

β protein interfaces: a networking story

Giovanni Feverati

LAPTH, Annecy

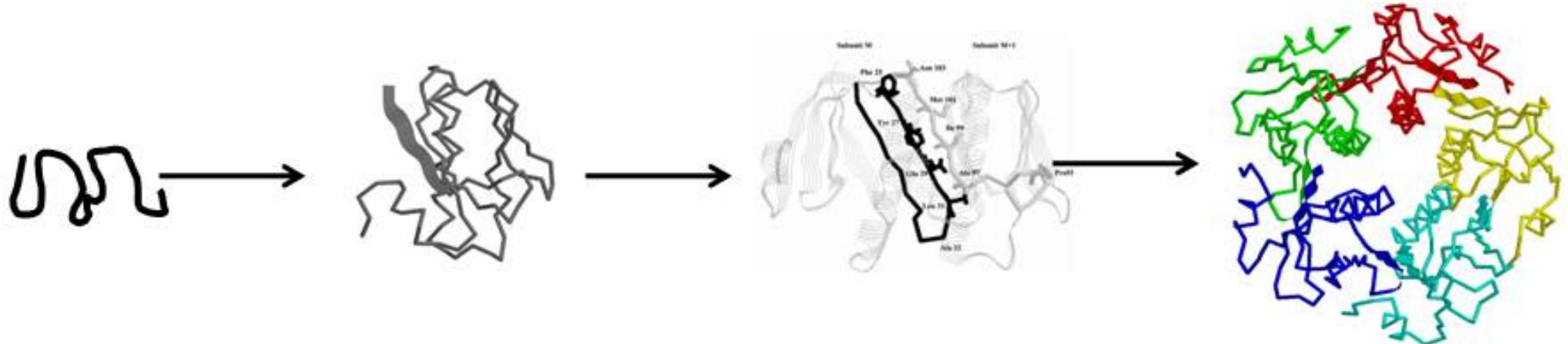
Equipe Gemini:

C. Lesieur, AGIM, Archamps

L. Vuillon, LAMA, Bourget du Lac

MiniTAGp2011: September, 15-16, 2011.

Assembly of oligomeric proteins



Estimations in *E. Coli* indicate that 20% only of proteins is strictly monomeric (D. Goodsell, 2000) the remaining ones are oligomeric, from dimeric on.

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

Public health: Alzheimer, Anthrax, Cholera ...

