



Novel computational tools for integrative structural biology

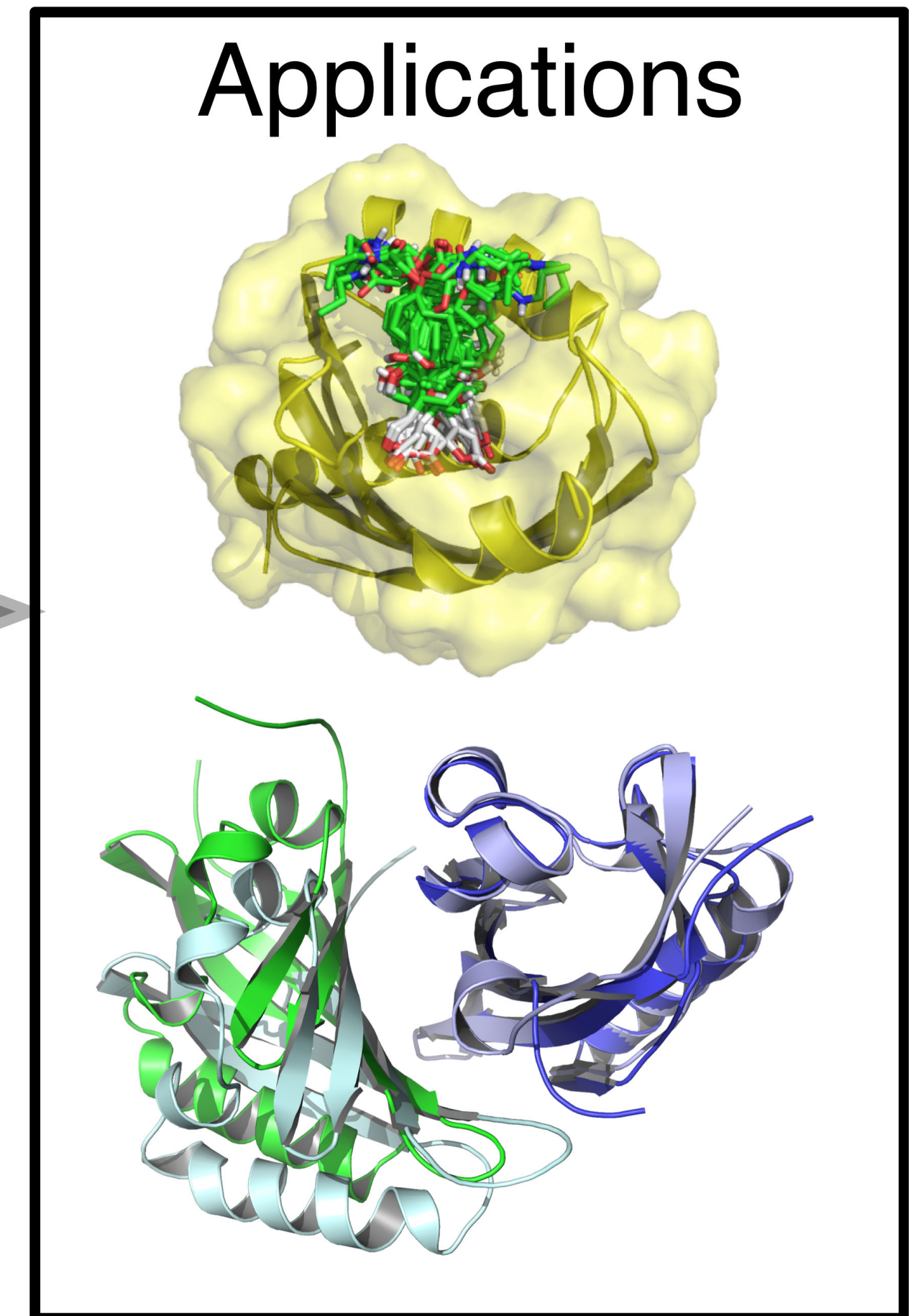
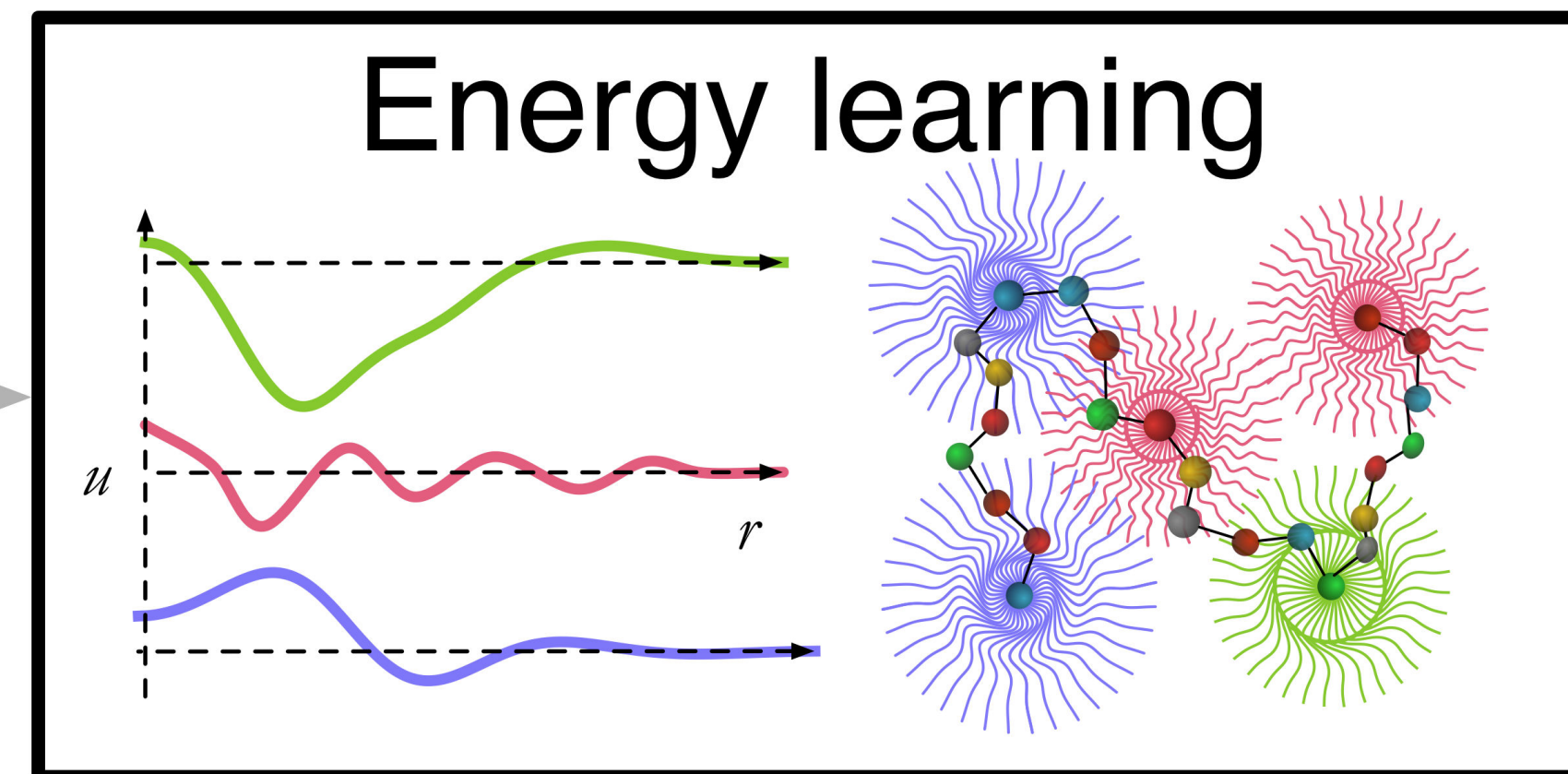
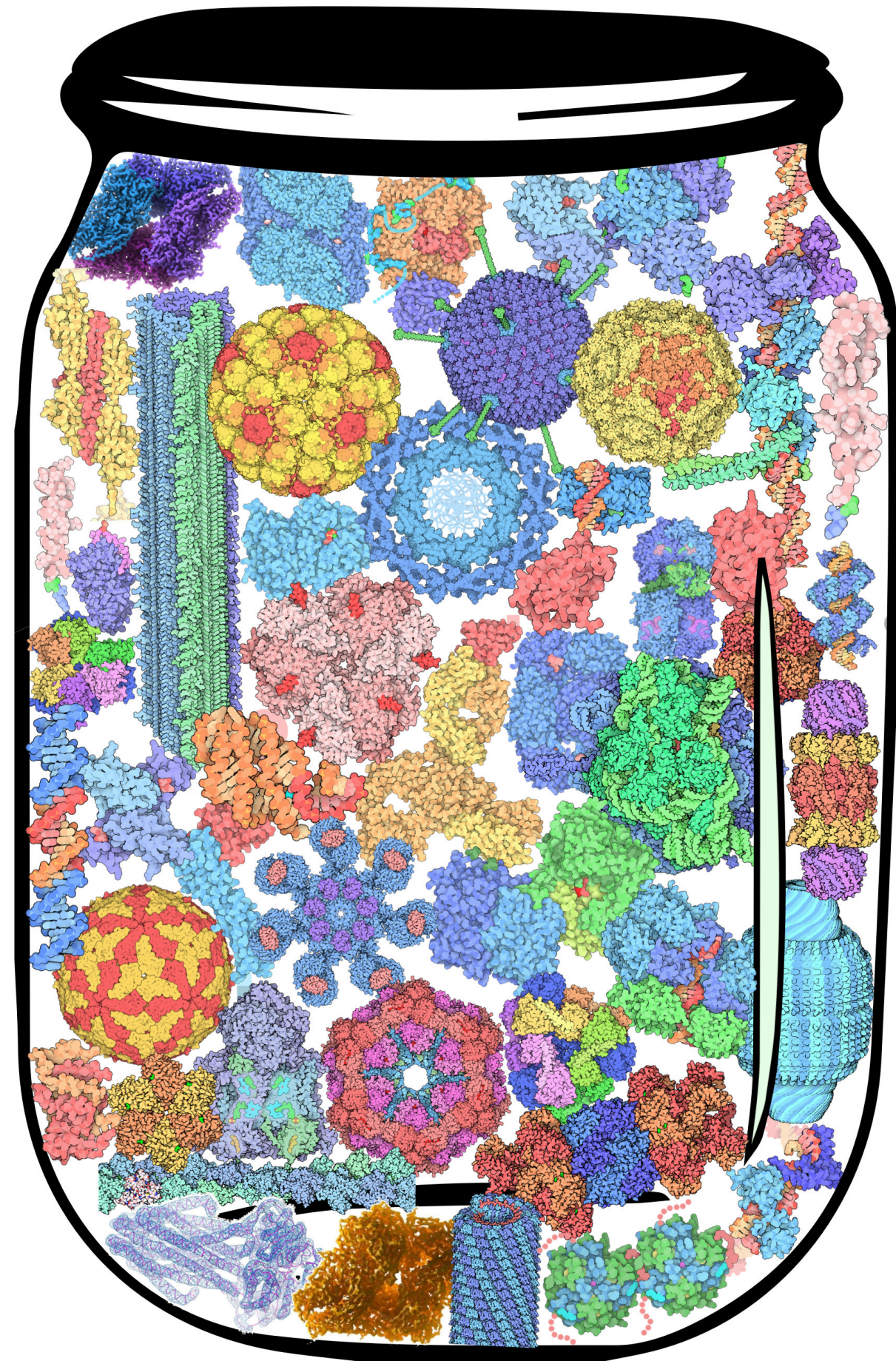
team.inria.fr/nano-d/

Sergei Grudinin, G. Pagès, M. Kadukova, M. Karasikov, D. Devaurs

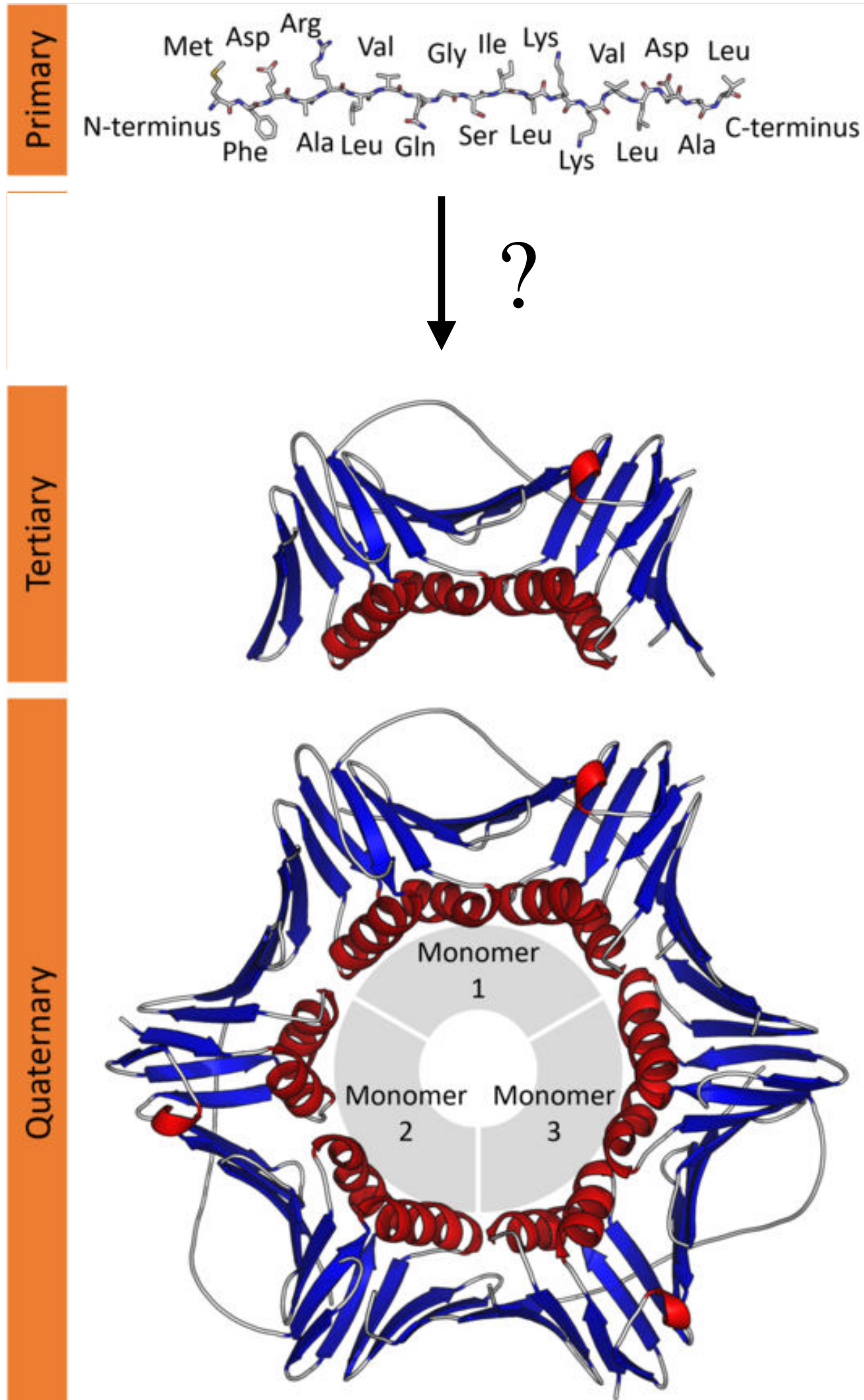
Nano-D – Inria – Grenoble

sergei.grudinin@inria.fr

Outline



CASP community challenge



All targets

[Regular](#) [Heteromers](#) [Refinement](#)
[All groups](#) | [Server](#)

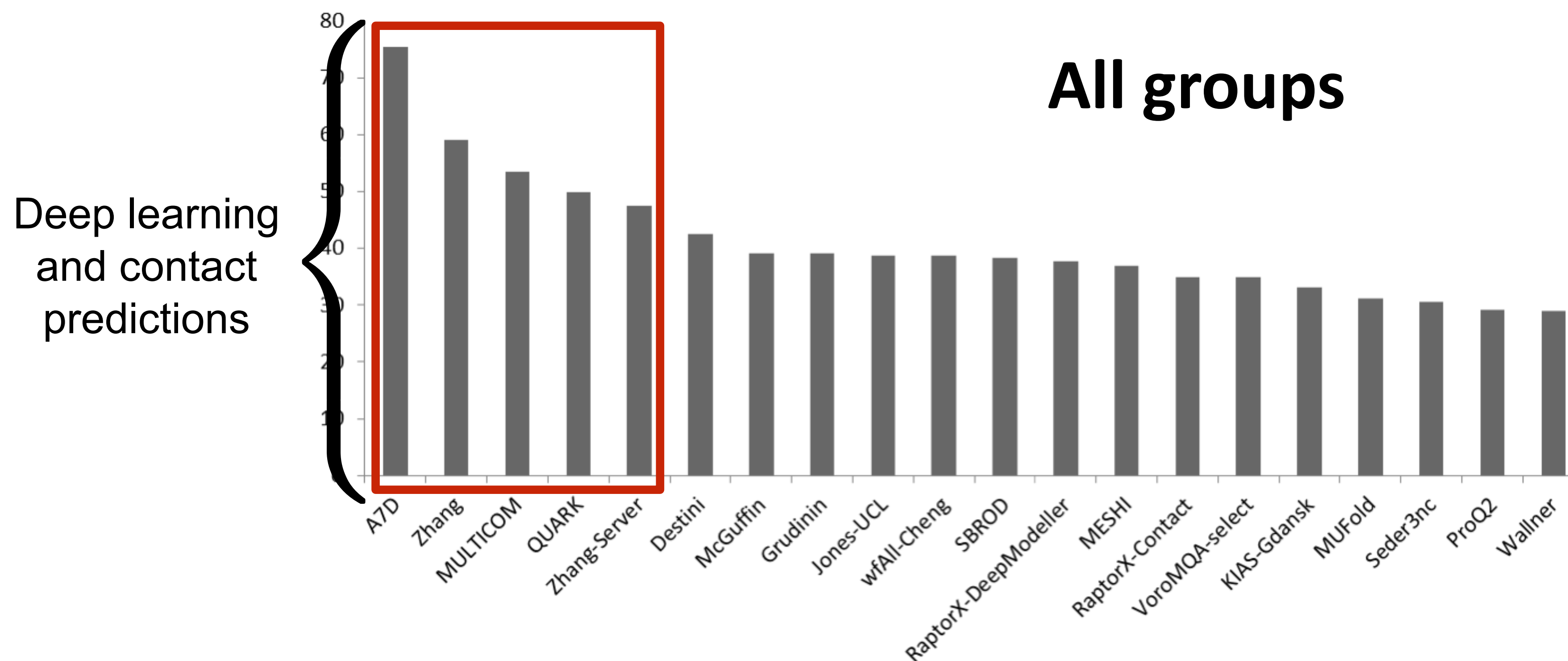
[Assisted structure prediction](#)
[SAXS](#) | [X-link](#) | [NMR](#) | [SANS](#) | [FRET](#)

