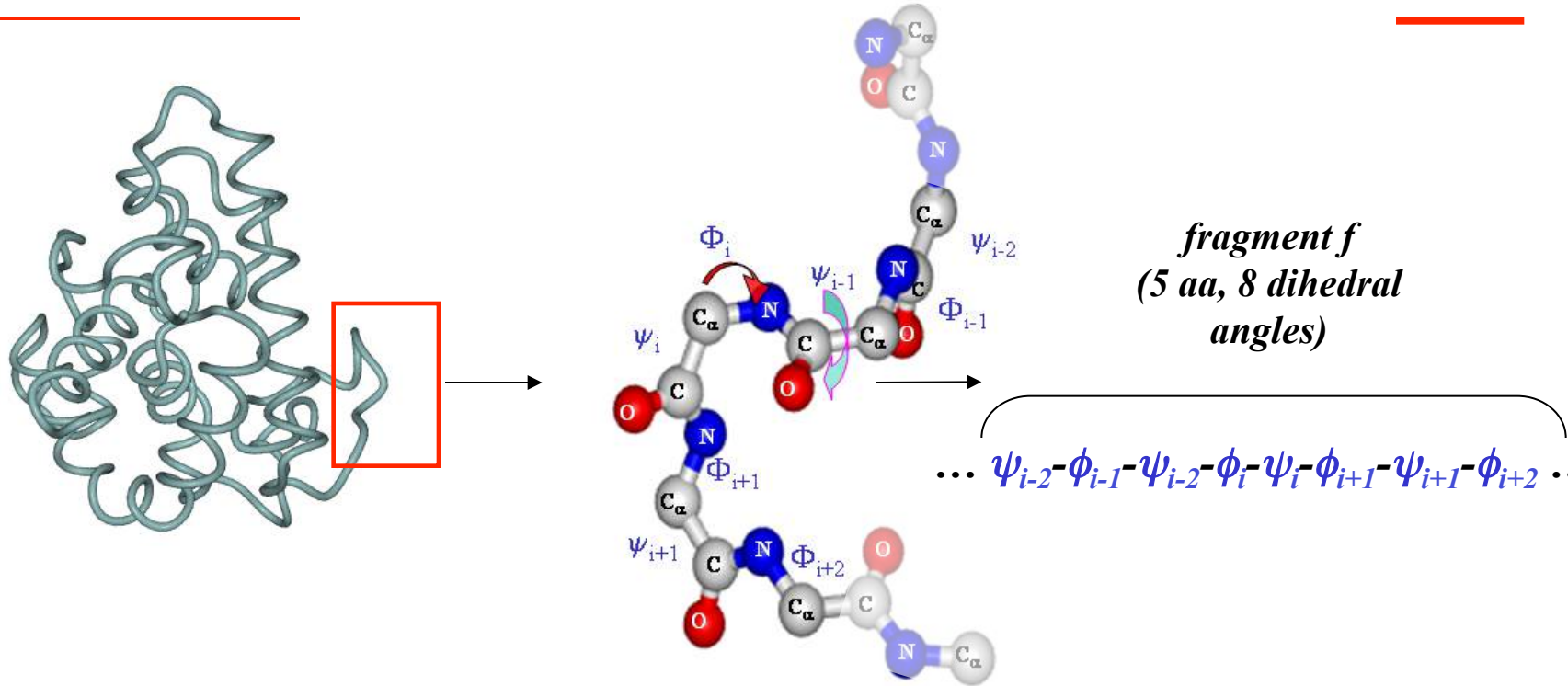


dihedral angles



*5 residues => 8 dihedral angles (φ, ψ)*

## Protein Blocks



# Protein Blocks

	$\psi_{i-2}$	$\phi_{i-1}$	$\psi_{i-1}$	$\phi_i$	$\psi_i$	$\phi_{i+1}$	$\psi_{i+1}$	$\phi_{i+2}$
PB a	41.14	75.53	13.92	-99.80	131.88	-96.27	122.08	-99.68
PB b	108.24	-90.12	119.54	-92.21	-18.06	-128.93	147.04	-99.90
PB c	-11.61	-105.66	94.81	-106.09	133.56	-106.93	135.97	-100.63
PB d	141.98	-112.79	132.20	-114.79	140.11	-111.05	139.54	-103.16
PB e	133.25	-112.37	137.64	-108.13	133.00	-87.30	120.54	77.40
PB f	116.40	-105.53	129.32	-96.68	140.72	-74.19	-26.65	-94.51
PB g	0.40	-81.83	4.91	-100.59	85.50	-71.65	130.78	84.98
PB h	119.14	-102.58	130.83	-67.91	121.55	76.25	-2.95	-90.88
PB i	130.68	-56.92	119.26	77.85	10.42	-99.43	141.40	-98.01
PB j	114.32	-121.47	118.14	82.88	-150.05	-83.81	23.35	-85.82
PB k	117.16	-95.41	140.40	-59.35	-29.23	-72.39	-25.08	-76.16
PB l	139.20	-55.96	-32.70	-68.51	-26.09	-74.44	-22.60	-71.74
PB m	-39.62	-64.73	-39.52	-65.54	-38.88	-66.89	-37.76	-70.19
PB n	-35.34	-65.03	-38.12	-66.34	-29.51	-89.10	-2.91	77.90
PB o	-45.29	-67.44	-27.72	-87.27	5.13	77.49	30.71	-93.23
PB p	-27.09	-86.14	0.30	59.85	21.51	-96.30	132.67	-92.91



*distance computation*  $\Rightarrow$  16 scores  $\longrightarrow$  *The smallest*

