# Graphs and statistics for protein interfaces

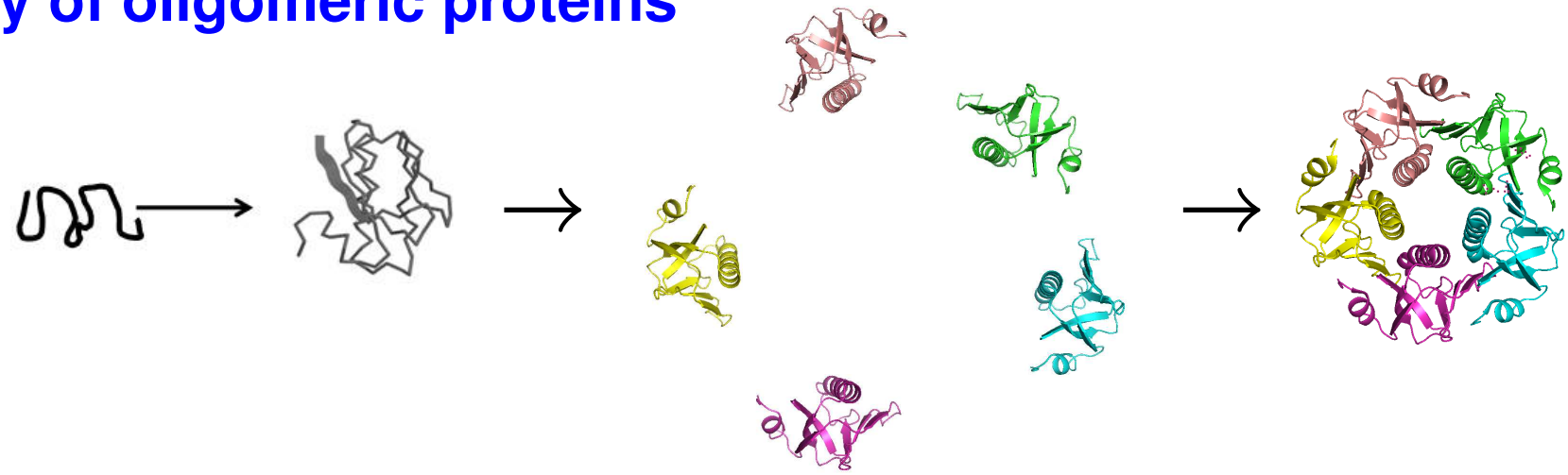
Giovanni Feverati

LAPTH, Annecy

Equipe Gemini

https://lapth.in2p3.fr/msif/fr/biophysique/

# Assembly of oligomeric proteins



From *E. Coli:* 20% only of proteins is strictly monomeric (D. Goodsell, 2000).

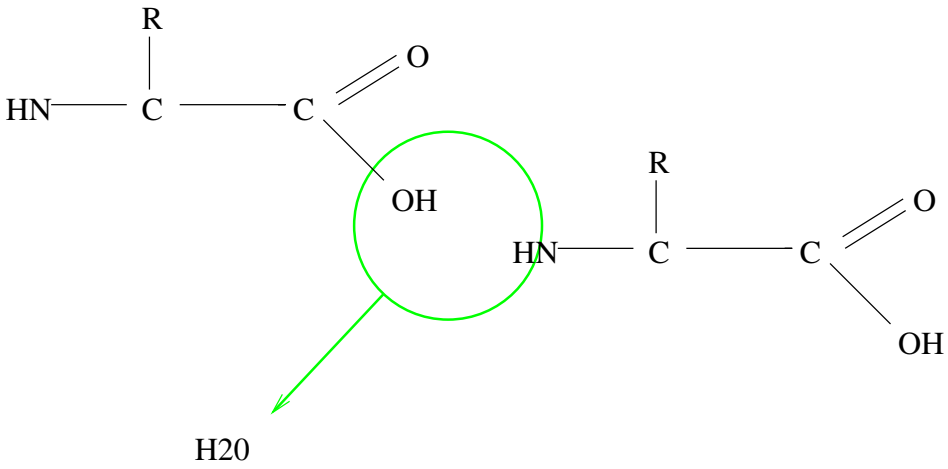
All proteins interact: formation of interfaces (permanent or temporary)

Public health: Alzheimer, anthrax, cholera ...