#	Tar-id	Type	Res	Stoi- chiom.	Entry Date	Server Expiration	QA Expiration	Human Expiration	Description
1.	T0949	All groups	183	A1	2018-05-01	2018-05-04	m1: 2018-05-08 m2: 2018-05-10	2018-05-21	B7JAQ5_ACIF2
2.	T0950	Server only	353	An	2018-05-02	2018-05-05	m1: 2018-05-09 m2: 2018-05-11	2018-05-15 canceled on 2018-05-14	PaxB PDB code 6ek4
3.	T0951	Server only	276	A1	2018-05-03	2018-05-09	m1: 2018-05-13 m2: 2018-05-15	2018-05-15	ShHTL7 PDB code 5z82
4.	T0952	All groups	35	A2	2018-05-03	2018-05-06 canceled on 2018-05-04	m1: 2018-05-10 m2: 2018-05-12	2018-05-24 canceled on 2018-05-04	O48503 PDB code 6fxa
5.	H0953	All groups	321	A3B1	2018-05-04	2018-05-07	-	2018-05-25	Adhesin tip PDB code 6f45
6.	T0953s1	All groups	72	A1	2018-05-04	2018-05-10	m1: 2018-05-14 m2: 2018-05-16	2018-05-25	Adhesin tip PDB code 6f45
7.	T0953s2	All groups	249	A1	2018-05-04	2018-05-10	m1: 2018-05-14 m2: 2018-05-16	2018-05-25	Adhesin tip PDB code 6f45
8.	T0954	All groups	350	A1	2018-05-07	2018-05-13	m1: 2018-05-17 m2: 2018-05-19	2018-05-28	RFWD3_HUMAN PDB code 6cvz

193.	R0996-D7	Refinement	-	-	2018-08-10	2018-08-13	-	2018-08-25	
194.	n1008	Assisted	-	A1	2018-09-18	2018-09-21	-	2018-09-30	

<http://predictioncenter.org>

Predictions ranking in CASP13 (Dec 2018)

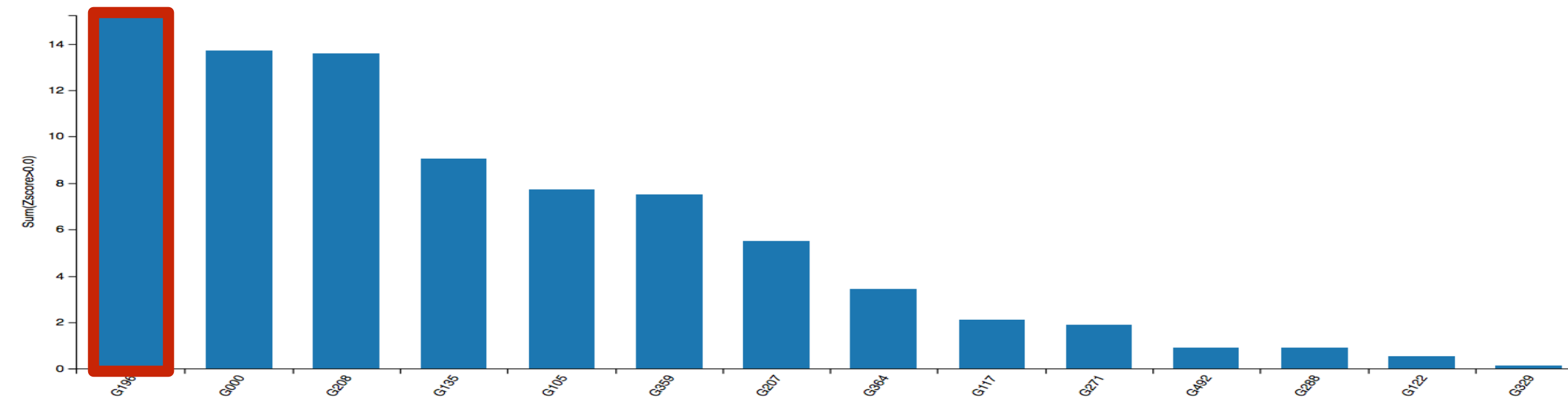


Ranking = **sum Z-scores** combined from **GDTTS & QCS** (as these are by far the two most informative scores to guide visual assessment) on all **models submitted as #1**, for **TBM/FM, FM and FM_sp** target EUs, and considering **sum of Z-score > -2**.

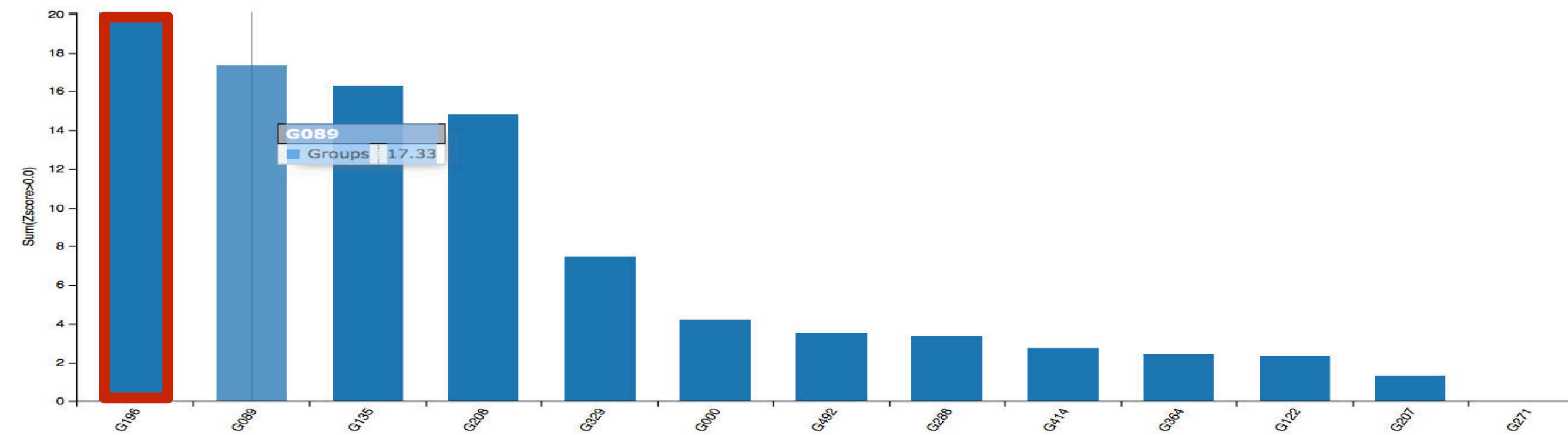
CASP13 free modeling ranking by Luciano A. Abriata and Matteo Dal Peraro, EPFL

CASP community challenge

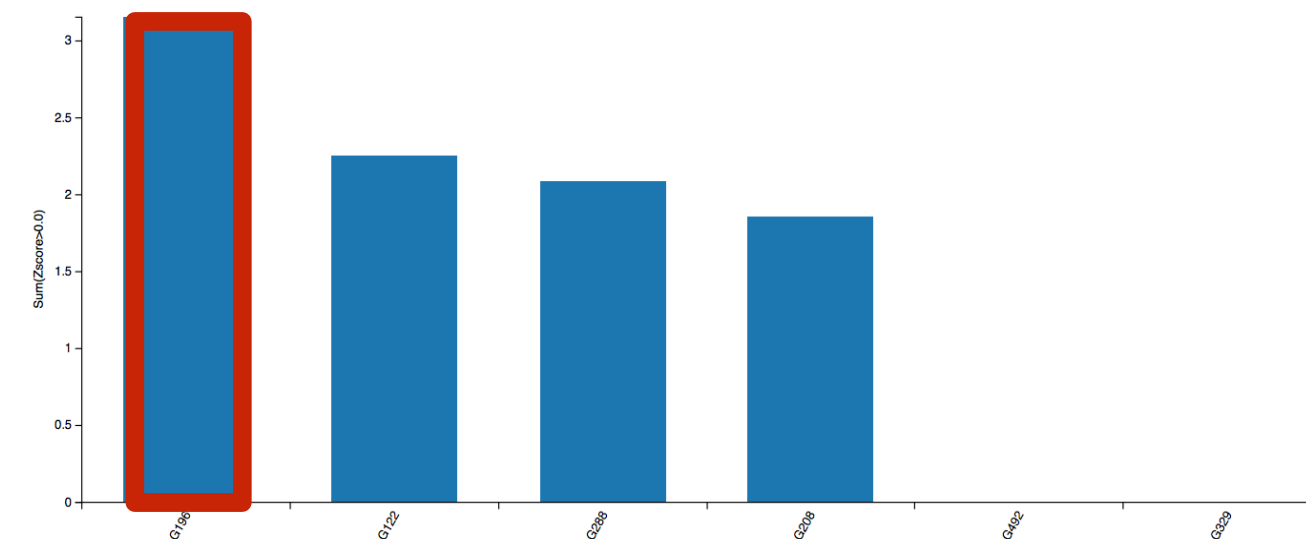
● Cross-link assisted modeling



● SAXS-assisted modeling

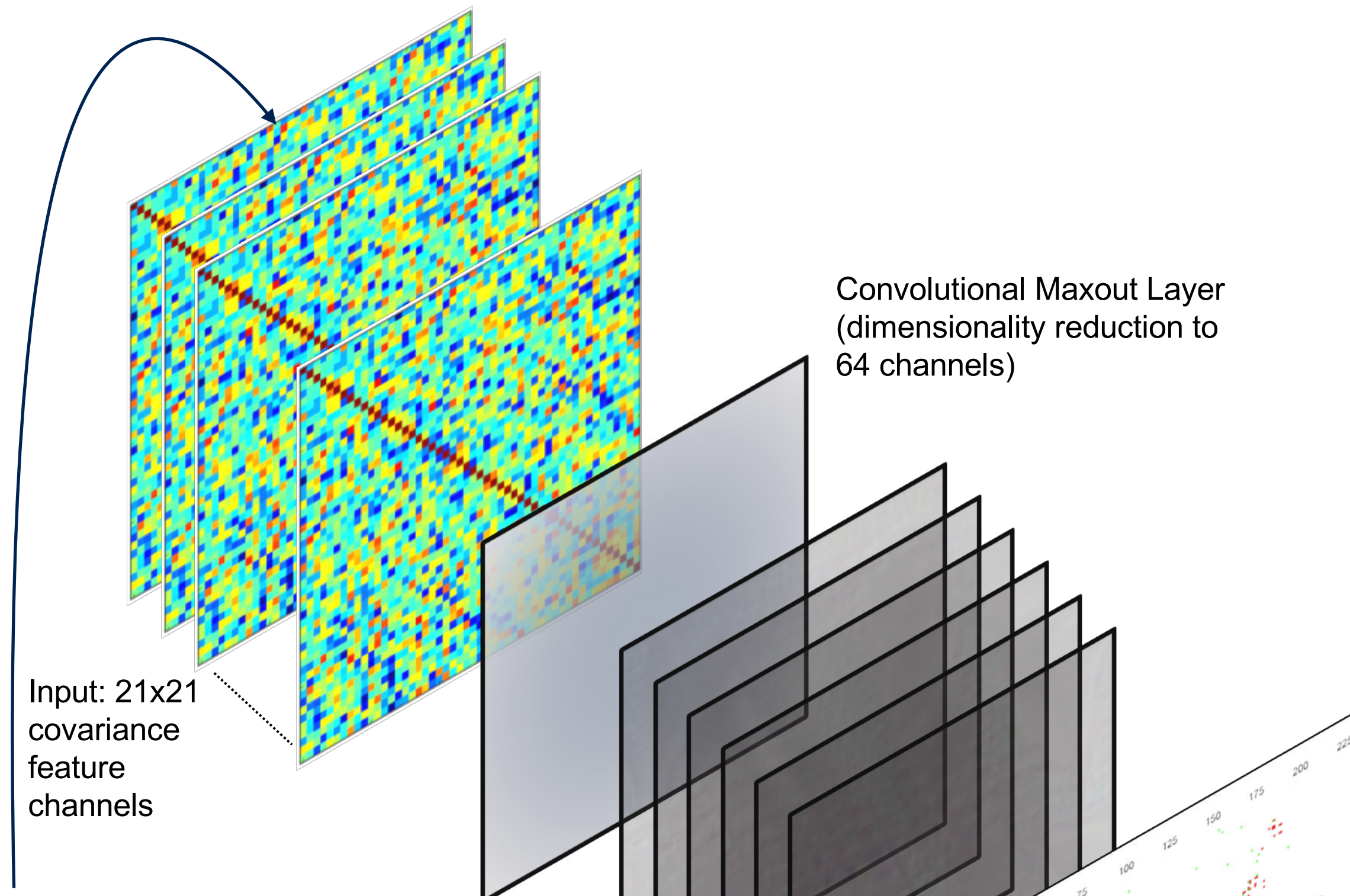
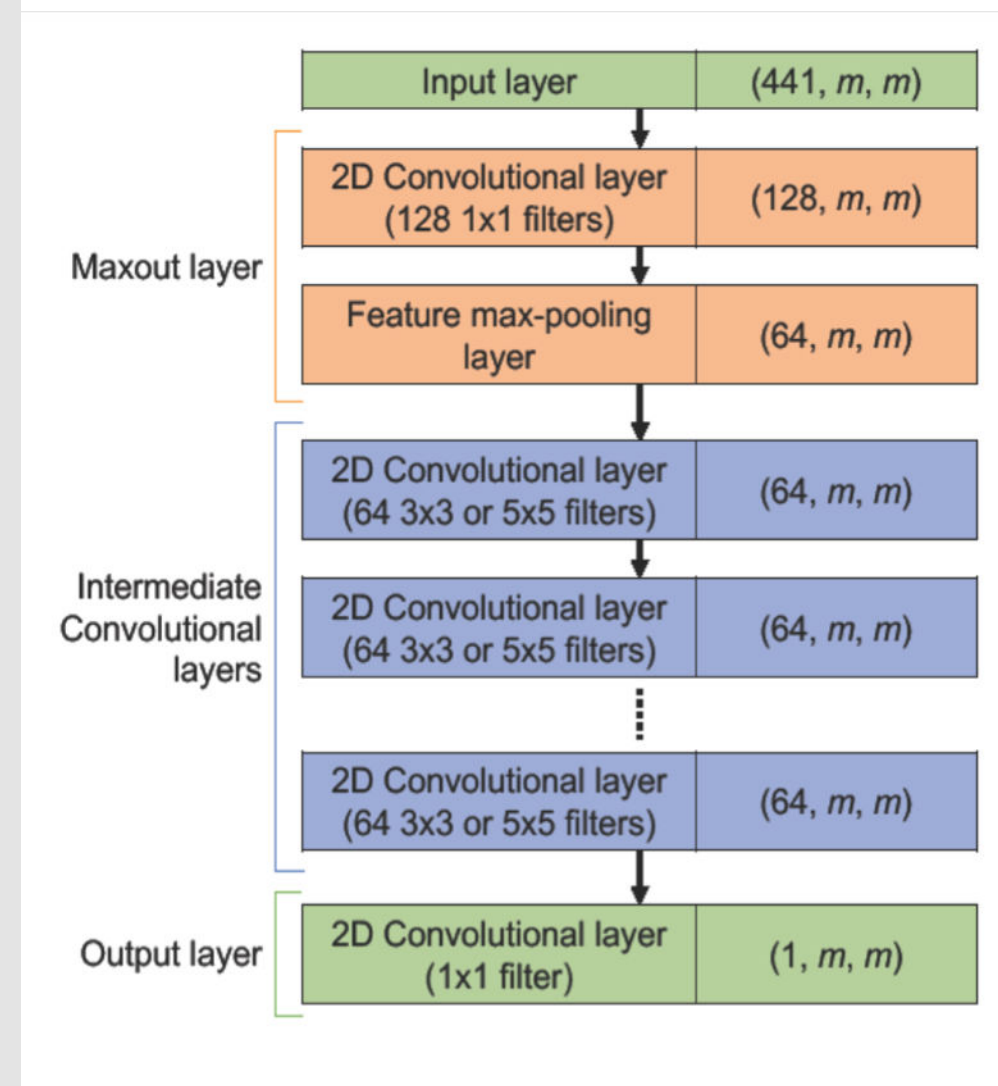


● SANS-assisted modeling



<http://predictioncenter.org>

DeepCOV: Analysing Residue Covariation using FCNs



Input: 21x21 covariance feature channels

Convolutional Maxout Layer (dimensionality reduction to 64 channels)

8-10 Padded Convolutional layers
5x5x64 Filters
ReLu Activation
BatchNorm

Output: Contact Probability Map

covariance between MSA columns

i	j
...EEANVVLGTVEEIMNVDPVHHTY...	...WRYLKGKDIVTHEILLDGG...
...EEANVVLGTVEEIMNVDPVHHTY...	...WRYLKGREIVTNDILLDGG...
...EEANVVLGTVEEIMNMDPVHNTY...	...WRYLKGKEVVTNDILVDGG...
...EEANVVLGTVEEILNVDPVQHTY...	...WRYLKGKDLVARESLLDGG...
...EEANIVLGTVEEILNMDPVHNTY...	...WRYLKGREVVTRDVLLDGG...

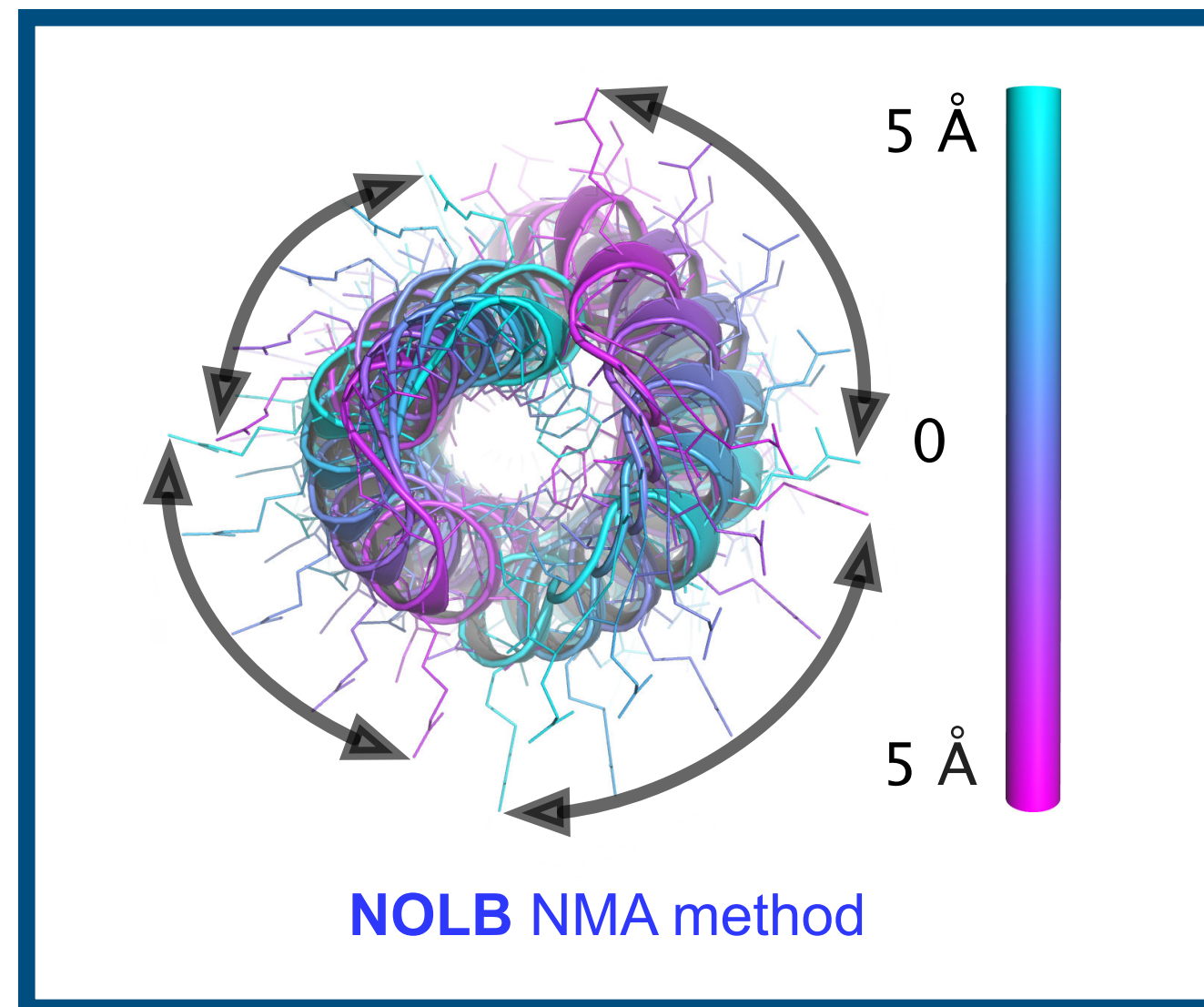
from Prof. David Jones, UCL Dept. of Computer Science & The Francis Crick Institute, <http://predictioncenter.org>