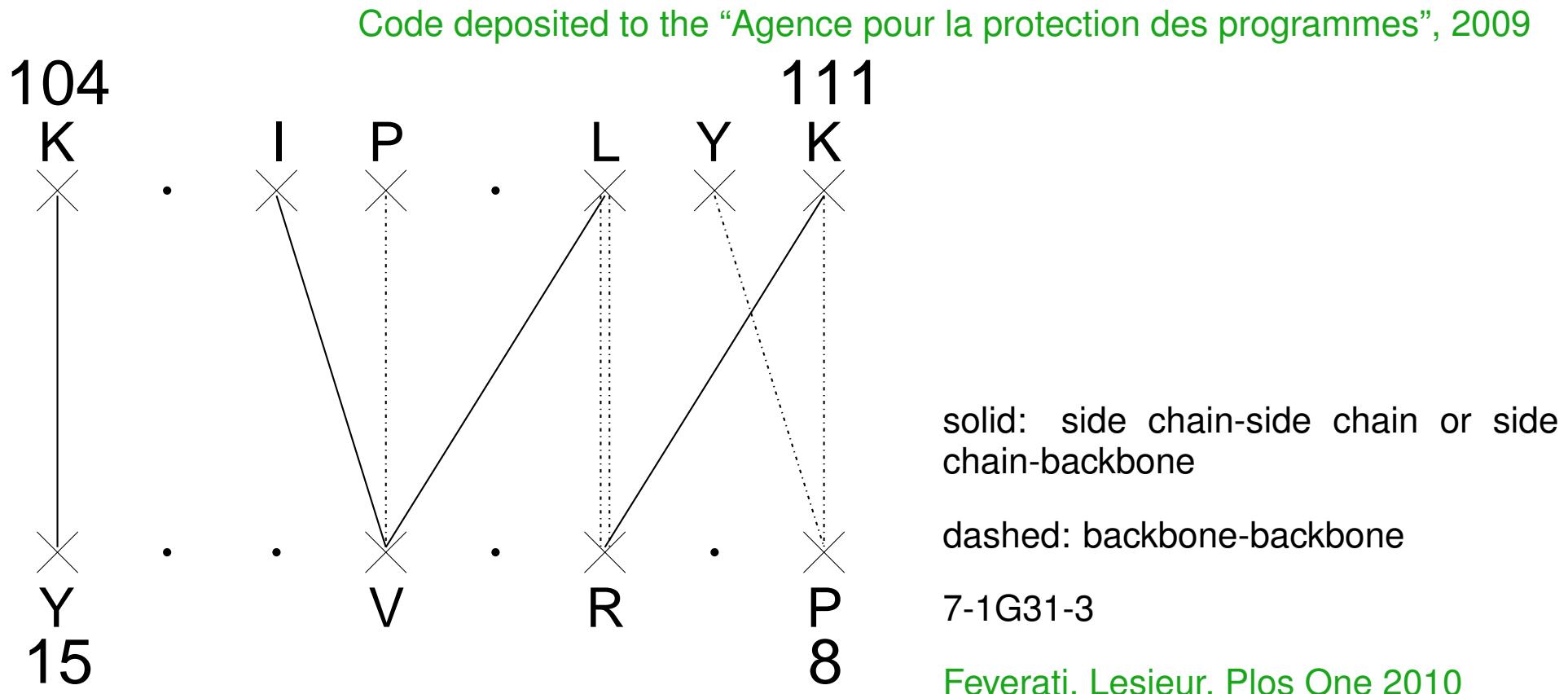
Interface = geometrical and chemical complementarity

Goal: systematic analysis of 3D structures to find the key of interface formation

Interfaces are specific (no erroneous mixing of protein interfaces), no “chimeric” interfaces, oligomeric proteins are chemically stable; still ...

Gemini: a tool to investigate interfaces

- ◆ GeminiDistances: from PDB 3D structure, detect the interface.
- ◆ GeminiRegions: interface is broken into non-interacting regions.
- ◆ GeminiGraph: representing the network of interactions.
- ◆ GeminiStat: statistics
- ◆ ...
- ◆ GeminiData: database



GeminiDistances: from PDB structure, detect the interface

- distance between atoms of adjacent subunits <0.5 nm
- symmetrization (the closest, for the two subunits)

Symmetrization makes the interface weakly cut-off dependent

Protein: 1G31 Program Version: 2 Number of monomers: 7 Symmetry:circular
Amino acids for each monomer: 107 107 107 107 107 107 107
Distance cutoffs: 20.0 5.0 Angstrom Algorithm: simmetrized interface
Bonded atoms (PDB atom serial numbers)

A:	181	197	268	300	473	486	500	753	767	775	776	785	788	790	794	804	808	8
	OE2	CG2	NE	OE2	CD2	OG	CG	0	0	0	CB	N	0	CG	CA	OH	0	C
G:	5545	5546	5148	5429	5491	5451	5426	4979	4954	4948	5610	4935	4932	4953	4920	5449	4918	49
	CE	NZ	CD	NH1	0	CG	CD	OH	CG2	N	OG1	0	N	CG1	0	0	CA	N
	3.48	4.25	3.92	3.18	3.82	3.28	3.72	2.66	3.15	2.91	3.69	2.94	2.90	3.61	3.28	2.58	3.22	3.