Interface = geometrical and chemical complementarity

Goal: from 3D structures to the required sequence features

PROTEIN: sequence of amino acids that folds and realises a biological function

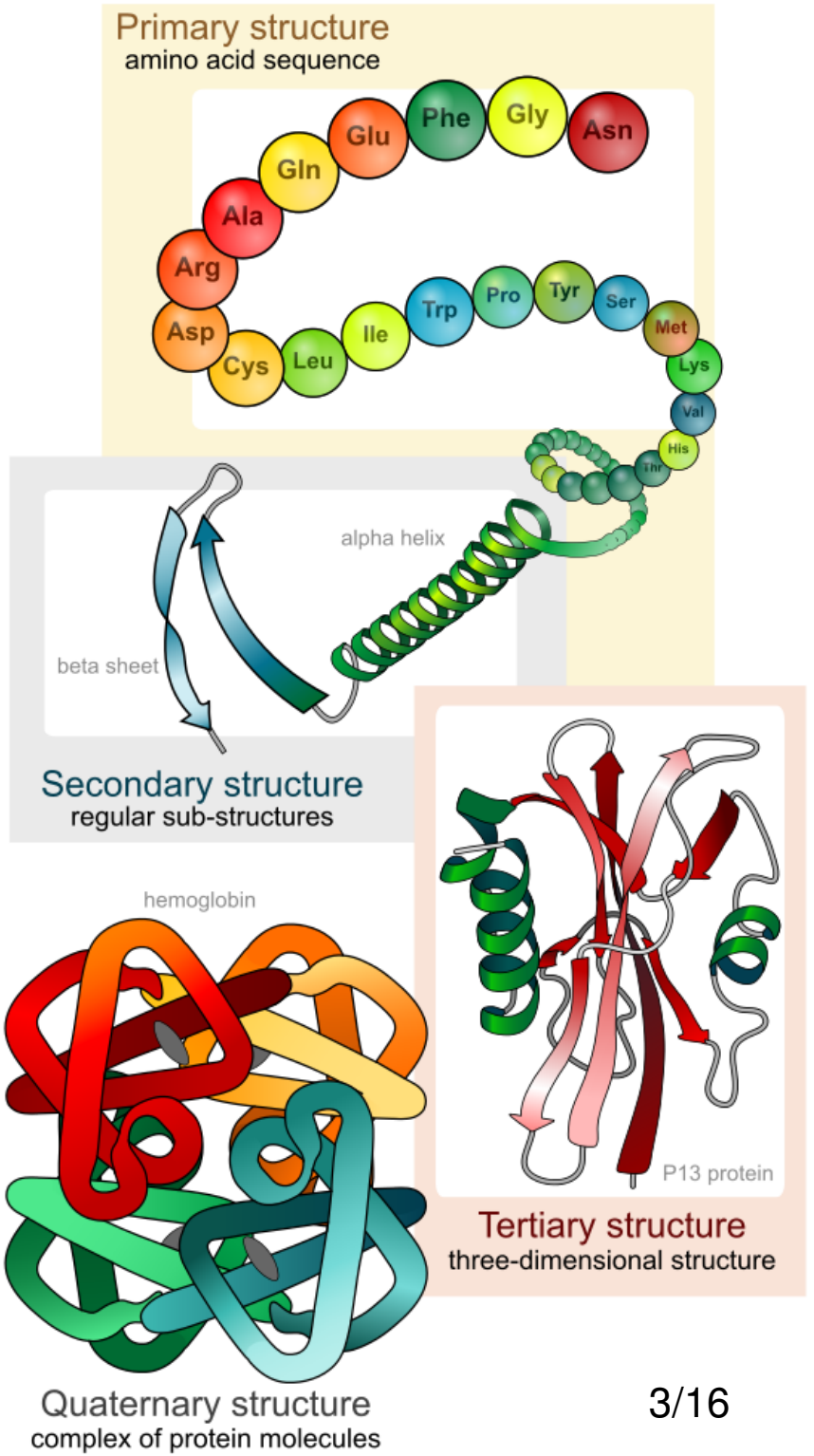



R: 20 standard amino acids = protein building blocks

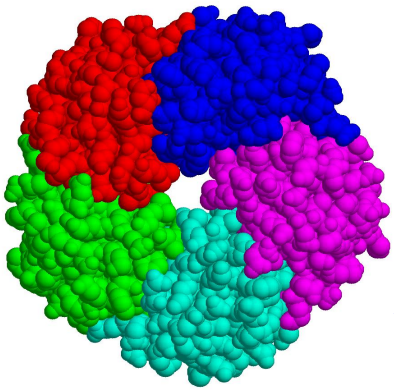
Main chain (backbone), side chain

Intra/inter molecular

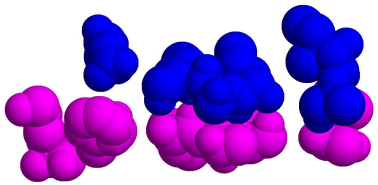
Folding/assembly

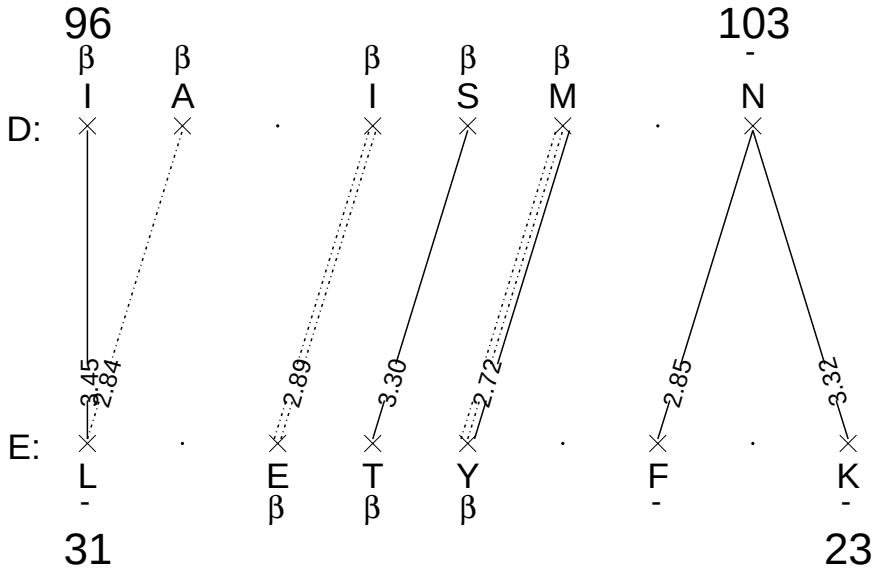


# GEMINI: a tool to investigate interfaces



**PDB**






**Interactions at the interface**

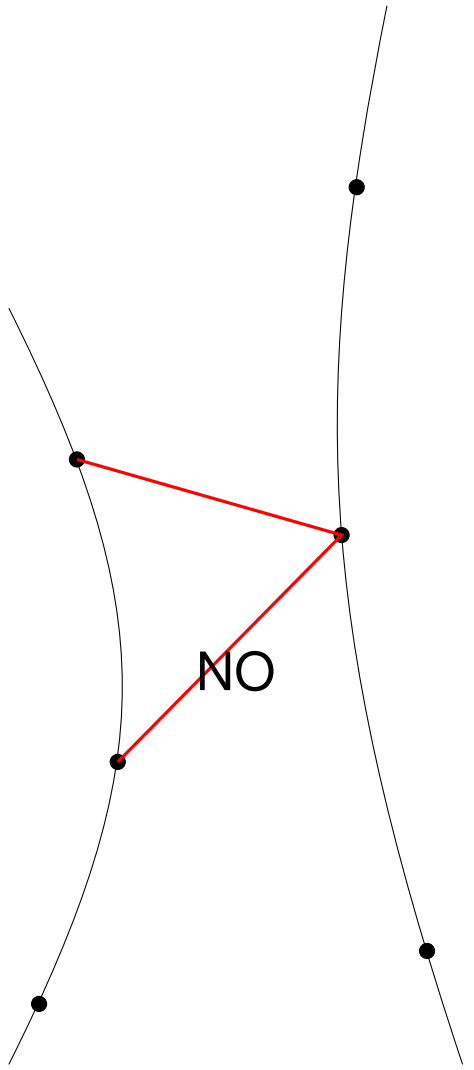
