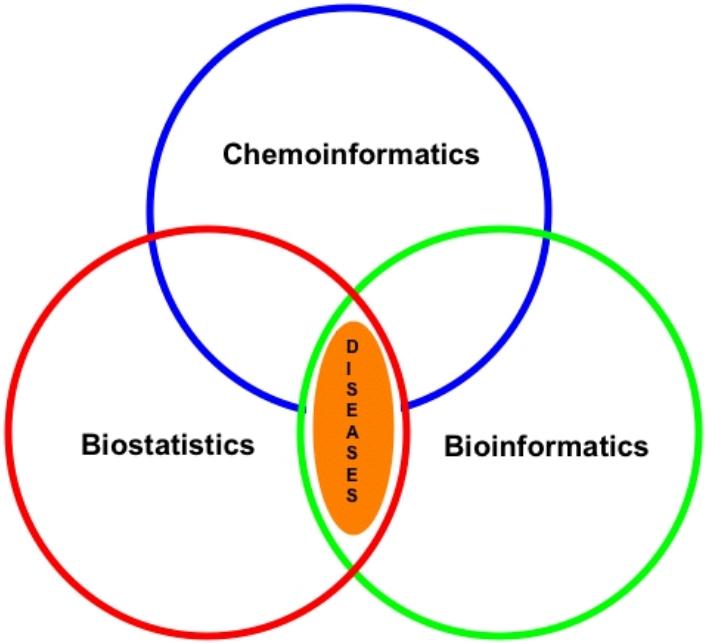
Peptide modeling in isolation and in interaction : steps towards rational peptide design

Pierre Tufféry

GT MASIM
Paris, 17 novembre 2017



Equipe 1 : "Structure-based peptide design"
(Dr P. Tuffery)

Equipe 2 : "Computational approaches applied to pharmacological profiling"
(Pr A-C. Camproux, Pr O. Taboureau)

Equipe 3 : "Virtual screening and rational design of protein-protein interaction modulators with balanced ADME-Tox properties"
(Dr M. Miteva)

docker

Virtualization

Galaxy

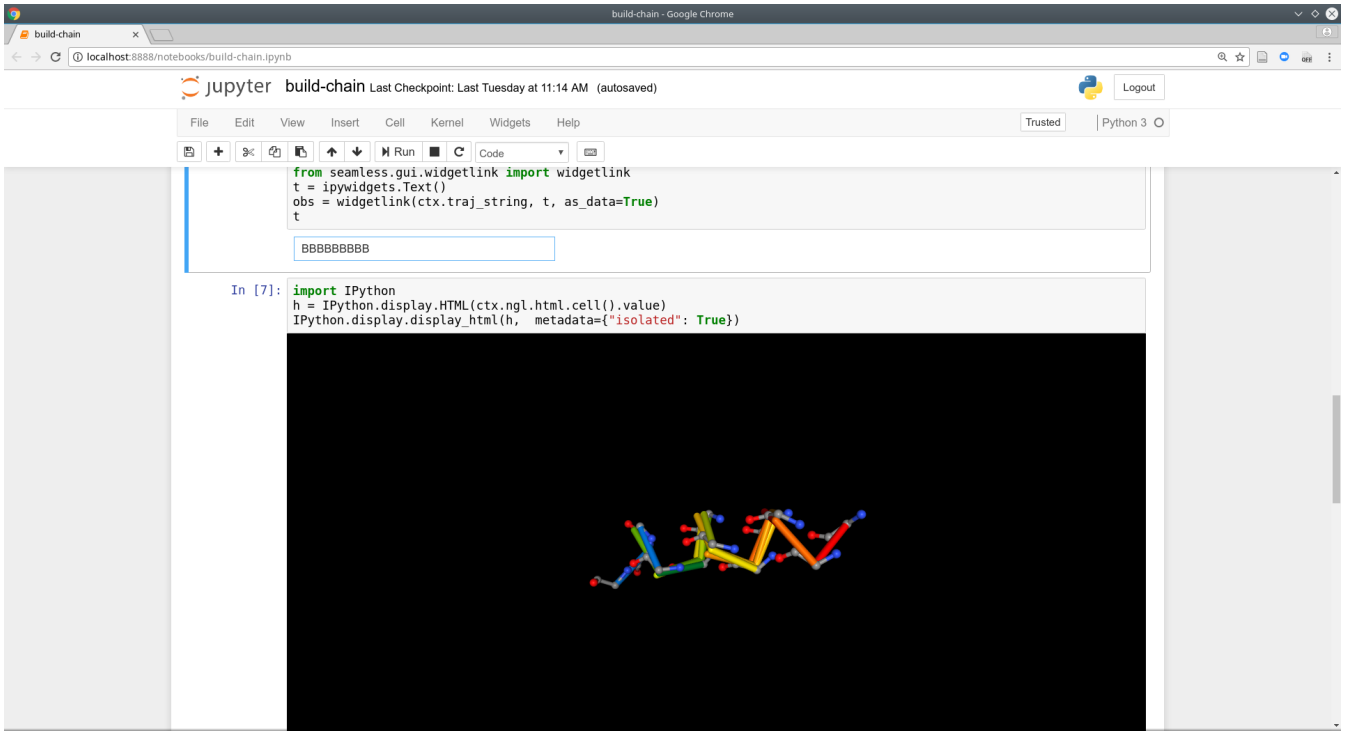
Mobyle

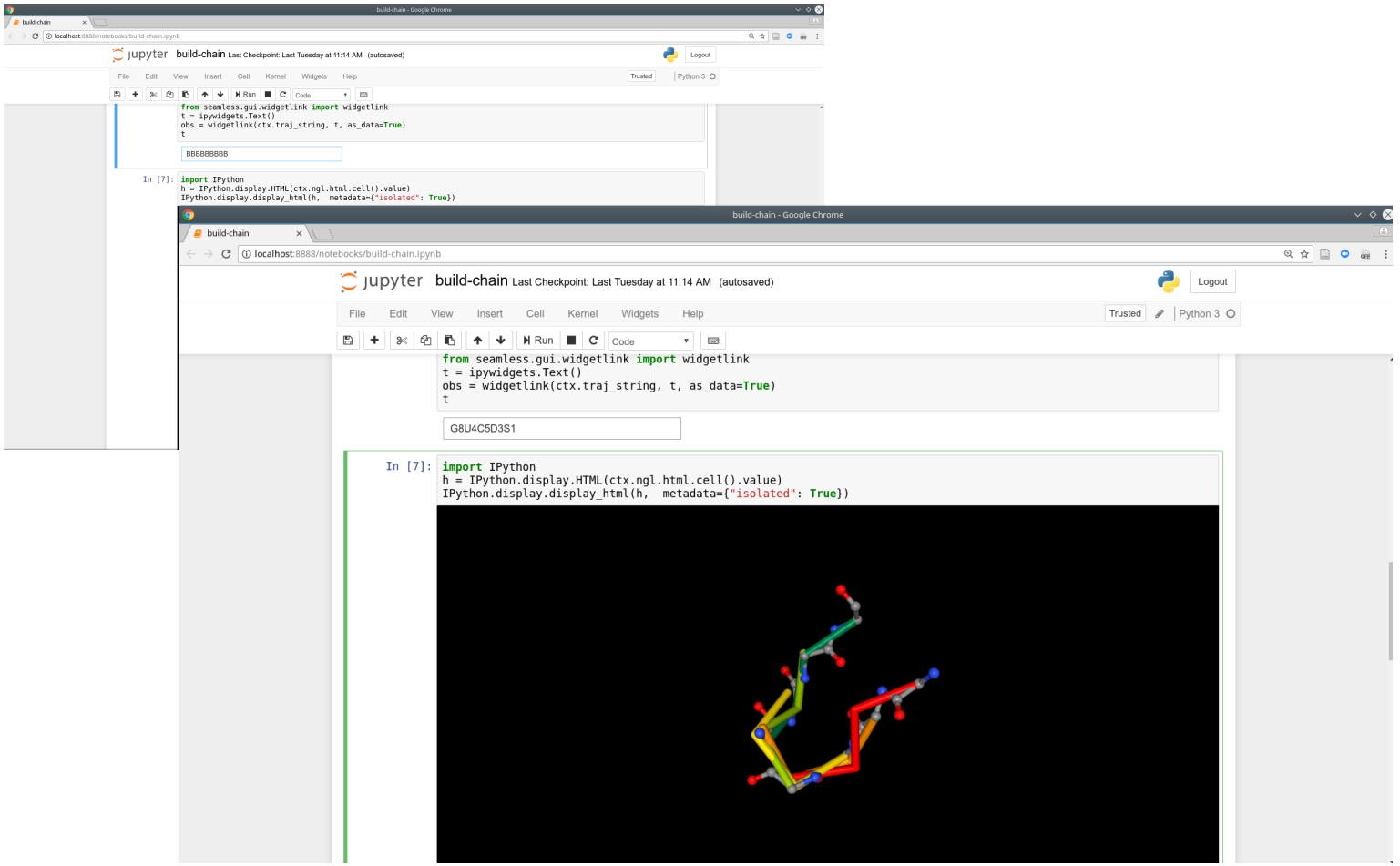
Service publication, workflows

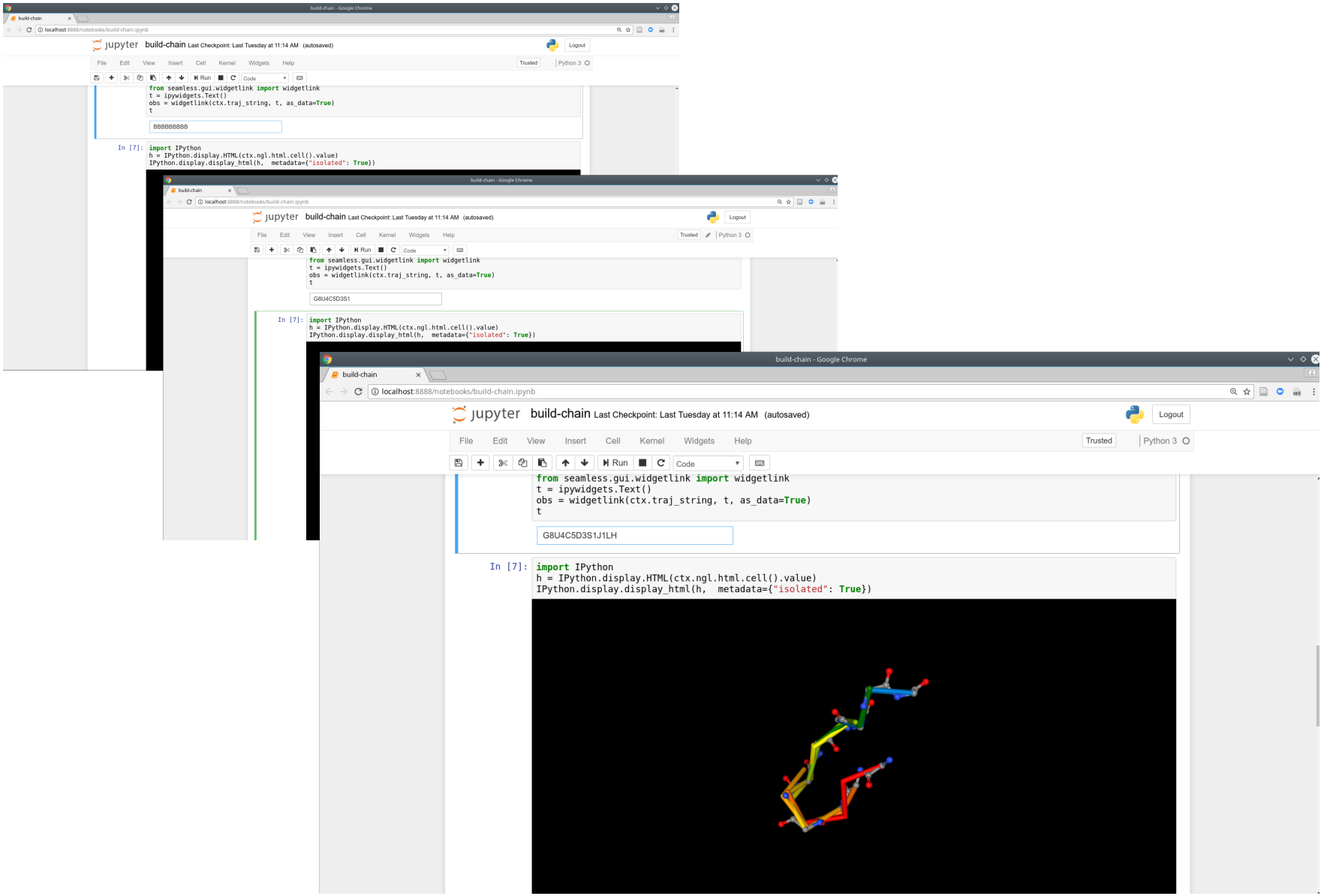
cloud

jupyter

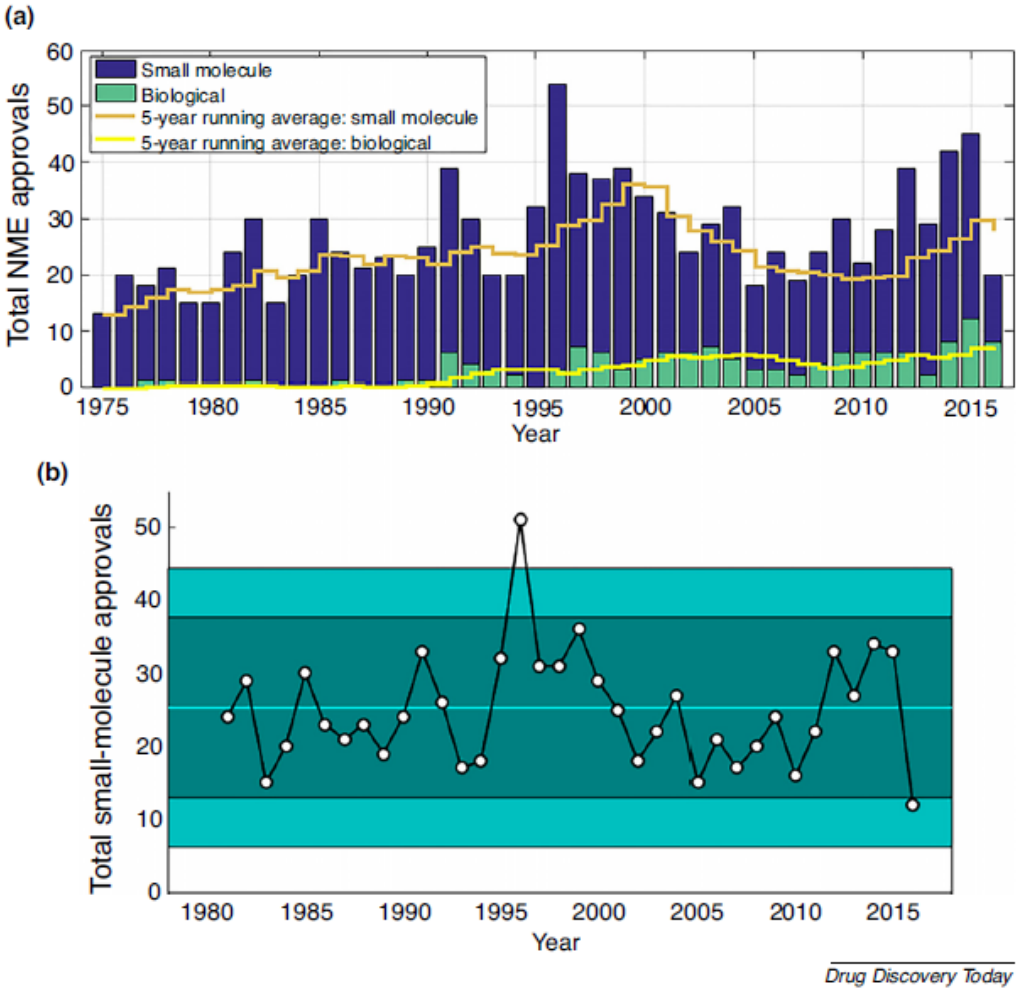
All-in-one !


build-chain - Google Chrome
build-chain
localhost:8888/notebooks/build-chain.ipynb
jupyter build-chain Last Checkpoint: Last Tuesday at 11:14 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
from seamless.gui.widgetlink import widgetlink
t = ipywidgets.Text()
obs = widgetlink(ctx.traj_string, t, as_data=True)
t
BBBBBBBBB
In [7]:
import IPython
h = IPython.display.HTML(ctx.ngl.html.cell().value)
IPython.display.display_html(h, metadata={"isolated": True})


build-chain - Google Chrome
build-chain
localhost:8888/notebooks/build-chain.ipynb
jupyter build-chain Last Checkpoint: Last Tuesday at 11:14 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
from seamless.gui.widgetlink import widgetlink
t = ipywidgets.Text()
obs = widgetlink(ctx.traj_string, t, as_data=True)
t
G8U4C5D3S1
In [7]:
import IPython
h = IPython.display.HTML(ctx.ngl.html.cell().value)
IPython.display.display_html(h, metadata={"isolated": True})


build-chain - Google Chrome
build-chain
localhost:8888/notebooks/build-chain.ipynb
jupyter build-chain Last Checkpoint: Last Tuesday at 11:14 AM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
from seamless.gui.widgetlink import widgetlink
t = ipywidgets.Text()
obs = widgetlink(ctx.traj_string, t, as_data=True)
t
G8U4C5D3S1J1LH
In [7]:
import IPython
h = IPython.display.HTML(ctx.ngl.html.cell().value)
IPython.display.display_html(h, metadata={"isolated": True})

Therapeutic peptides: why ?