Bonded amino acids (PDB serial numbers)

A:	28	30	40	44	68	70	72	104	106	107	107	109	109	109	110	110	111	1
	E	V	R	E	L	S	P	K	I	P	P	L	L	L	Y	Y	K	
	1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	1	
	-	-	a	a	b	b	b	a	-	-	-	-	-	-	-	-	-	
G:	92	92	39	77	85	80	77	15	12	12	100	10	10	12	8	80	8	
	K	K	K	R	A	P	R	Y	V	V	T	R	R	V	P	P	P	
	1	1	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0	
	a	a	a	b	a	-	b	b	-	-	b	-	-	-	-	-	-	

#####

GeminiRegions: interface is broken into non-interacting regions;

Each region = network of interactions = bipartite graph

“1d distance” along the polypeptidic chain: emphasis on the sequence

Other criteria: connected component, emphasis on the geometry.

Protein: 1G31 Program Version: 2 Number of monomers: 7 Symmetry:circular
Amino acids for each monomer: 107 107 107 107 107 107 107
Distance cutoffs: 20.0 5.0 Angstrom Algorithm: simmetrized interface
Bonded atoms (PDB atom serial numbers)

A:	181	197	268	300	473	486	500	753	767	775	776	785	788	790	794	804	808	8
	OE2	CG2	NE	OE2	CD2	OG	CG	0	0	0	CB	N	0	CG	CA	OH	0	C
G:	5545	5546	5148	5429	5491	5451	5426	4979	4954	4948	5610	4935	4932	4953	4920	5449	4918	49
	CE	NZ	CD	NH1	0	CG	CD	OH	CG2	N	0G1	0	N	CG1	0	0	CA	N
	3.48	4.25	3.92	3.18	3.82	3.28	3.72	2.66	3.15	2.91	3.69	2.94	2.90	3.61	3.28	2.58	3.22	3.

Bonded amino acids (PDB serial numbers)

A:	28	30	40	44	68	70	72	104	106	107	107	109	109	109	110	110	111	1
	E	V	R	E	L	S	P	K	I	P	P	L	L	L	Y	Y	K	
	1	0	1	1	0	1	0	1	0	0	0	0	0	0	1	1	1	-
	-	-	a	a	b	b	b	a	-	-	-	-	-	-	-	-	-	-
G:	92	92	39	77	85	80	77	15	12	12	100	10	10	12	8	80	8	
	K	K	K	R	A	P	R	Y	V	V	T	R	R	V	P	P	P	
	1	1	1	1	0	0	1	1	0	0	1	1	1	0	0	0	0	
	a	a	a	b	a	-	b	b	-	-	b	-	-	-	-	-	-	

#####

Symmetrization: “the closest, for the two subunits”

Goal: framework of interactions

One atom, one interaction.

Based on atoms $\neq H$.

Chemical bonds often decay as $1/r^6$

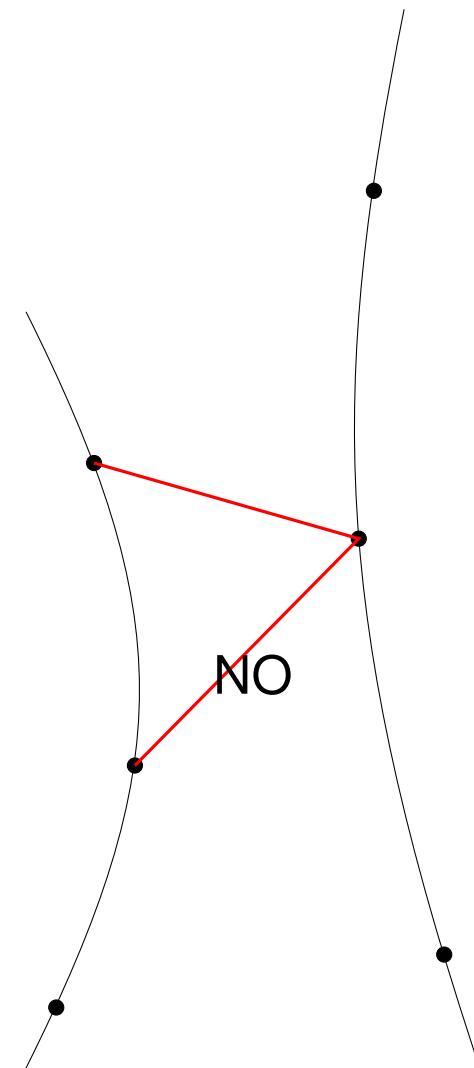
(Lennard-Jones)

