

# Modélisation des structures protéiques

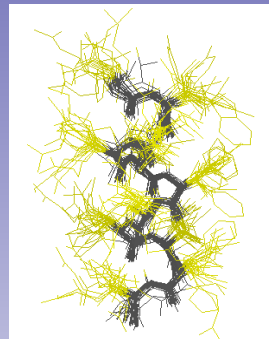
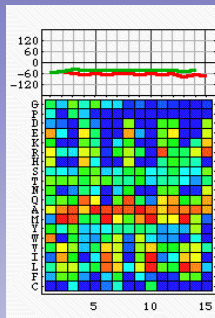
- “ab initio”

- Par simulation du repliement

Saunders et al., Pac Symp Biocomput. 2002;:601-12.

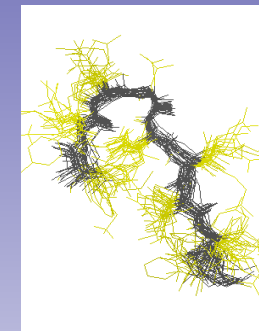
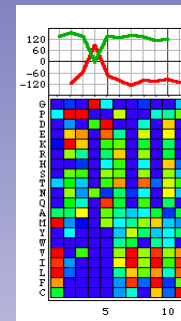
Derreumaux Ph., Phys Rev Lett. 2000 Jul 3;85(1):206-9.

- Par assemblage de “briques structurales”



$\alpha$ -helical

Isites

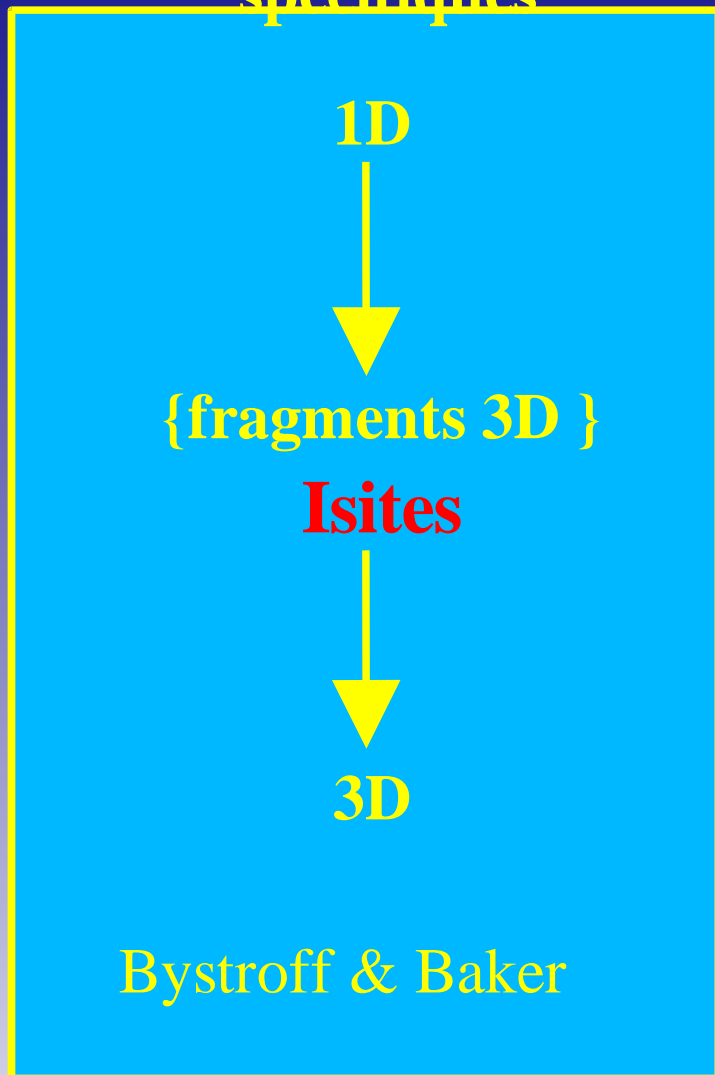


$\beta$ -turn (type II)