- Large ensembles of data
- Time: 0.1 s/protein
- Interaction patterns
- Statistics
- Interface geometry
- Interface prediction

Agence pour la protection des programmes, 2009
G. Feverati, C. Lesieur, Plos ONE 2010
G. Feverati et al., Plos ONE 2012

# Symmetric minimization 1

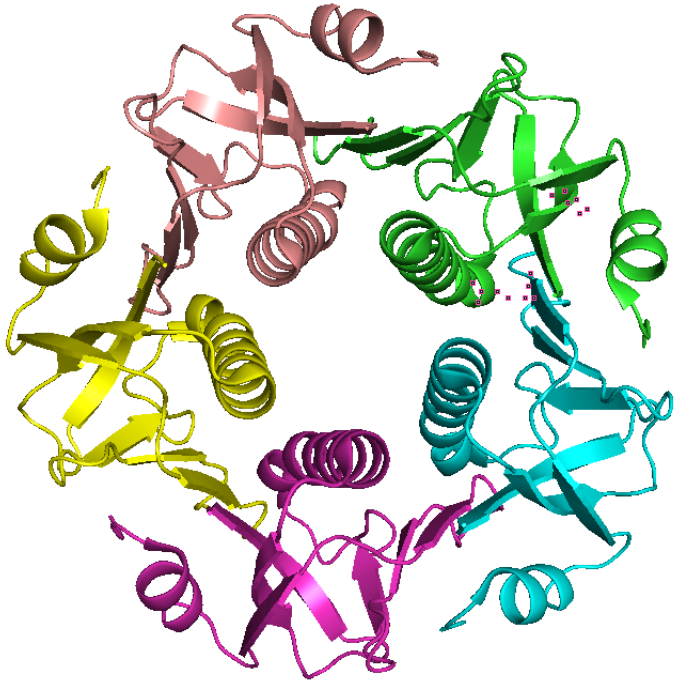
- Interface: $R_0 = \{\text{bonds} \leq 5 \text{ Angström}\}$
  Interface = strongest bonds $\longrightarrow$ framework of the interface
- For each atom: the shortest bond, from both sides $\longrightarrow S_0$
- $S_0$: comparison with amino acids and bonds of known interfaces $> 85\%$
- $R_0 - S_0$