“IN FAVOUR”

- ★ Simple and accurate: comparison with H bond data obtained with other means is excellent
- ★ Captures the strongest interaction
- ★ H atoms are located by heavier atoms

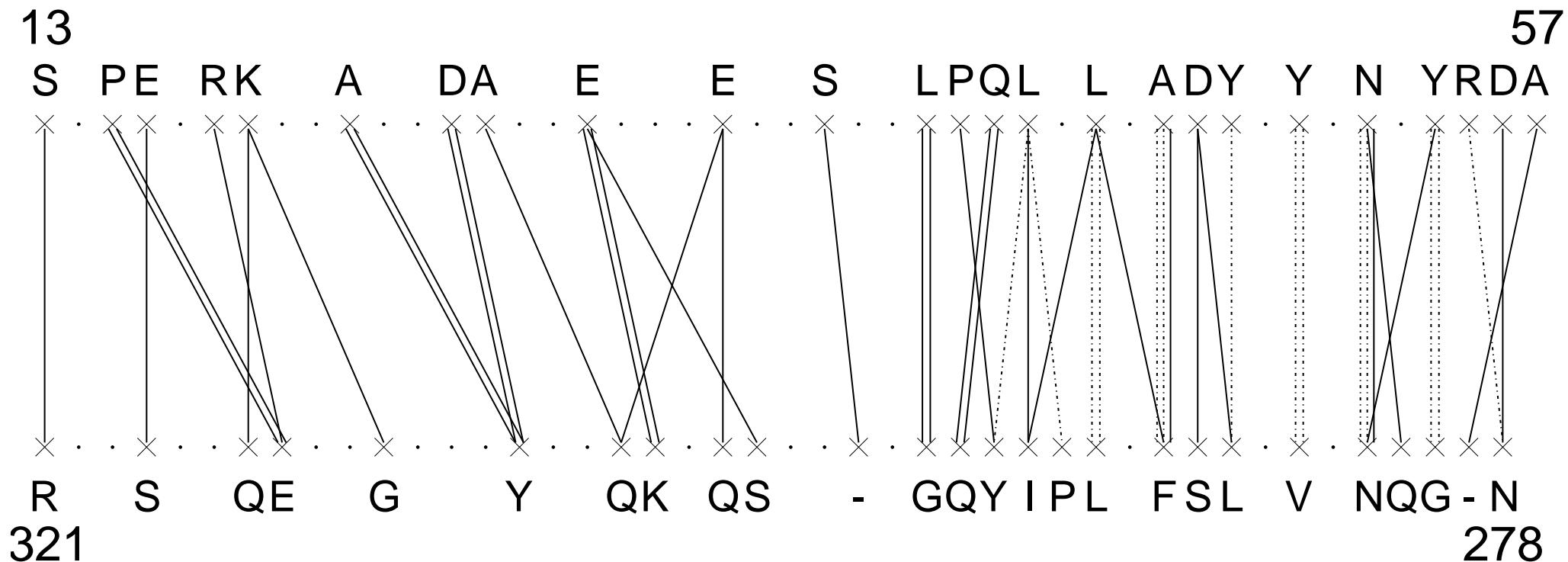
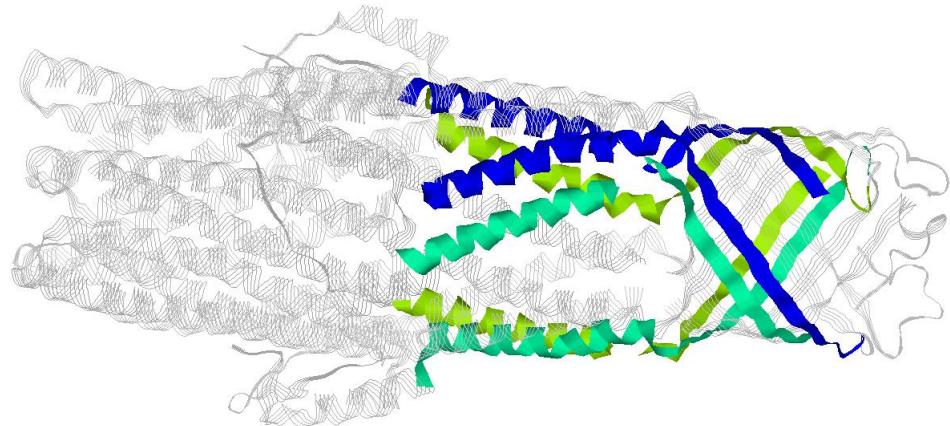
“AGAINST”