## Protein Blocks



*Example PB a*



>153L

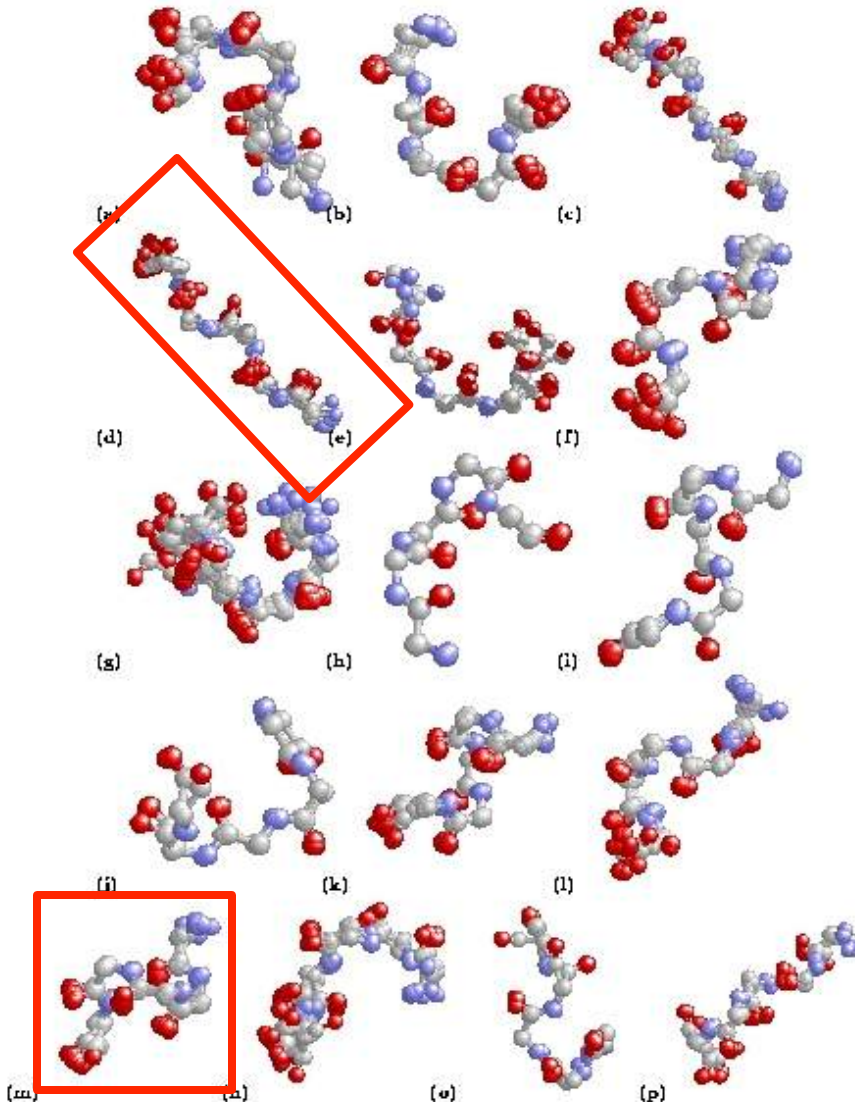
*ZZmnopfk1pccebjafklmmnnoꞑ*

l  
l  
l  
l

*The smallest*



## Structural alphabet



Pr. Serge Hazout

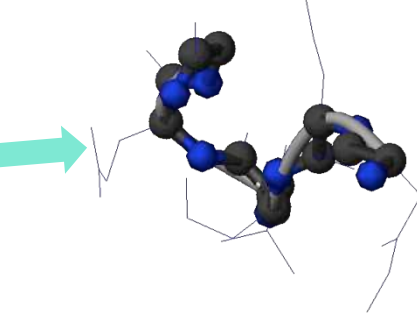