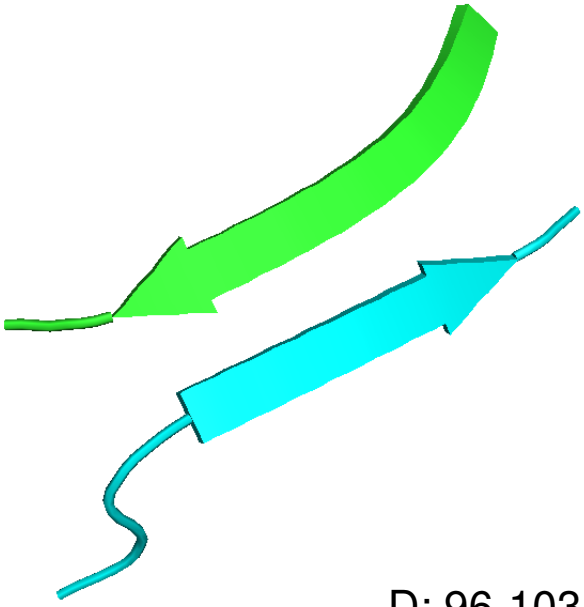
Interaction decreases with distance




G. Feverati, C. Lesieur, L. Vuillon, International Journal "Information Technologies and Knowledge", 2012

# Focus on one interface geometry: two aligned $\beta$-strands





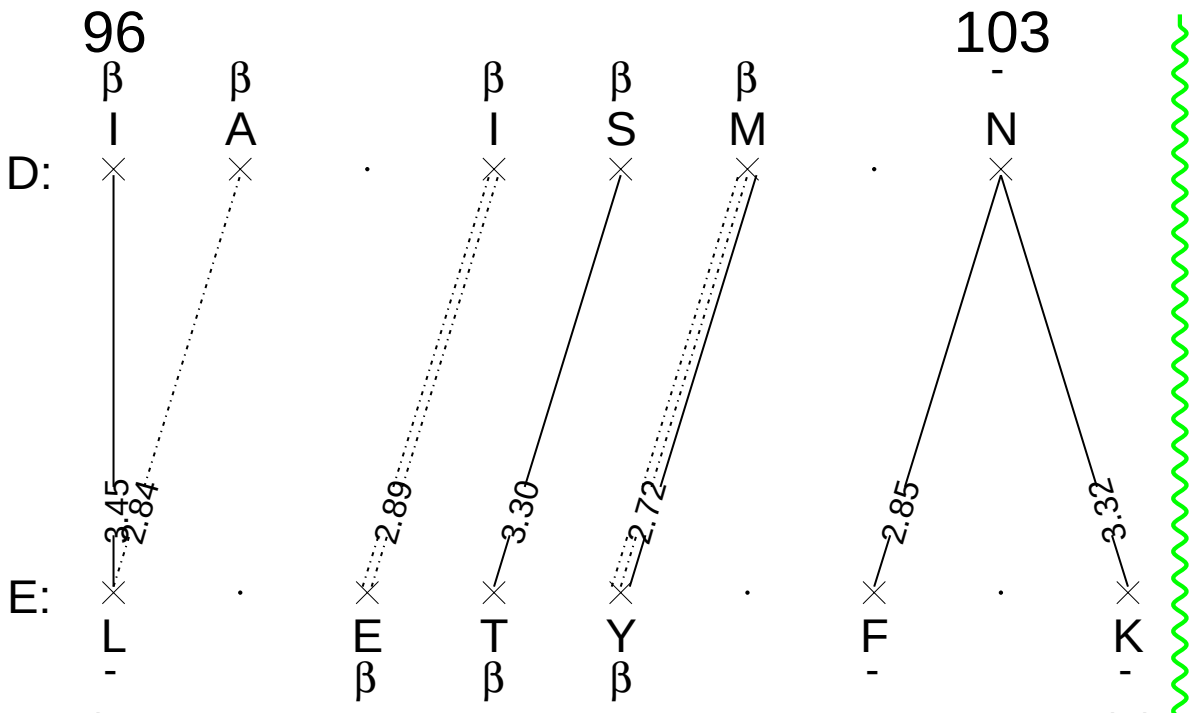
D: 96-103, E: 23-31

- ★ geometry $\longrightarrow$ chemistry ($\alpha$-coil interfaces, Crick 1952)
- ★ less understood than $\alpha$-coil; "planar" geometry: much less constrained than $\alpha$
- ★ many pathologies (Alzheimer, Parkinson, cholera, ...)
- ★ stoichiometry