- ★ Electrostatic forces are not directional
- ★ H atoms not considered
- ★ Errors in the atomic structure can lead to a wrong choice



Geometrical validation

similar geometry \rightarrow similar graph
graph topology \rightarrow geometric constraints
 α, β structures



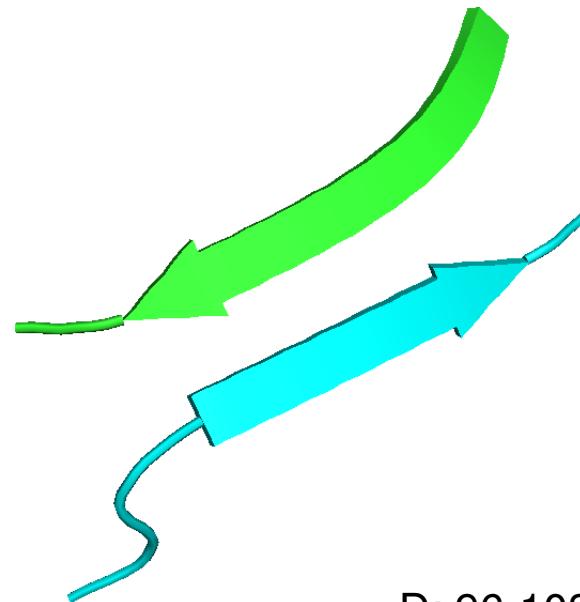
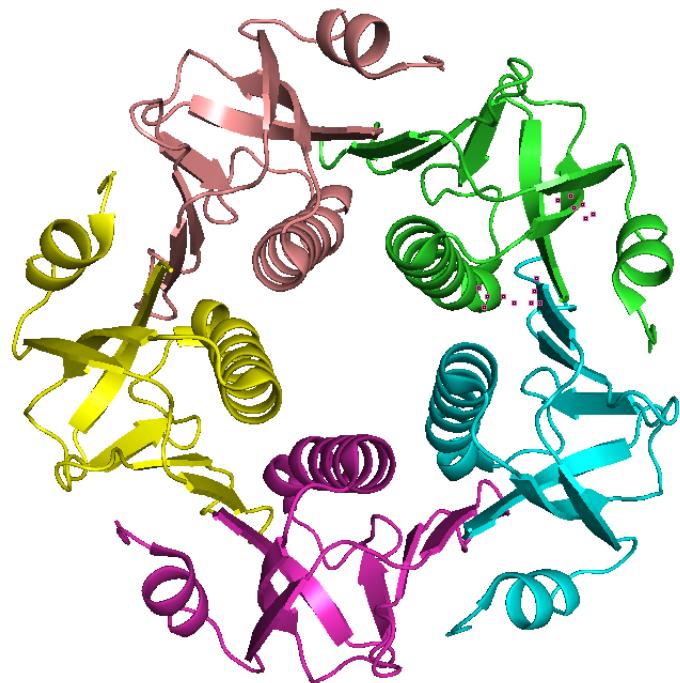
trimeric membrane protein 1ek9