de Brevern A.G., Etchebest C. & Hazout, S. (*Proteins*, 2000).

16 Protein Blocks

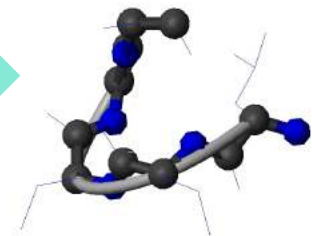
# Structural alphabet

➤ PBs & secondary structure

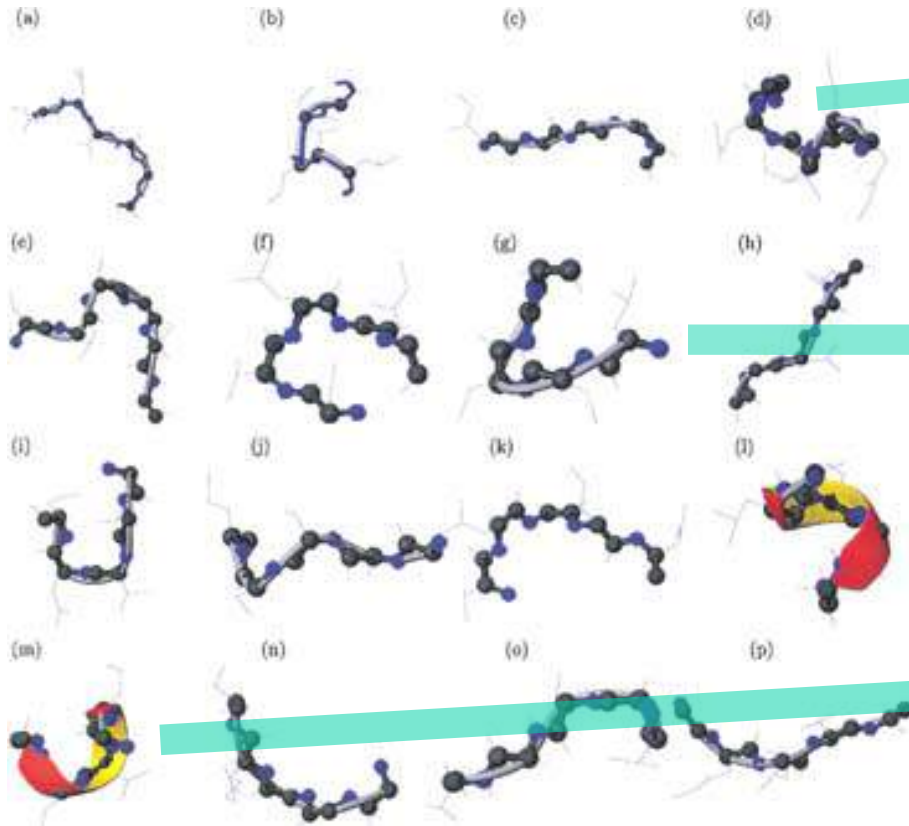
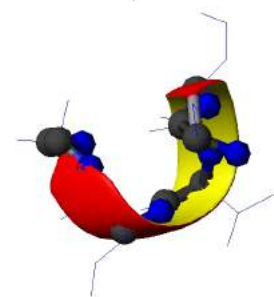
*PB d* 29.0% coil 71.0  $\beta$