classes: continuous $\beta$-sheet, $\beta$-sandwich

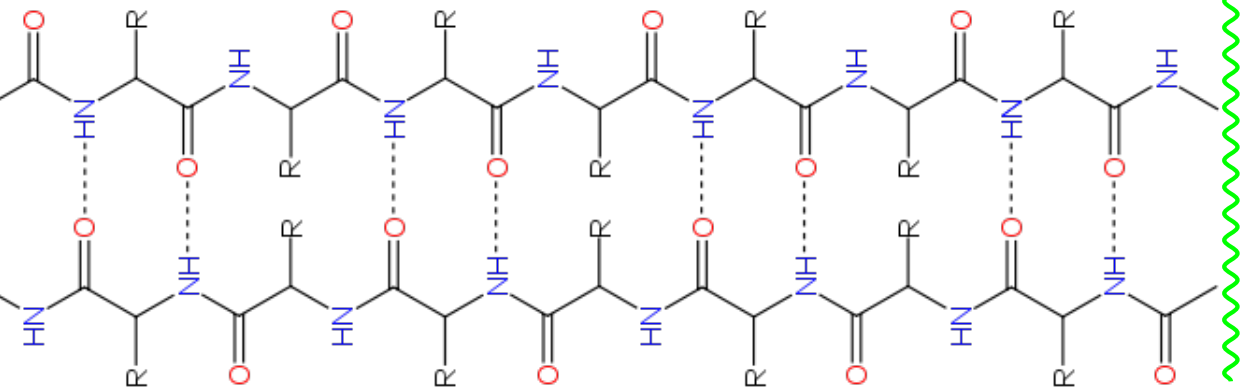
Guharoy, Chakrabarti 2007

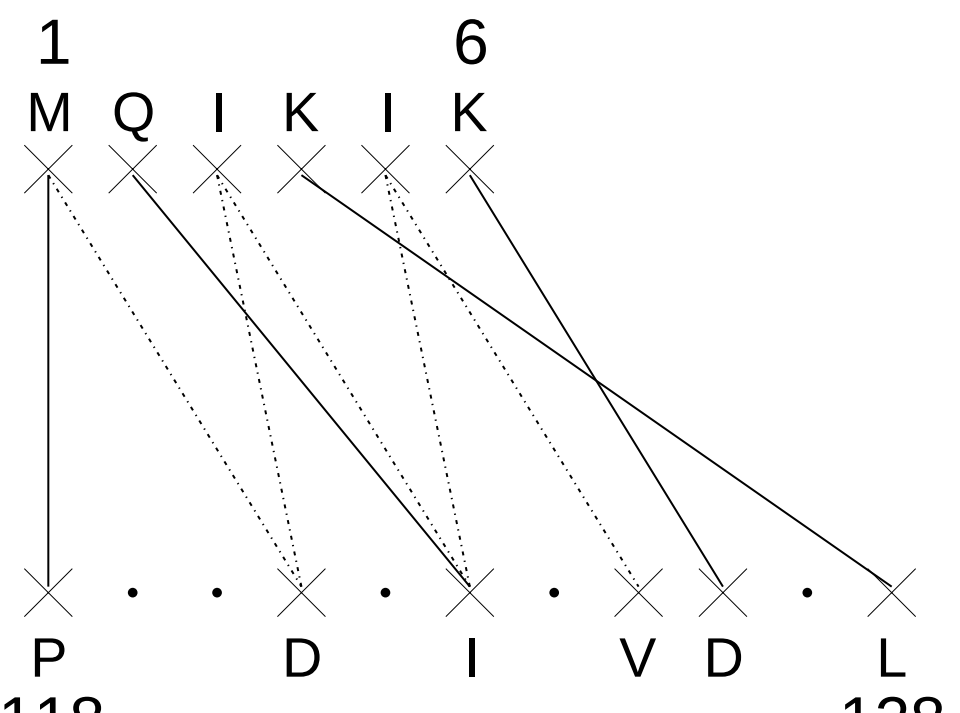
40 cases = the data set: circular → two $\beta$-strands → well separated segments on the sequence
Two subnetworks: BB network (geometry), SC network




5_1EEI

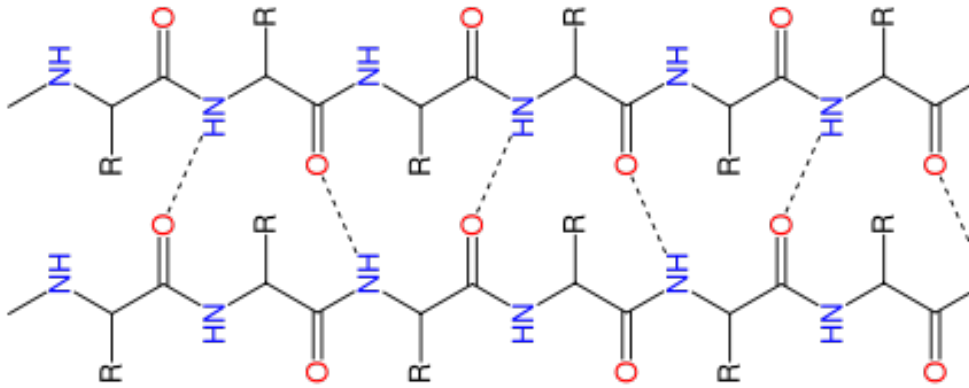
"ladder"

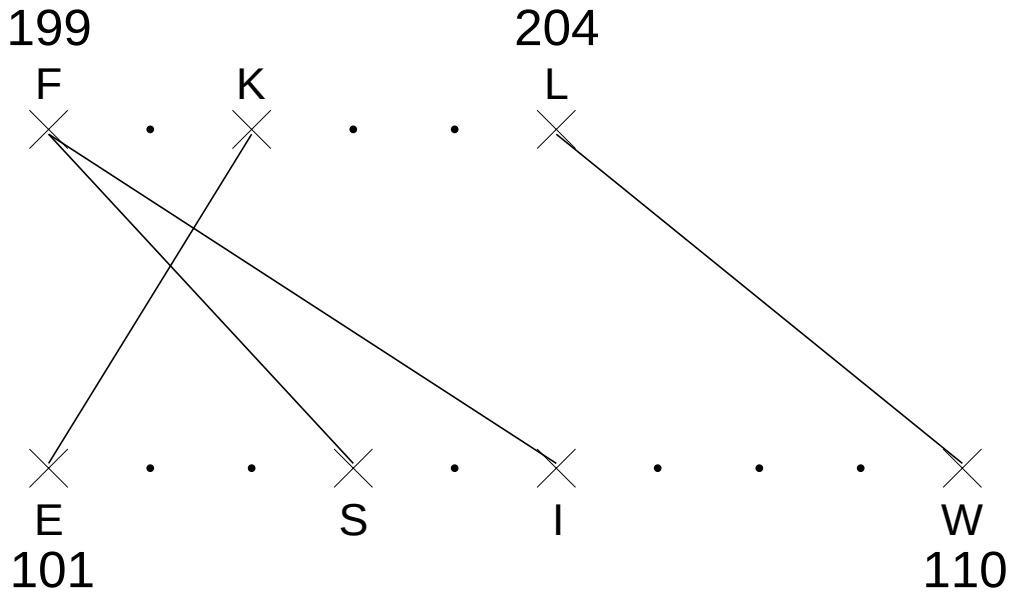





3_2BAZ

"zig-zag"