The current state

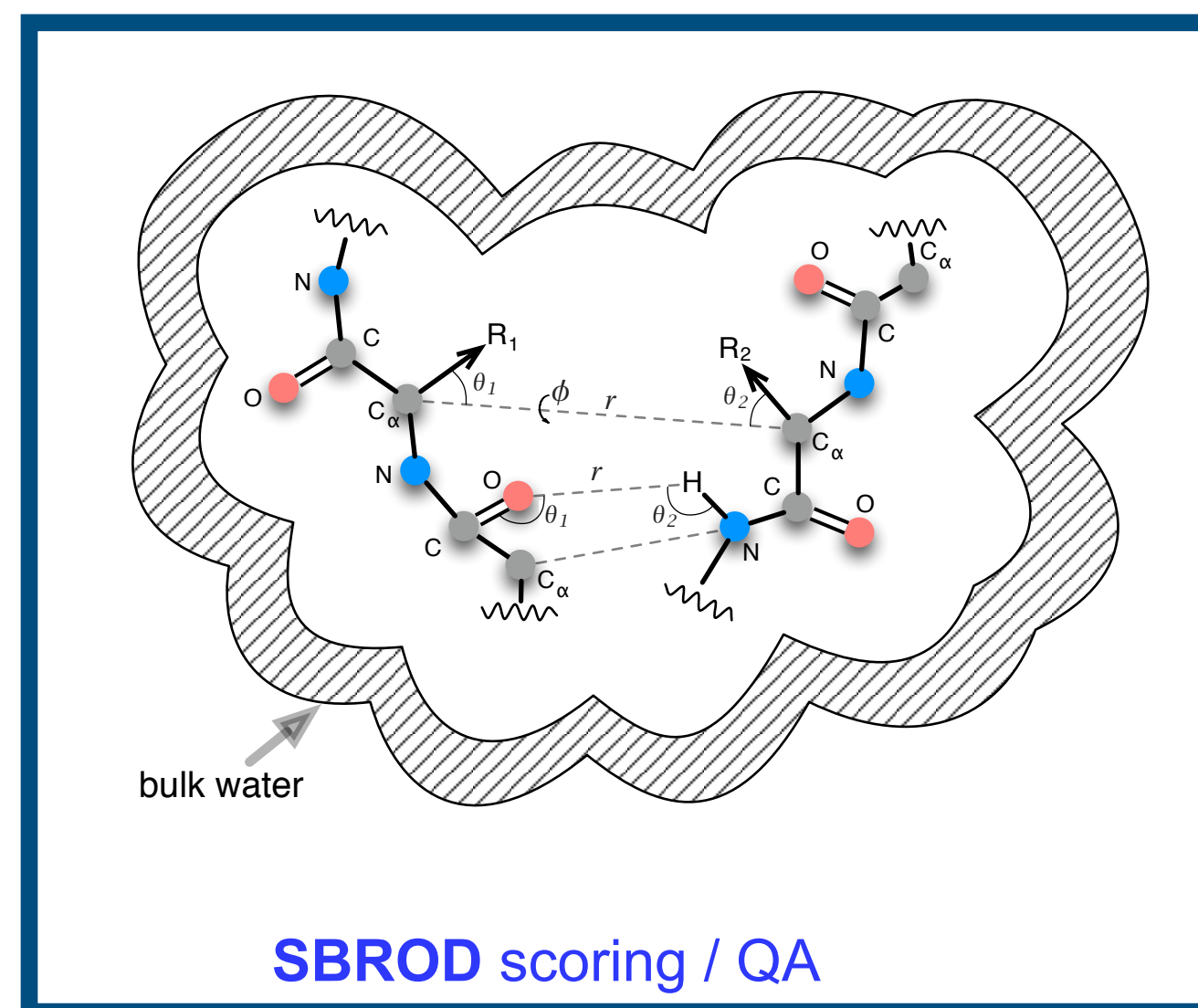
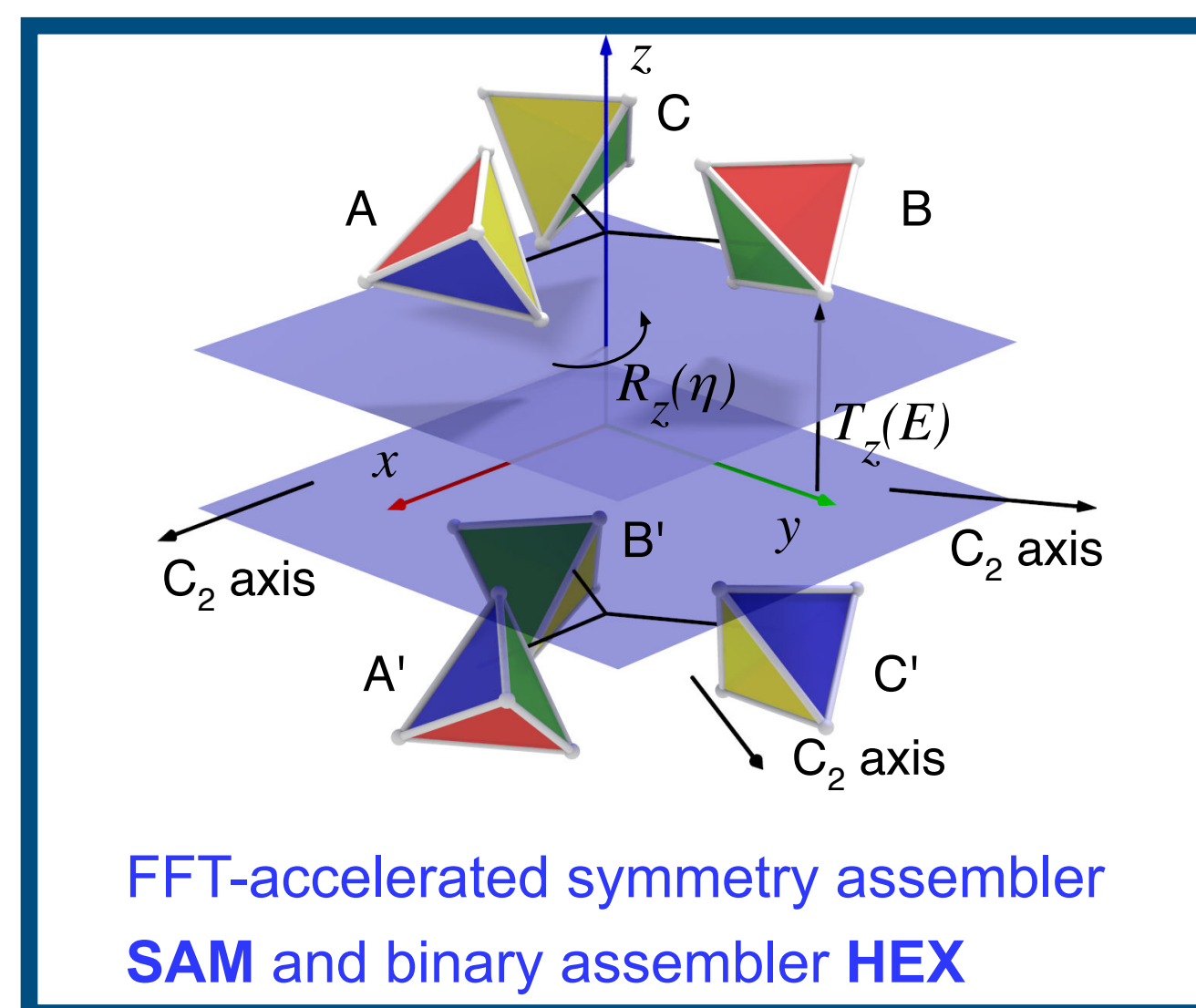
- According to John Moult (the organizer of CASP), the classical folding problem is essentially solved.
- The future challenges include
 - Prediction of protein assemblies (the docking problem)
 - Prediction of protein flexibility
 - Prediction of multiple conformational states and conformational transitions
 - Protein design

Integrative Structural Biology

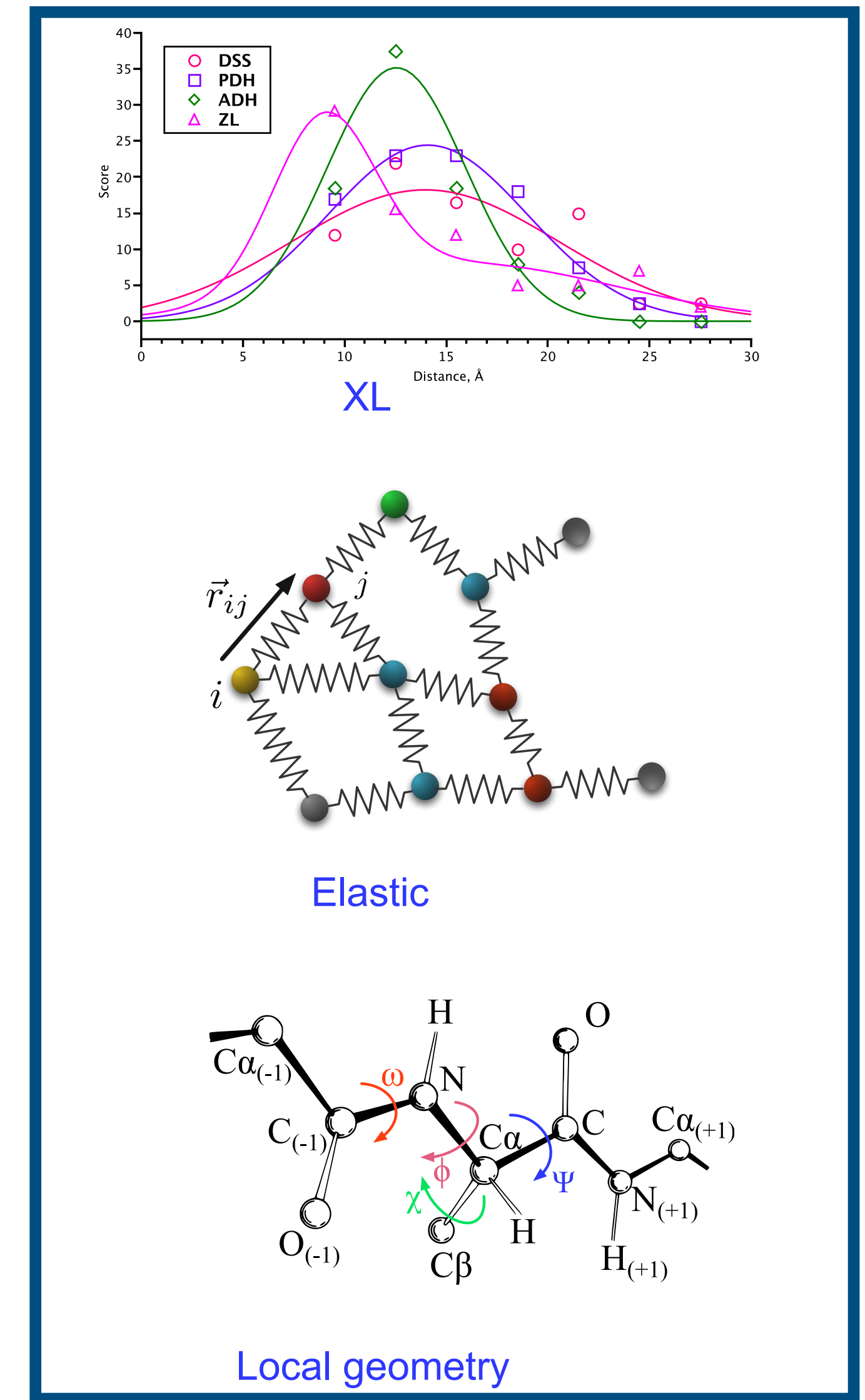
SAXS/SANS-assisted modelling



Docking / Model Selection

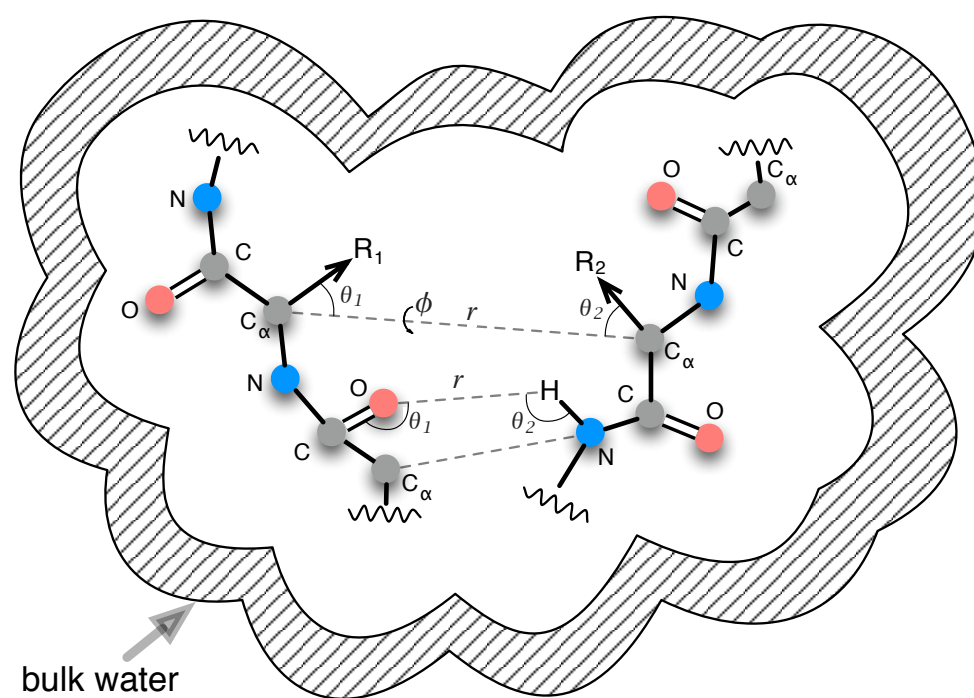
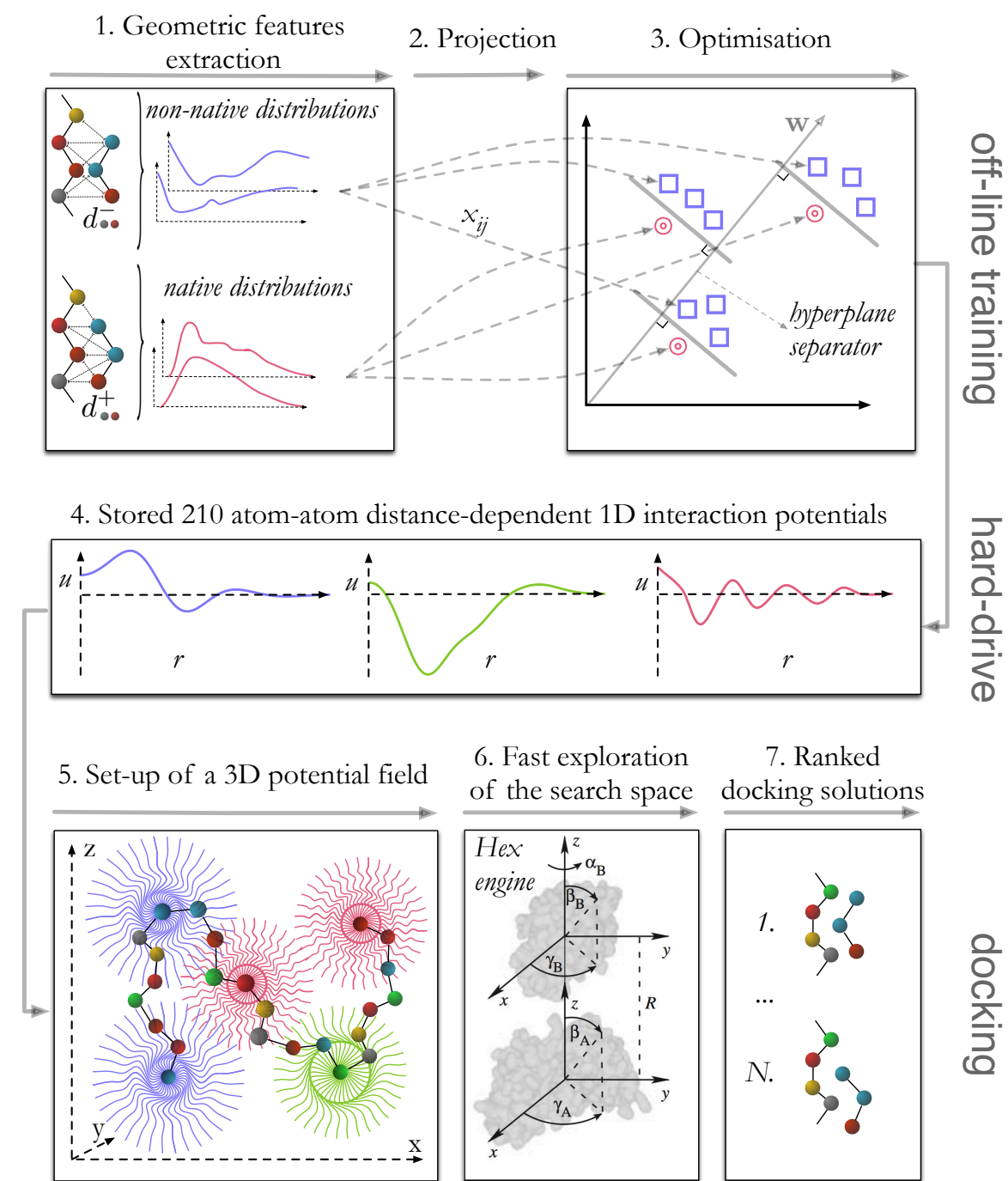


Crosslinks-assisted modelling



Artificial Intelligence

Descriptor-based :