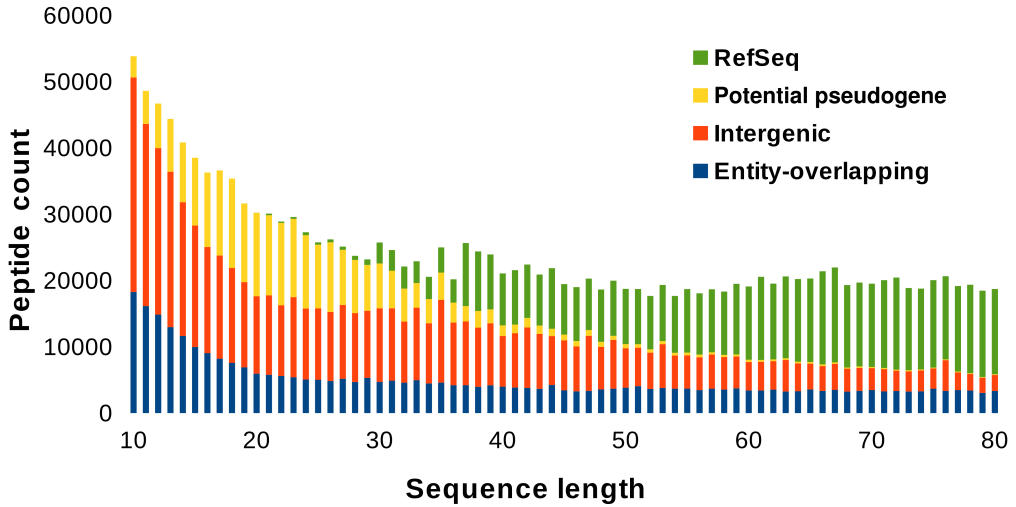
Griesenauer et al., Drug. Discov. Today, 2017

BactPepDB

Bacteria Peptide Database

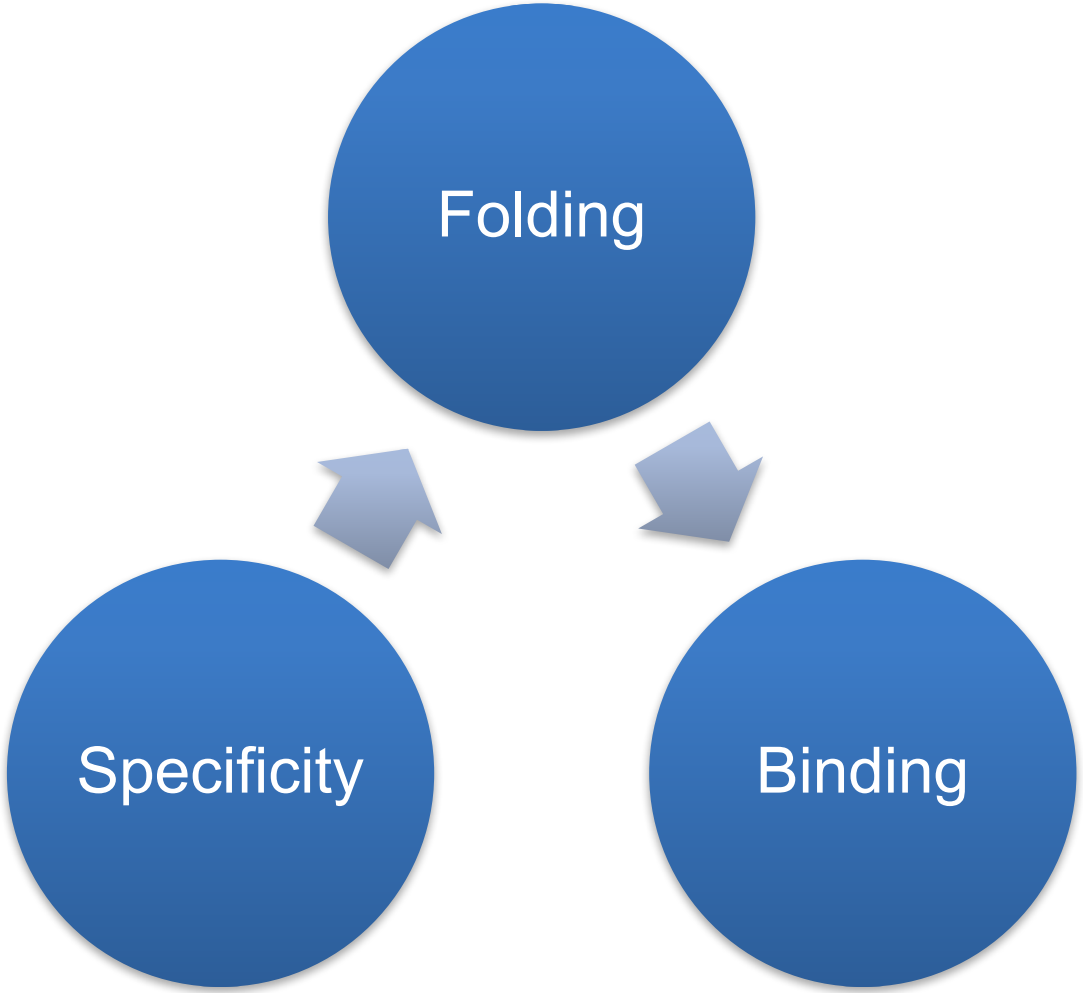
Last production update (15/10/2015) :

- Prokaryotes genomes :
  - ~ 100 orders
  - ~ 200 families
  - ~ 700 genera
  - ~ 1,500 species
  - ~ 2,700 strains
- Total : ~ 2,000,000 peptides
  - ~ 70 % of newcomers
  - ~ 200,000 (~ 20 %) of new intergenic SCSs are conserved to some extent : consistent with genes found in RefSeq



Rey et al., Database, 2014

Towards *rational* peptide design

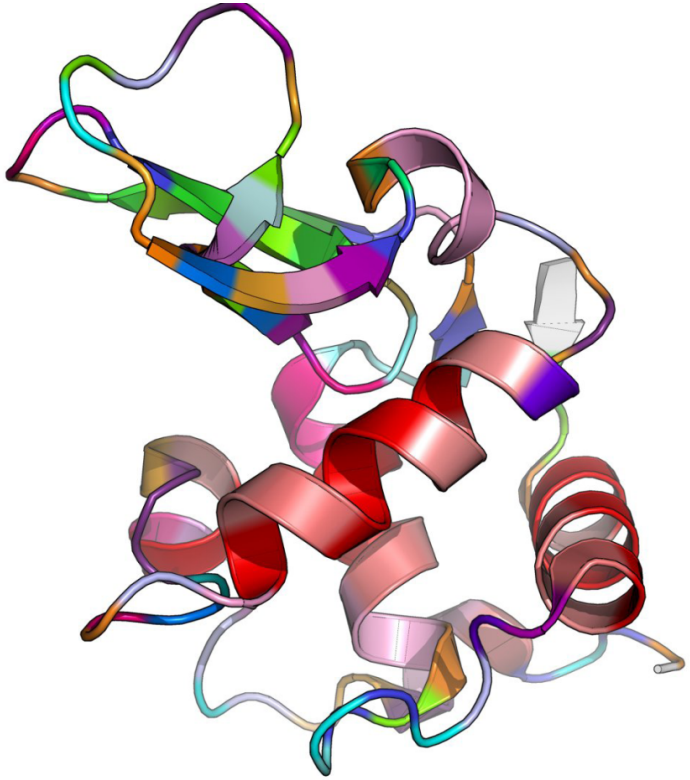


De novo peptide structure modeling

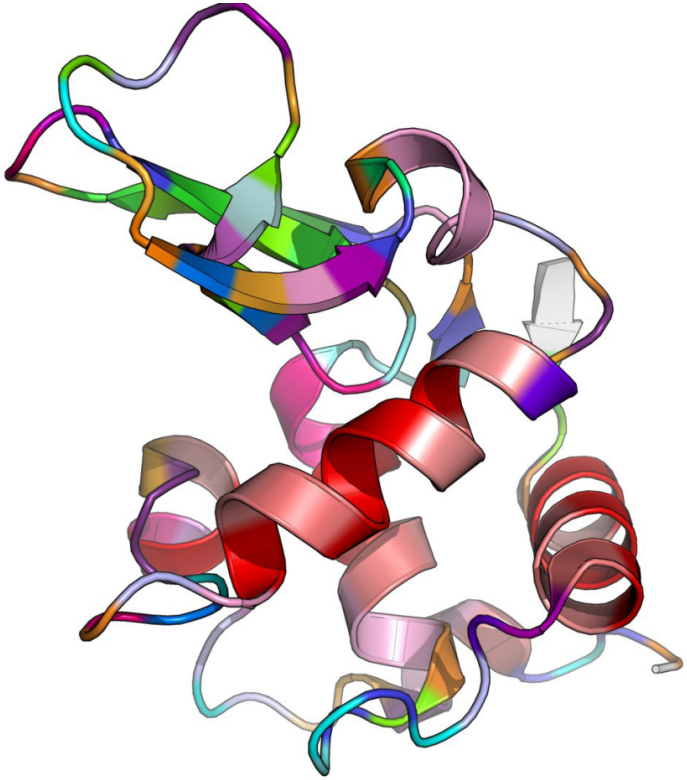
Probabilistic model of polypeptidic chain

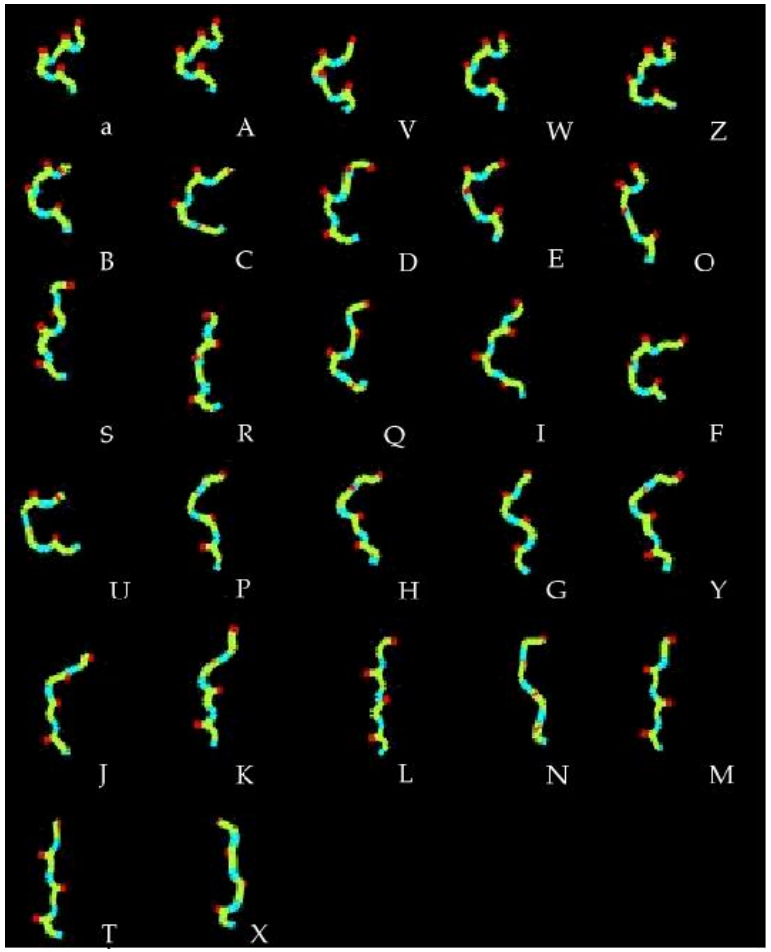
Scoring

Conformational Sampling



HMM derived Structural alphabet

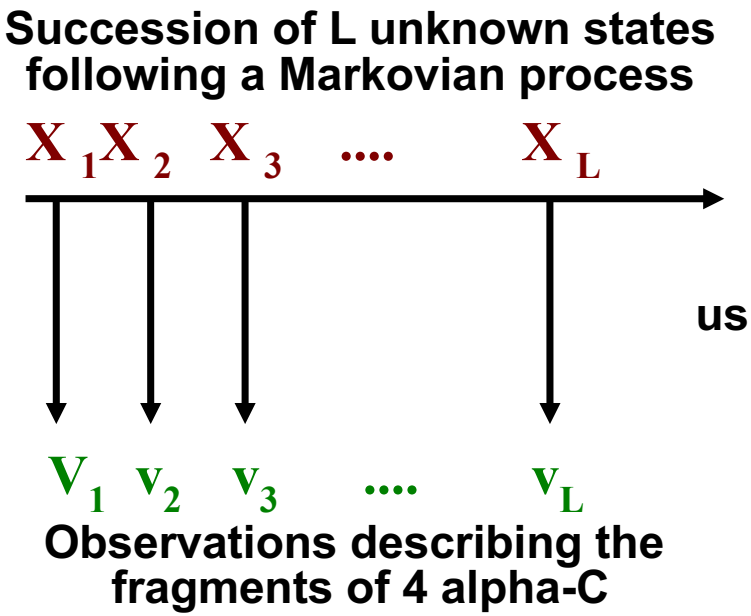




>HMM-SA
NLHWAAAAAVWAVDOQUSUFSLHBBVWAAAVZZFFFSPBS
XTLNHZDSNLNJFZDRLPECCILGDEQLUGPRGBDSKHBB
BBQHEGOWAVWAAAVWABQHZRUEEEGWAAZCCQUQYGEB
BVSUSP

A. C. Camproux et al., Prot. Eng, 1999, *J Mol Biol* 339, 591–605 (2004)

HMM derived Structural alphabet



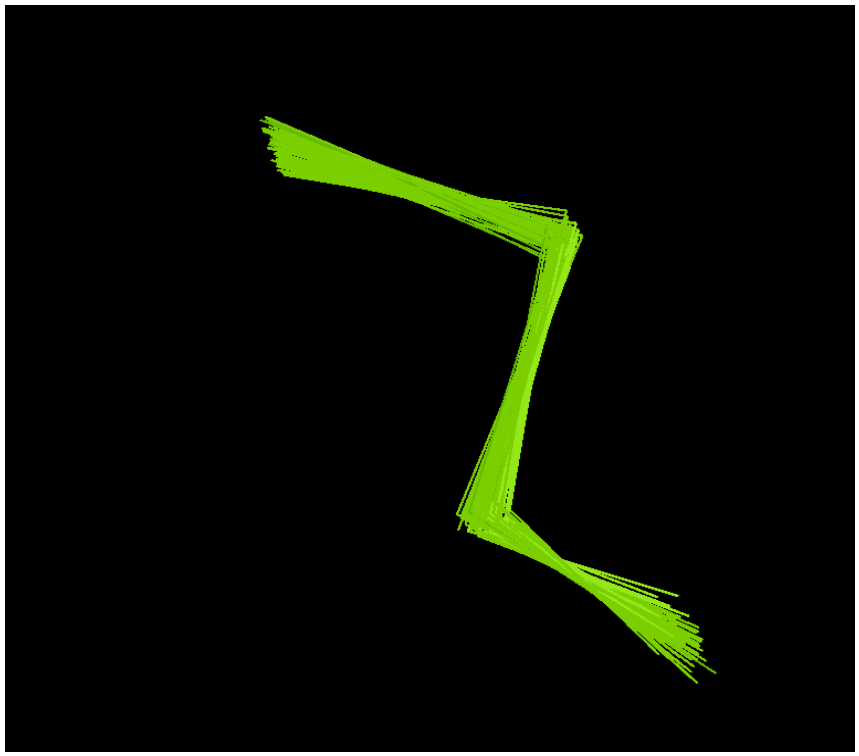
***R* hidden states $\{S_1,...,S_R\}$ emitting the vectors of descriptors of each fragment ~ R multi-gaussian densities** $f_{xi}(y,\theta_i)$ *with* $\theta_i = (\mu_i, \sigma_i^2)$

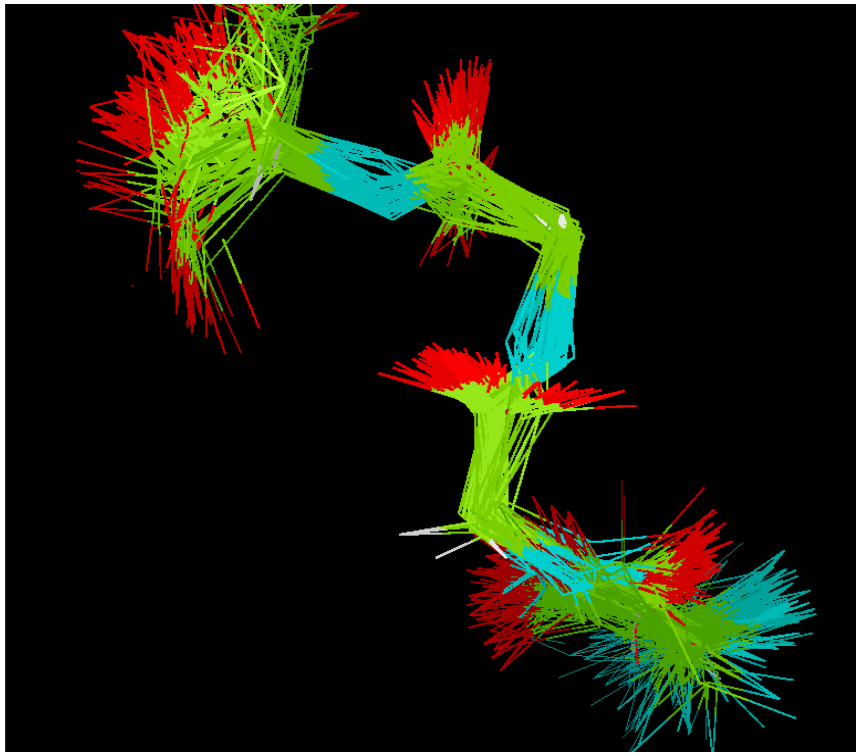
N states of a protein $(X_1,..,X_L)$ **~ R states Markov chain (order 1)**

Transition Matrix (R*R) : $\Pi_{jl} = P(X_i=S_j \mid X_{i-1}=S_l),\ 1<j,l<R$

Initial law of the chain (R) : $\upsilon_j = P(X_1=S_j),\ 1<j<R$

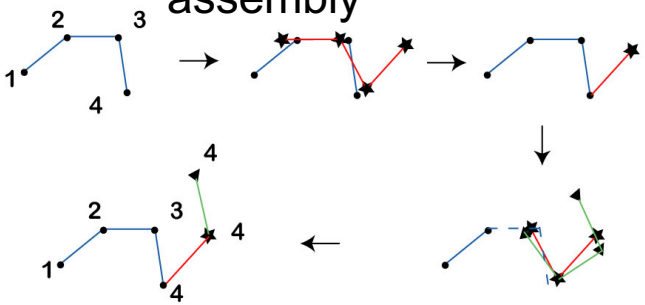
SA letters Phi/Psi

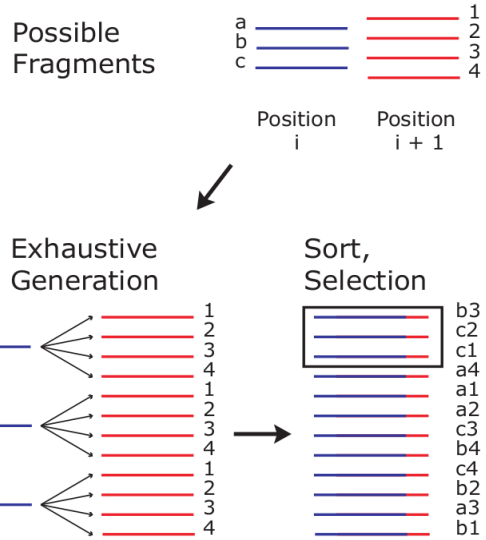




HMM-SA letter “O”