Chemical validation: chemical properties of interacting amino acids

- ★ H bonds from PDB viewer: 87% recognized by Gemini, for 40 proteins
If backbone considered: 98% of H bonds identified (RING / Cytoscape)
- ★ Selected amino acids:
ToIC interface compared with PPIDB (geom) and SCOWLP (chem. geom)

overlap	SCOWLP 55	PPIDB 65
Gemini 51	47	41
SCOWLP 55		45
- ★ Interactions: ToIC, 87% of interactions detected by Gemini
- ★ β -spiral (1QIU): matching pattern; 85% of a.a. detected, 77% of bonds detected
- ★ Experimental validation: mutants (Heptameric co-chaperons cpn10)

Focus on one interface geometry: two aligned β -strands

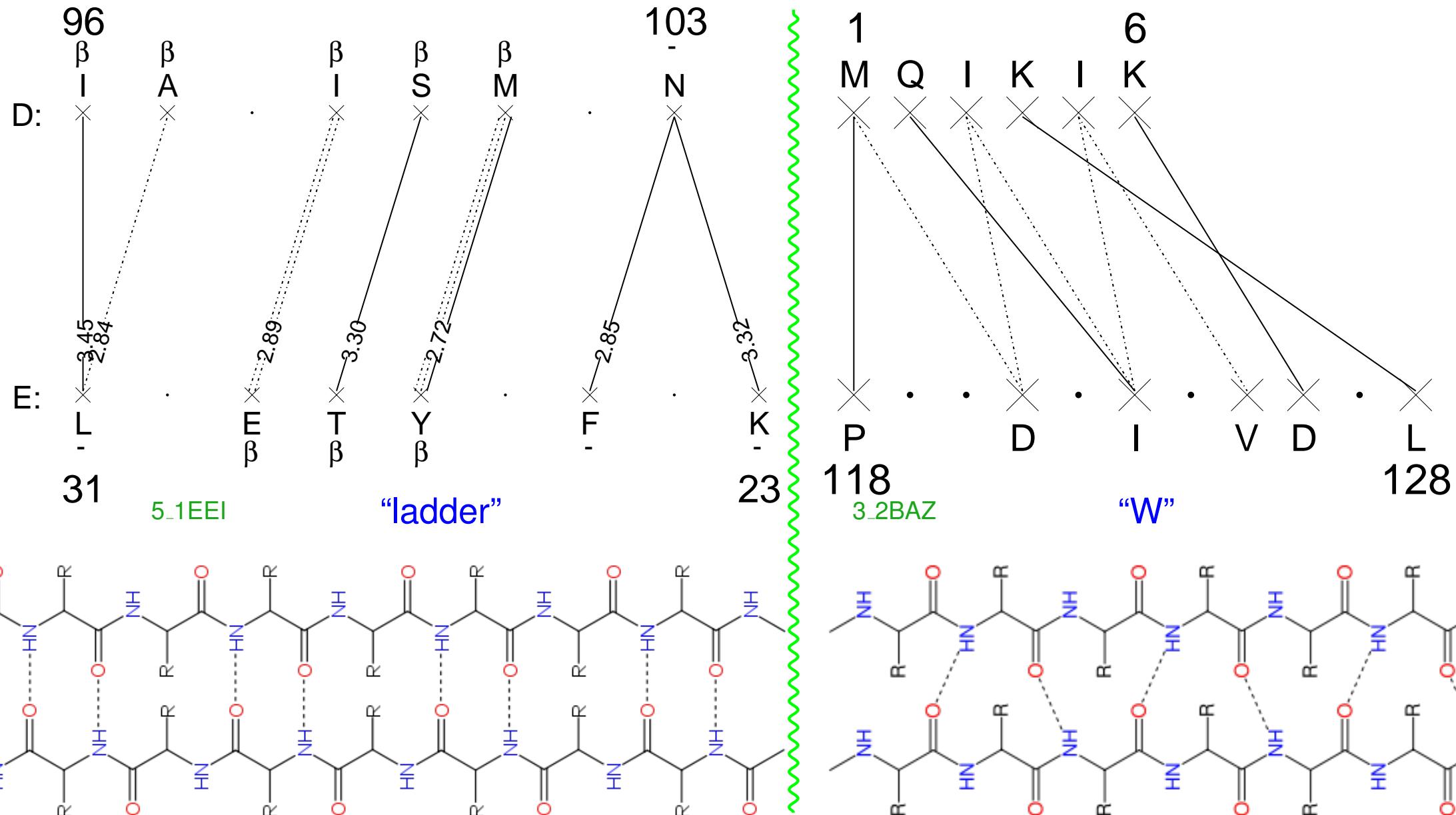