5_1B09: no BB subnetwork

$\beta$-strands in "sandwich" or "orthogonal"

Properties of the two subnetworks:

- ★ BB subnetwork represents backbone-backbone H bonds: N-0. RING (Cytoscape): 99% of matches (J. Zrimi)
- ★ BB central within SC
- ★ BB enriched in hydrophobic
- ★ hydrophobic residues central in BB
- ★ SC network contains BB network
- ★ SC network enriched in charged amino acids
- ★ SC network has no standard topology

## GLOBAL amino acid propensity

$\beta$-interfaces

V+: MIVWHF

V-: PAGLEKD

N: RC

$\beta$-interfaces in dimers

V+: WHRFMVIL / WHRMP

V-: ED / AVI

N: LCDE

## LOCAL amino acid propensity

$\beta$-interfaces, N: KD

| | | |
|---|---|---|
| R | HEVF | R |

$\beta$-intramolecular

| | | |
|---|---|---|
| GPKD | ILVMWFC | GPKD |

$\beta$-intramolecular edge

| | | |
|---|---|---|
| GPM | IVFCW charged | GPM |

$\beta$-fiber

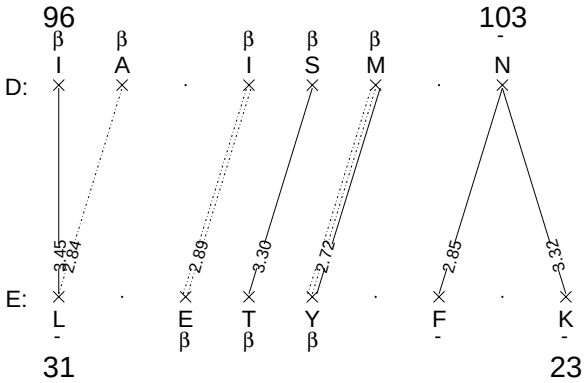
| | | |
|---|---|---|
| GAEKRMD | LIF | GAVMFED |

# Translational application

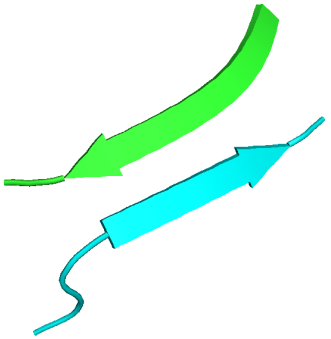
## Inhibition of the assembly by preventing the formation of $\beta$ interfaces

symmetric

minimization




=



Statistics +

pattern research

**R . Φ D Φ . Φ . R**

**ou**

**R . Φ H Φ . Φ . R**

**with Φ = I, M, C, W, V**

$20^6 = 64$ milions

250 sequences

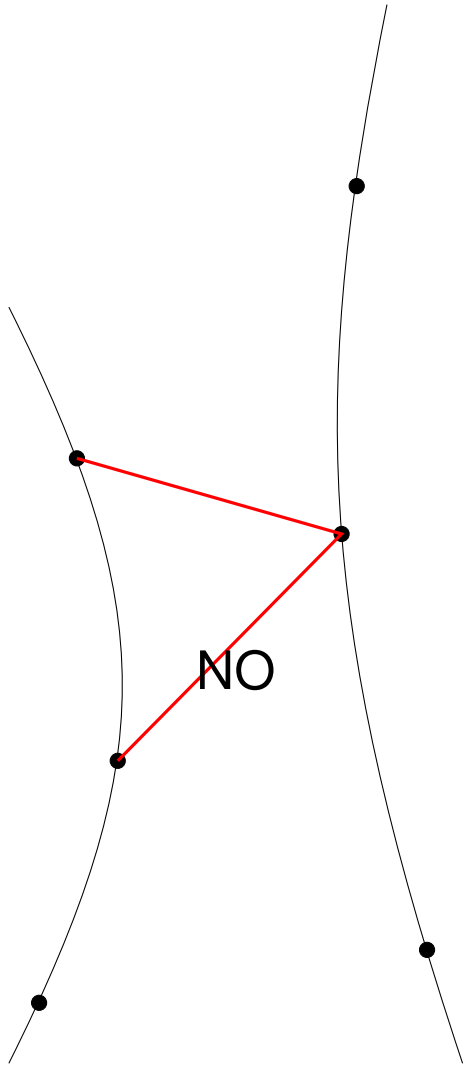
Generic properties / specific properties