Generating models





1. Grow peptide
2. Monte-Carlo (~30 000 steps)

Conformer
heap

Maupetit et al., J Comput Chem., 2010
Maupetit et al.,, Nucleic acids Res., 2009

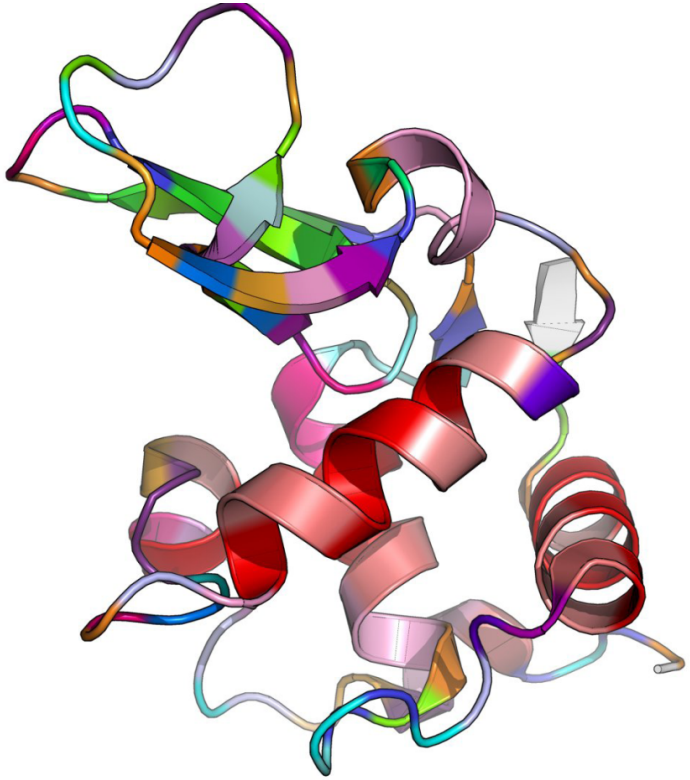
Tuffery et al., J Comput Chem., 2005
Tuffery and Derreumaux, Proteins., 2005

De novo peptide structure modeling

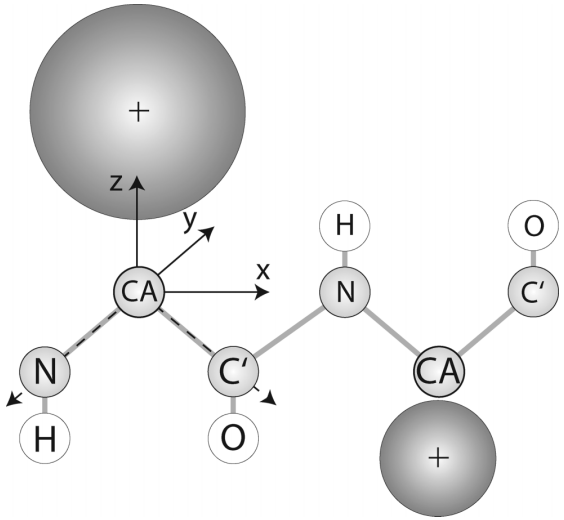
Probabilistic model of polypeptidic chain

Scoring

Conformational Sampling



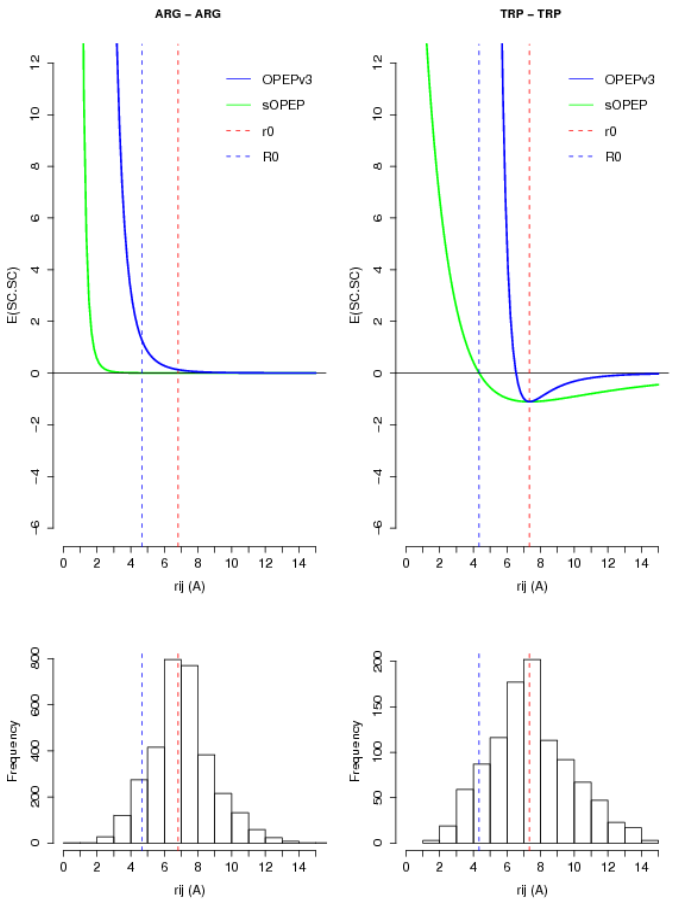
Generating models



$$E = E_{\text{local}} + E_{\text{nonbonded}} + E_{\text{H-bond}}$$

$$E_{\text{HB1}} = w_{\text{hb1-4}} \sum_{ij,j=i+4} \varepsilon_{\text{hb1-4}} \mu(r_{ij}) \nu(\alpha_{ij}) + w_{\text{hb1>4}} \sum_{ij,j>i+4} \varepsilon_{\text{hb1>4}} \mu(r_{ij}) \nu(\alpha_{ij}) \qquad (7)$$

$$E_{\text{HB2}} = \sum \varepsilon_{\alpha}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \exp(-(r_{kl} - \sigma)^2/2) \times \Delta(ijkl) + \sum \varepsilon_{\beta}^{\text{coop}} \exp(-(r_{ij} - \sigma)^2/2) \times \exp(-(r_{kl} - \sigma)^2/2) \Delta'(ijkl) \qquad (10)$$



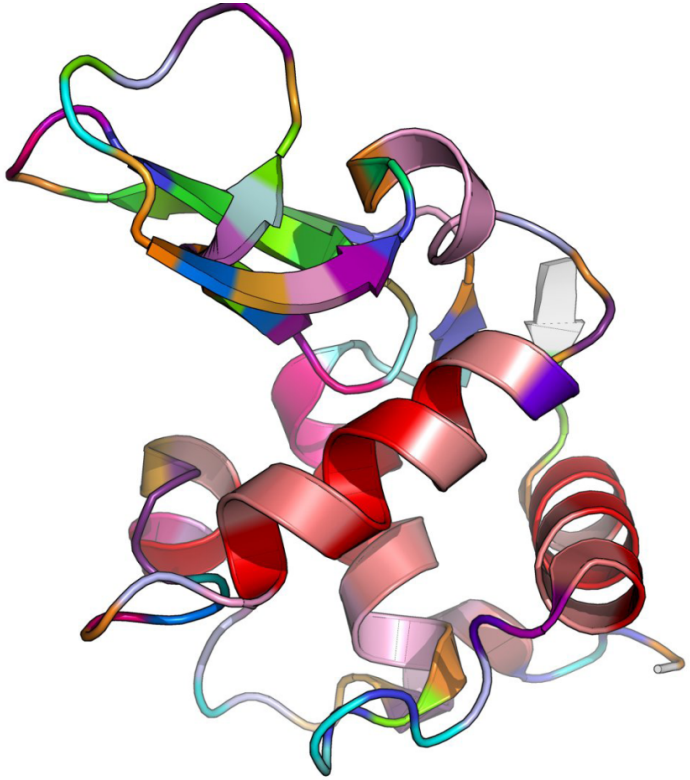
Maupetit et al., Proteins, 2007
Maupetit et al., J. Comput Chem., 2009

De novo peptide structure modeling

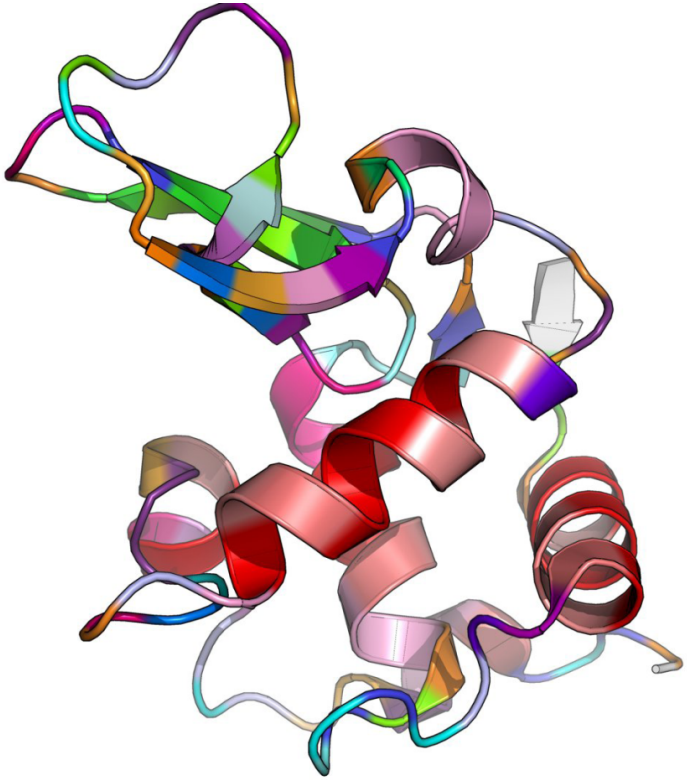
Probabilistic model of polypeptidic chain

Scoring

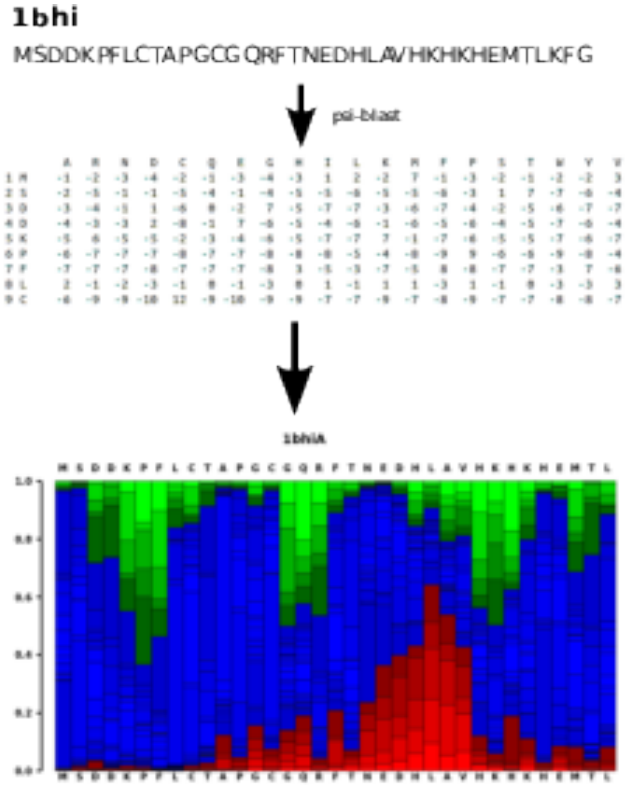
Conformational Sampling



Peptide structure de novo modelling (PEP-FOLD)



>HMM-SA
NLHWAAAAAVWAVDOQUSUFSLHBBVWAAAVZZFFFSPBS
XTLNHZDSNLNJFZDRLPECCILGDEQLUGPRGBDSKHBB
BBQHEGOWAVWAAAVWABQHZRUEEEGWAAZCCQUQYGEB
BVSUSP



Amino acid sequence

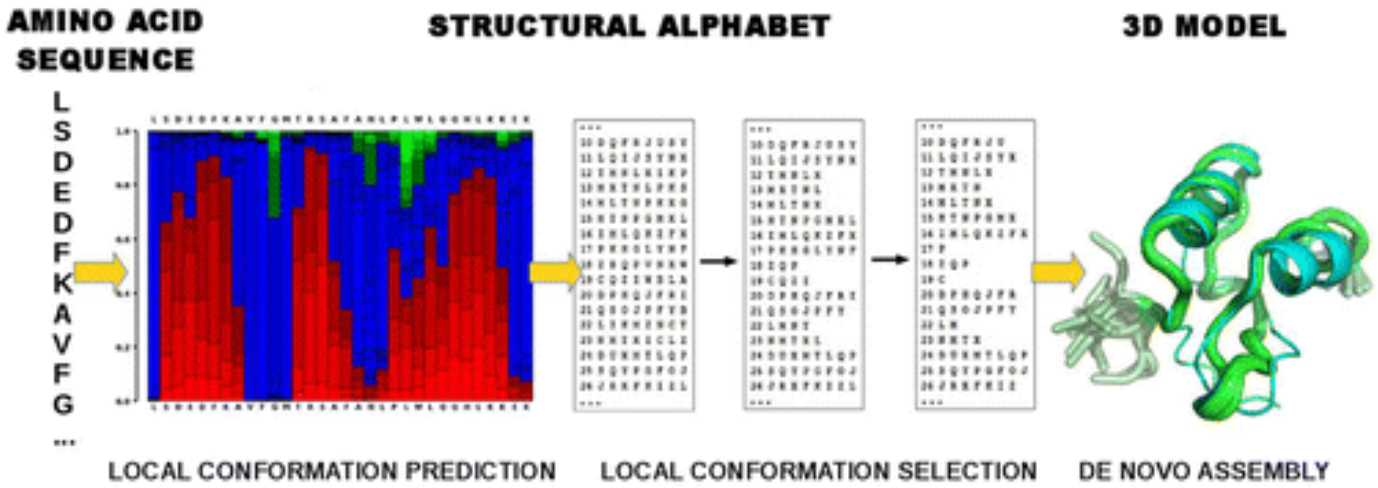
PSSM

Local structure profile

P(SA|Obs)

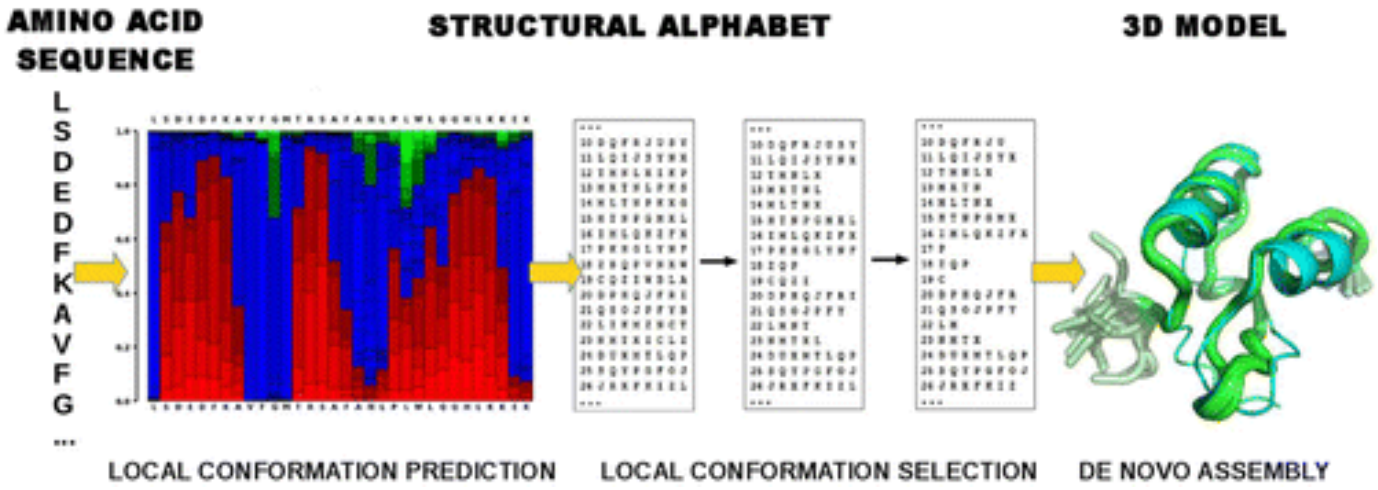
Camproux et al., J. Mol. Biol., 2004

Peptide structure de novo modelling (PEP-FOLD1-2)

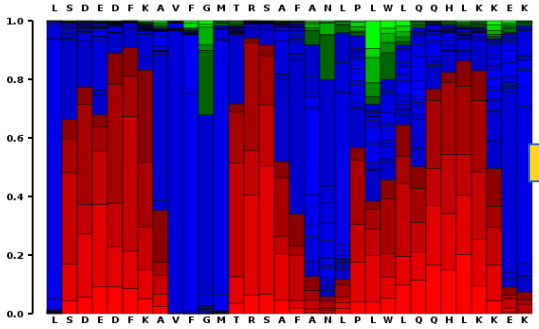


Maupetit et al. J. Comput Chem., 2009; Shen et al. *J. Chem. Theor. Comput., 2014*

Peptide structure de novo modelling (PEP-FOLD3)

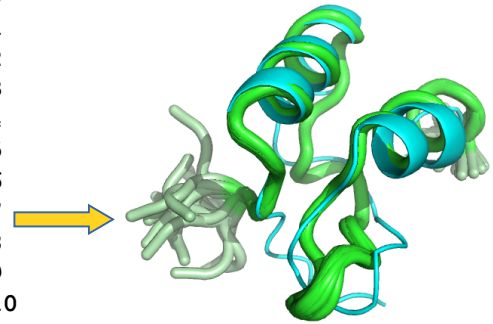


Complexity ~4.8



P(Obs|SA) = P(SA|Obs) * P(Obs) / P(SA)

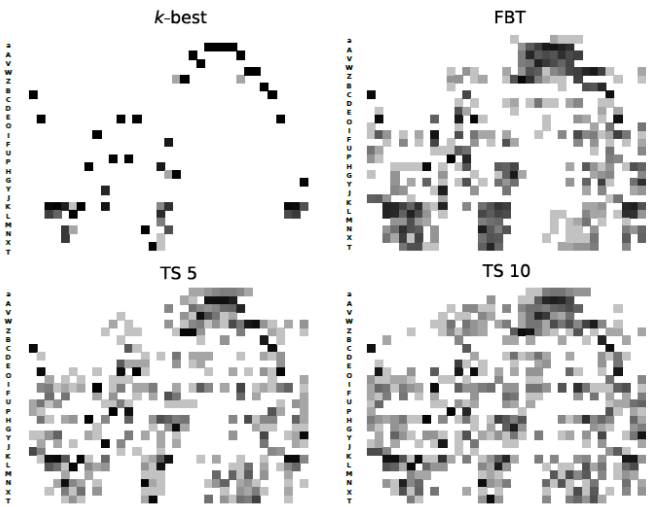
```
DSLLNLYKKKUODSKTKLHVZWAAVWESLGGSNKR   0
DSLKNLXPXKUOBQNTLHBBZEWAAAZNTPPQXKK   1
DSXKNMNPQYUSUSXTLNHBZWAAAAVQPSPSKKK   2
DSLKXKKPQKUSGSTXMHBBEBQMNMNNPZDSLLU   3
JPIKXKYPSPZFDSNTLHBBBAABVWVZZCDSKPG   4
DSLKNMXPQYUSUSNKLNPBZVWAAAVZZCDSKKG   5
DSKHQLXPIFUSGSTTPIHBVWAAVWEGZCDSLKN   6
DSKLNMNPQKUSUSLTKPQPVWAAVZIPQGDSKLG   7
DSLLNLXPIYUSUSLPPIHVZWAAAAVZZCDFQFZ   8
USKKLLLGIKUSUSNTMNHZWAAAVSKGZZDSKLN   9
DSXKKNXKNYUEGITMXLHBBVWAAAVZZCDSLKK   10
DSKKXLNPQYUSUSNTMLHBBVWAAABZZCDQKLK   11
```