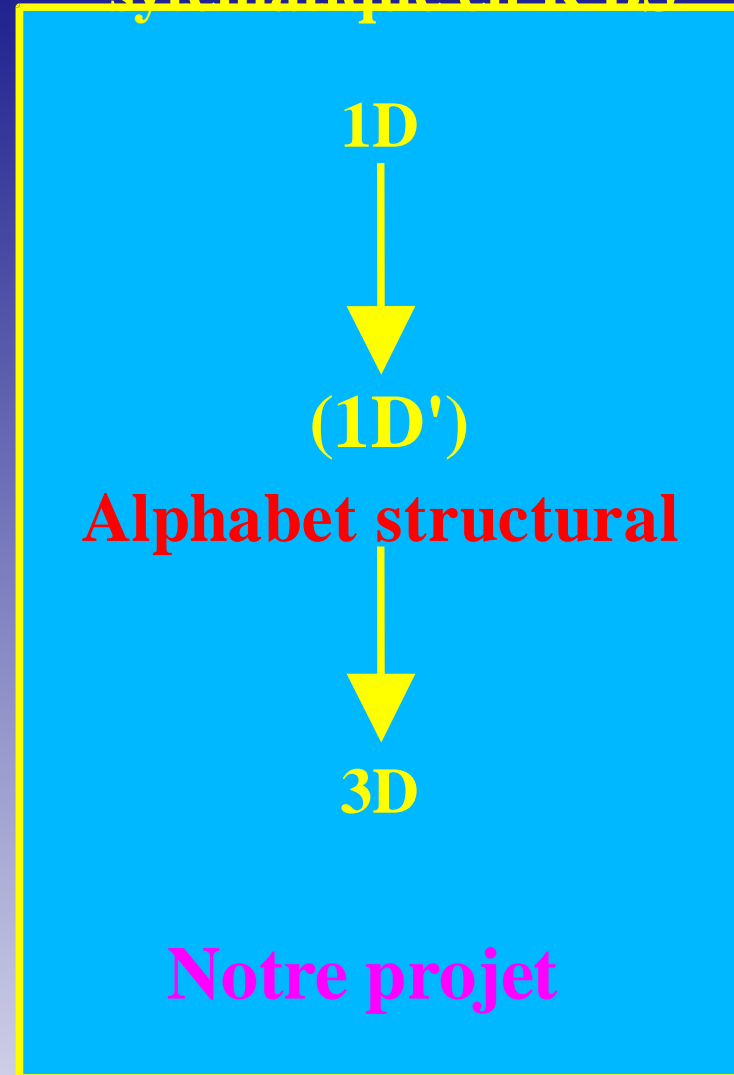
Bonneau et al., J Mol Biol. 2002 Sep 6;322(1):65.

[www.bioinfo.rpi.edu/~bystrc/hmmstr/server.html](http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.html)

# Fragments 3D spécifiques



# Décomposition sytématique en R BS

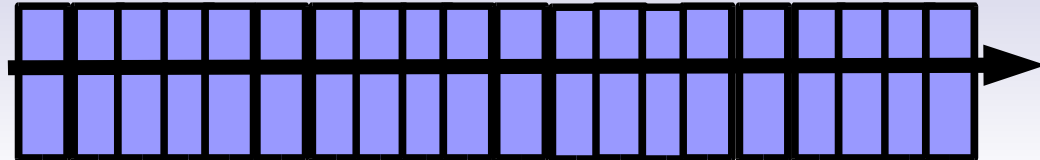
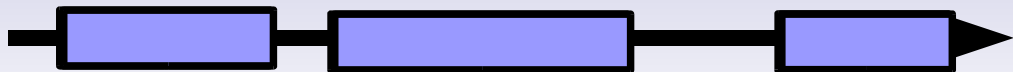