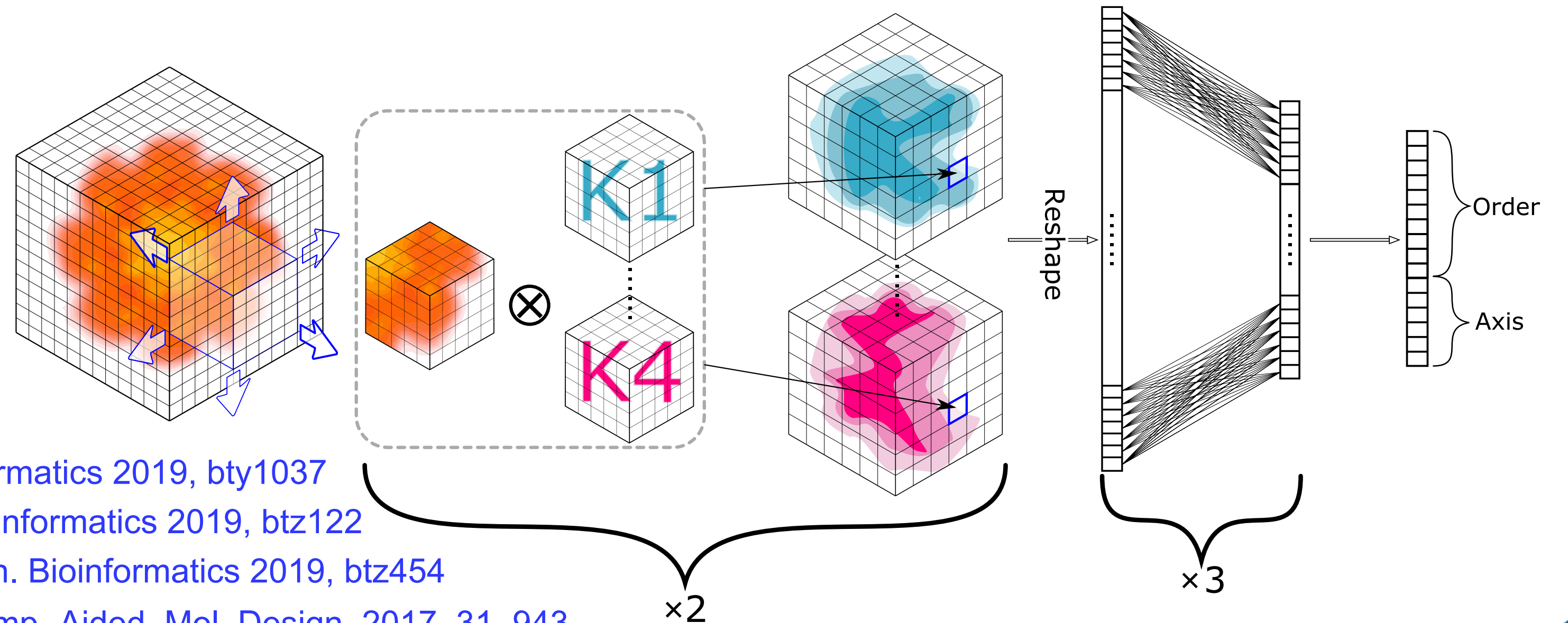
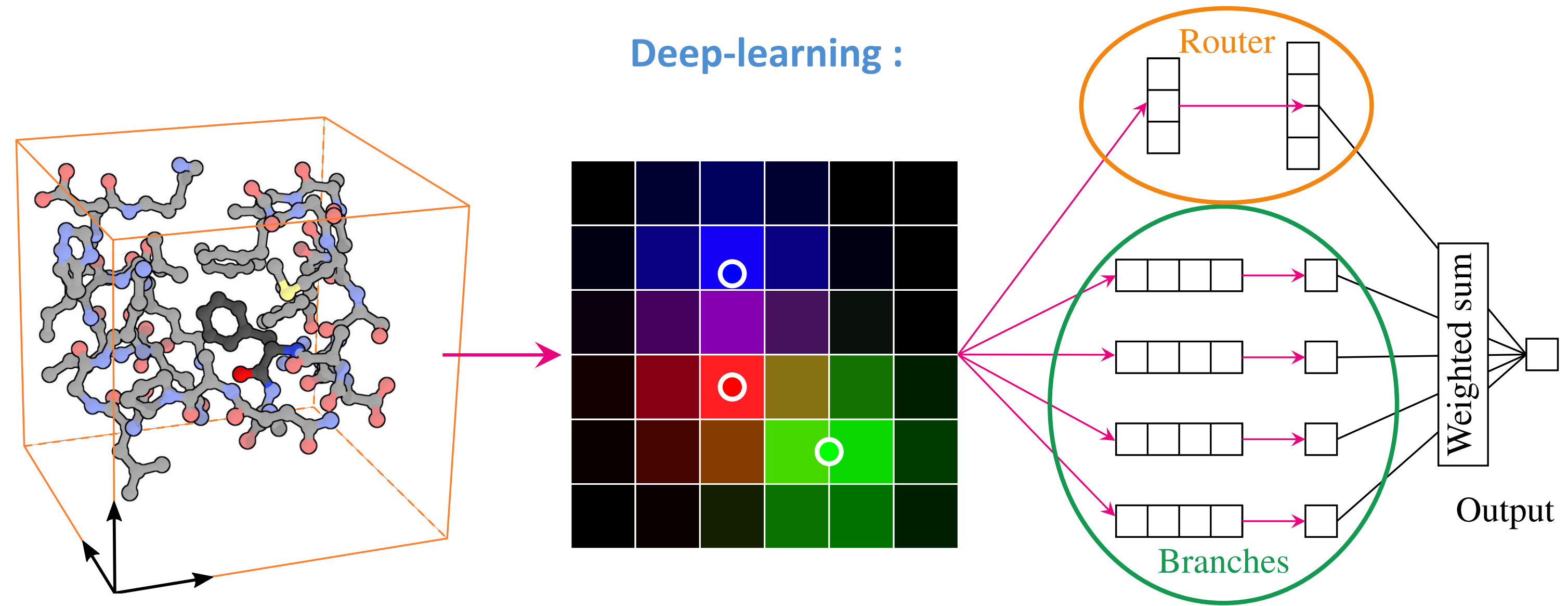
G. Pages et al. *Bioinformatics* 2019, *bty1037*

M. Karasikov et al. *Bioinformatics* 2019, *btz122*

G. Pages & S. Grudinin. *Bioinformatics* 2019, *btz454*

M. Kadukova & S. Grudinin, *J. Comp.-Aided. Mol. Design*, 2017, 31, 943.

Deep-learning :



Docking & Molecular Symmetries

Docking :

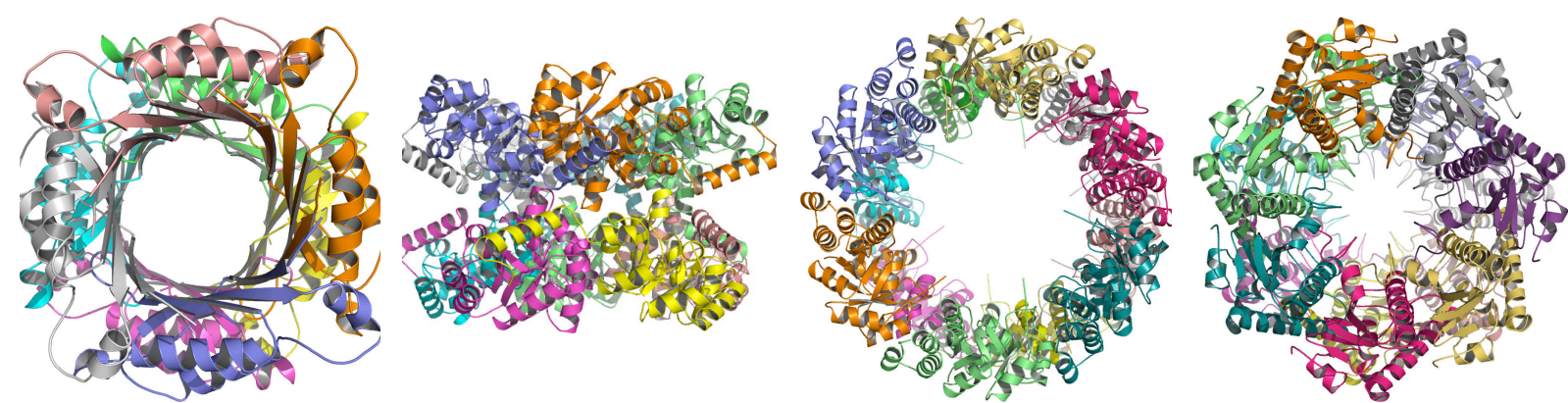
$$\hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(\underline{r}) \longleftrightarrow \hat{R}_y(\omega) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(\underline{r})$$

↑ ↑ ↑ ↑
↑
↑
↑ ↑ ↑ ↑

free
fixed
free

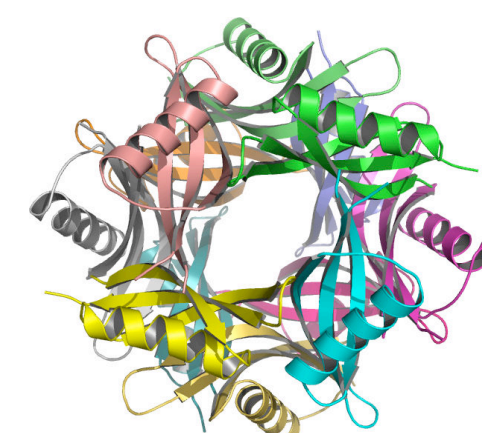
parameters

parameters

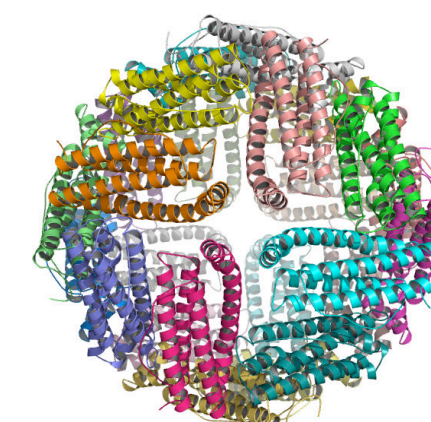


$D_4 / 1b9l$ $D_5 / 1l6w$ $D_6 / 1znn$ $D_7 / 1yg6$

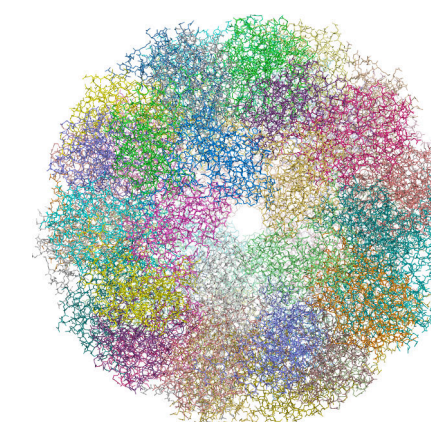
<http://sam.loria.fr>



$T / 1cc9$

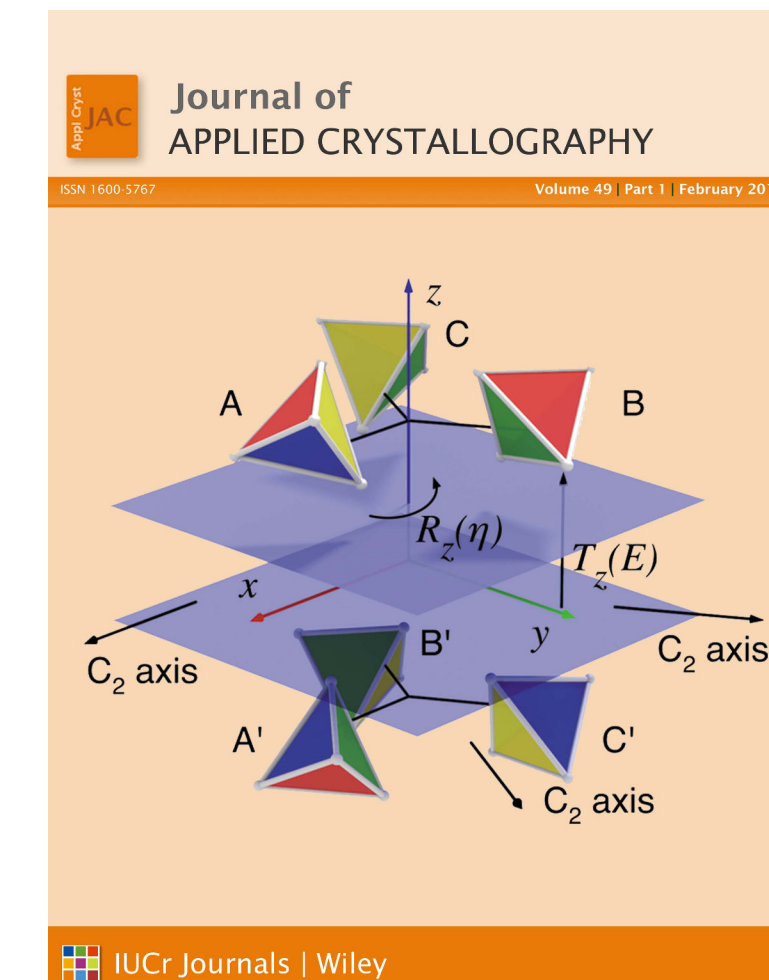


$O / 1ies$



$I / 1hqk$

D.W. Ritchie and S. Grudinin. J. Appl. Cryst. 2016

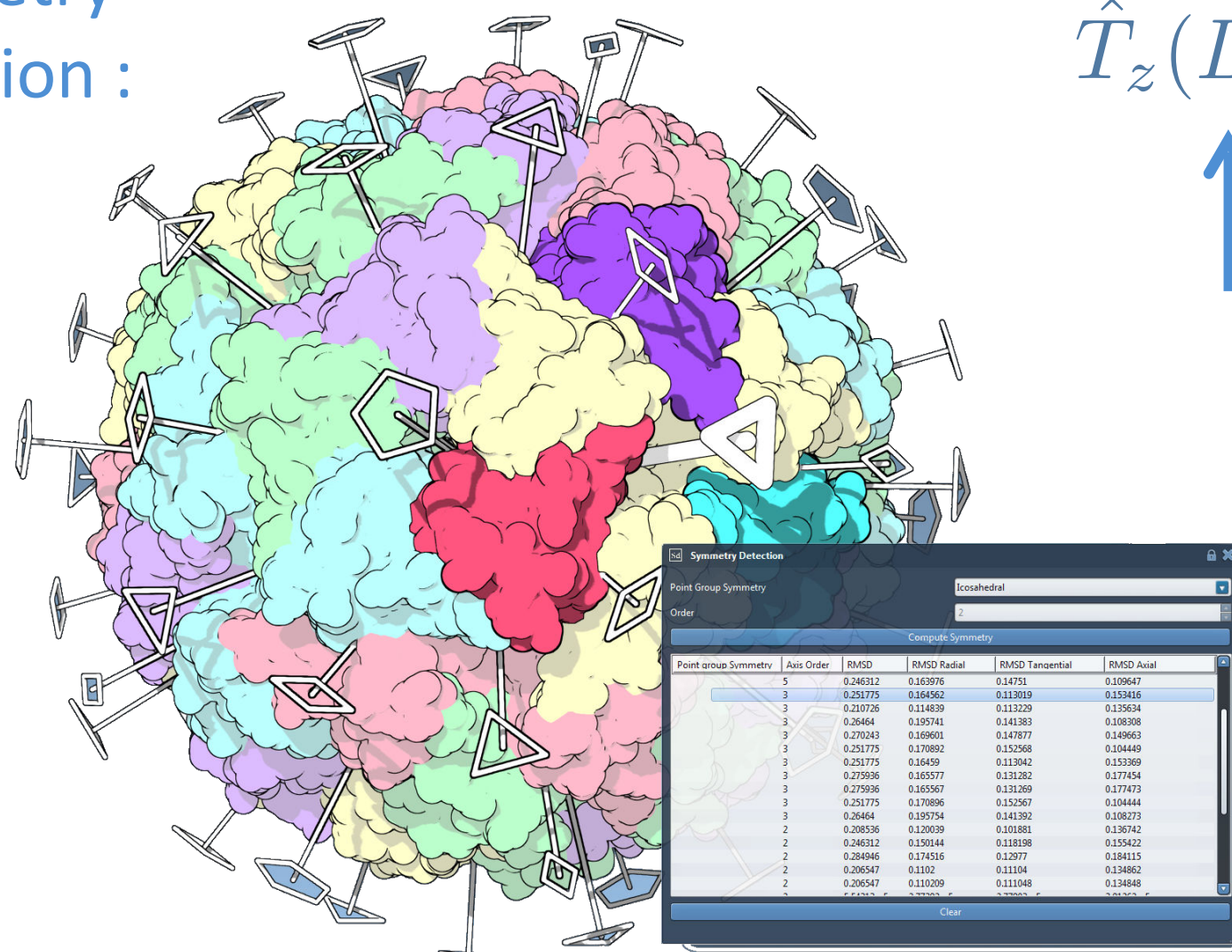


Symmetry detection :

$$\hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(\underline{r}) \longleftrightarrow \hat{R}_y(\omega) \hat{T}_z(D) \hat{R}(\alpha, \beta, \gamma) A(\underline{r})$$