Feverati et al., Plos One 2012

# Symmetric minimization 2

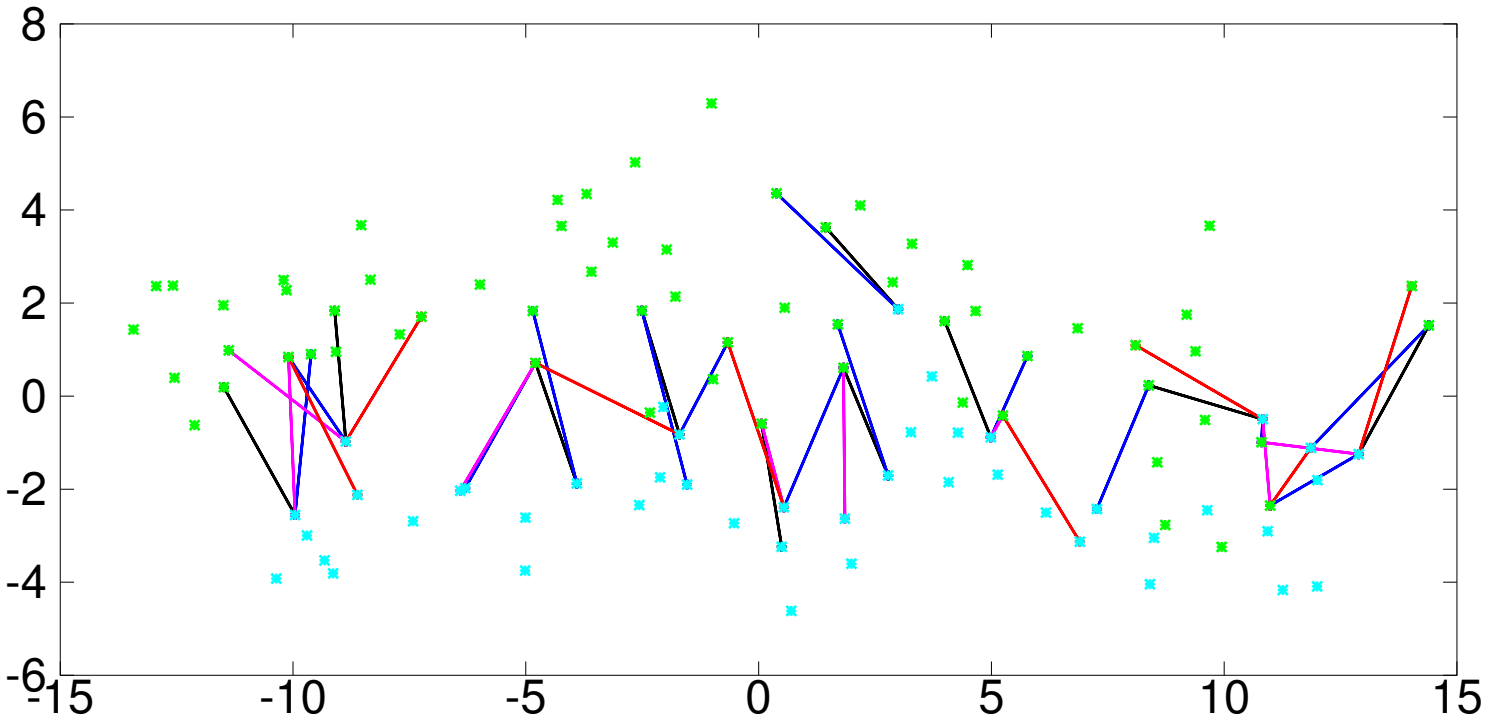
- Interface: $R_0 = \{\text{bonds} \leq 5 \text{ Angström}\}$
  Interface = strongest bonds $\longrightarrow$ framework of the interface
- For each atom: the shortest bond, from both sides $\longrightarrow S_0$
- $S_0$: comparison with amino acids and bonds of known interfaces $> 85\%$
- On $R_0 - S_0$: repeat the symmetric minimization $\longrightarrow S_1$
- Iterating the procedure
- Rank: $S_0, \quad S_1, \quad S_2, \quad S_3 \ \ldots$