**Isite 1**

**Isite 2**

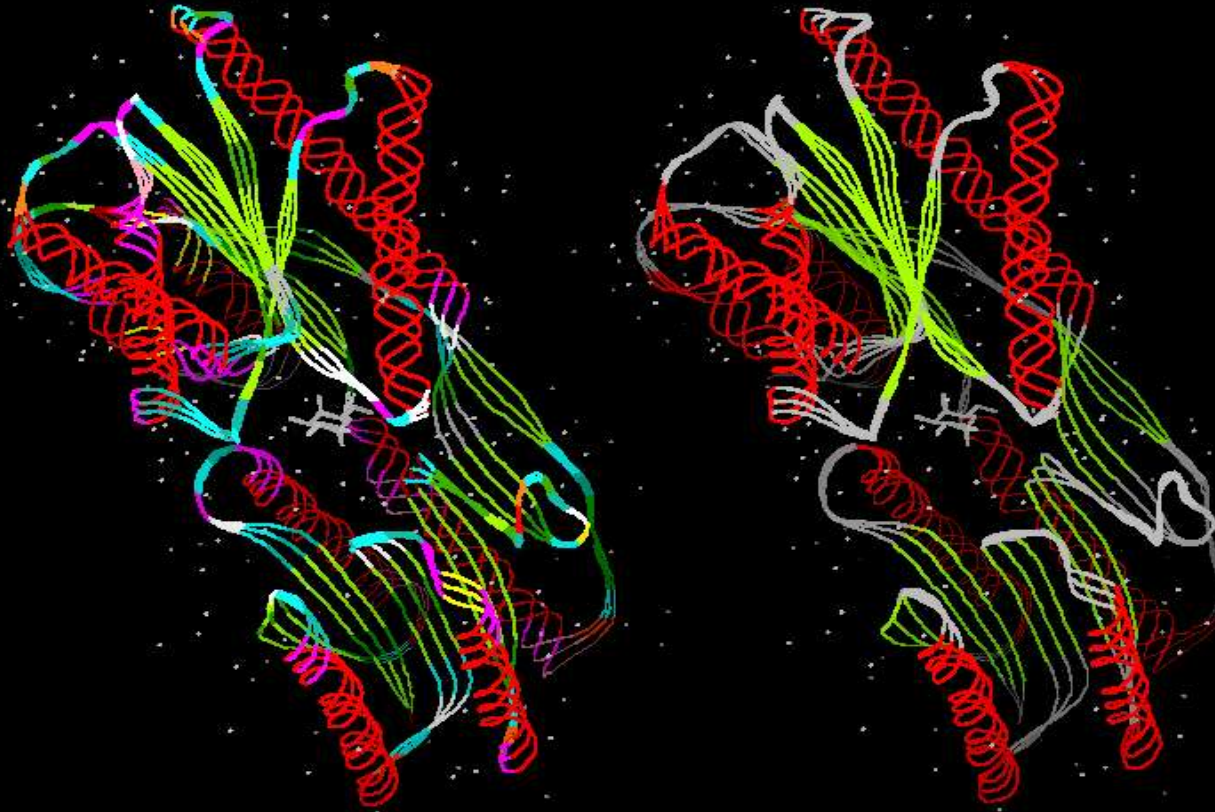
**Isite 3**

R RRRRR RRRRRR RRRRRR RRRR



# Alphabet Structural

L-arabinose binding protein  
(8abp)



- Structure secondaire :

Hélices  $\alpha$  (30%)

Feuillets  $\beta$  (20%)

Boucles

(Pauling et al., 1951)