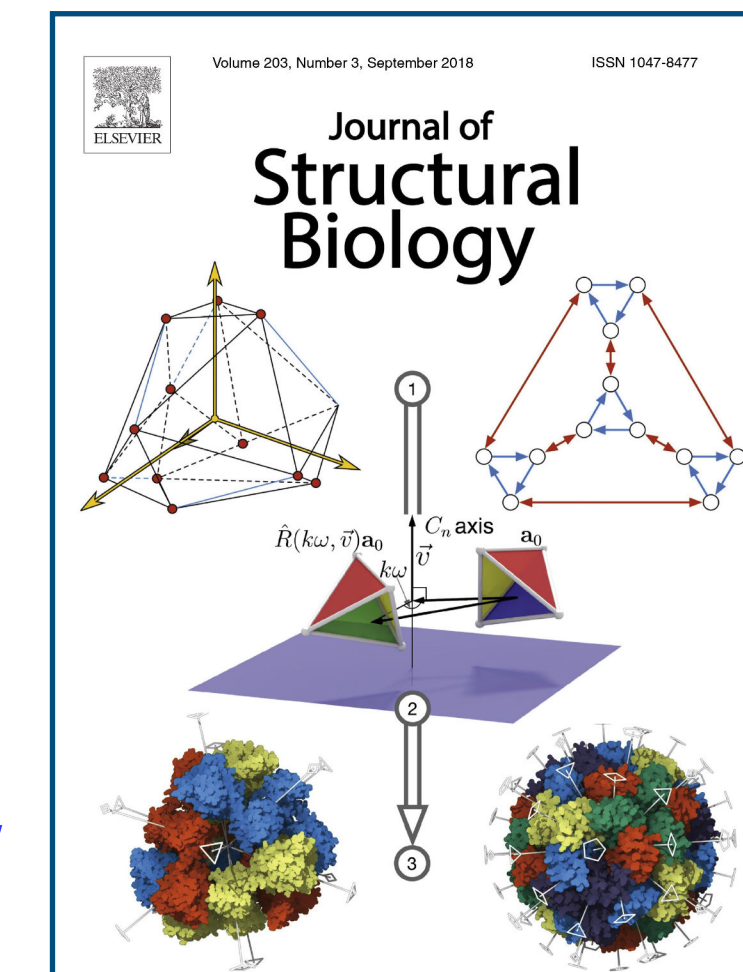
↑
↑ ↑ ↑ ↑
↑
↑
↑ ↑ ↑ ↑

fixed
free
fixed

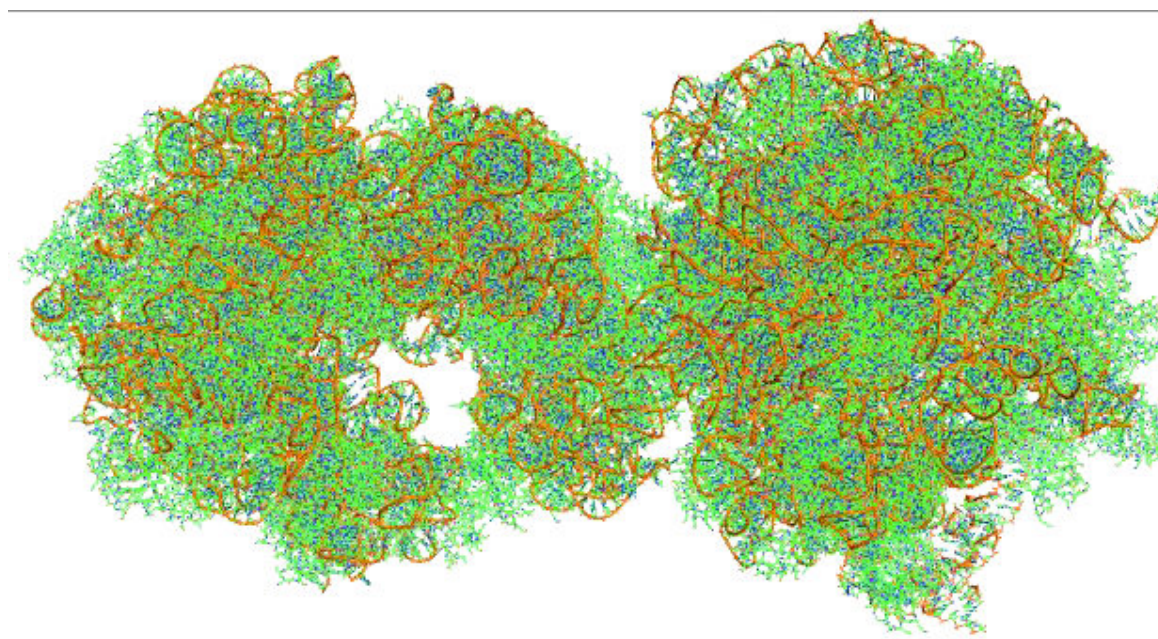
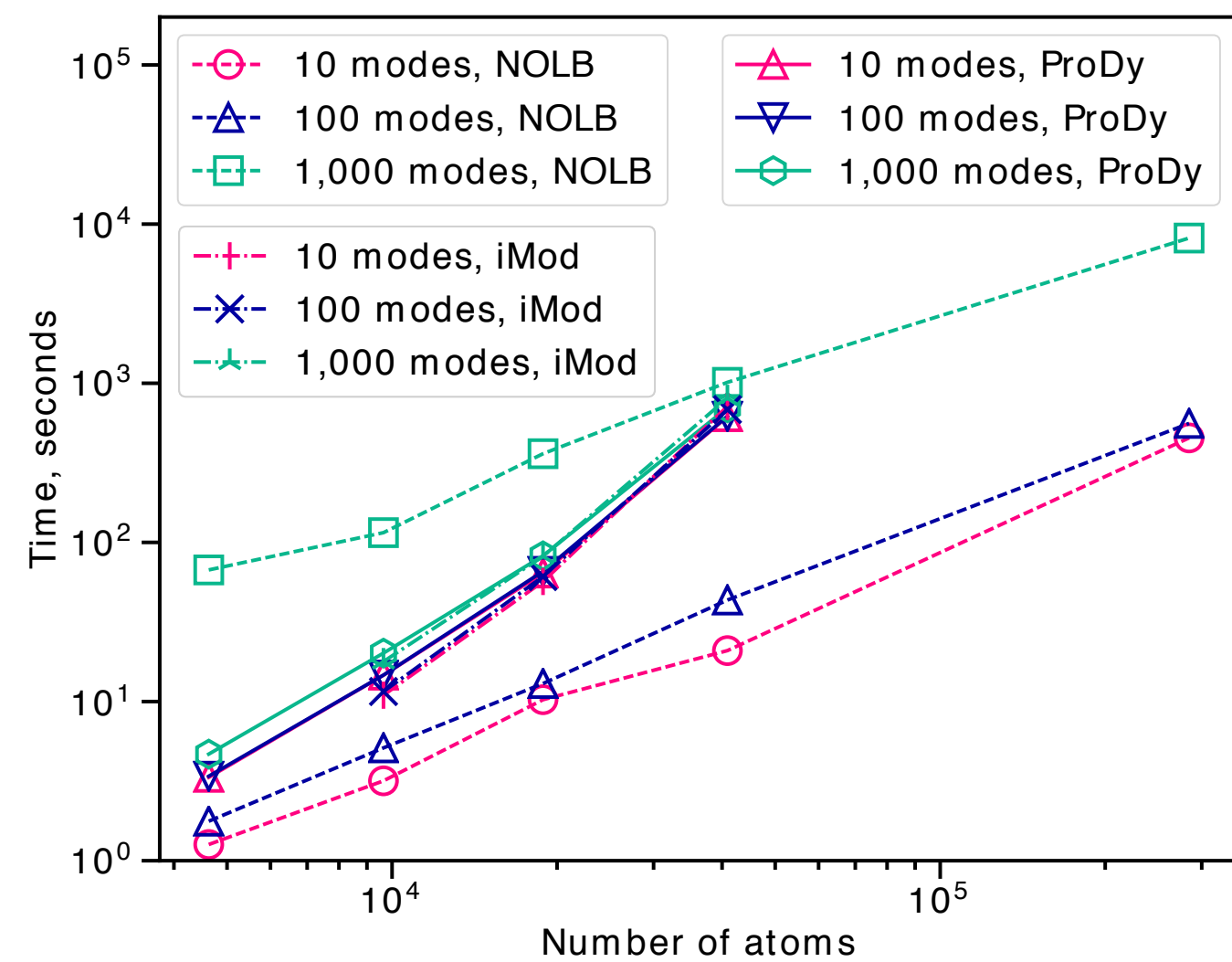
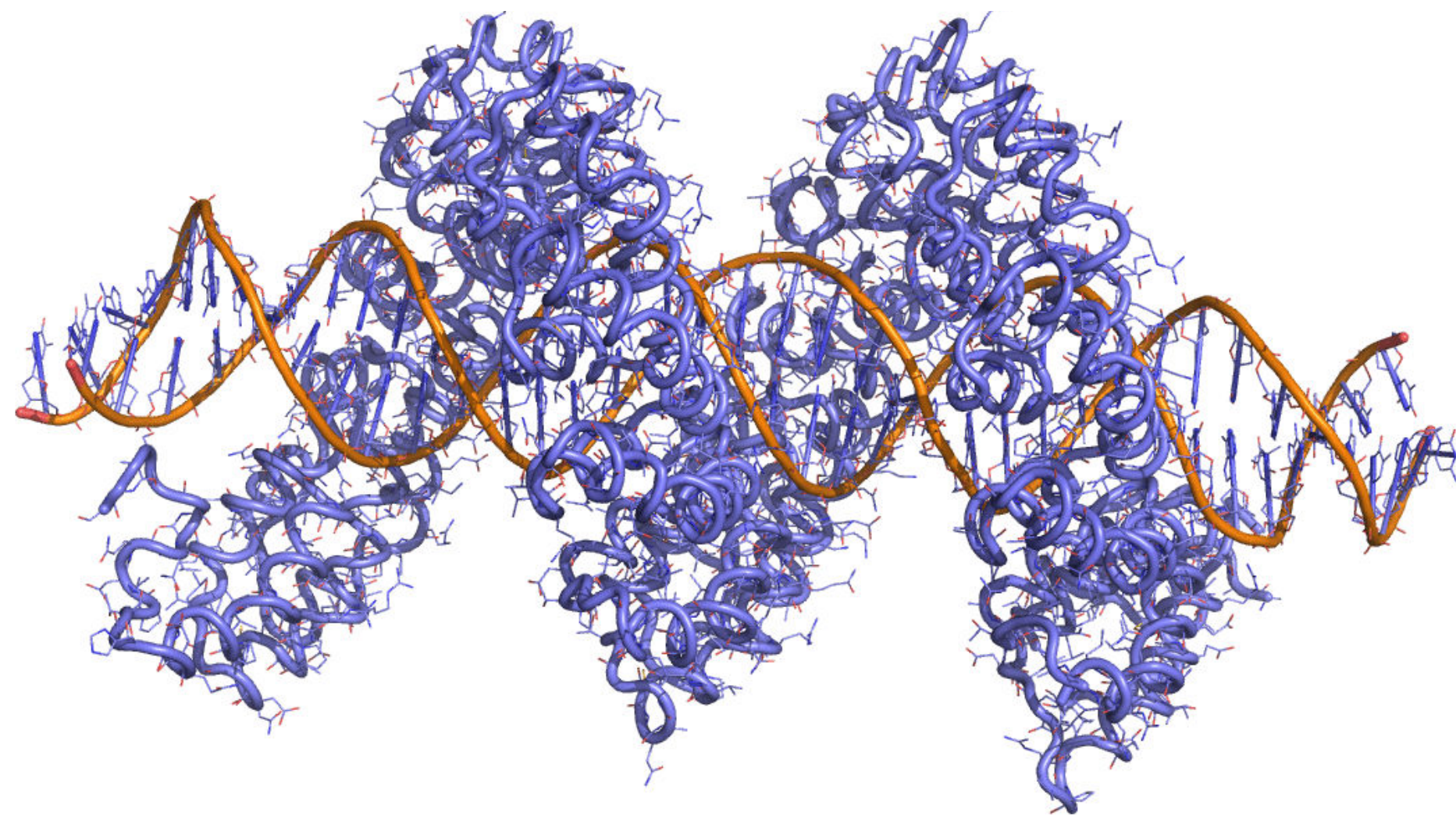


- Detects the best axes of symmetry
- Computes per-axis RMSD
- Deals with missing subunits

<https://team.inria.fr/nano-d/software/ananas/>
 G. Pages et al. J. Struc. Biol. 2018
 G. Pages & S. Grudinin. J. Struc. Biol. 2018



NOLB – Nonlinear NMA



two 70S Ribosomes

team.inria.fr/nano-d/software/nolb-normal-modes/

JCTC

Journal of Chemical Theory and Computation
May 2017 Volume 13 Number 5 pubs.acs.org/JCTC

NOLB NMA

$$M(\ddot{q} + \dot{q}_0) + \nabla V(q_0 + q) \approx M\ddot{q} + Kq = 0$$

$$K_w = M^{-1/2} K M^{-1/2} = L \Lambda L^T$$

$$\eta = L^T M^{1/2} q$$

$$\ddot{\eta}_i + \lambda_i \eta_i = 0 \quad i = 1 \dots 3N_a$$

nonlinear normal modes

linear normal modes

$$\vec{A}' = R(\Delta\phi, \vec{n})(\vec{A} - \vec{r}_0) + \vec{r}_0 + \Delta\vec{x}_{||}$$

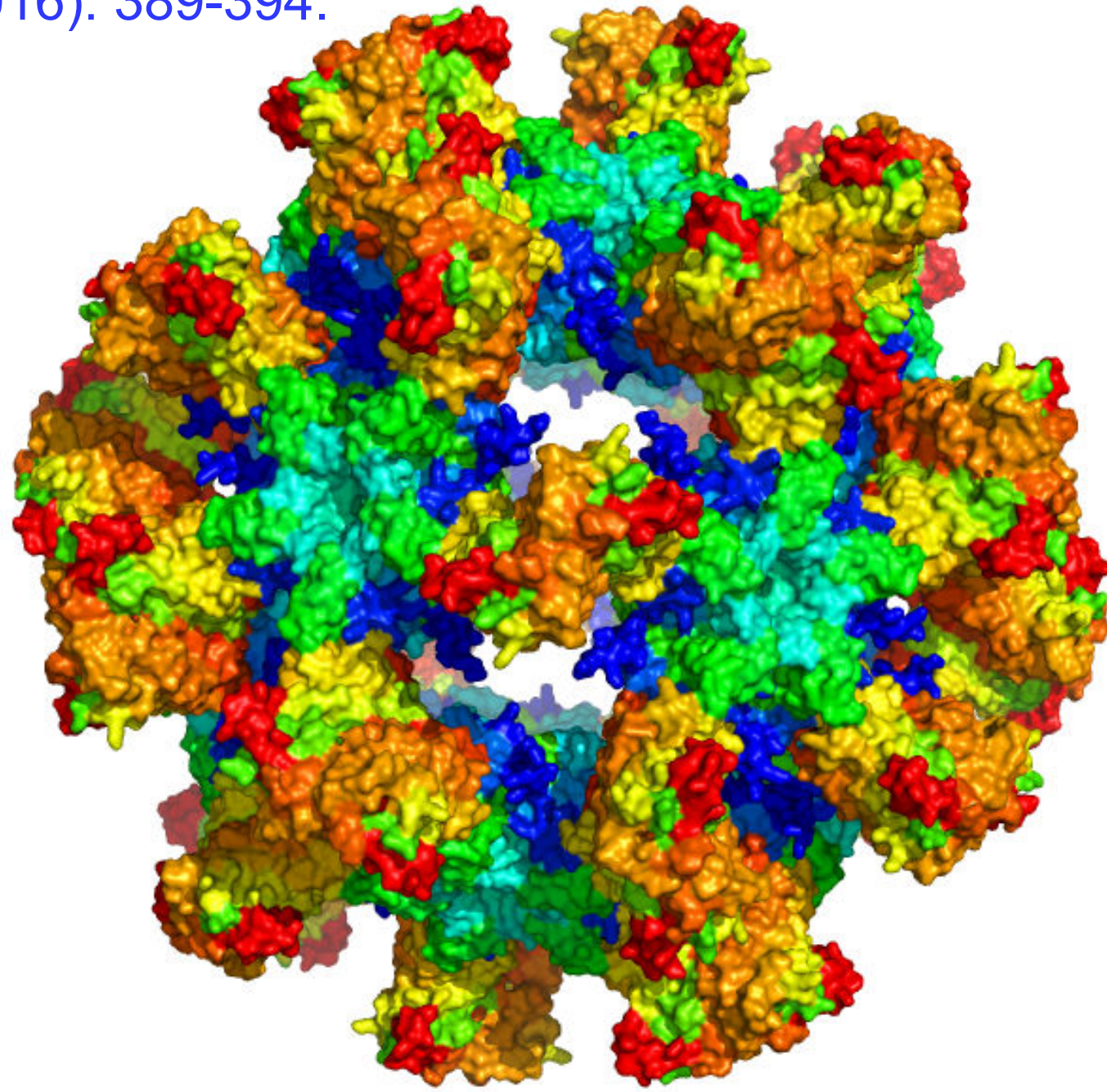
Most Trusted. Most Cited. Most Read.

www.acs.org

A. Hoffmann and S. Grudinin. JCTC 2017

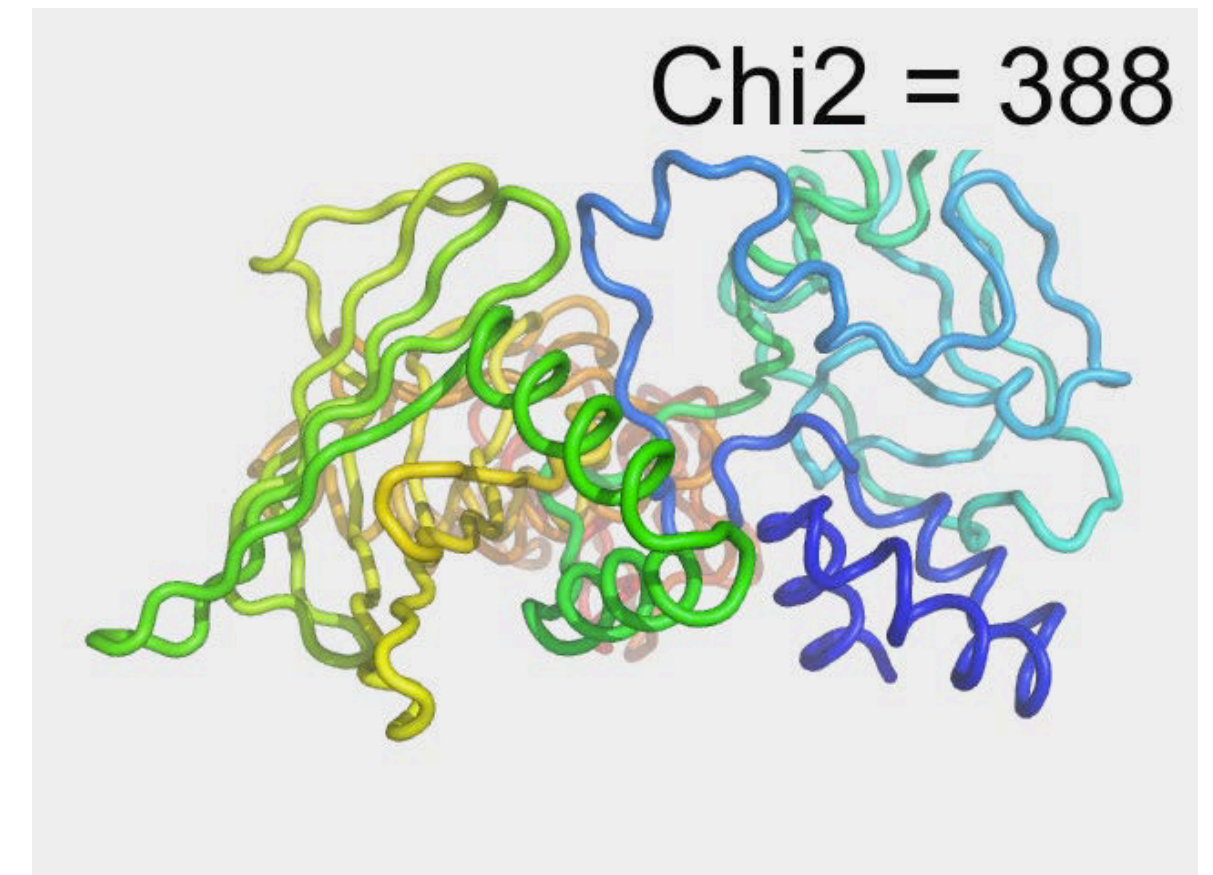
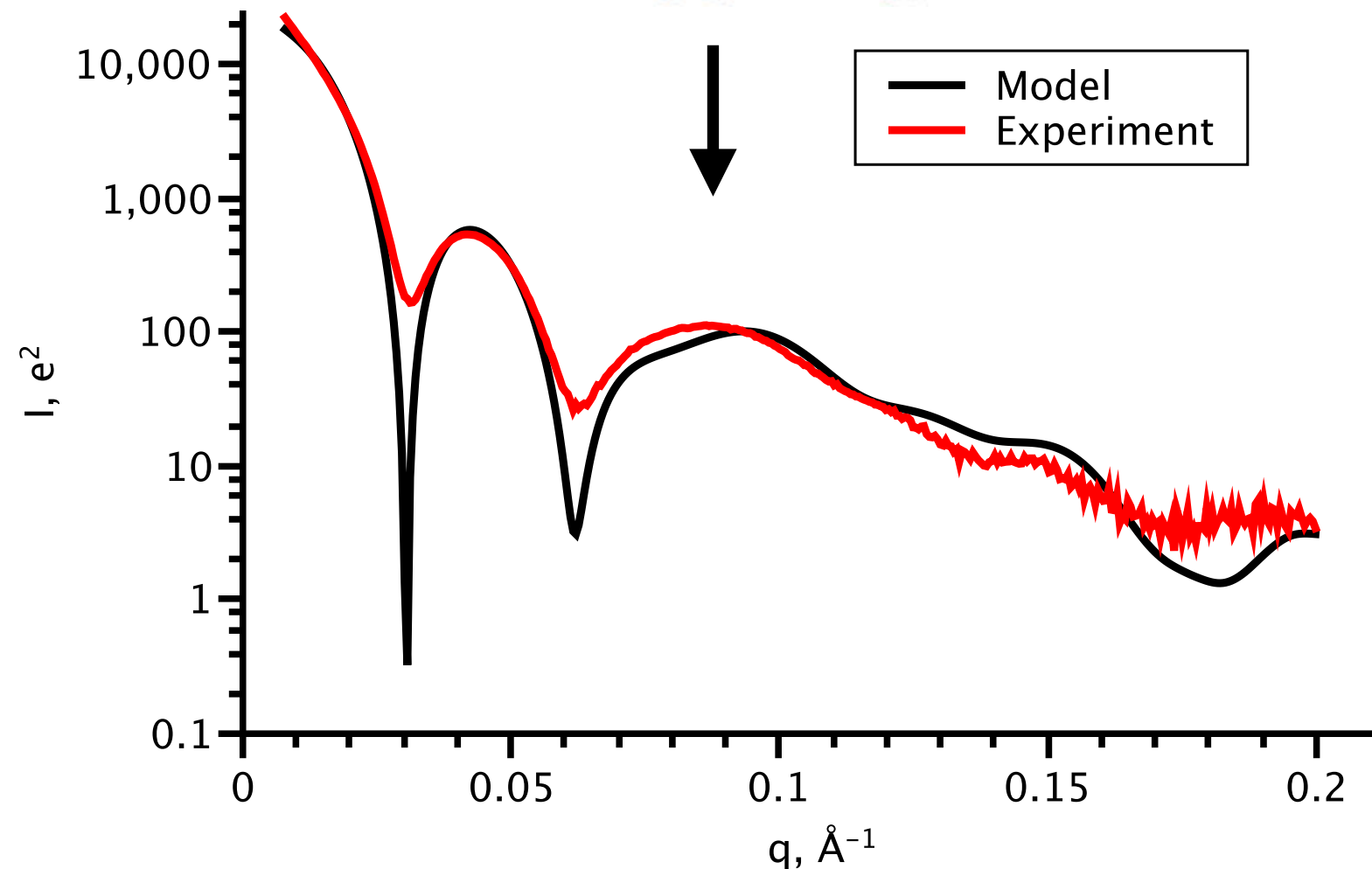
Bale, Jacob B., et al. *Science*
353.6297 (2016): 389-394.