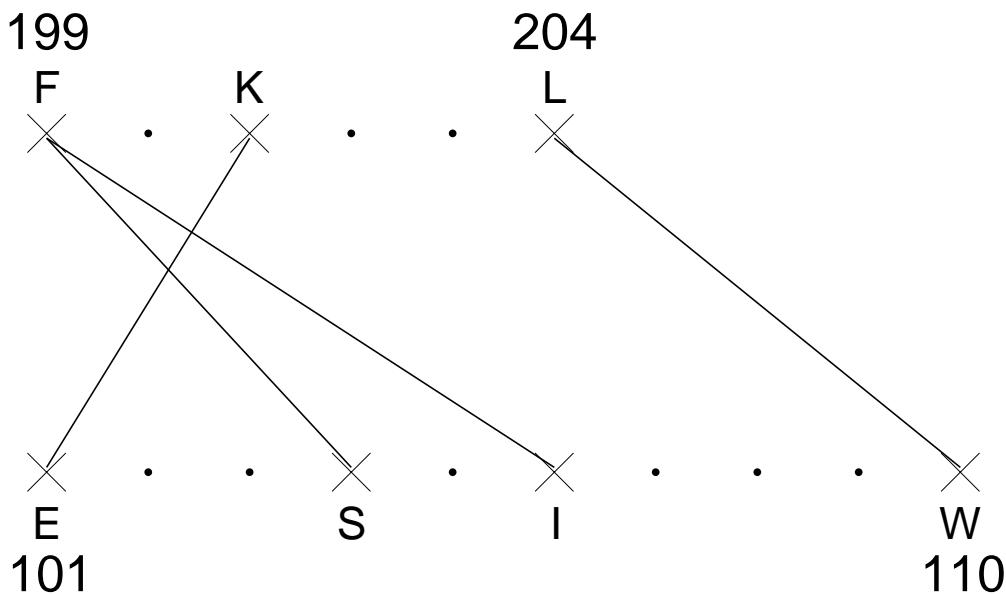
- ★ geometry → chemistry (α -coil interfaces, Crick 1952)
- ★ less understood than α -coil; “planar” geometry: much less constrained than α
- ★ many pathologies (Alzheimer, Parkinson, cholera, ...)
- ★ stoichiometry

classes: continuous β -sheet, β -sandwich

Guharoy, Chakrabarti 2007

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





5_1B09: no BB subnetwork

β -strands in “sandwich” or “orthogonal”

Properties of the two subnetworks:

- ★ BB subnetwork represents backbone-backbone H bonds: N-O. 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 charges amino acids
- ★ SC network has no standard topology