-AS-12

GIMMXLPQXKUCNYWAAAAAAAAAVQPZCOZ **Camroux et al., Prot. Eng., 1999**

SKNMNNTPIPTPWAADVWVWAADSPIMTLFF

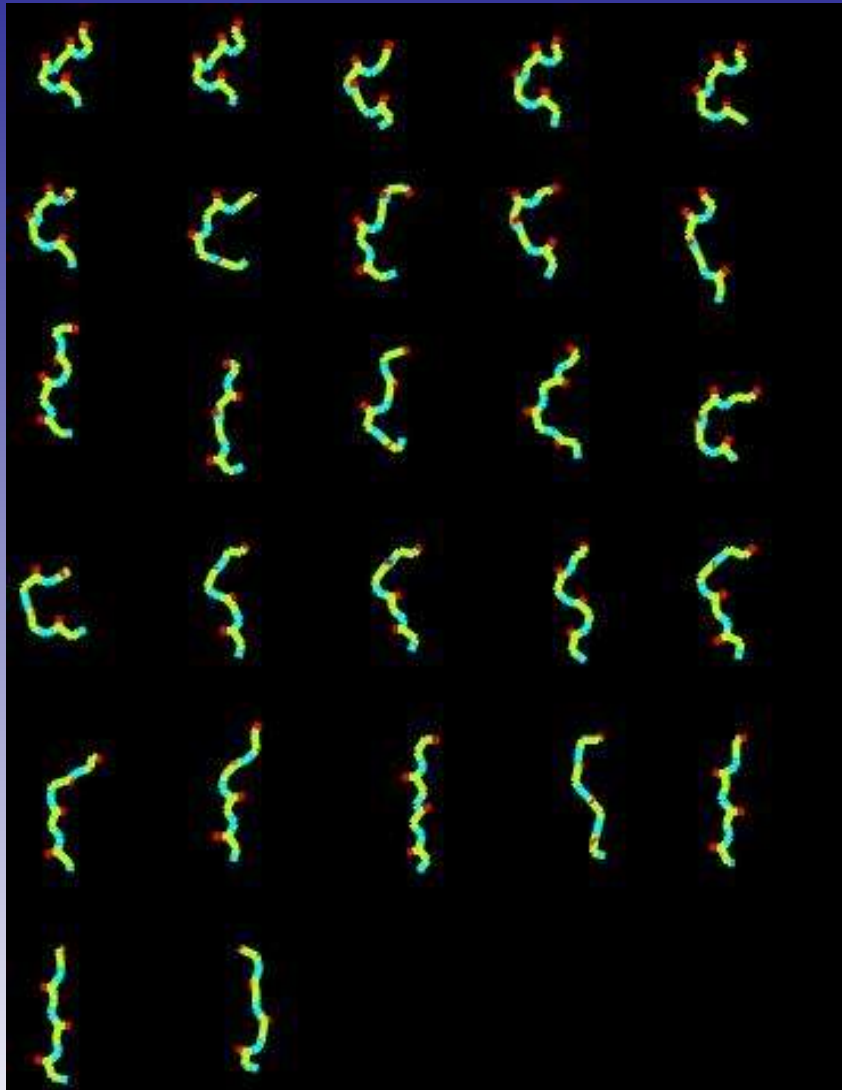
RLHAAAAAAAAAAVWVWPIMMNMPQXLUIK

PBAAAAAVWVQPVCBQKNMNTXPQGS

AAAAVWAADGPIMTLLLPQPWAAAAAA

AAV77

# ALPHABET OPTIMAL : 27 SBBs

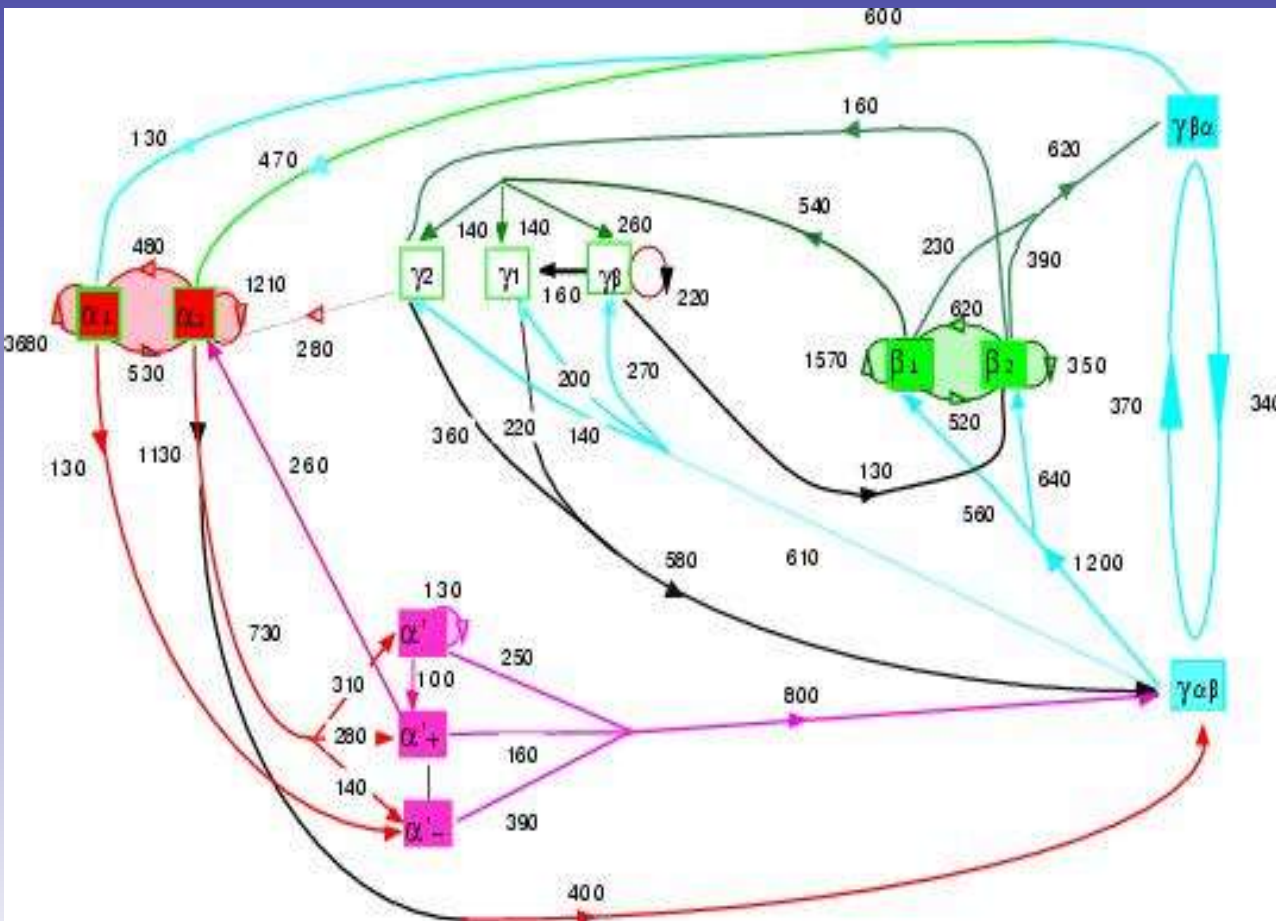


SA-27 SBBs	RMSd in (Å)		% (N) (56167)	K Set 1
	Within	Between		
a	0.13	0.15 (aA)	2.6 (1460)	0.19****
A	0.07	0.15 (V/a)	12.6 (7077)	0.40****
V	0.10	0.15 (VA)	5.6 (3145)	0.32****
W	0.09	0.15 (WA)	5.3 (2977)	0.23****
Z	0.19	0.27 (ZV)	4.5 (2528)	0.15****
B	0.13	0.27 (BA)	4.7 (2640)	0.17****
C	0.23	0.42 (CZ)	1.8 (1011)	0.32****
D	0.25	0.63 (DH)	2.0 (1123)	1.17****
E	0.26	0.53 (EB)	2.0 (1123)	0.10****
O	0.25	0.64 (OS)	1.5 (843)	0.20****