Pepsi-SAXS/SANS



```

=====Reading PDB file=====
Read PDB file..... : I52-33.pdb
Number of atoms read..... : 162120
=====Analysing input PDB data=====
Radius of gyration..... : 106.919 A
Maximum extension from the center..... : 143.247 A
Number of residues..... : 346
Molecular weight..... : 38.2875 kDa
=====Adaptive memory allocation=====
Estimated minimal expansion order of spherical harmonics..... : 23
===== Timing : =====
Reading PDB file..... : 0.131131 s
Computing the hydration layer..... : 0.120579 s
Computing amplitudes for the hydration shell..... : 0.379462 s
Computing atomic and excluded volume amplitudes..... : 2.82286 s
Fitting to the experimental curve..... : 0.010125 s
.....
Total time : ..... : 3.47689 s
=====
  
```

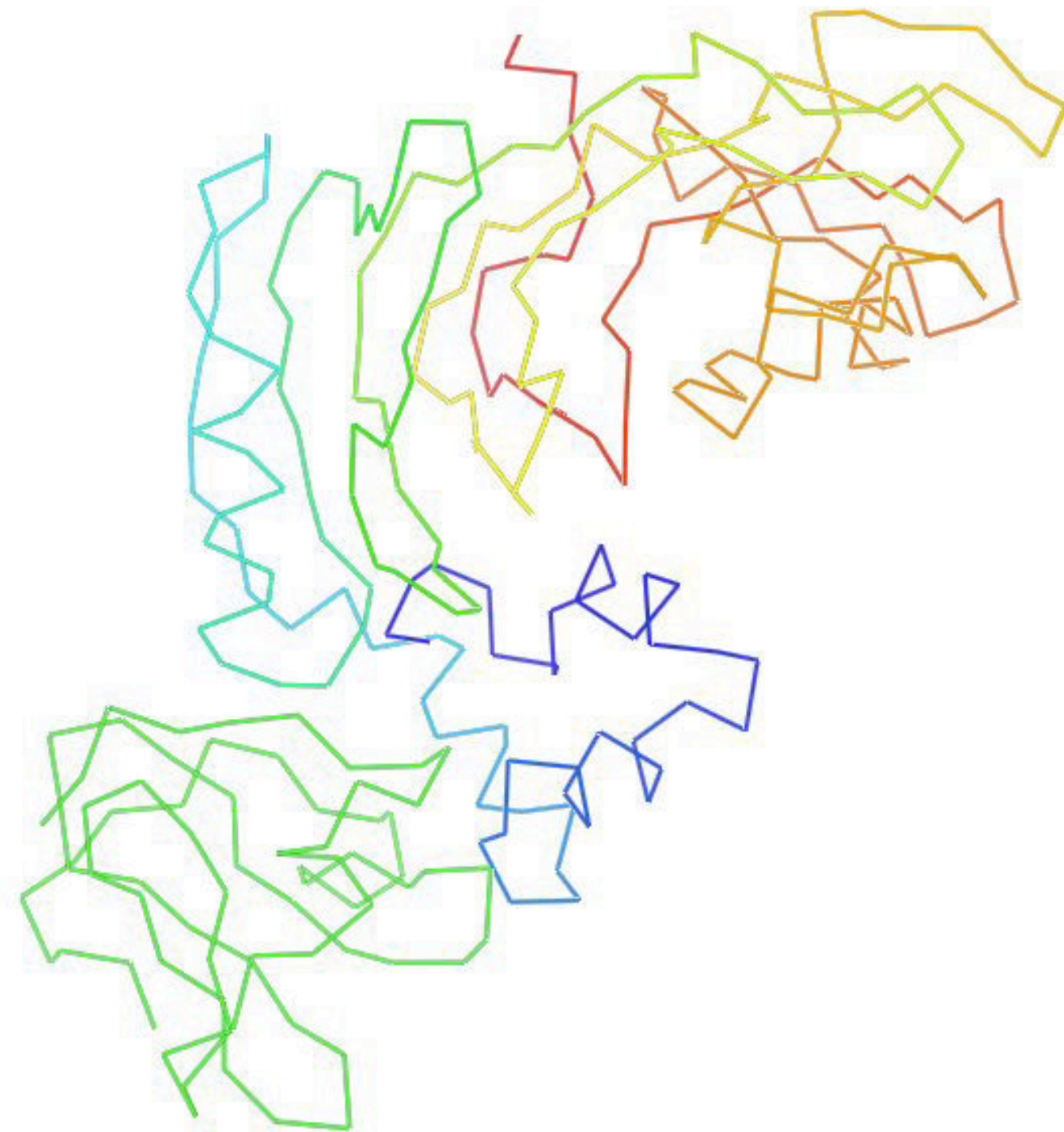


<https://team.inria.fr/nano-d/software/Pepsi-Saxs/>

Grudin et al (2017). *Acta Cryst*, D73, pp.449 – 464

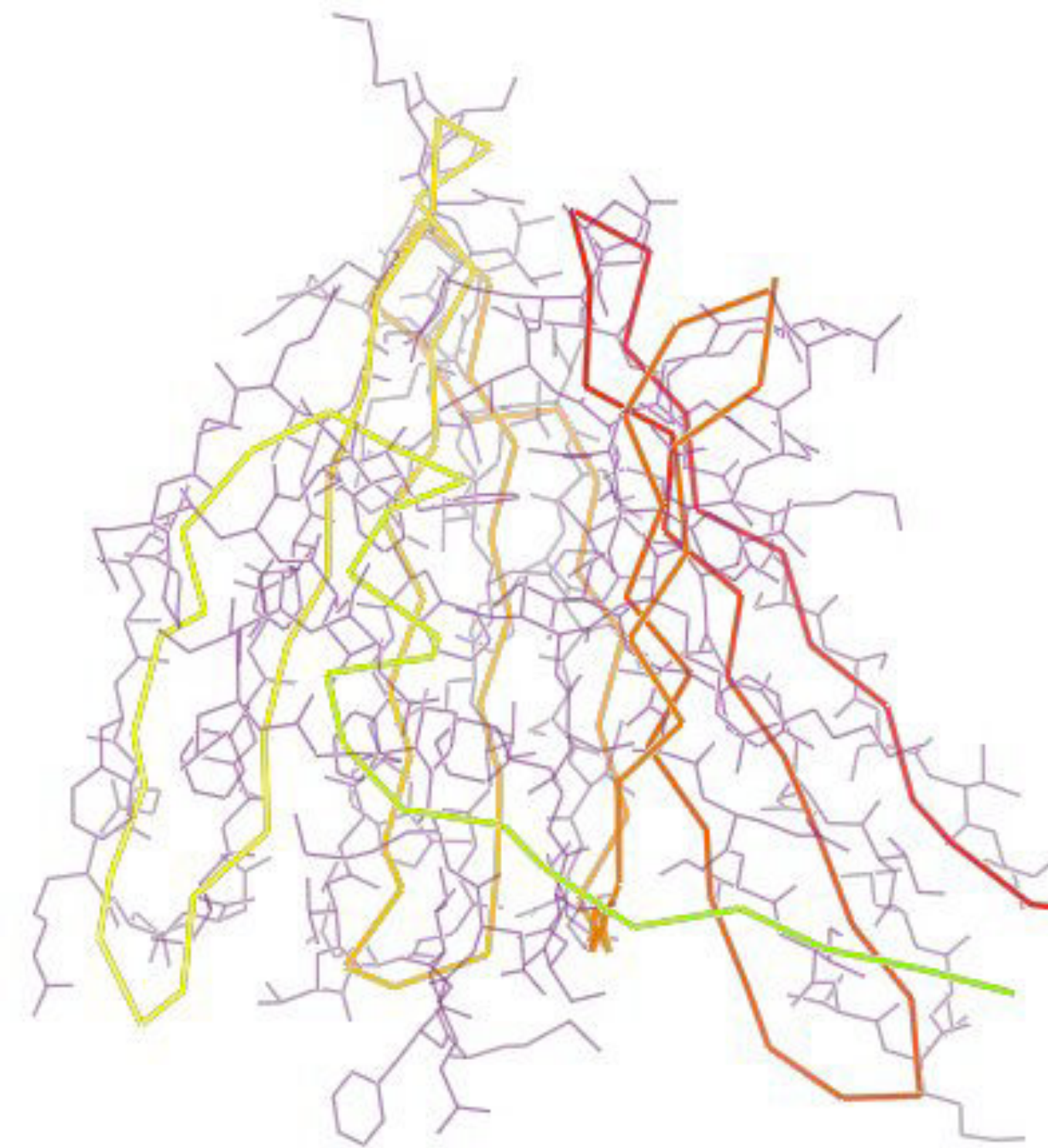
Pepsi-SAXS in CASP13

S0953



Chi2 dropped from 153 to 5.37

S0968s2



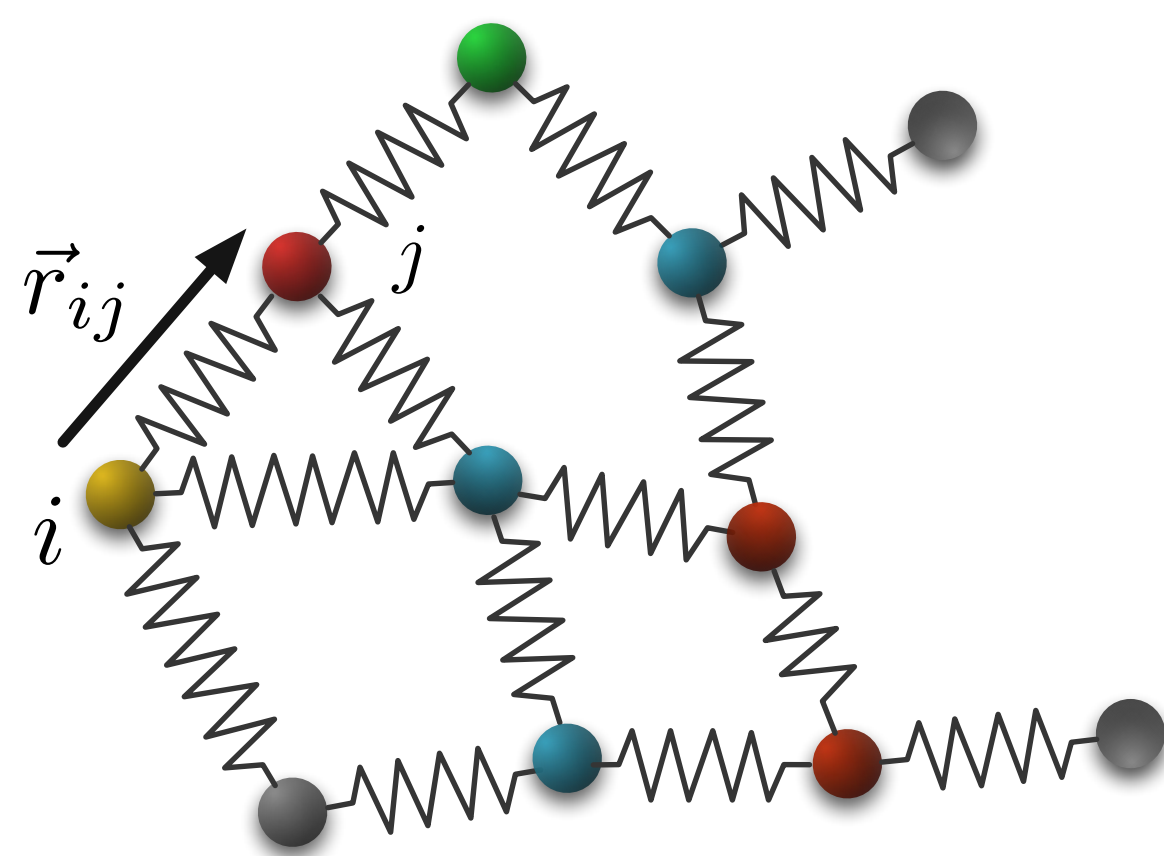
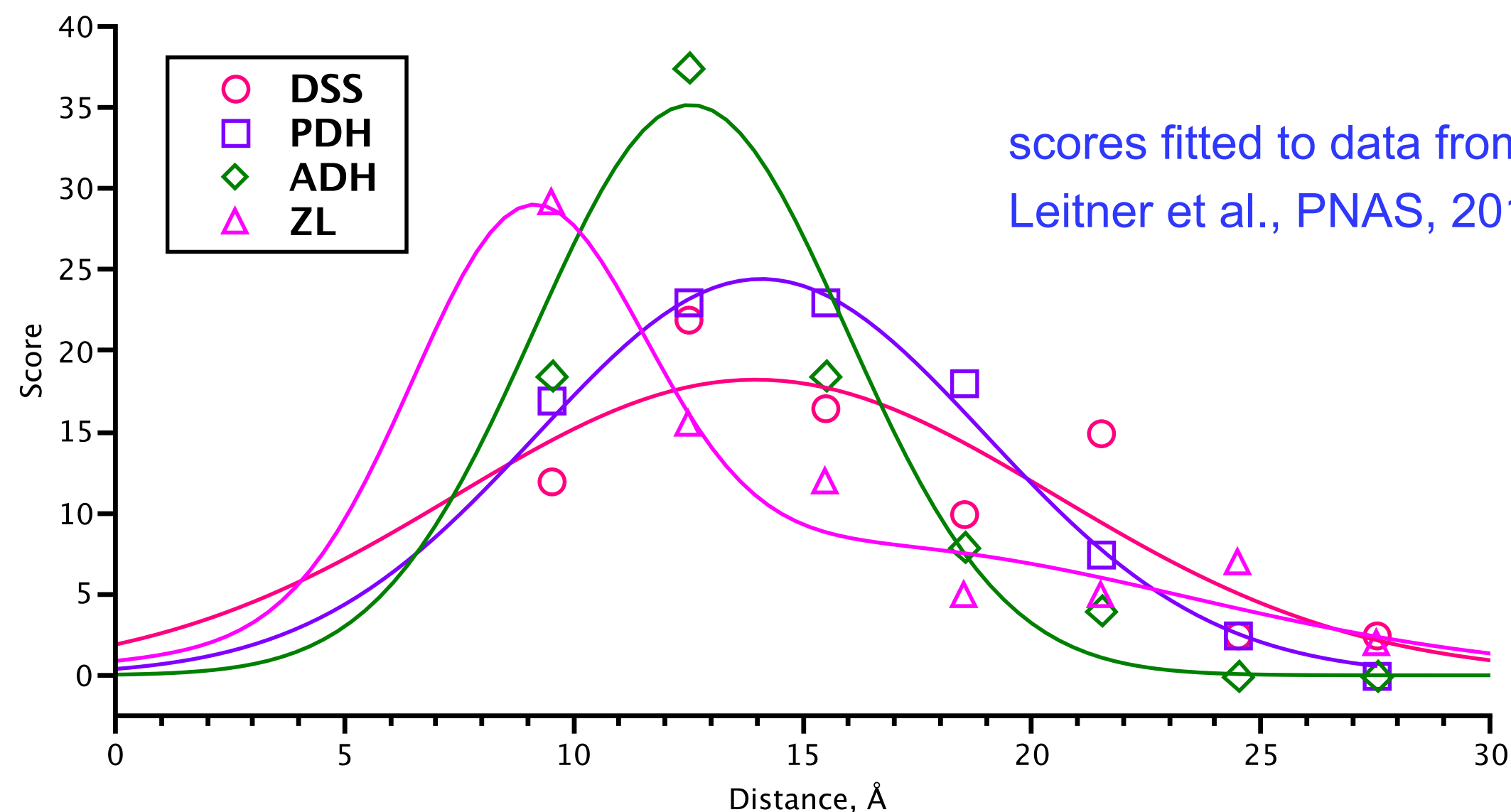
Chi2 dropped from 0.28 to 0.18

Hura, Greg L., et al. "Small angle X-ray scattering-assisted protein structure prediction in CASP13 and emergence of solution structure differences." *Proteins: Structure, Function, and Bioinformatics* (2019).