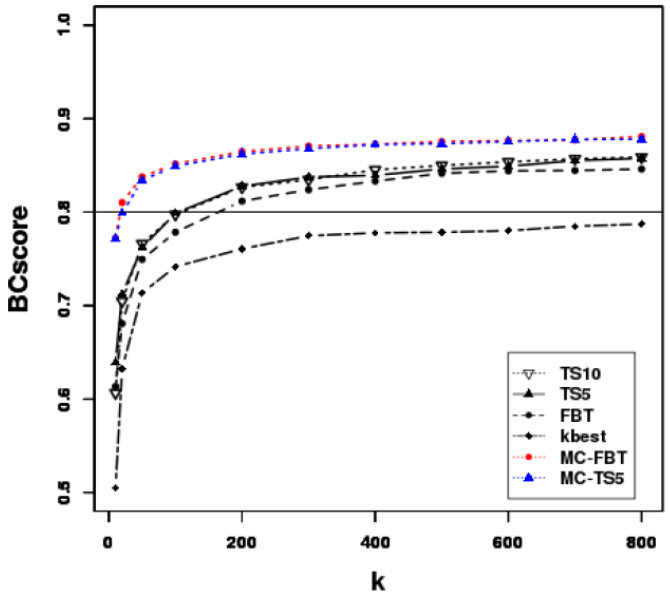


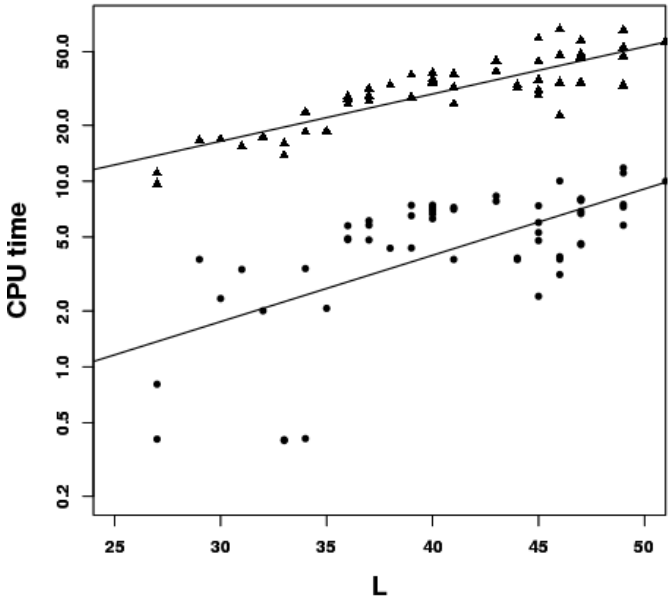
Complexity 1

Lamiable et al., J. Comput Chem, 2016 ; Nucleic Acids Res., 2016

Peptide structure de novo modelling (PEP-FOLD3)

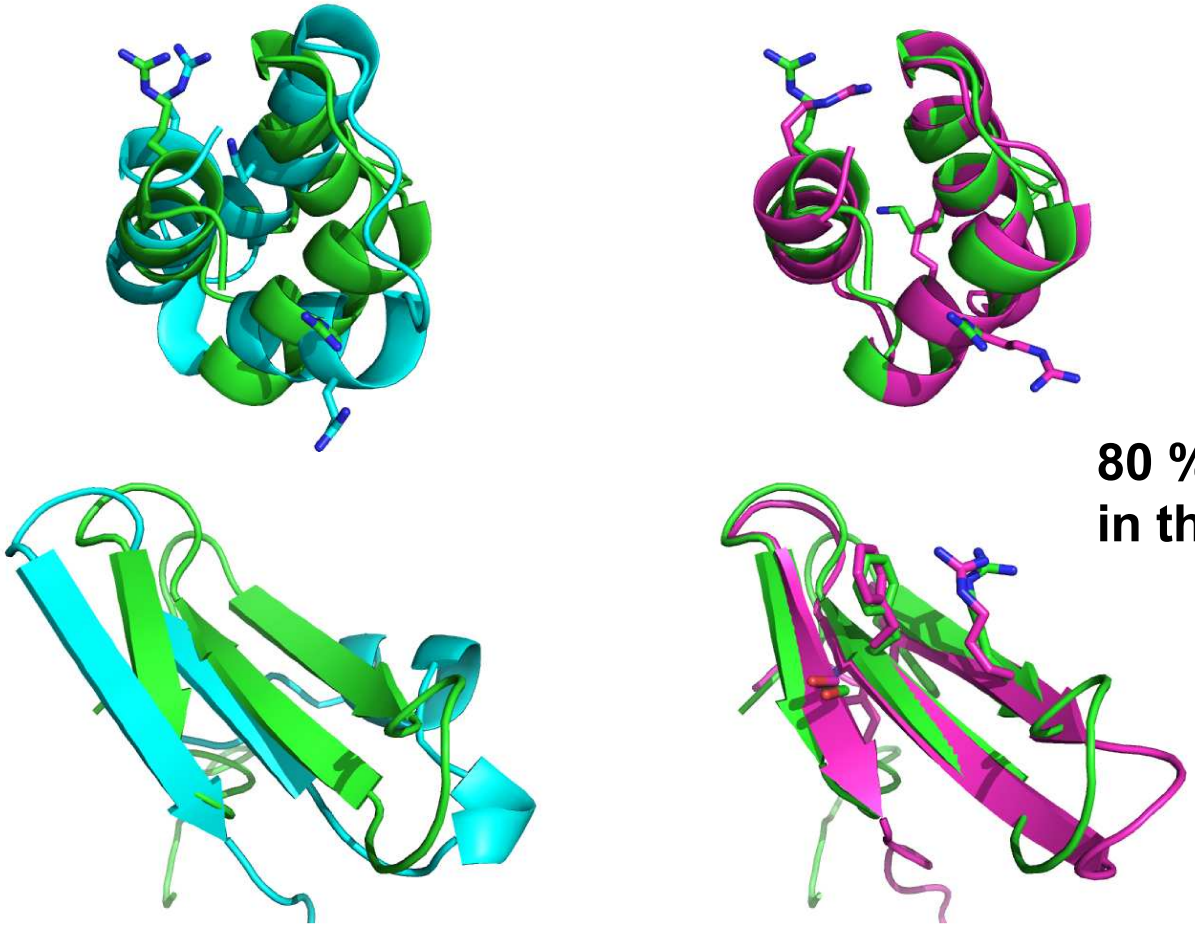






Lamiable et al., J. Comput Chem, 2016 ; Nucleic Acids Res., 2016

Peptide structure de novo modelling (PEP-FOLD)

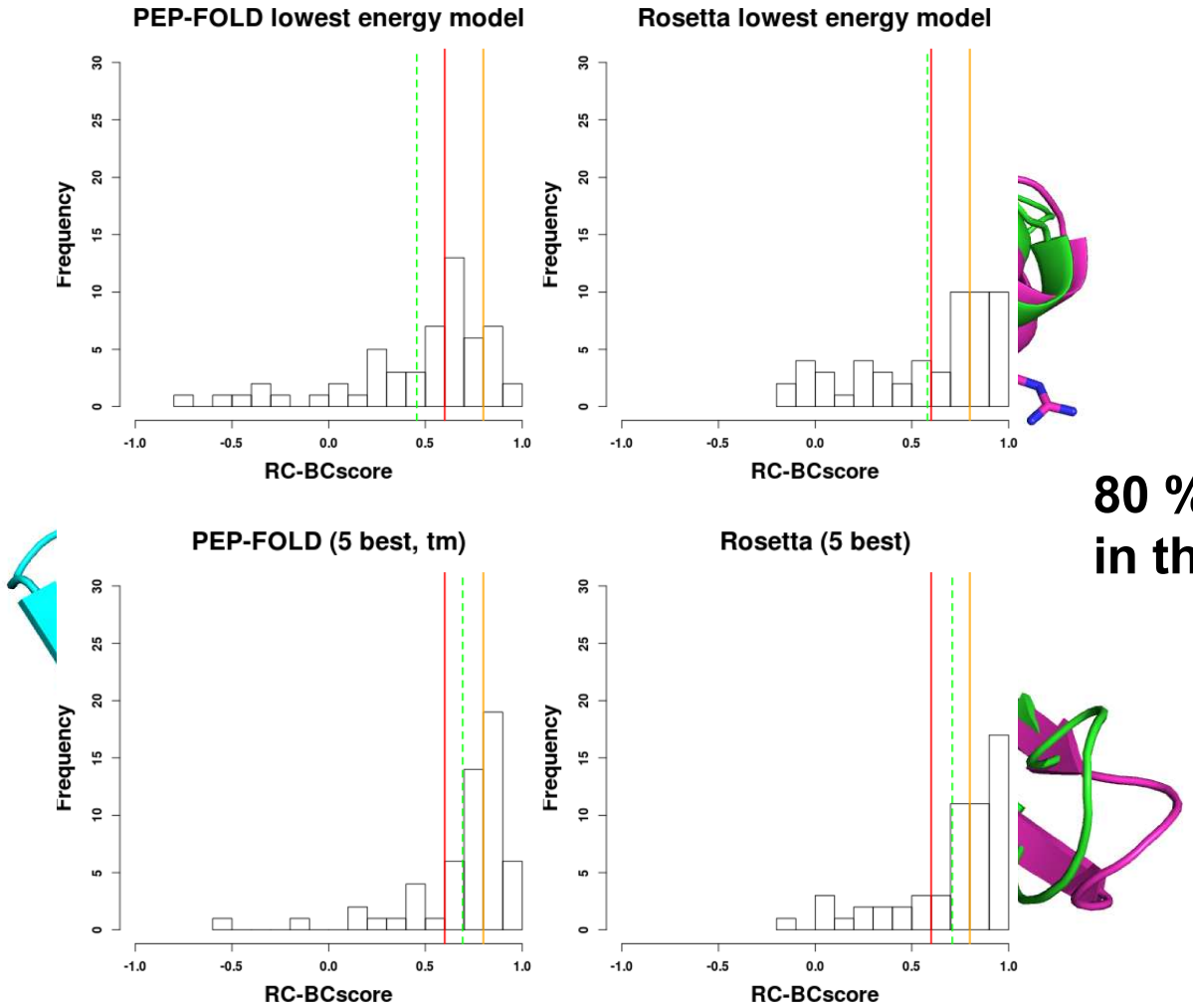


80 % native models in the top 5 models

PEP-FOLD1 (cyan) and PEP-FOLD2 (magenta)
compared to the experimental conformation (green) of 2jnh (top) and 1i6c (bottom)

Shen et al., J. Chem. Theor. Comput., 2014

Peptide structure de novo modelling (PEP-FOLD)

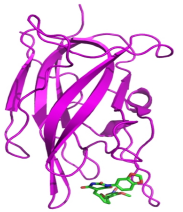


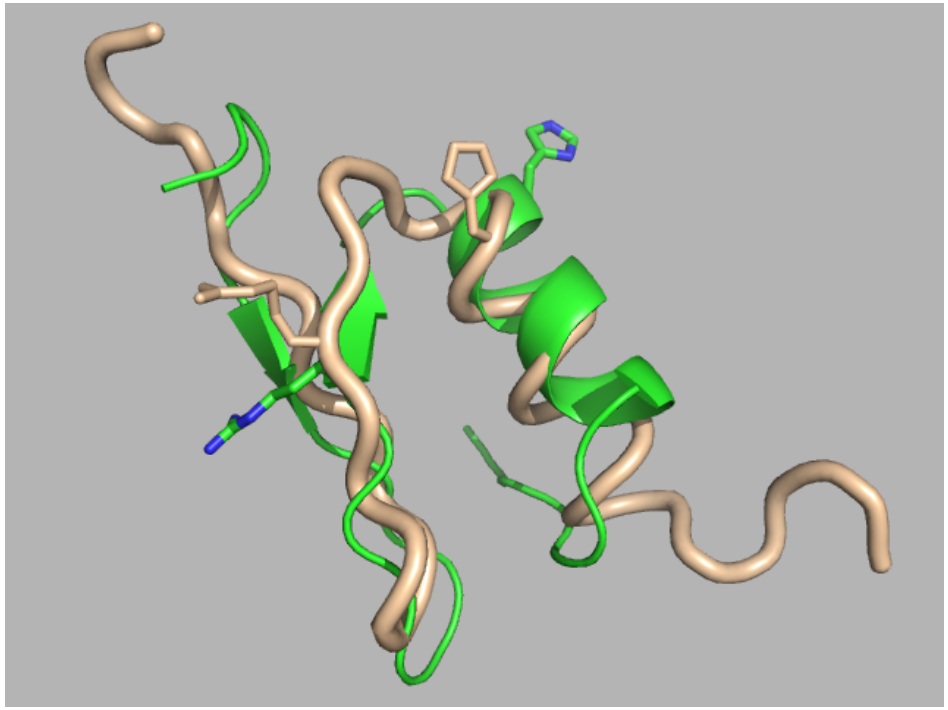
80 % native models in the top 5 models

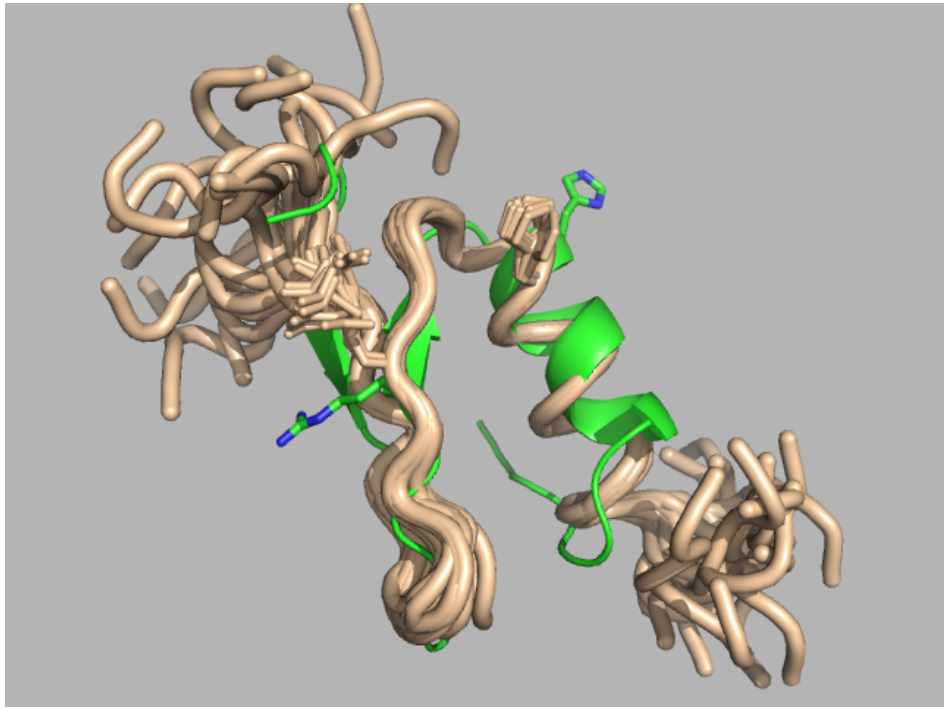
PEP-FOLD1 (cyan) and PEP-FOLD2 (magenta)
compared to the experimental conformation (green) of 2jnh (top) and 1i6c (bottom)

Shen et al., J. Chem. Theor. Comput., 2014

Peptide structure de novo modelling (PEP-FOLD)





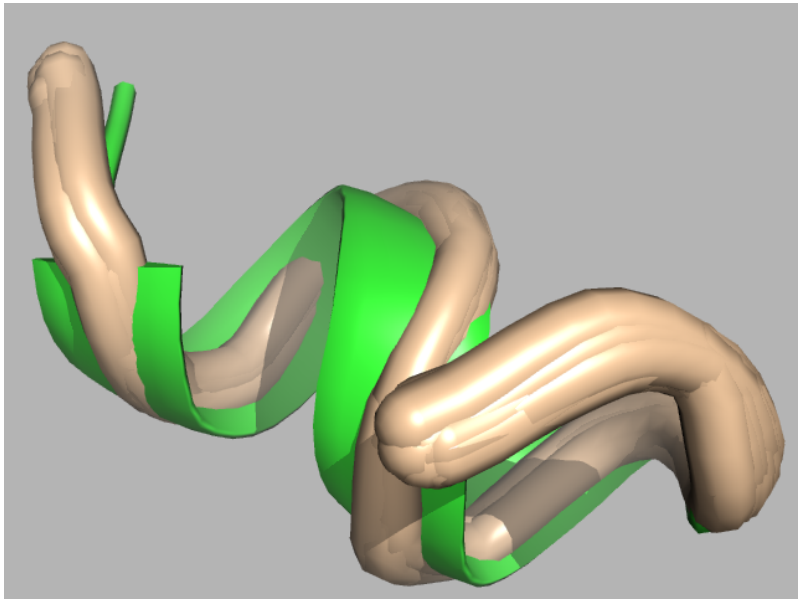


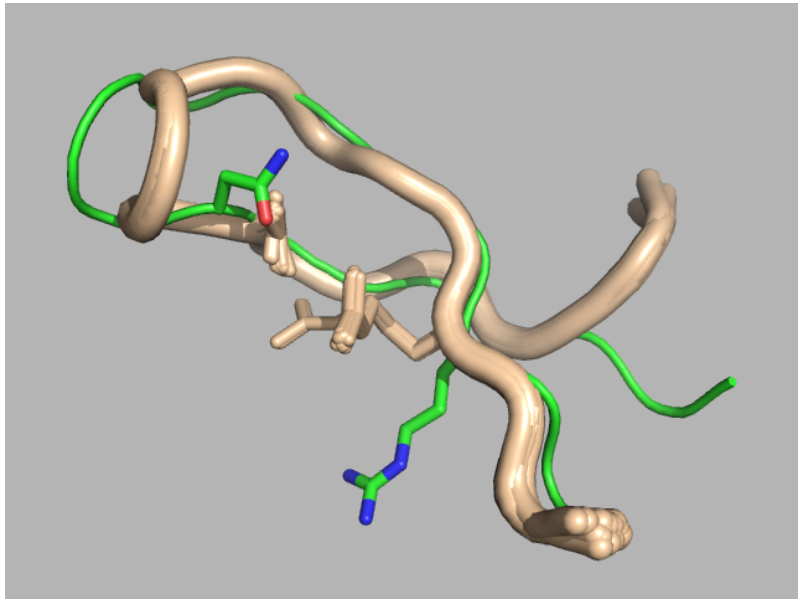
1bhi (zinc finger like)
Green : model rank2
Wheat : NMR