GLOBAL amino acid propensity

β -interfaces

V+: MICVWH

V-: PAGL

N: FREKD

β -interfaces in dimers

V+: WHRFMVIL / WHRMP

V-: ED / AVI

N: LCDE

LOCAL amino acid propensity

β -interfaces, N: KD

R	HEVF	R
---	------	---

β -intramolecular

GPKD	ILVMWFC	GPKD
------	---------	------

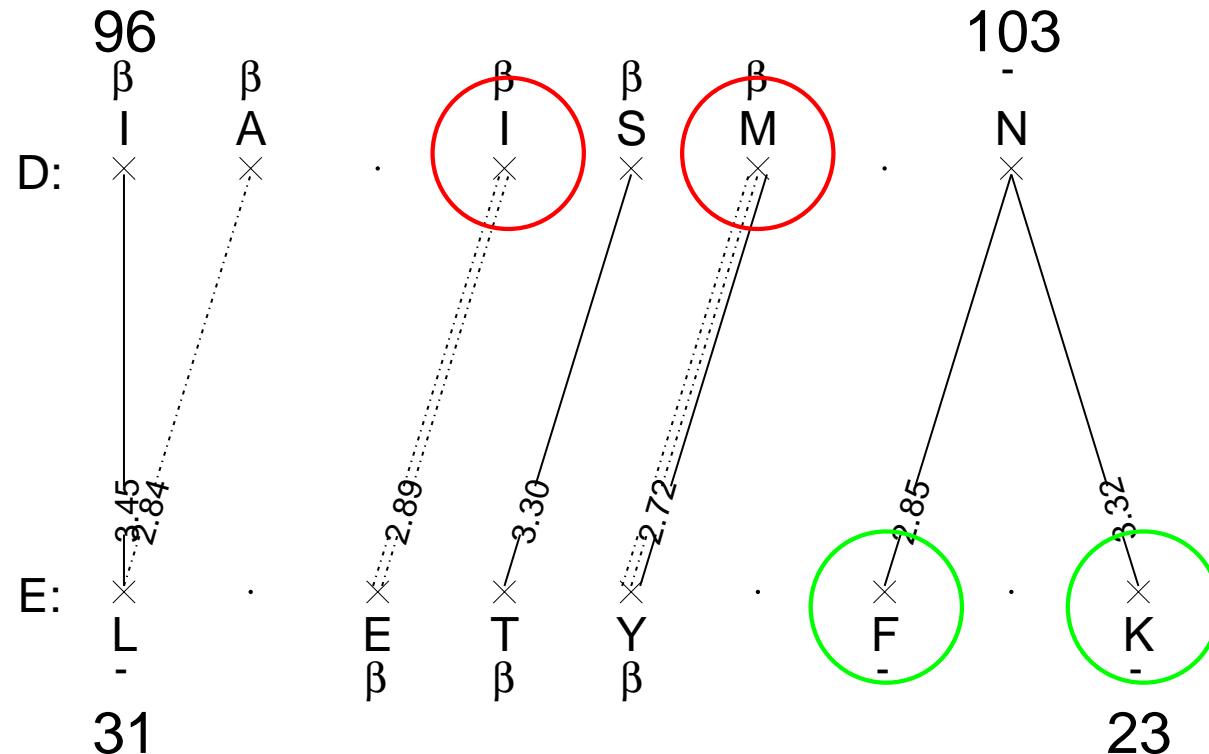
β -intramolecular edge

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

β -fiber

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

FOLD X



GOR IV central hydrophobic
V+: VLF
V-: PAGWC
N: M

Fold X mutation effects I99
V+: VLM
V-: PAGCWF

Corner charged
V+= R
V- = HE
N= KD

Fold X mutation effects K23
V+= RDH
N= E

Effects of mutations:
reduced complex stability:
I99 and M101 ("volume")
reduced interaction energy:
K23 and F25 (aromatic)

Consistent with effect on interface if SC mutated but no effect if BB mutated

Conclusions

- Enlarging the analysis to 1500 cases
- Testing propensities against random graphs
(exemple: charged corner, statistical effect or signal?)
- Prediction tool based on the differences among the two subnetworks
- In vitro association: peptides keep their “interface” properties in the absence of the protein (C. Lesieur, A. Mounia)
- Synthetic peptides: inhibition