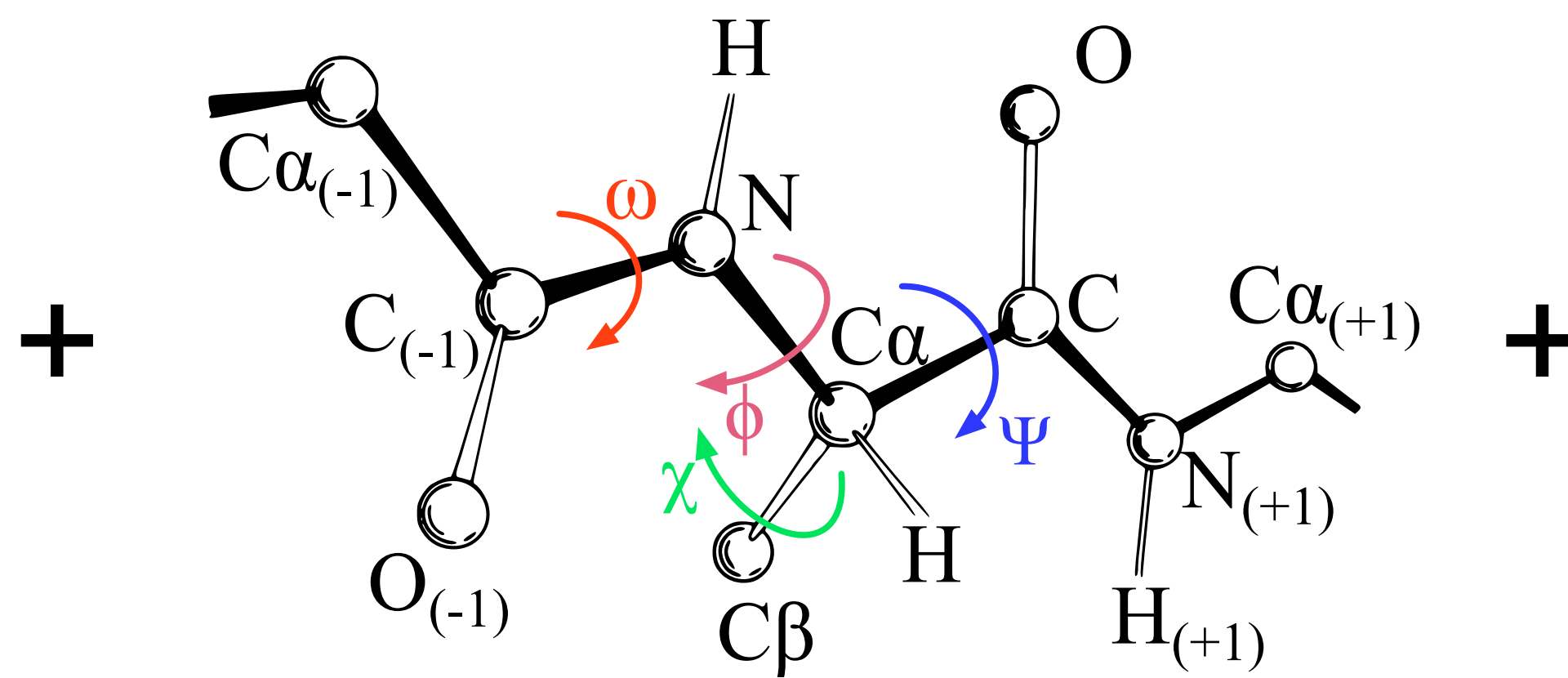
Cross-links-assisted modelling

XL Energy function :

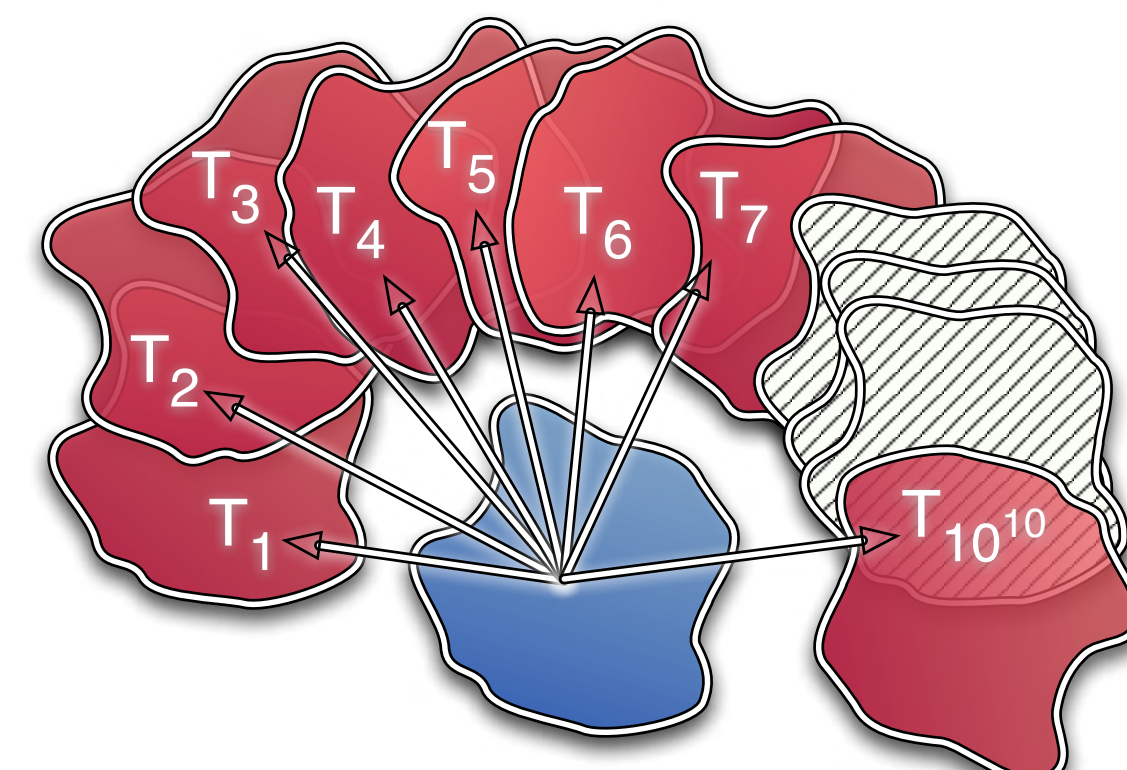
- DSS : one Gaussian
- PDH : one gaussian
- ADH : one Gaussian
- ZL : double Gaussian



Elastic network



Local optimization



T	Score
1	2.34
2	4.56
3	9.01
4	1.71
5	6.48
6	3.25
7	7.15
10 ¹⁰	8.18

Exhaustive rescoring of docking poses

Fajardo, J. Eduardo, et al. Proteins: Structure, Function, and Bioinformatics (2019).

Cross-links-assisted modelling

- ⦿ Solve

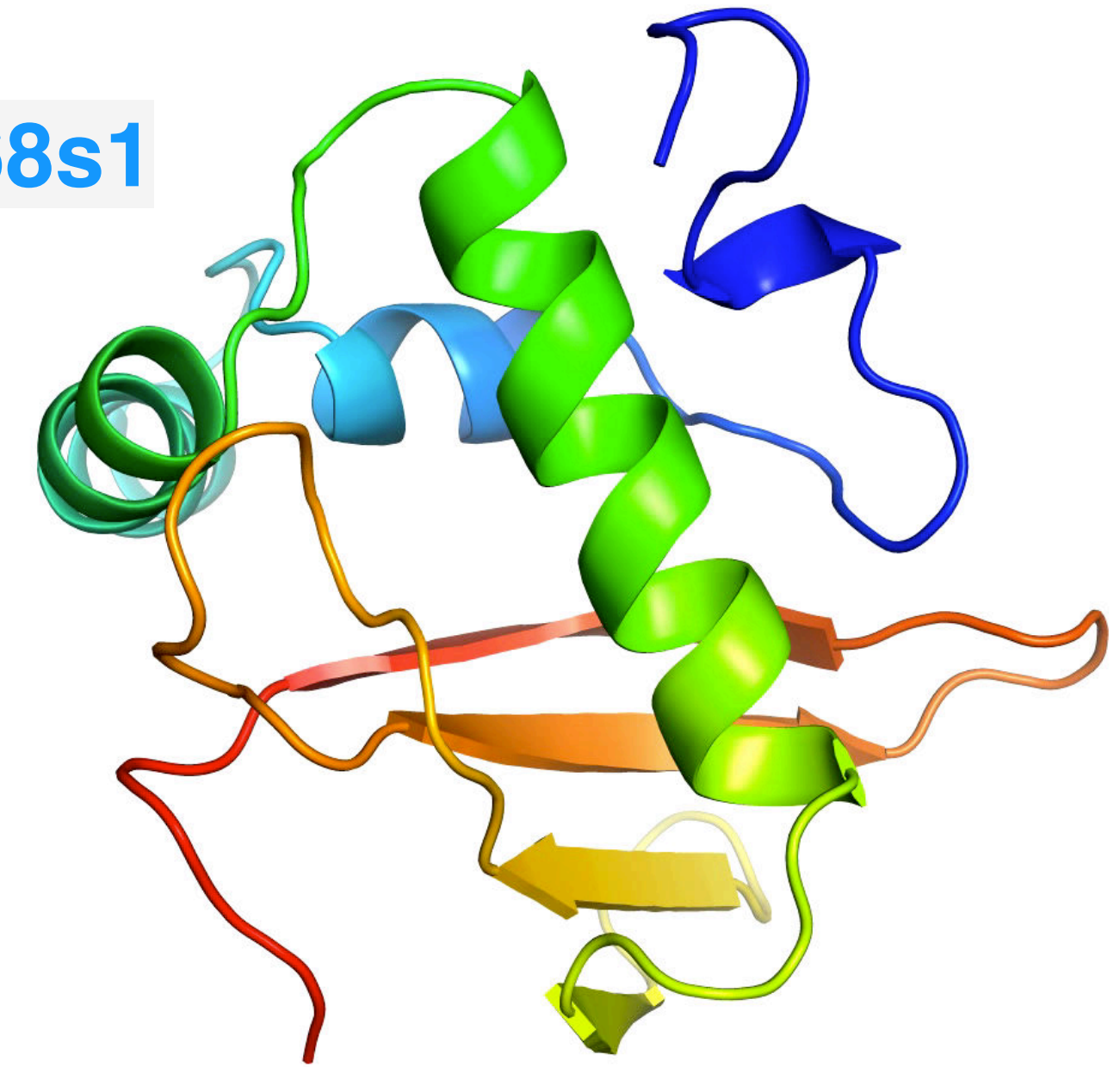
$$\frac{1}{2}x^T Hx + x^T f \rightarrow \min$$

thus

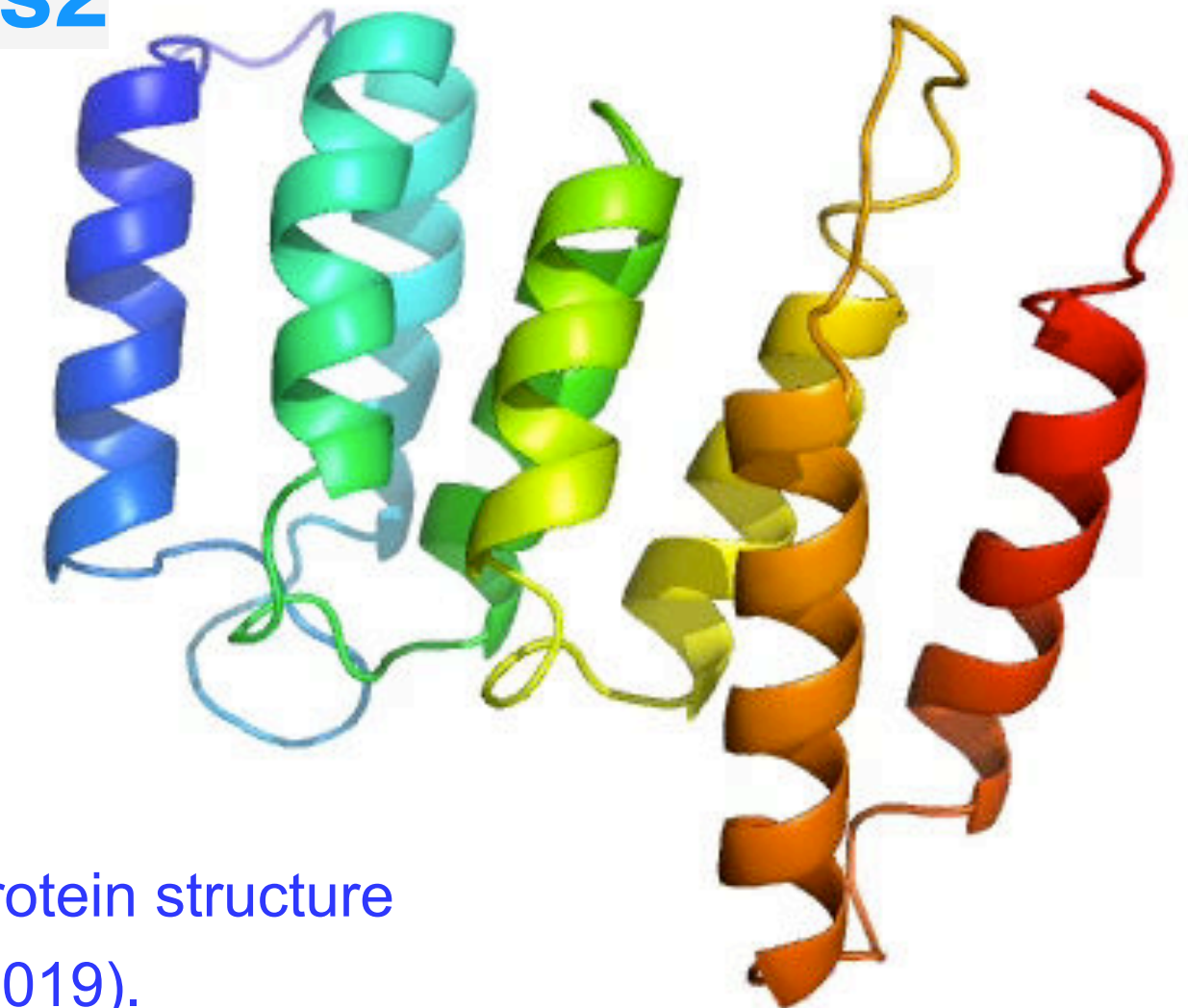
$$Hx = -f$$

- ⦿ Compute x
- ⦿ Apply displacement x
- ⦿ Minimize the structure locally

X0968s1

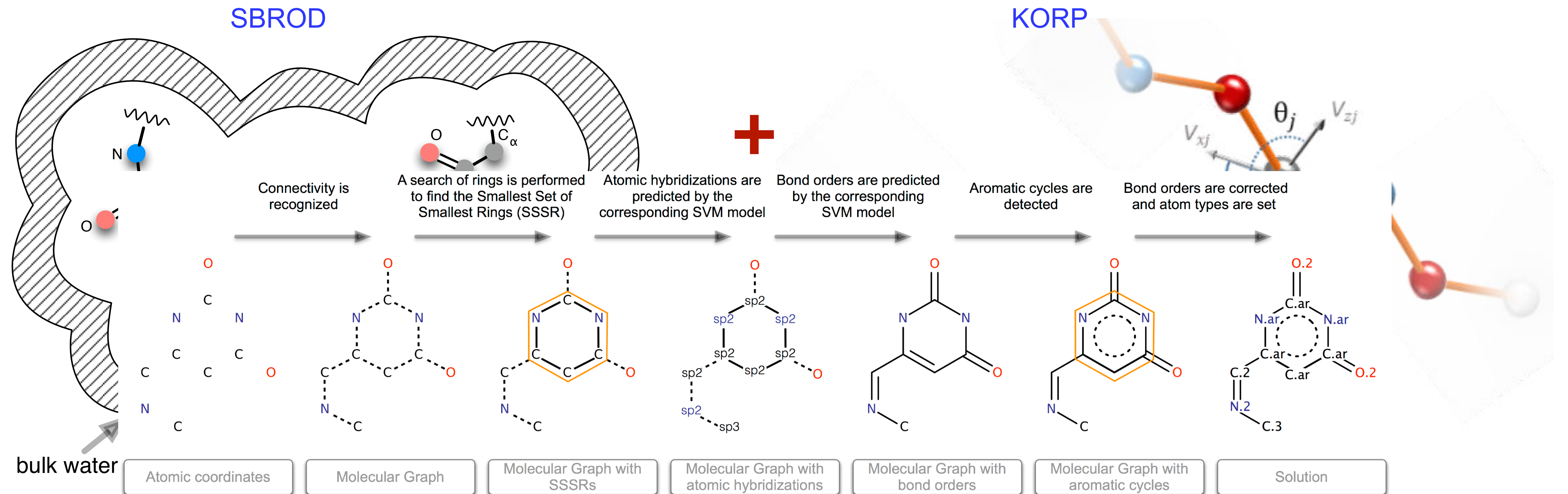


X0957s2



Fajardo, J. Eduardo, et al. "Assessment of chemical-crosslink-assisted protein structure modeling in CASP13." *Proteins: Structure, Function, and Bioinformatics* (2019).

Under development : Orientation-based coarse-grained potentials



- 4 types of features

- 4D representation of orientations

- Ridge-regression model

<https://team.inria.fr/nano-d/software/sbrod/>

M Karasikov et al. Bioinformatics 2019.

Sergei Grudin

- 1 type of features

- 6D representation of orientations

- Boltzmann inversion

<http://chaconlab.org/modeling/korp>

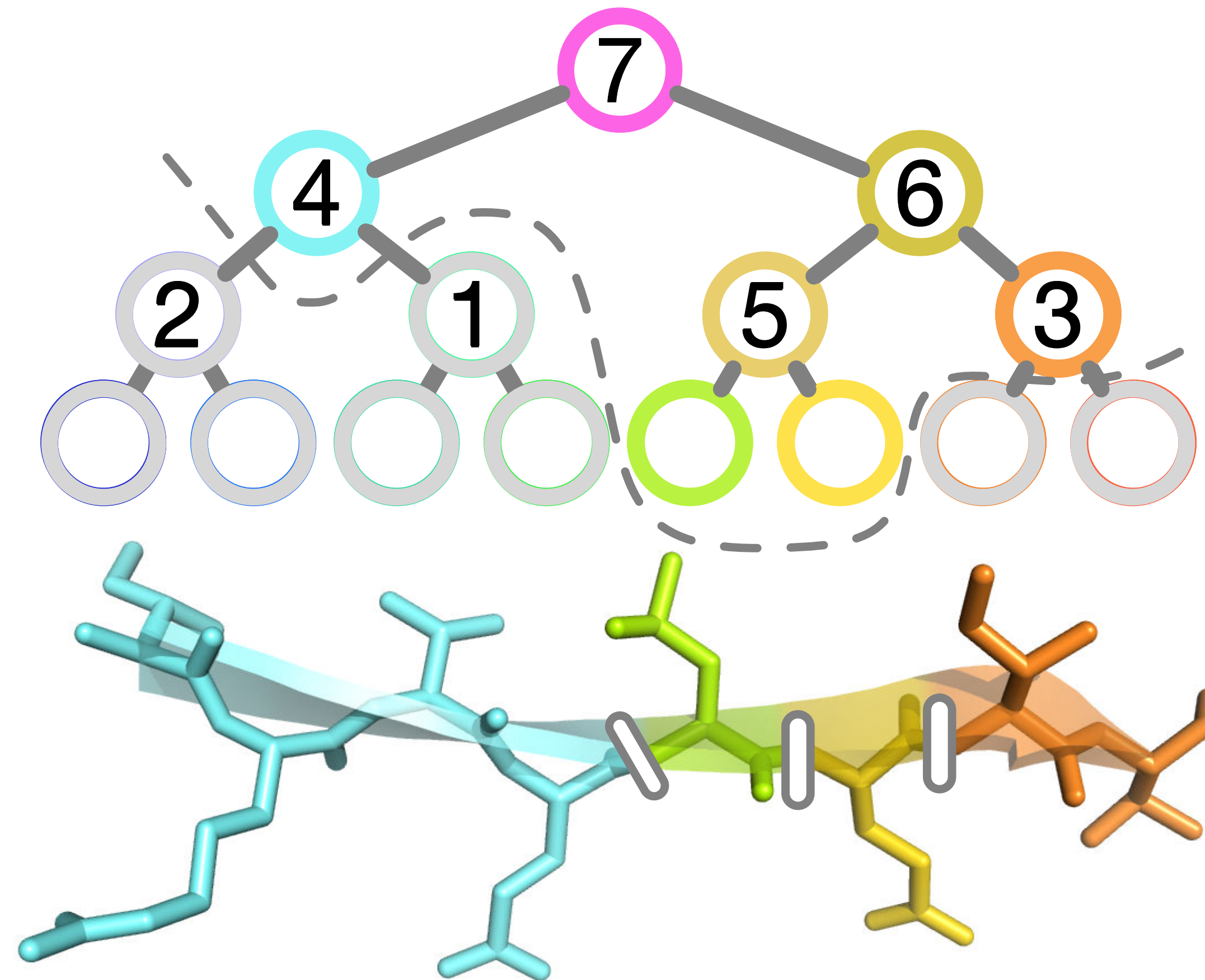
López-Blanco, J. R., & Chacón, P. Bioinformatics 2019.