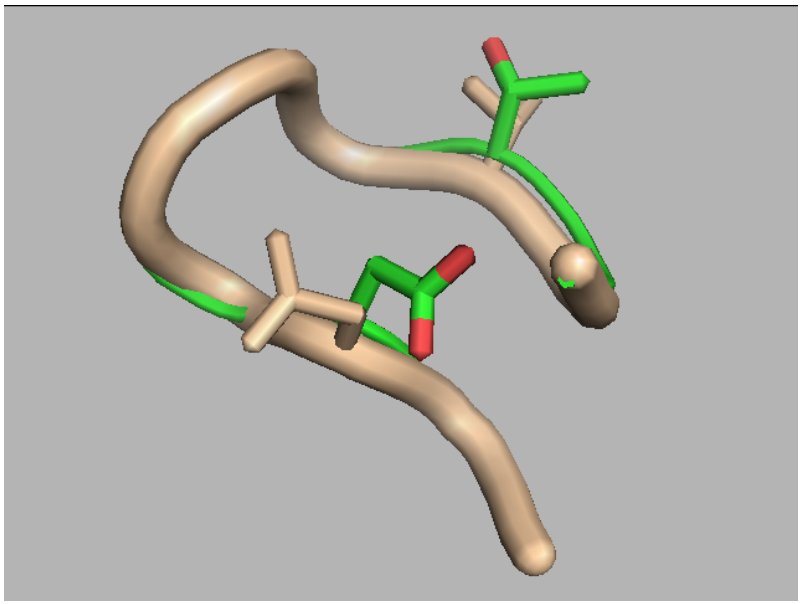
RMSd : 6.3A

RC-RMSd : 1.4A

Peptide structure de novo modelling (PEP-FOLD)







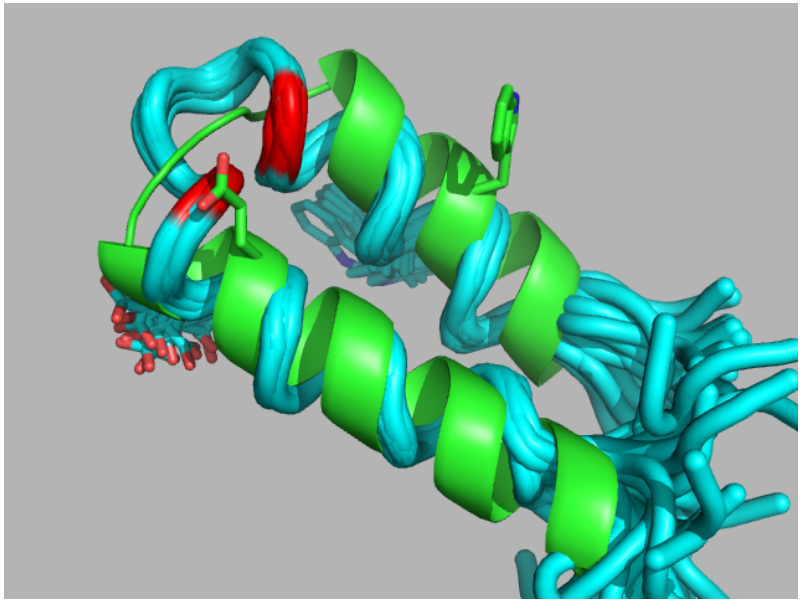
1aqg(11 aas)
rank 1
RC-RMSd : 0.9A

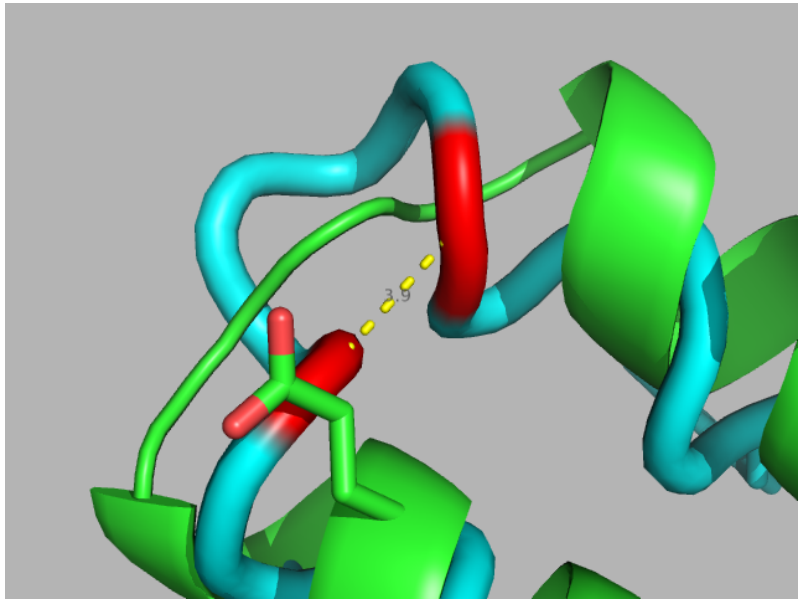
2oru(20 aas)
rank 2
RC-RMSd : 2.8A

Green : model
Wheat : NMR model

1uao (10 aas)
rank 1
RC-RMSd : 0.9A

Peptide structure de novo modelling (PEP-FOLD)

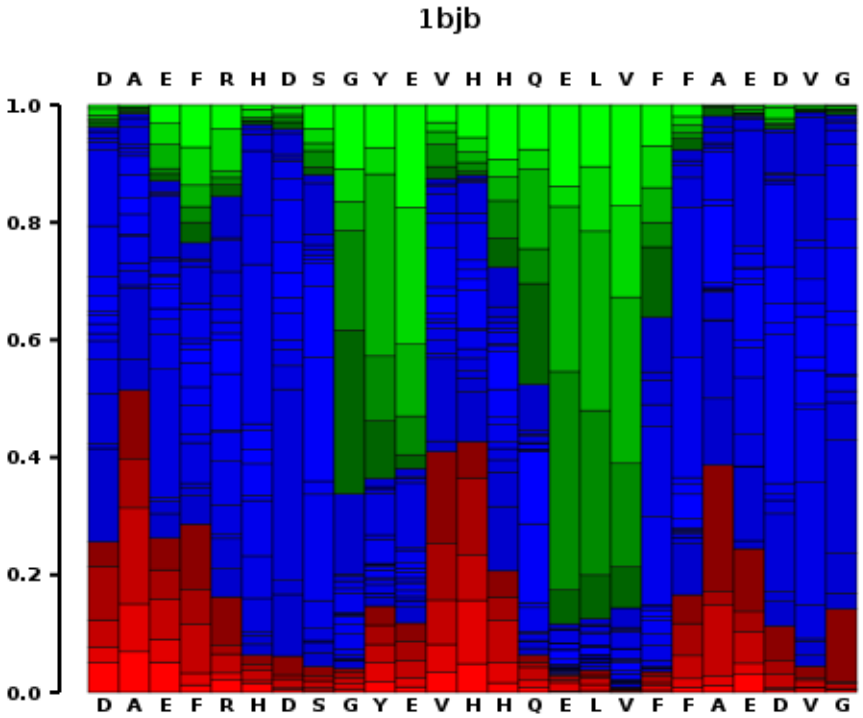


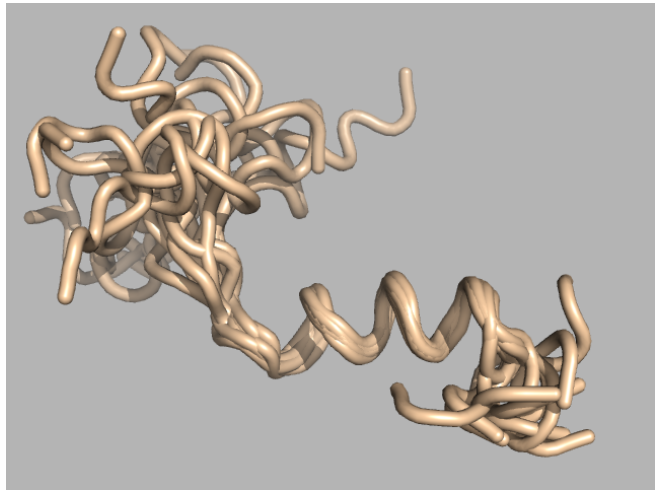


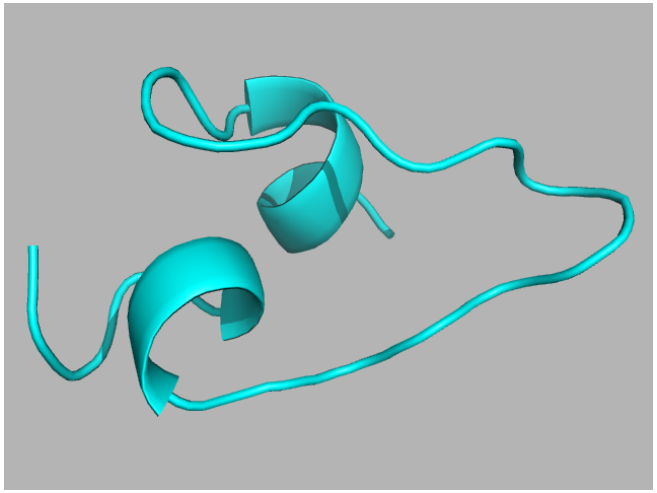
2bn6 (P-element somatic inhibitor)
Green : model rank 1
Cyan : NMR model

RC-RMSd : 4.3A

Peptide structure de novo modelling (PEP-FOLD)

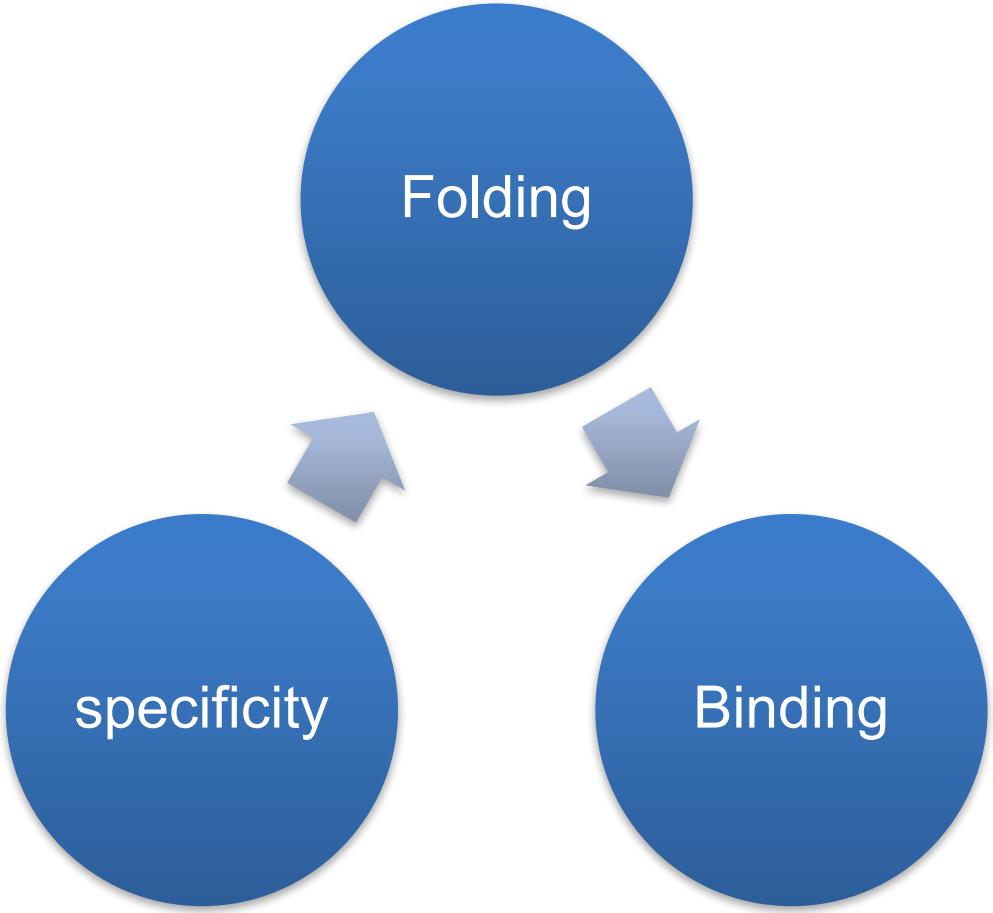




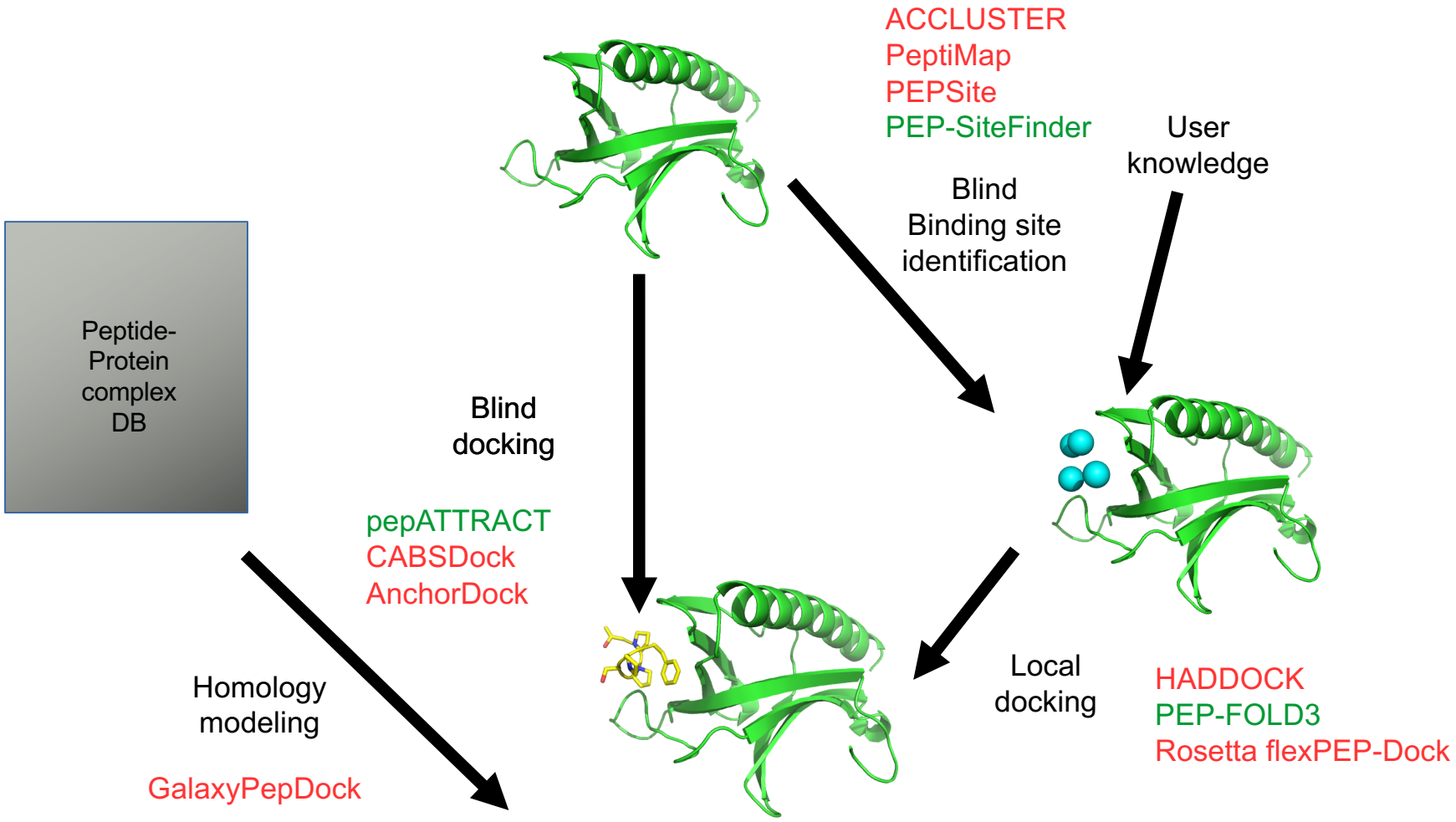


1bjb (Amyloid beta [e16], res.1-28)
Green : model rank 1
Cyan : NMR model

Towards *rational* peptide design

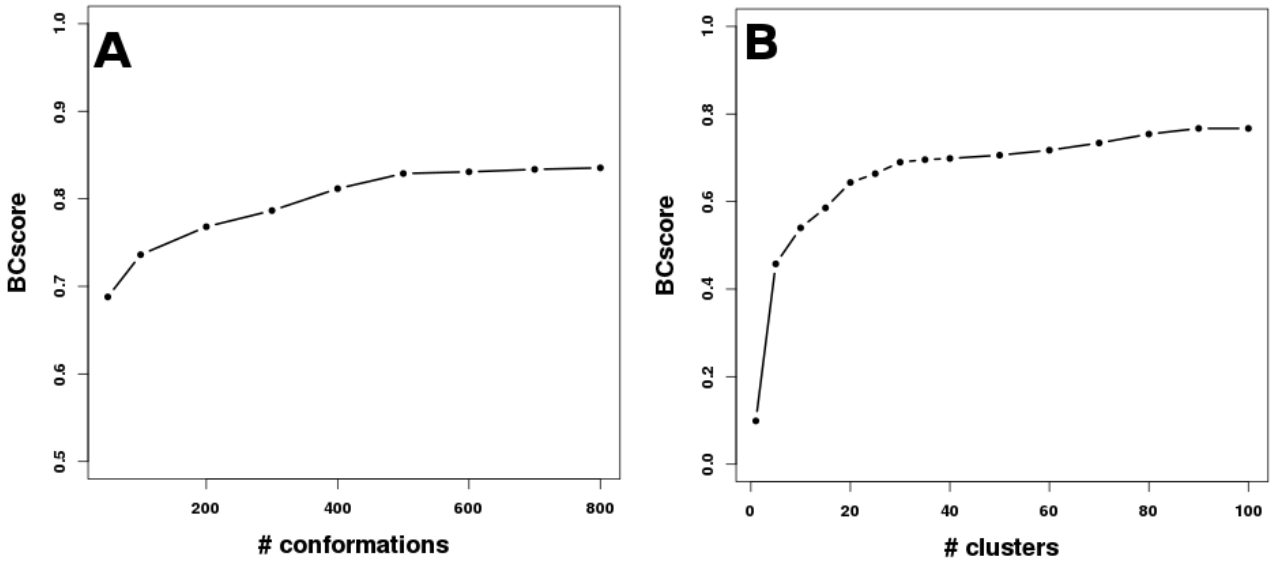


Peptide-protein interactions : an increasing panel of on-line tools



Accessing peptide conformation in complexes ?

Can suboptimal conformations of peptides in isolation
approximate conformation of peptides in interactionwith proteins ?



Peptidb collection *(London & al., Structure, 2010)* (100 peptides)

Lamiable et al., Methods Mol. Biol.,2017.