S	0.27	0.67 (SO)	3.2 (1797)	0.59****
R	0.48	0.67 (RN)	1.7 (955)	0.38****
Q	0.30	0.69 (QO)	4.1 (2303)	0.31****
I	0.32	0.75 (IC)	2.9 (1629)	0.23****
F	0.94	1.08 (FU)	1.9 (1067)	0.51****
U	0.33	0.54 (UC)	2.0 (1123)	0.60****
P	0.29	0.53 (PL)	4.4 (2471)	0.26****
H	0.23	0.53 (HY)	2.7 (1517)	0.68****
G	0.29	0.69 (GM)	3.4 (1910)	0.28****
Y	0.28	0.53 (YH)	2.0 (1123)	0.40****
J	0.55	0.83 (JY)	2.0 (1123)	0.75****
K	0.23	0.49 (KL)	4.1 (2303)	0.41****
L	0.25	0.49 (LK)	5.1 (2865)	0.13****
N	0.21	0.43 (NT)	4.9 (2752)	0.32****
M	0.23	0.46 (MN)	5.3 (2977)	0.41****
T	0.22	0.43 (TN)	3.0 (1685)	0.25****
X	0.24	0.53 (XT)	4.7 (2640)	0.18****
mean	0.23		3.7	0.347

# Graphe de connexions d'un AS-12

## Matrice de transition creuse

=>Schéma simplifié des trajectoires entre BS



Cluster A:

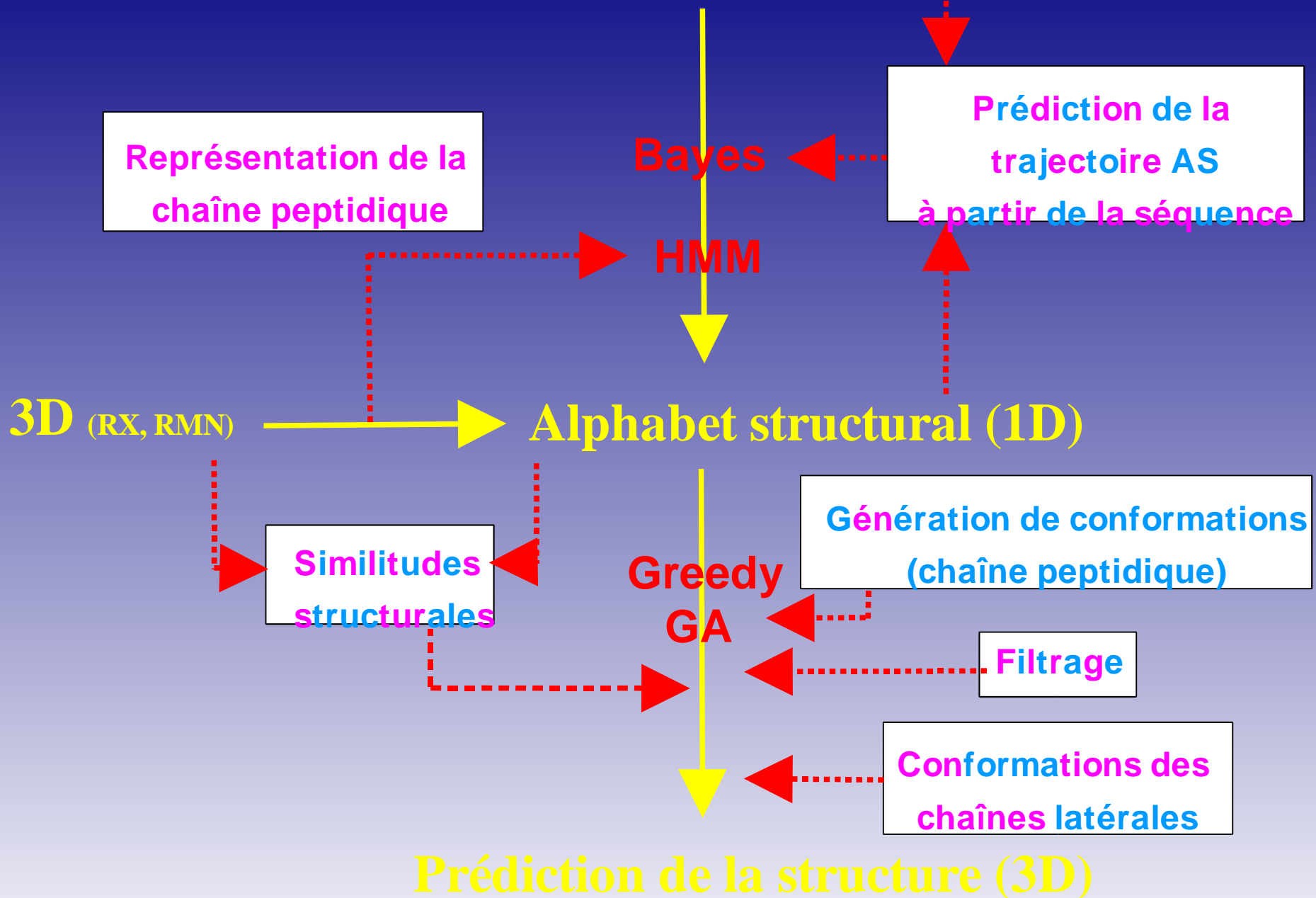
$(\alpha_1, \alpha_2)$  décrit les  $\alpha$ -helices  
 $\alpha_1 \sim$  coeur des helices,  
 $\alpha_2 \sim$  extrémités

Cluster B:

$(\beta_1, \beta_2) \sim$   $\beta$ -strands + ou -  
réguliers

Autres: décomposent les régions  
en boucles

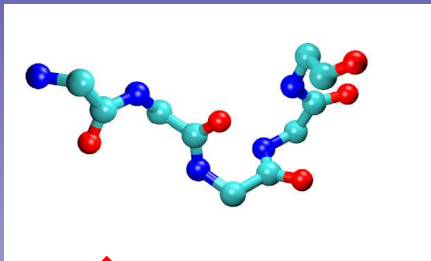
# Séquence protéique (1D)



# Prédiction de la structure locale à partir de la séquence

## Approche statistique Bayésienne

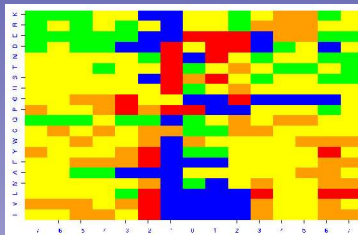
### Repliement local



SYARMDIGTTHDDYA  
RMDIGTTHDDYANDV  
THEVLAPGCLDAFPL  
GRDTSVEGSEMVPVK  
VIGLLEPMKKSMPV  
CVMLKSRGSRGHVRF  
GRLGLGEGAEKSI  
HLWVHQEGIYRDEYQ  
LMWQLYPEERYMDNN  
MWQLYPEERYMDNNS  
QHMLMVRGVQELLLK  
NLADVAPGKSIGGDI  
APGKSIGGDIFSNRE  
QIAKYFDRKQIGNAM ...