*PB g* 2.0%  $\alpha$   
 76.2% coil  
 21.9%  $\beta$

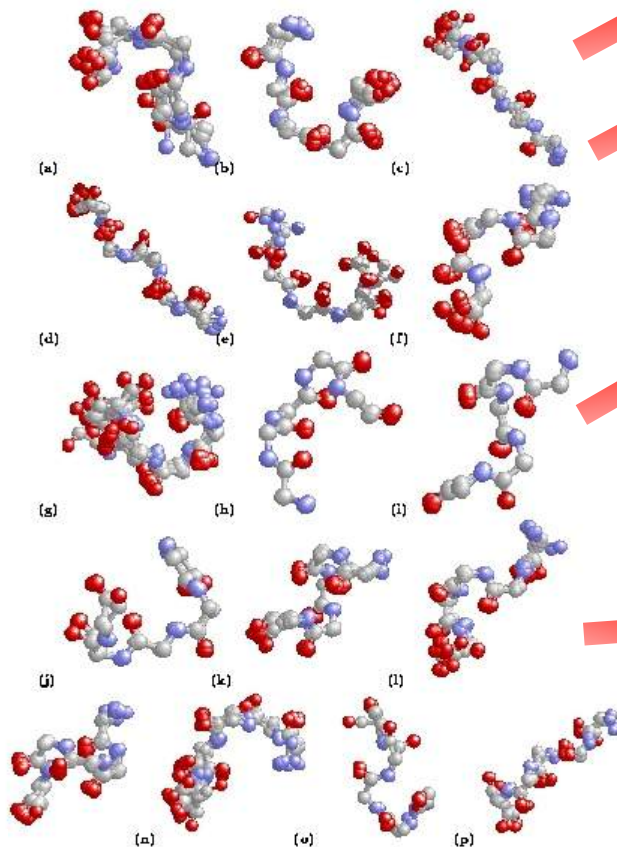


*PB m* 92.3%  $\alpha$  7.6% coil



# A structural alphabet

- <sup>1</sup> (*Proteins*, 2000)
- <sup>2</sup> (*ISB*, 2005)
- <sup>3</sup> (*CBIO*, 2007)
- <sup>4</sup> (*Biophys Rev*, 2010)
- <sup>5</sup> (*Proteins* 2006)
- <sup>6</sup> (*NAR*, 2006)
- <sup>7</sup> (*Proteins*, 2008)
- <sup>8</sup> (*PlosOne*, 2011)
- <sup>9</sup> (*Biochimie*, 2011)
- <sup>10</sup> (*NAR*, 2011)
- <sup>11</sup> (*Biochimie*, 2012)
- <sup>12</sup> (*JBSD*, 2013a)
- <sup>13</sup> (*Amino Acids*, 2013)
- <sup>14</sup> (*JBSD*, 2013b)
- <sup>15</sup> (*Prot Sci*, 2002)
- <sup>16</sup> (*Theo Chem Ac*, 2001)
- <sup>17</sup> (*Bioinf*, 2003)
- <sup>18</sup> (*Proteins*, 2005)
- <sup>19</sup> (*ISB*, 2004)
- <sup>20</sup> (*J Theo Biol*, 2011)
- <sup>21</sup> (*J Biosc.*, 2007)
- <sup>22</sup> (*Proteins*, 2005)
- <sup>23</sup> (*J Theo Biol*, 2009)
- <sup>24</sup> (*Proteins*, 2009)
- <sup>25</sup> (*BMC Bioinfo*, 2004)
- <sup>26</sup> (*CBAC*, 2009)
- <sup>27</sup> (*Protein Sci*, 2009)
- <sup>28</sup> (*BBA*, 2005)
- <sup>29</sup> (*IDDT*, 2009)
- <sup>30</sup> (*EBJ*, 2007)
- <sup>31</sup> (*Biochimie*, 2009)