PEP-SiteFinder : a protocol for binding site identificati

De novo modeling from sequence

generation of a
library of peptide conformations

PepFold, 200 to 400 models

Clustering

Select representatives of
20 to 40 lowest energy clusters

Blind rigid docking

Iterate
for each peptide conformation

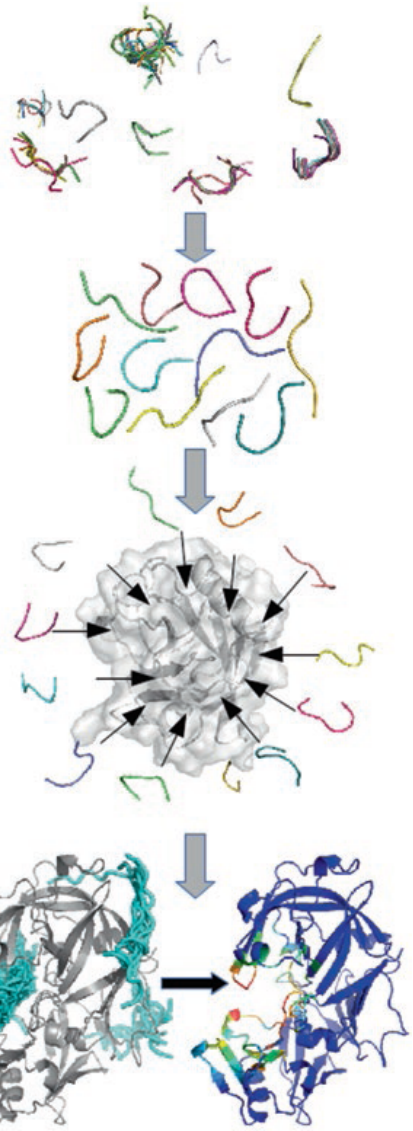
Systematically explore
protein surface

Ptools / ATTRACT2

Result compilation

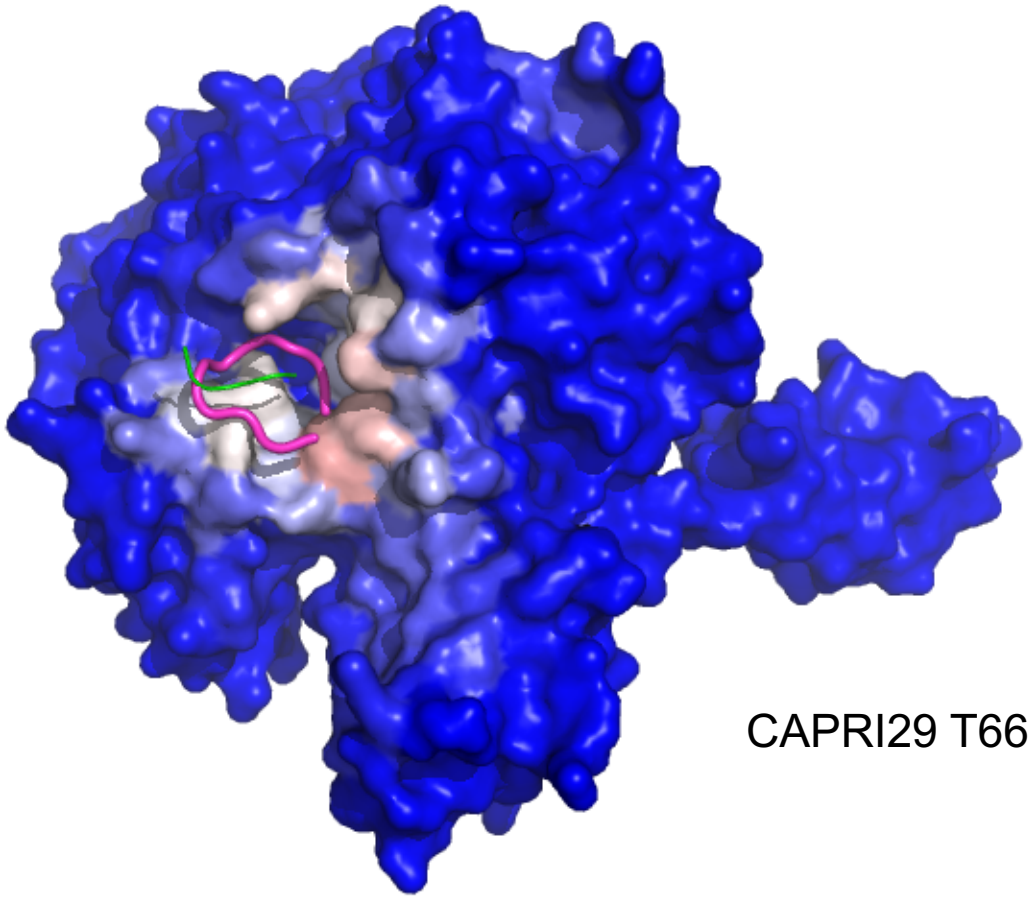
Merge results of all dockings

Calculate protein residue
propensities to bind peptide



Saladin et al., Nucleic Acids Res.,2014.

Peptide sub-optimal conformations to search for interaction sites ?

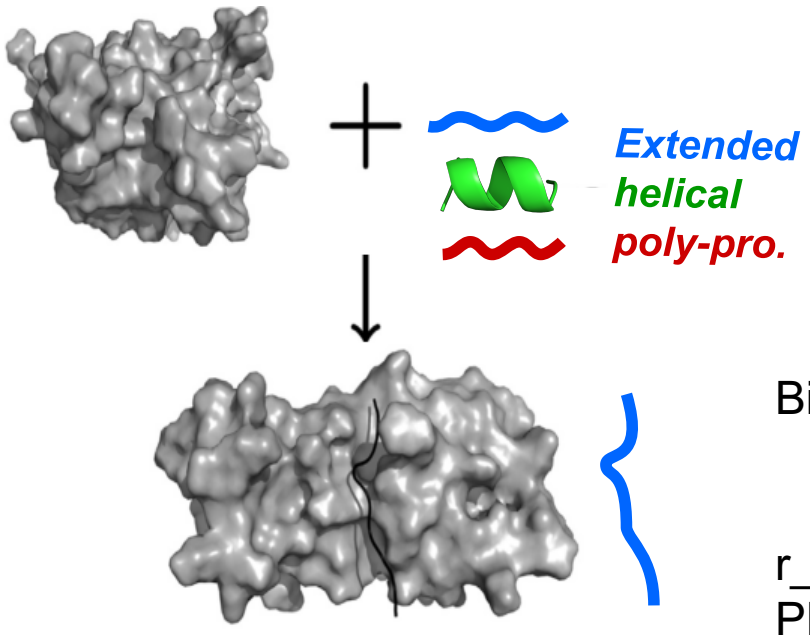


CAPRI29 T66

PriA Helicase Bound to SSB C-terminal Tail Peptide (PDB code: 4NL8)

http://bioserv.rpbs.univ-paris-diderot.fr/PEP-SiteFinder/

PepATTRACT : blind rigid docking step



Docking performance
(50 best models)

iRMSd < 2 : 34 %

Binding site identification performance

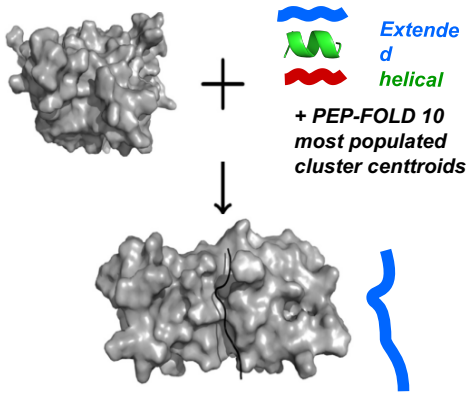
	Sens.	Spec.
r_pepATTRACT	37.2	37.2
PEP-SiteFinder	27.3	27.3
PepSite	13.4	26.6

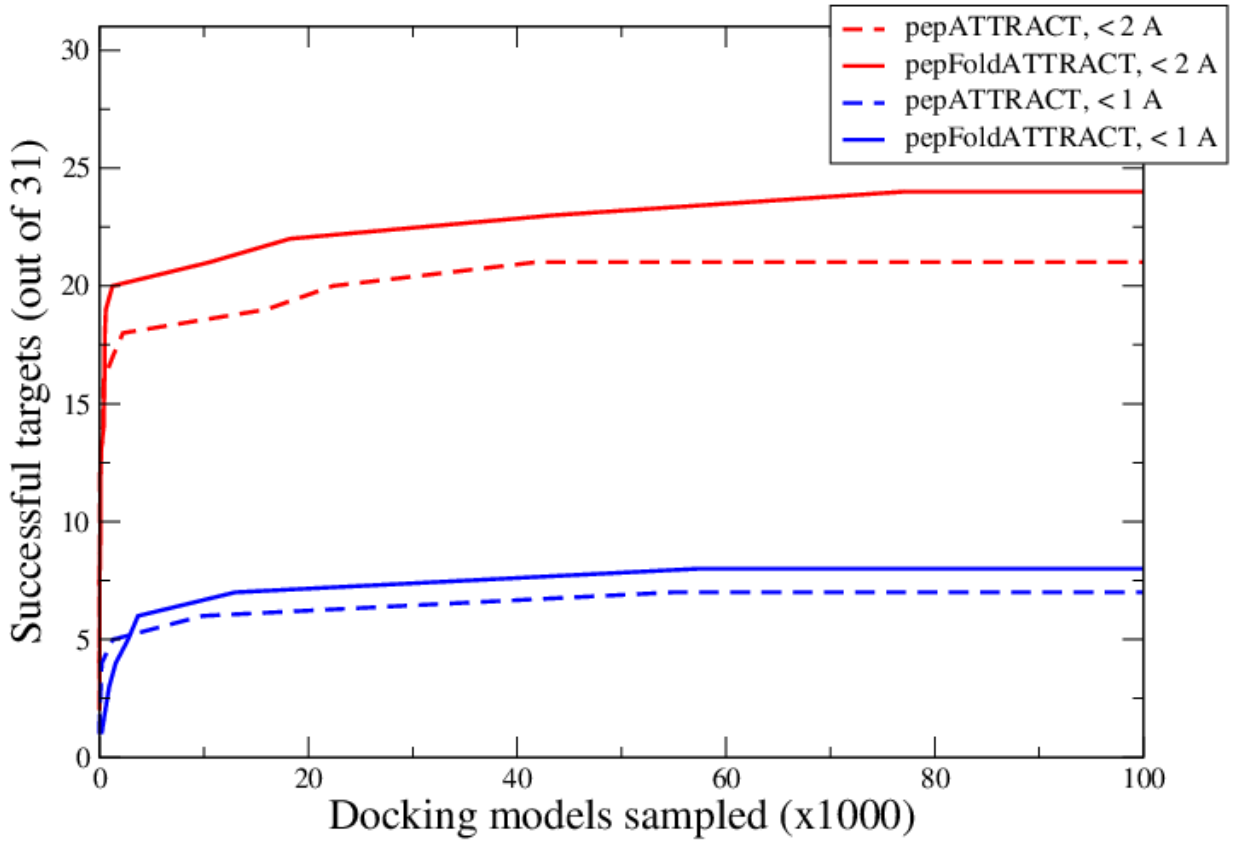
http://bioserv.rpbs.univ-paris-diderot.fr/services/pepATTRACT/

deVries et al., Nucleic Acids Res., 2017

Direct blind search for peptide-protein complex conformat

Combining PEP-FOLD and pepATTRACT

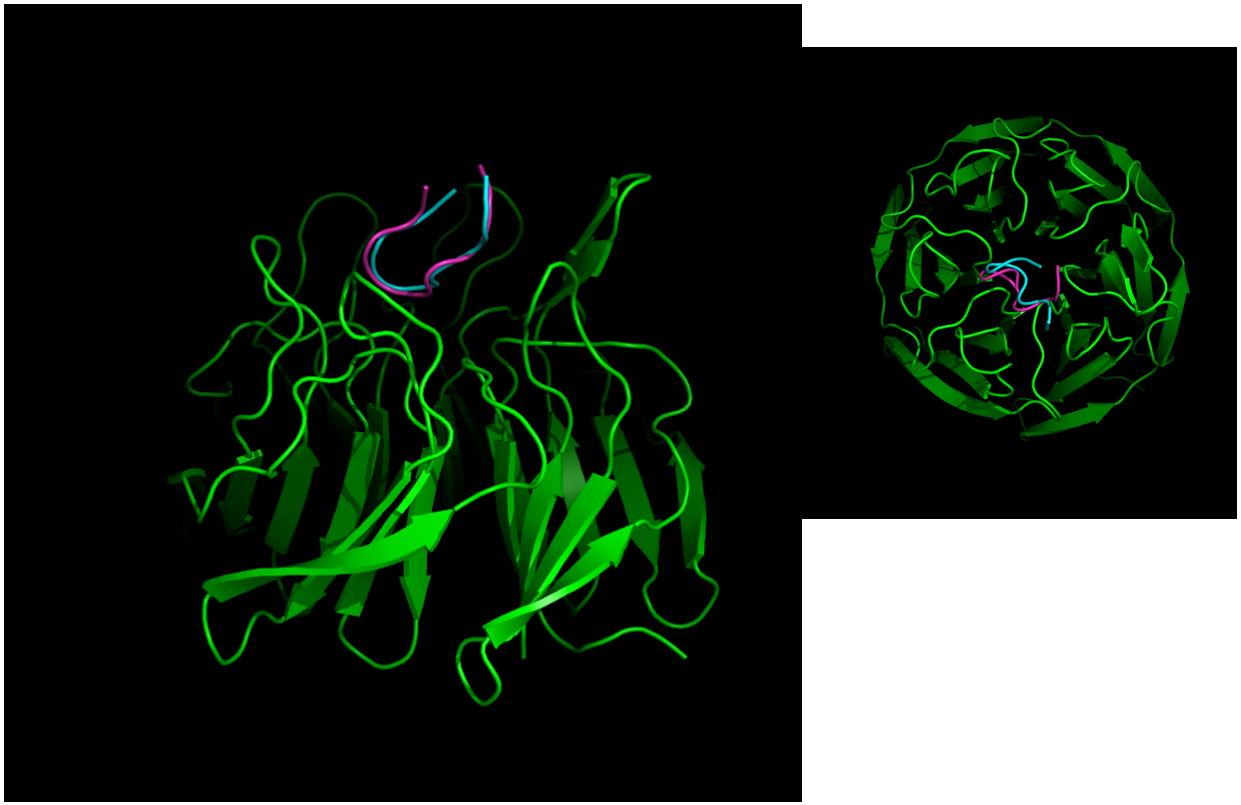




Peptide-protein complex is better sampled

Direct blind search for peptide-protein complex conformat

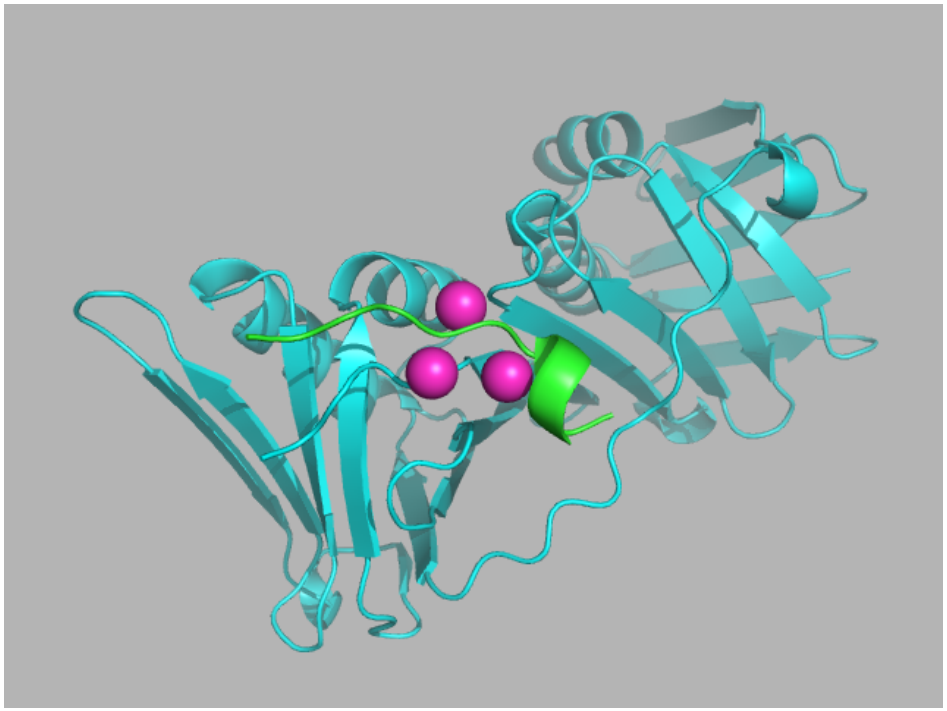
PEP-FOLD-ATTRACT for failed pepATTRACT

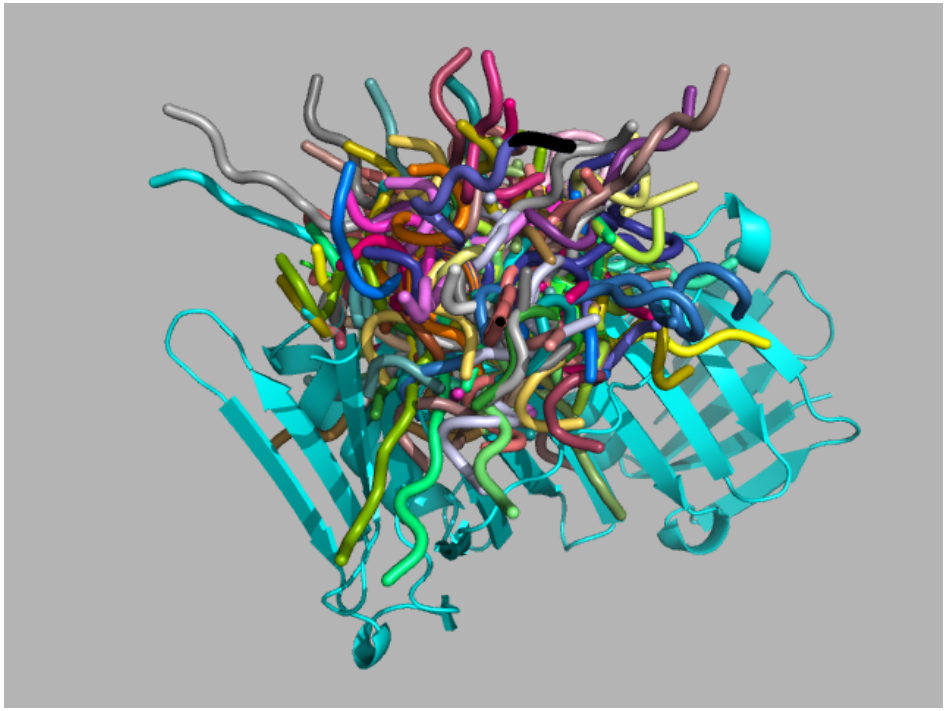


KELCH-LIKE ECH-ASSOCIATED PROTEIN 1 / 9-mer NRF2
PDB : 1X2J (unbound) 1X2R (bound)

The first correct (iRMSD < 2) structure is at rank #6, with iRMSD = 1.1 A

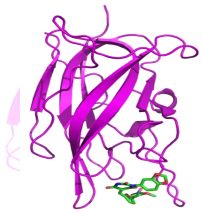
Local docking by folding peptide at protein vicinity ?