### Matrice de contingence

Acides  
aminés



Positions

### Séquence Nouvelle

SFITPVPGGVGPMTVFLEMDLTNKNVIFVA  
DKRKGGPGGI IANICVHTFNWSWLDVEPRVA  
IEANKNGAIWKLDLAIWKLDLGTLEAIEWW  
DSHIGAFLDKPKMENAQQQGNGLRYGLSSD  
AHTAVIGLPSGLES AVIGLPSGLESWSFFF  
AVYDGHAGSQVAKY...

*Théorème de  
Bayes*

Prédiction

*indice de confiance de la  
prédiction*

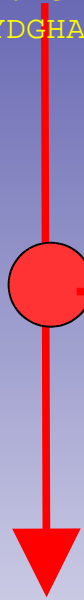
# Travaux en cours

## Approches utilisant les transitions entre les Blocs Structuraux :

### • Utilisation des Chaînes de Markov Cachées

Séquence

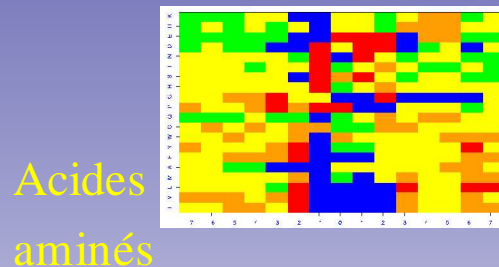
SFITPVPGGVGPMTVFLEMDLTNKNVIFV  
ADKRKGGPGGI IANICVHTFNSWLDVEPR  
VAIEANKNGAIWKLDLAIWKLDLGTLEAI  
EWWDSHIGAFLDKPKMENAQQQGNGLYG  
LSSDAHTAVIGLPSGLES AVIGLPSGLES  
WSFFFAVYDGHAGSQVAKY . . .



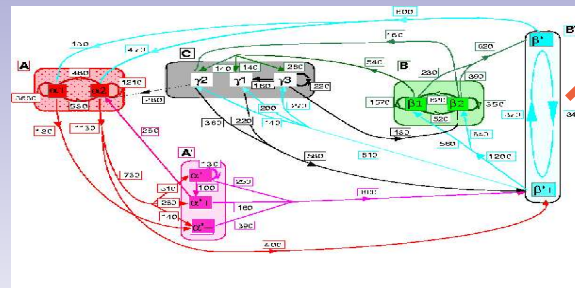
Prédiction

*indice de confiance de la  
prédiction*

Matrice de contingence



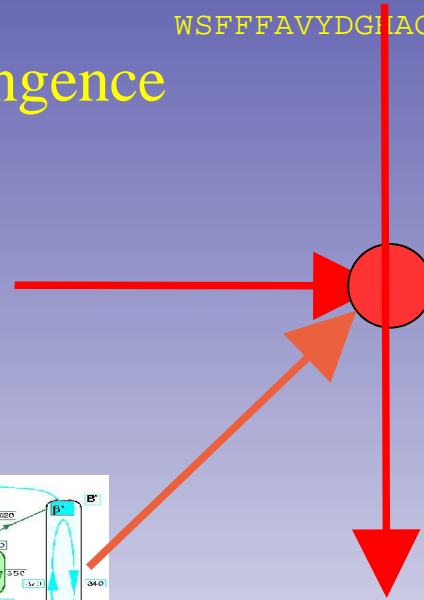
Acides  
aminés



Matrice de transitions

Séquence

SFITPVPGGVGPMTVFLEMDLTNKNVIFV  
ADKRKGGPGGI IANICVHTFNSWLDVEPR  
VAIEANKNGAIWKLDLAIWKLDLGTLEAI  
EWWDSHIGAFLDKPKMENAQQQGNGLYG  
LSSDAHTAVIGLPSGLES AVIGLPSGLES  
WSFFFAVYDGHAGSQVAKY . . .



*Viterbi  
(programmation  
dynamique)*

Prédiction