**Strength:** local, intrinsic, self-adapting.
**Weakness:** sensitive to errors (regularization).




G. Feverati, C. Lesieur, L. Vuillon, International Journal "Information Technologies and Knowledge", 2012

# and clusters

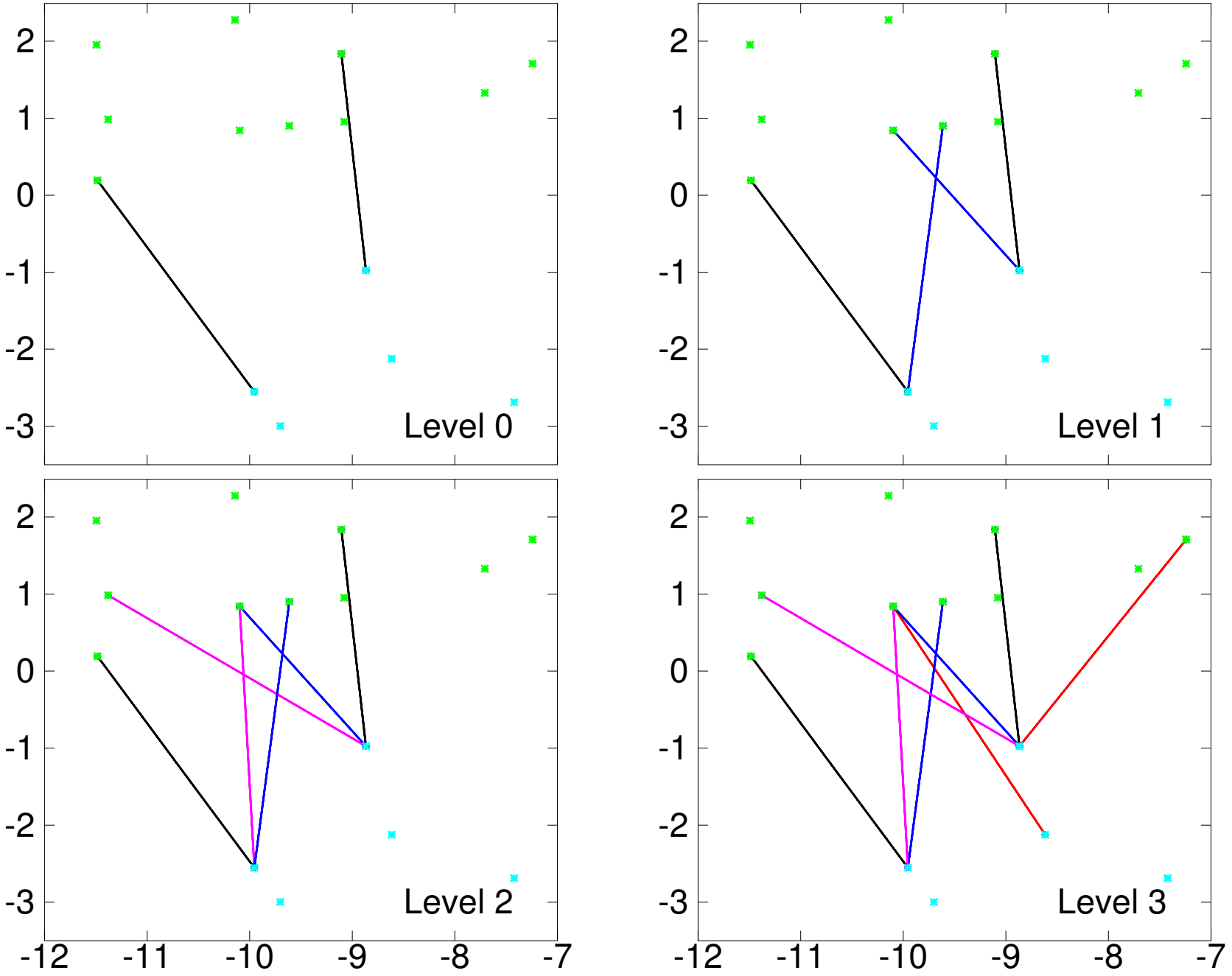



Planar projection of the interface. CtxB protein (axis coordinates in Angström).

Range $[96, 103]$ in chain D and $[23, 31]$ in chain E

Four lower levels: $S_0$ black, $S_1$ blue, $S_2$ purple, $S_3$ red.

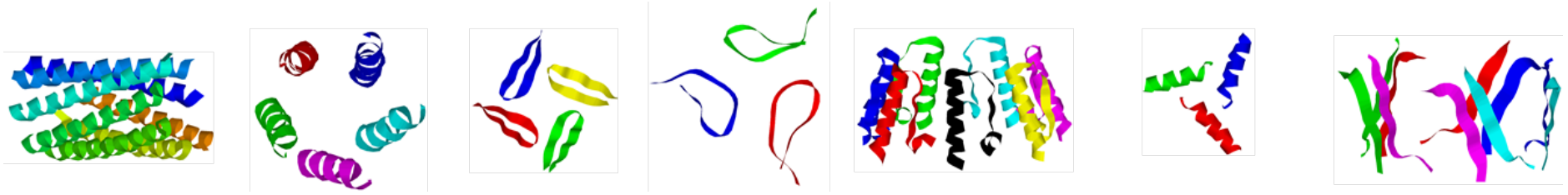
Cyan for the subunit D, green for the subunit E


Level 0
Level 1
Level 2
Level 3
2
1
0
-1
-2
-3
-12
-11
-10
-9
-8
-7

# Informatic development

GEMINI ⟶ software suite for global interface treatment, web portal (ANR)

- Data base
- Interface interactions (force field)
- Statistics
- Geometry (graphs, manifold learning)
- Interface predictions



- The user ⟶ GEMINI

## Conclusions

- Prediction tool based on the differences among the two subnetworks
- Classification of the interface geometry: spectral graph theory (laplacian)
- In vitro association: peptides keep their “interface” properties in the absence of the protein $\longrightarrow$ assembly inhibition

## The team

Giovanni Feverati, theoretical physics, postdoc CNRS, LAPTH, Annecy

Claire Lesieur, biology, CNRS, AGIM, Archamps

Kavé Salamatian, informatics, Université de Savoie, LISTIC, Annecy-le-Vieux

Laurent Vuillon, mathematics, Université de Savoie, LAMA, Bourget du Lac

Students: 7 (licence, master) + 1 thesis (automn 2012)

`https://lapth.in2p3.fr/msif/fr/biophysique/`