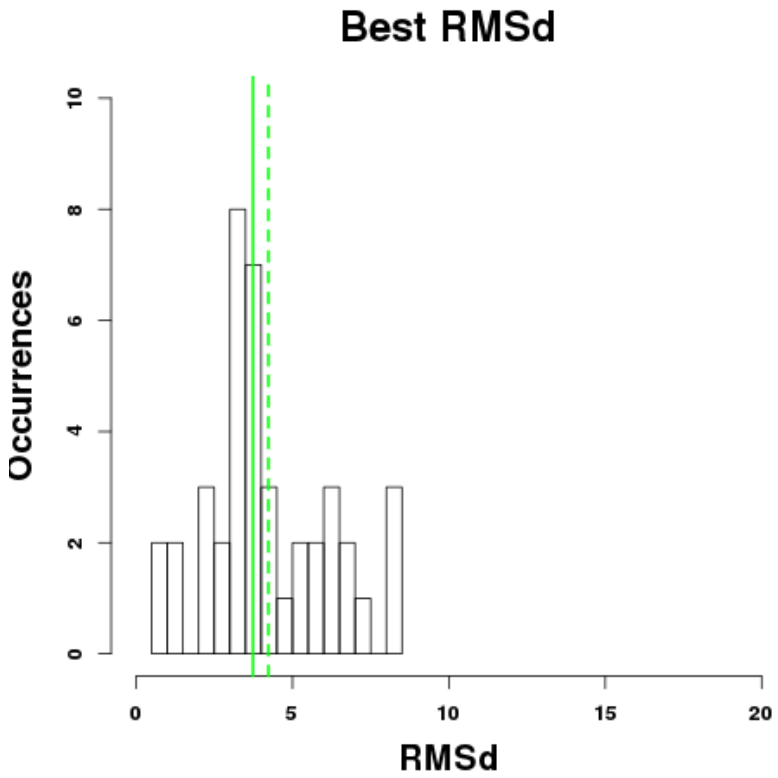


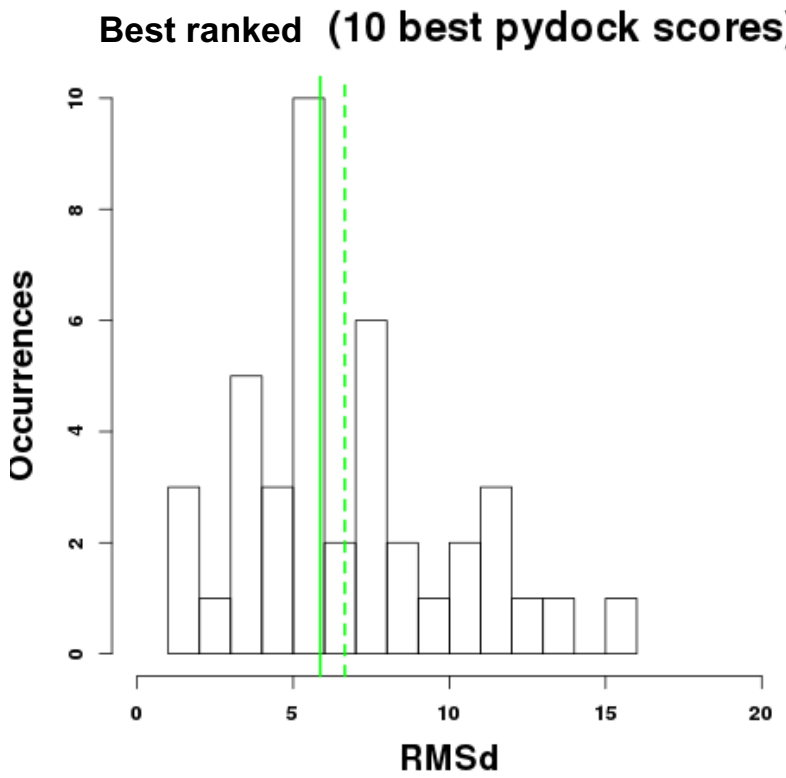


Lamiable et al., Nucleic Acids Res.,2016.

Local docking by folding peptide at protein vicinity ?





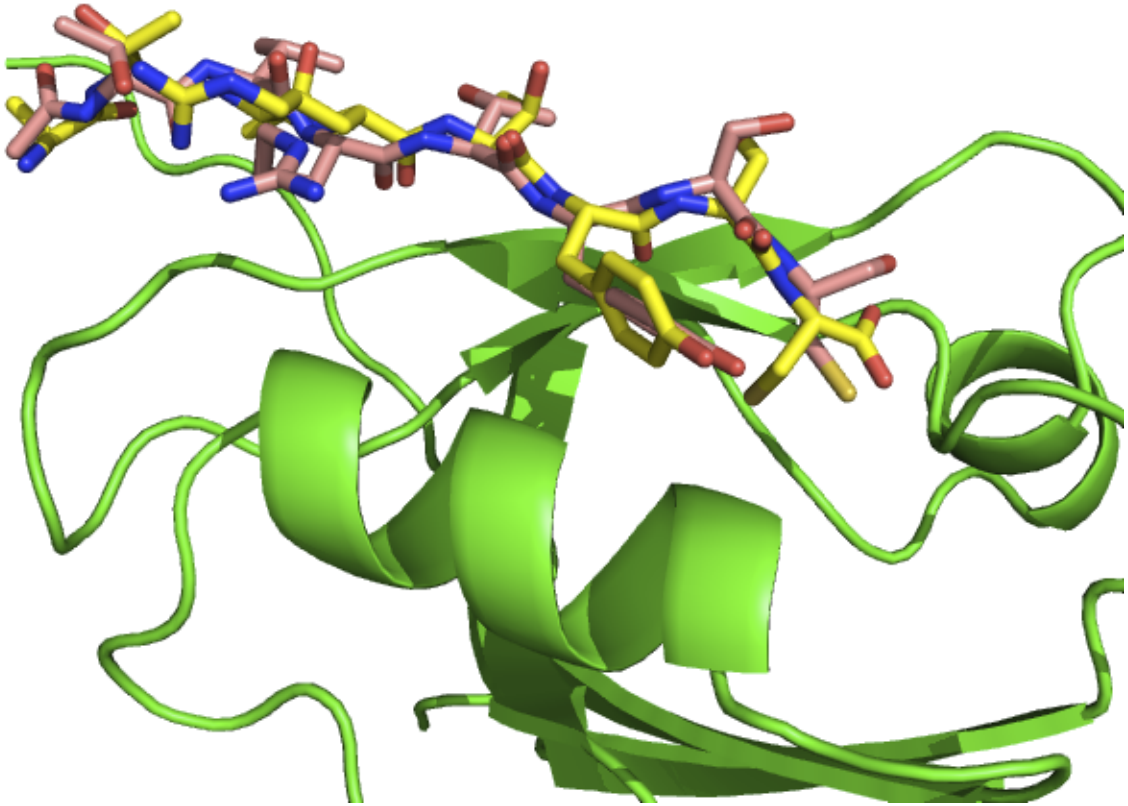


Dashed: mean
Plain: median

PeptiDB 41 complexes, APO protein conformation

PEP-FOLD3: protein-peptide interactions

http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD3/

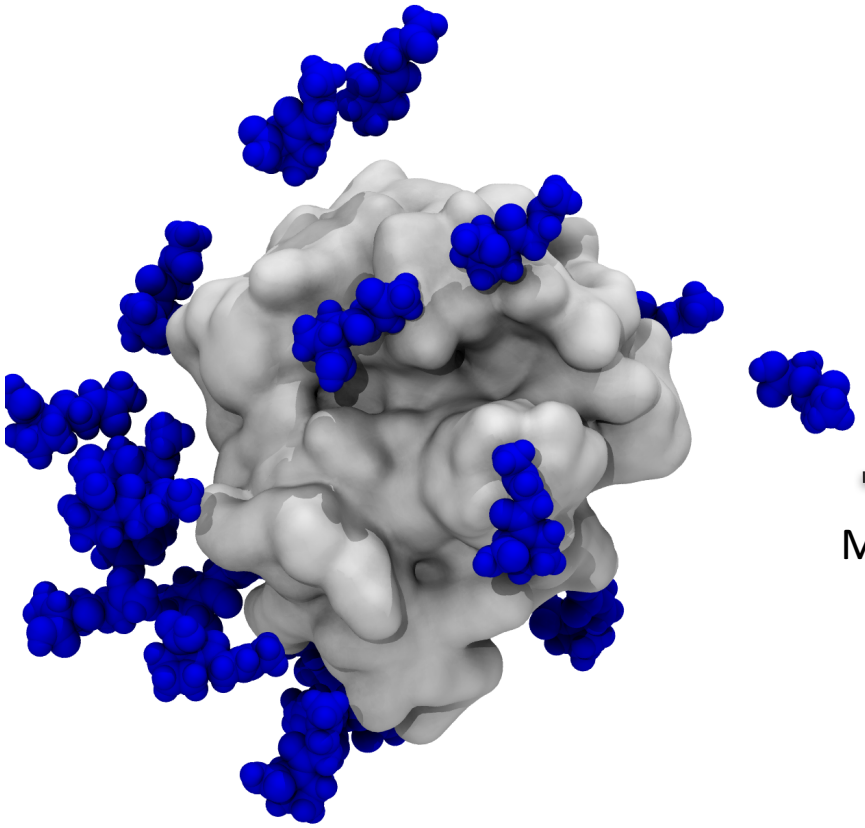


Generation ~ OK

Improve scoring
At the coarse grained level

Sampling issue ?

Hot spots of interaction

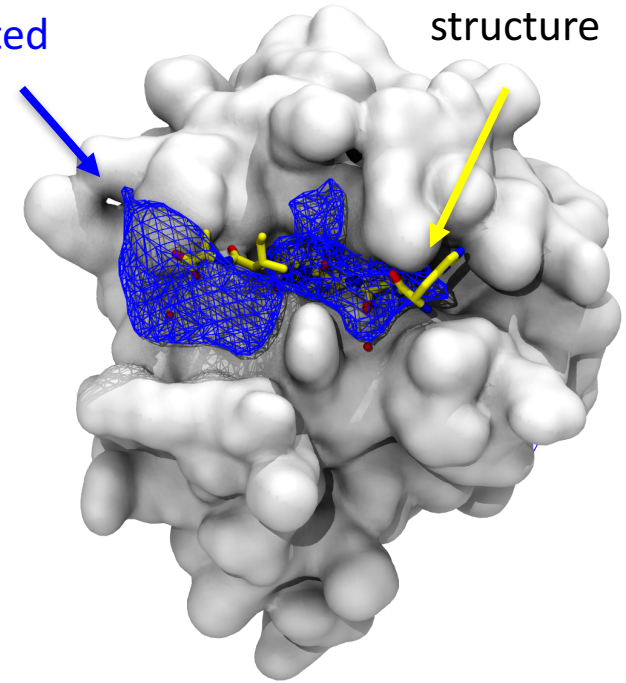


CYPA + 20 * GP

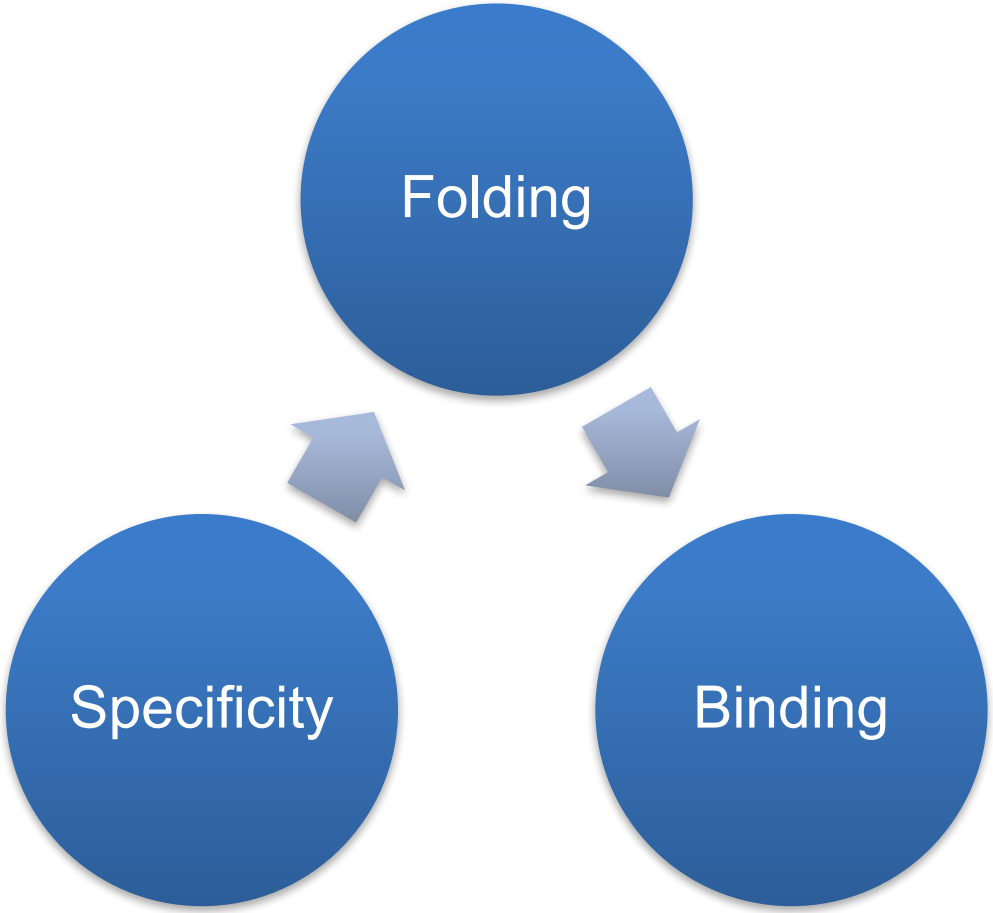
MD simulation
~300 ns

computed
density

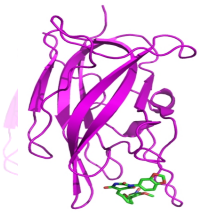
crystal
structure



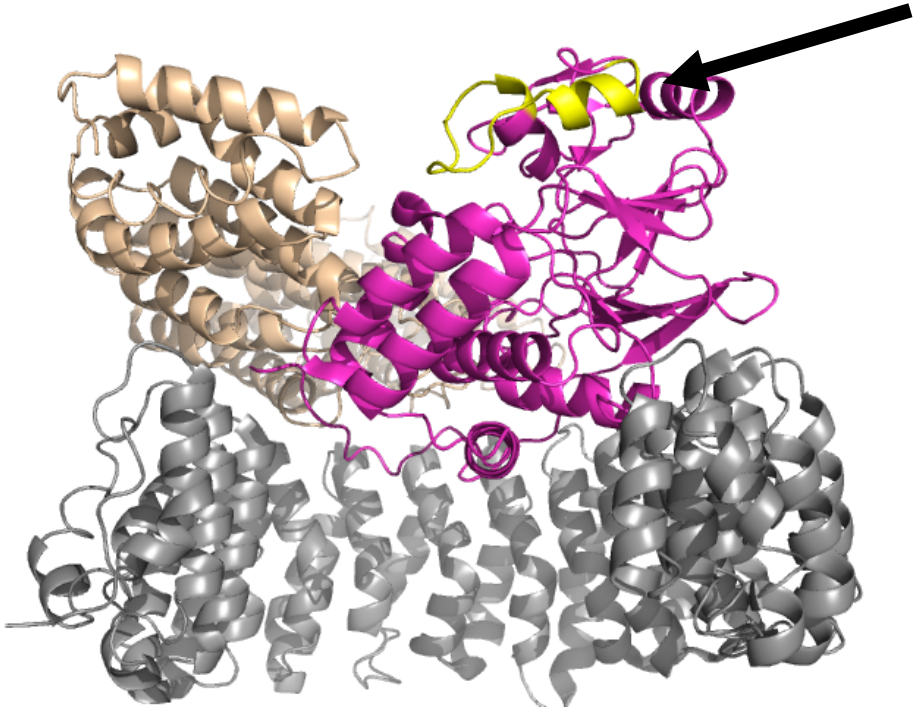
Towards *rational* peptide design

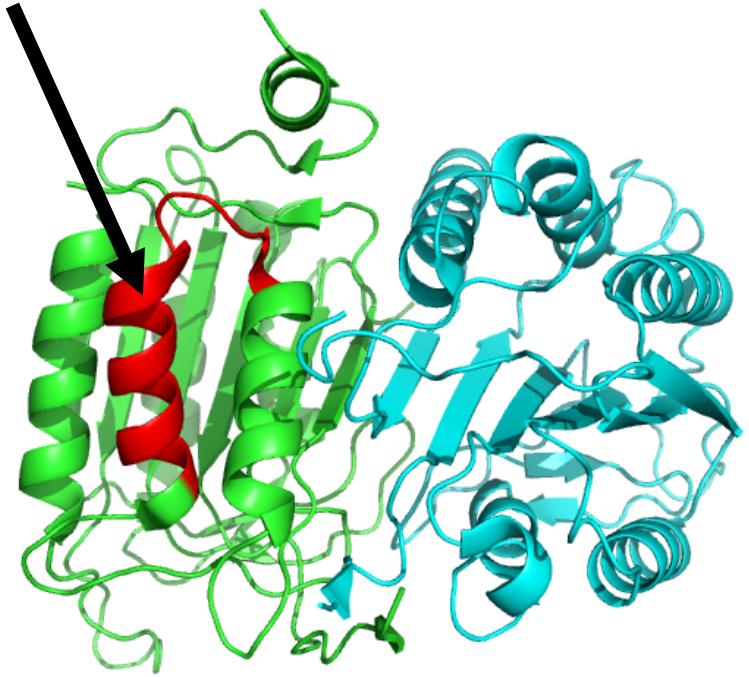


Example of CASPASE9-PP2A interaction



Interfering peptides identified using pepSCAN
(thanks to A. Rebollo)



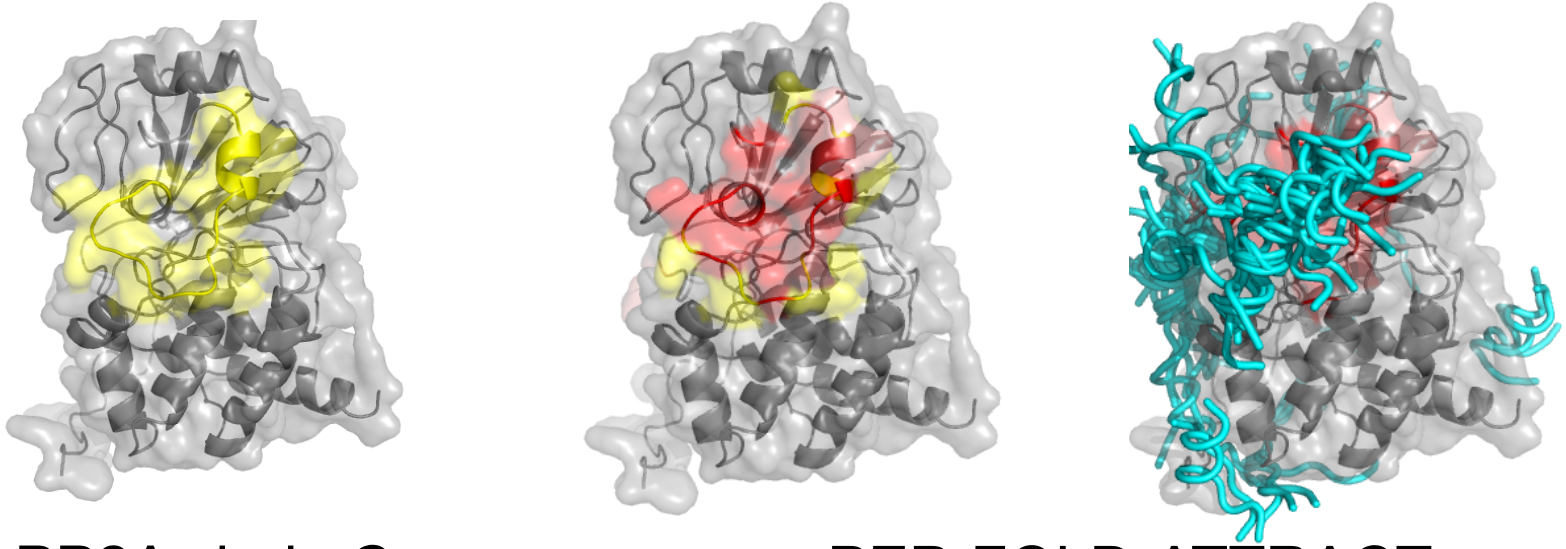


PP2A
PDB : 2IAE (ABC)