**Approximation** of local structure <sup>1-4</sup>

**Structural alignments** <sup>5-12</sup>

1. Omega angles <sup>13</sup>
2. NMR <sup>14</sup>

Approximation of **longs fragments**

1. Structural Words <sup>15</sup>
2. HPM <sup>16,17</sup>

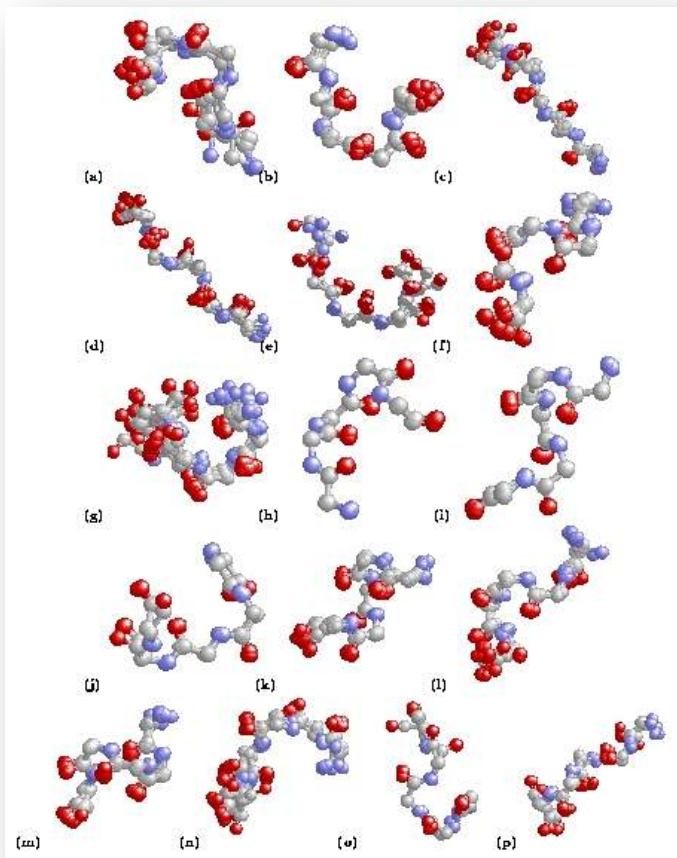
**Prediction & Analyses**

1. PBs <sup>1,18-20</sup>
2. Structural Words <sup>21</sup>
3. HPM <sup>22-24</sup>
4. Short loops <sup>25-27</sup>
5. DARC <sup>28-29</sup>
6. Mutations <sup>30</sup>
7. Contacts <sup>31</sup>
8. **Flexibility** <sup>32</sup>
9. Allostery <sup>33</sup>
10. protein domain linker <sup>34</sup> 90

<sup>32</sup> (*Proteins*, 2011), <sup>33</sup> (*BMC Struct Biol*, 2012), <sup>34</sup> (*JBSD*, 2013c)

+ book chapters (2001, 2005, 2007, 2010 & 2010).

➤ Original Proteins 2000 paper :



Web Of Sciences 148

Google scholar 280