CASPASE9
PDB : 1JXQ (AB)

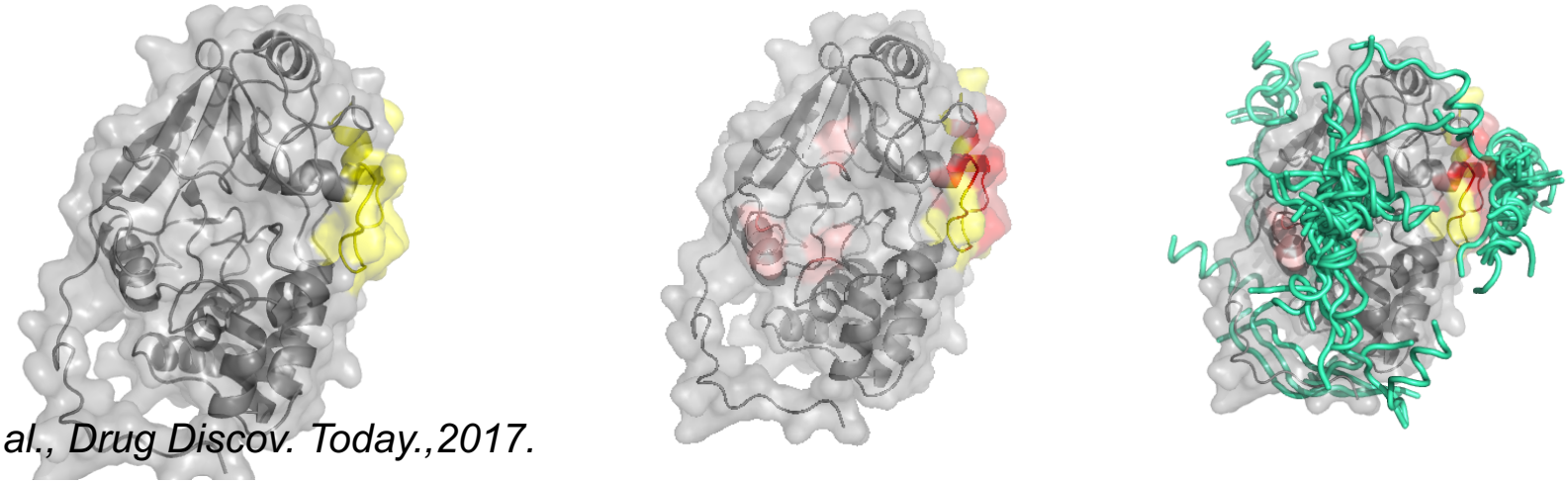
Example of CASPASE9-PP2A interaction PEP-FOLD-ATTRACT in agreement with pepSCAN ?

Q1 : CASPASE9 peptide to bind identified patch of PP2A ?



PP2A chain C

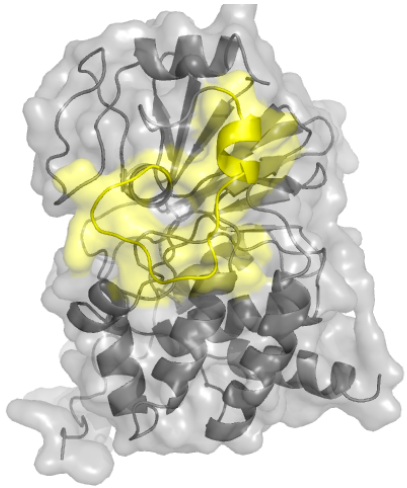
PEP-FOLD-ATTRACT



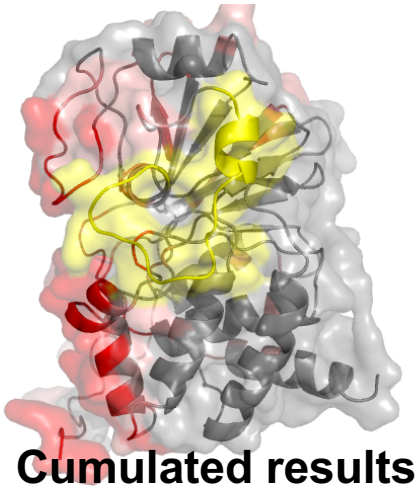
Bruzzoni et al., Drug Discov. Today.,2017.

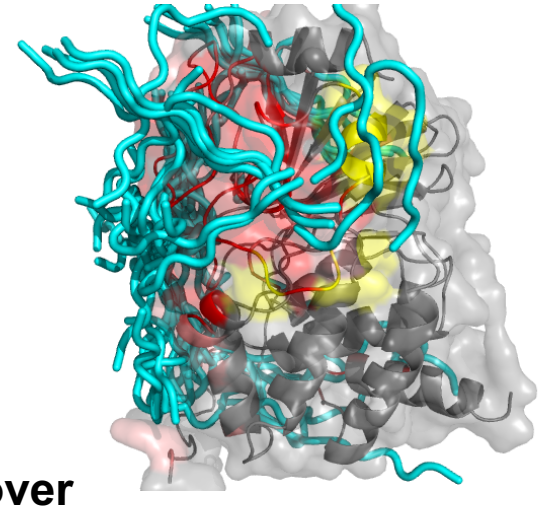
Example of CASPASE9-PP2A interaction PEP-FOLD-ATTRACT in agreement with pepSCAN ?

Q2 : CASPASE9 peptide specifically binds identified patch of PP2A ?

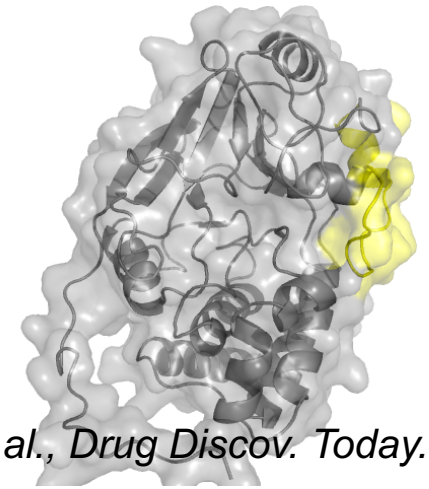


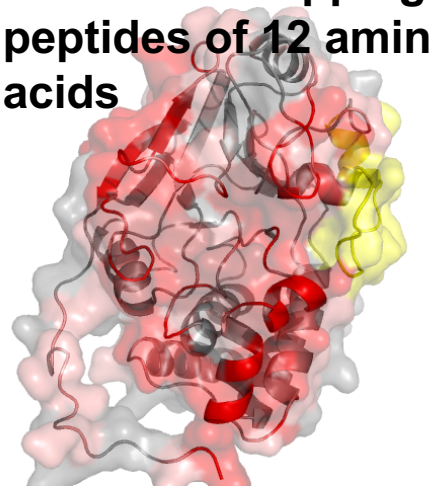
PP2A chain C

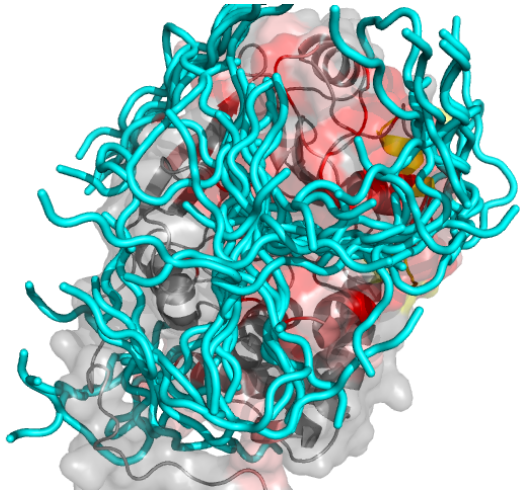




Cumulated results over 18 non-overlapping peptides of 12 amino acids

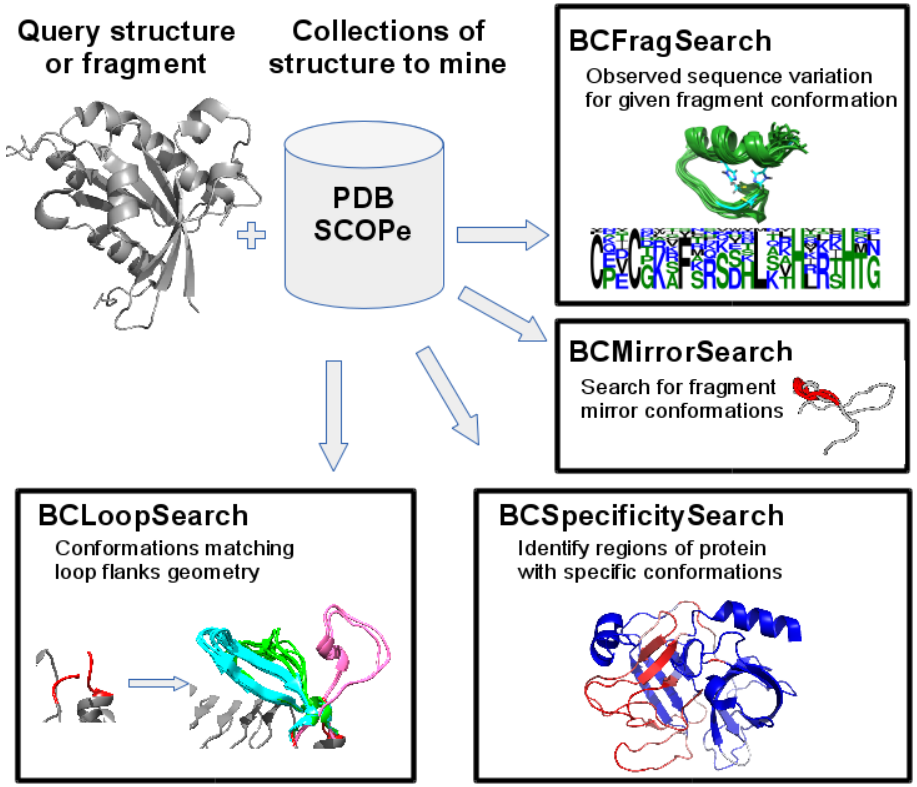






Bruzzoni et al., Drug Discov. Today.,2017.

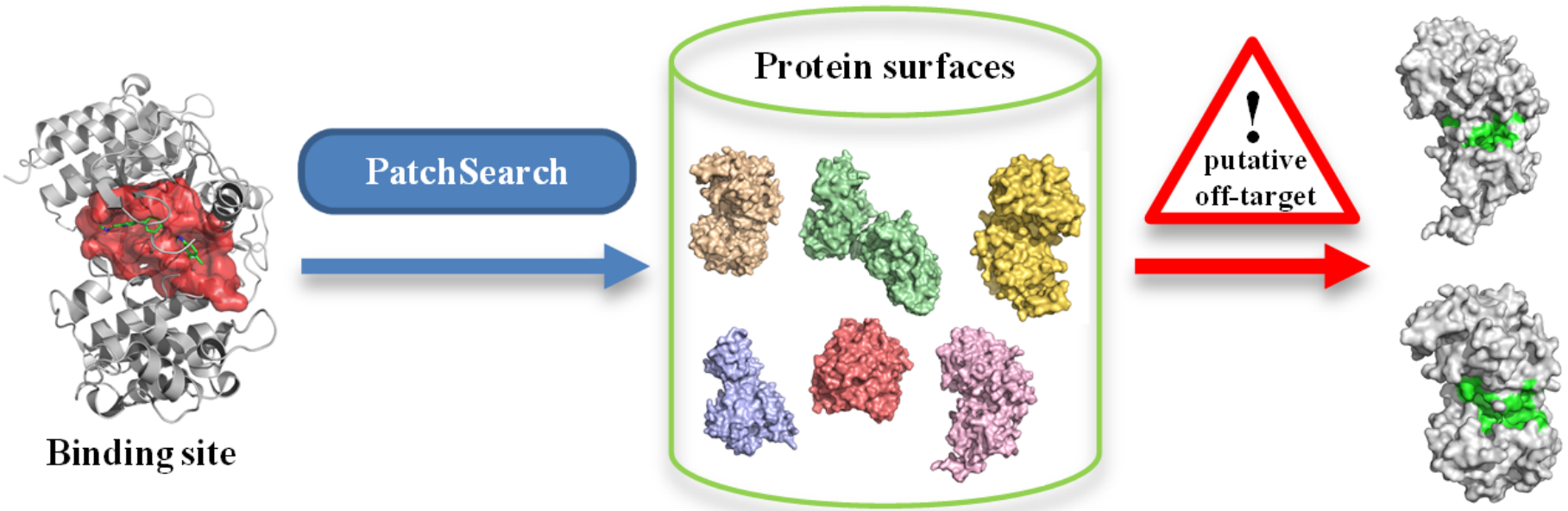
Mining protein structures to get information about candidate peptides.



bioserv.rpbs.univ-paris-diderot.fr/services/BCSearch

Guyon et al., Bioinformatics 2014,
Guyon et al., Nucleic Acids Res., 2015.

PatchSearch: A Fast Computational Method for Off-Target Detection.



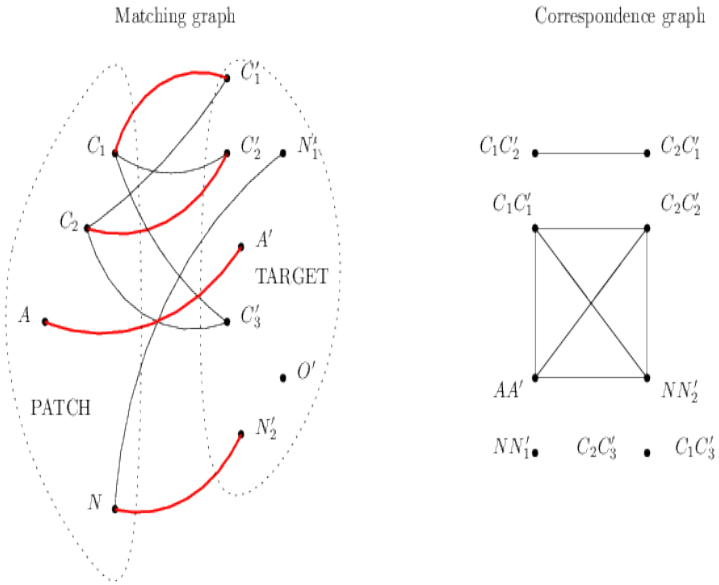
Rasolohery I. Moroy G. and Guyon F.
J. Chem. Inf. Model. 2017

PatchSearch: A Fast Computational Method for Off-Target Detection.

Method:
Similarities searches based on conservation of internal distances between atoms a binding site.

1) Stringent clique searching in correspondence graph
→ rigid core of binding sites

2) Enlargement of the cliques to construct quasi-cliques in correspondence graph
→ flexible parts of binding sites.



Construction of a correspondence graph.

Correspondence or matching between atoms C_1C_1' is linked to correspondence C_2C_2' as distance between C_1 and C_2 is equivalent to distance between C_1' and C_2'.

PatchSearch: A Fast Computational Method for Off-Target Detection.

Comparaison with other approaches (AUC):

methods	Kahraman data set		Homogeneous data set	
	shape	shape + size	shape	shape + size
PatchSearch: Clique	0.52	0.60	0.67	0.74
PatchSearch: Quasi-Clique	0.78	0.82	0.74	0.77
Spherical Harmonics (a)	0.64	0.77	NA	NA
sup-CK (a)	0.86	0.89	0.71	0.72
MultiBind (a)	0.71	NA	0.69	NA
PSIM (b)	0.79	NA	0.76	NA
PatchSurfer (c)	0.81	0.84	NA	NA

Rasolohery I. Moroy G. and Guyon F.
J. Chem. Inf. Model. 2017

« Some » directions

Clustering vs partitioning

Force field optimization vs optimality

Quick binding affinity estimates

Thanks to

MTi
INSERM UMR-S 973

J. Maupetit
Y. Shen
J. Rey
F. Guyon
A. Saladin
G. Moroy
P. Thévenet
A. Lamiable
M. Vavrusa
S. deVries

U. Montreal, Canada

N. Mousseau
S. Coté

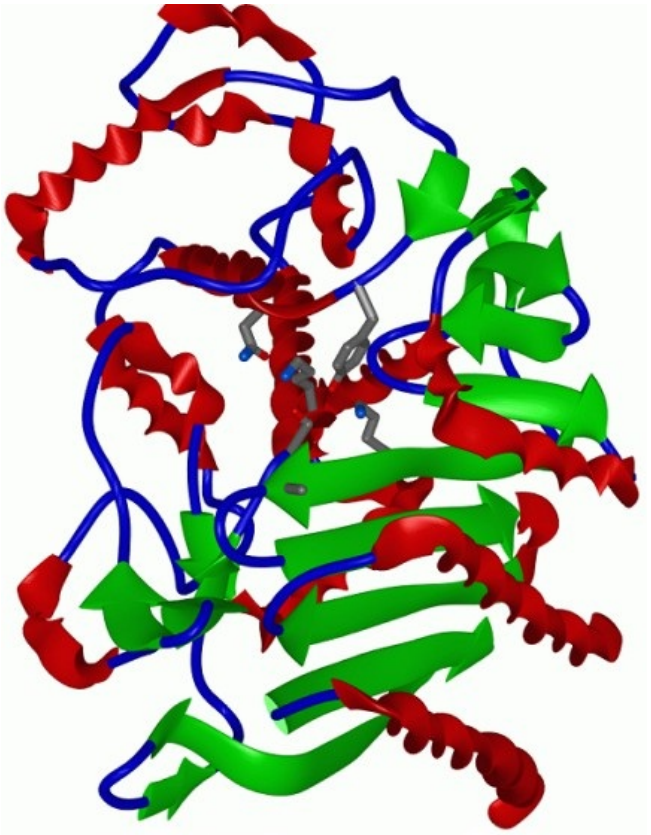
IBPC
CNRS UPR 9080

Ph. Derreumaux

TUM Munchen, Germany

M. Zacharias

http://bioserv.rpbs.univ-paris-diderot.fr



Ceci n'est pas une protéine

From
http://shoichetlab.compbio.ucsf.edu/