MASIM, November 2019

Under development : Domain identification

- Bottom-up tree construction based on motion similarity

- Top-down pass



team.inria.fr/nano-d/software/nolb-normal-modes/

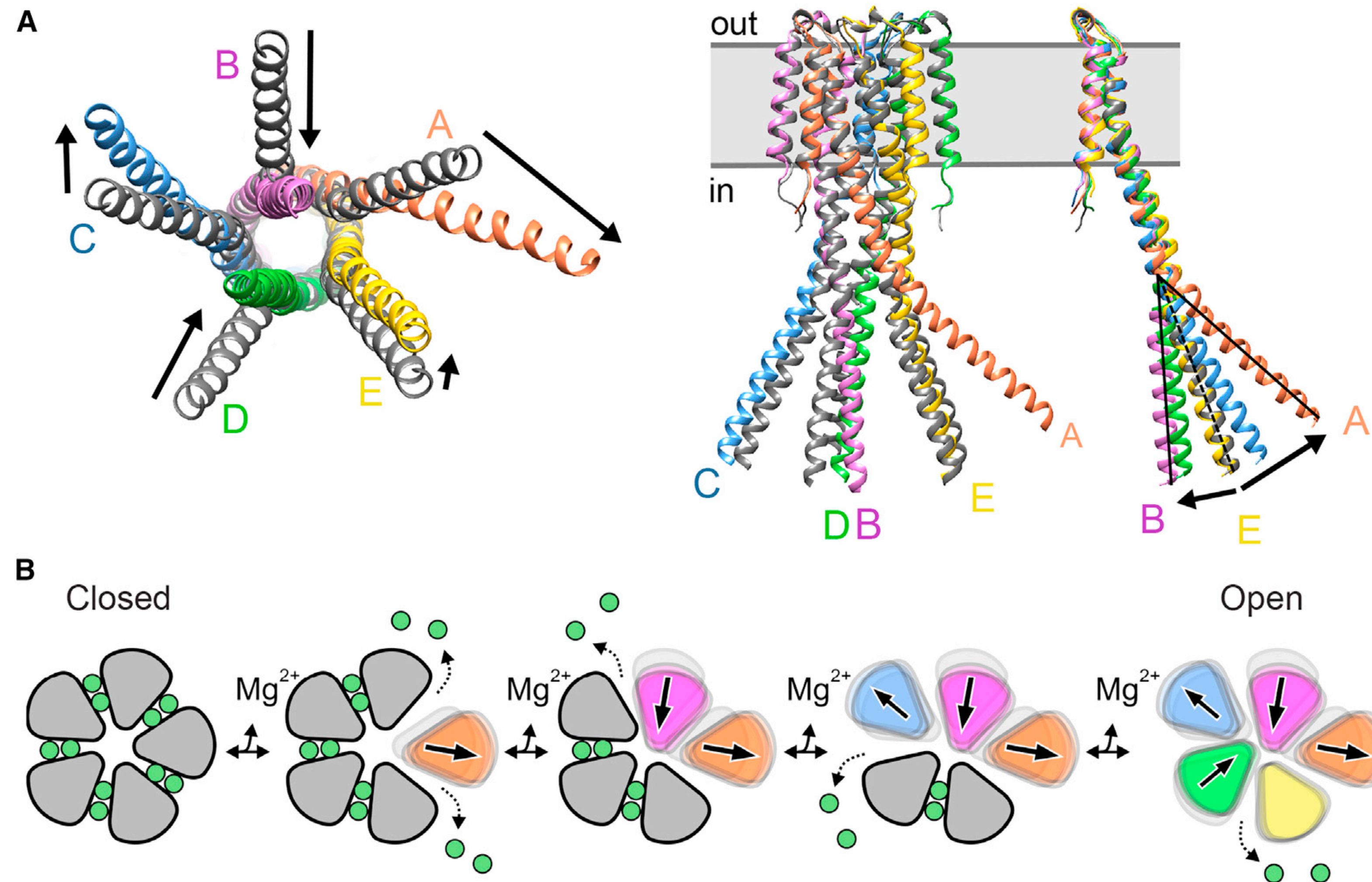
Inria
INVENTEURS DU MONDE NUMÉRIQUE



AGENCE NATIONALE DE LA RECHERCHE
ANR

Examples

Magnesium Channel CorA : Symmetry Break upon Gating

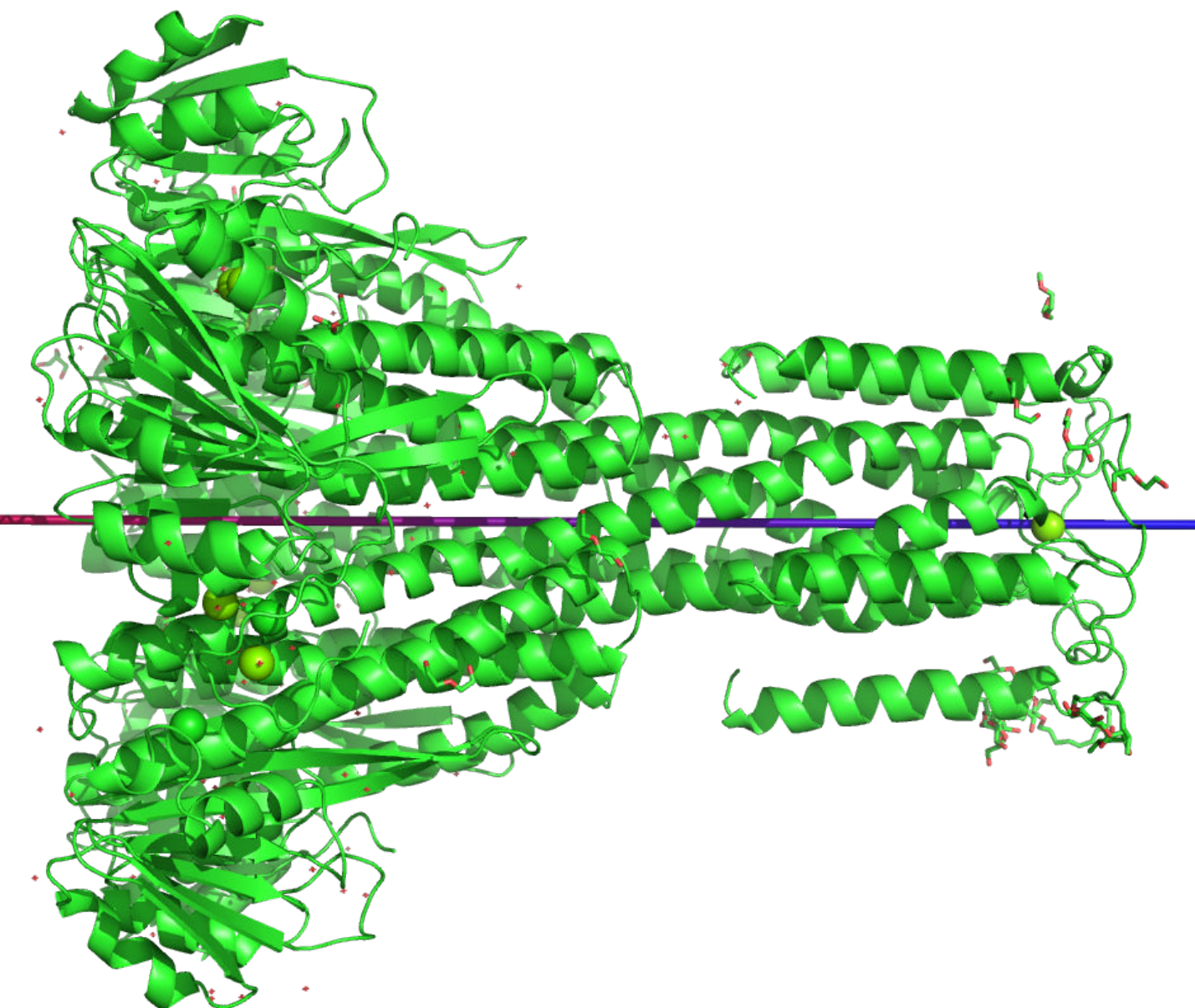
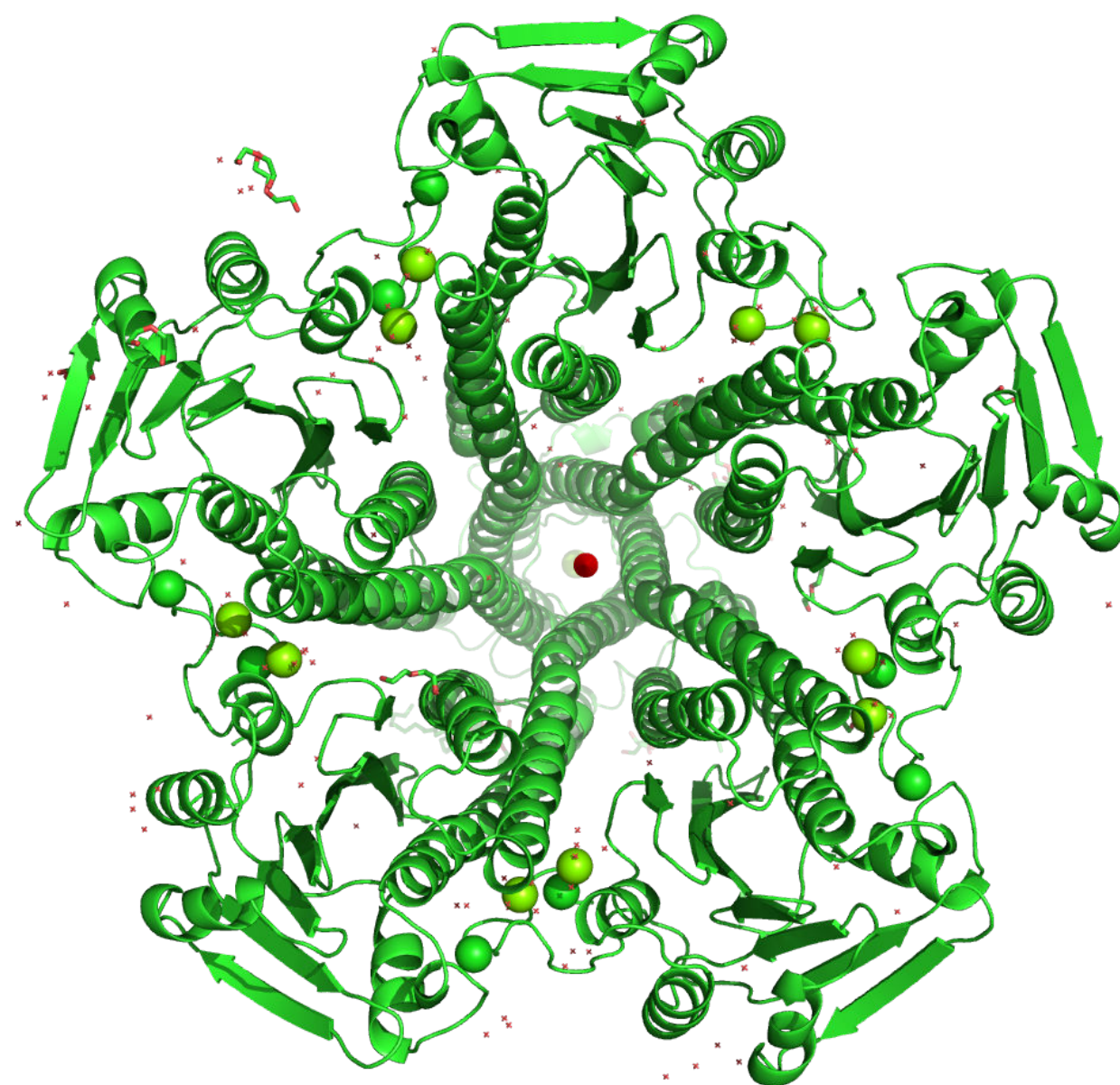


Proposed Schematic Model for CorA Conformational Changes. Accession PDB codes : 3JCF, 3JCG, 3JCH.

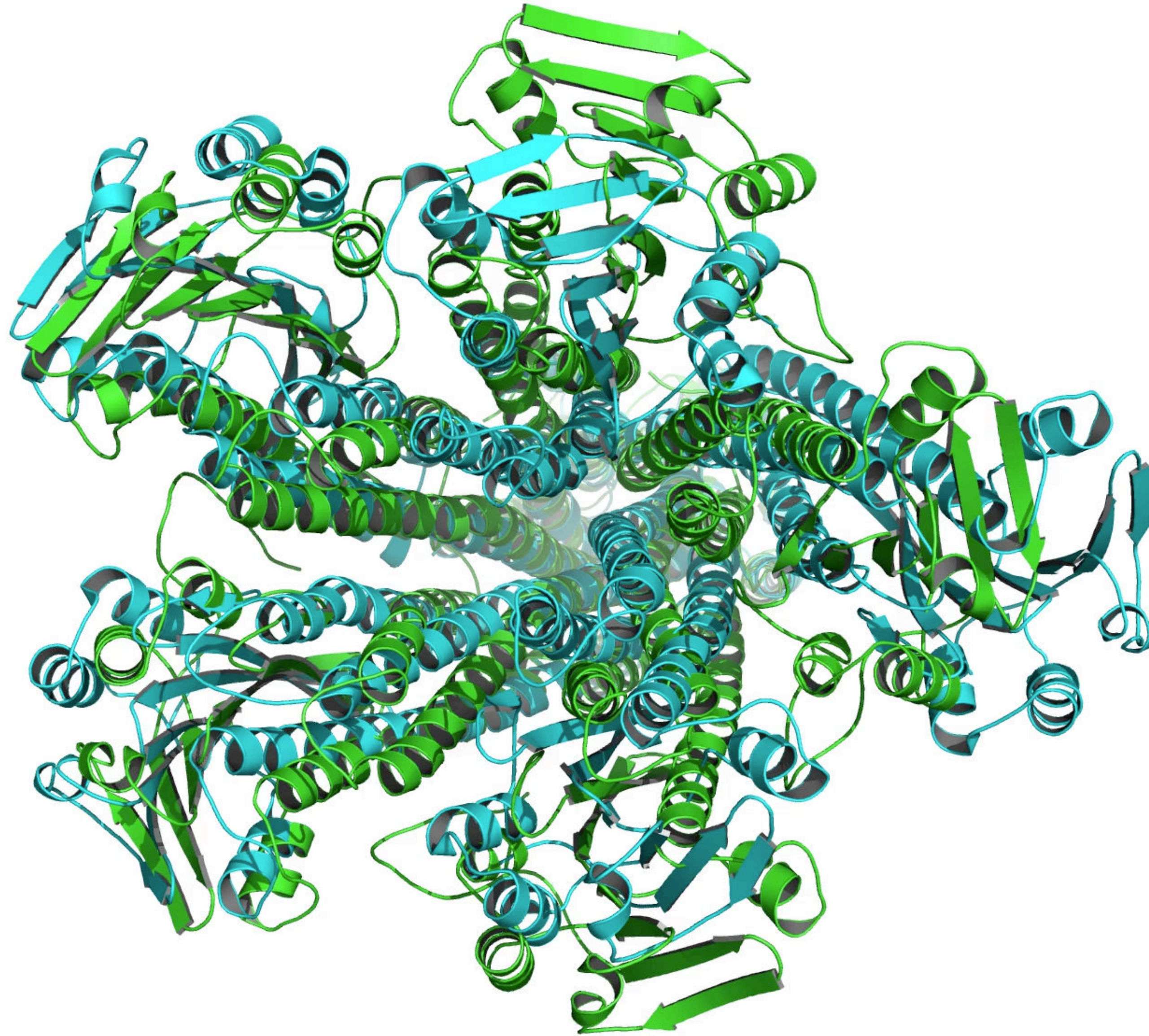
Matthies, Doreen, et al. "Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating." *Cell* 164.4 (2016): 747-756.

Symmetry analysis

```
====Parsing the Command Line====
No symmetry groups specified, will be detected automatically..... :
Command-line arguments..... : ananas corA.pdb -y
====Reading Input file====
Interpreted as a PDB file..... : corA.pdb
Number of chains read..... : 5
Number of atoms read..... : 14331
====Auto-detecting Compatible Groups====
Required similarity for chains..... : 20 %
Detected groups..... : c5 c3 c2
====Detecting Symmetry====
Cutoff for symmetry measure..... : 7 A
Symmetry group : c5
  RMSD RMSD_R RMSD_T RMSD_Z RADGYR ORDER  AXIS X  AXIS Y  AXIS Z  CENTER X  CENTER Y  CENTER Z
  1.272  0.764  0.670  0.766 29.590    5  0.968  0.216  0.124  55.753  18.438  1.706
Average RMSD : 1.281689
```

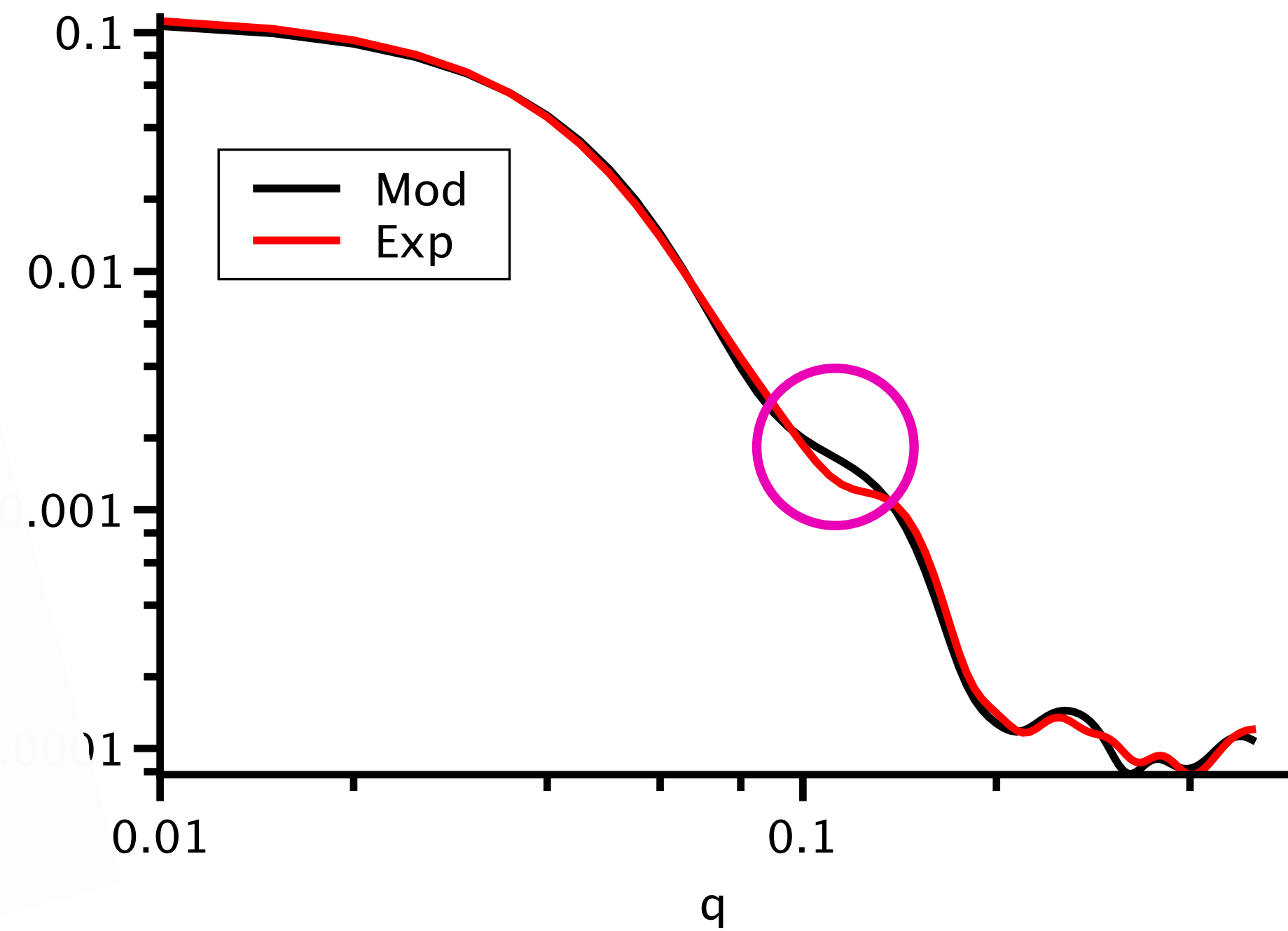
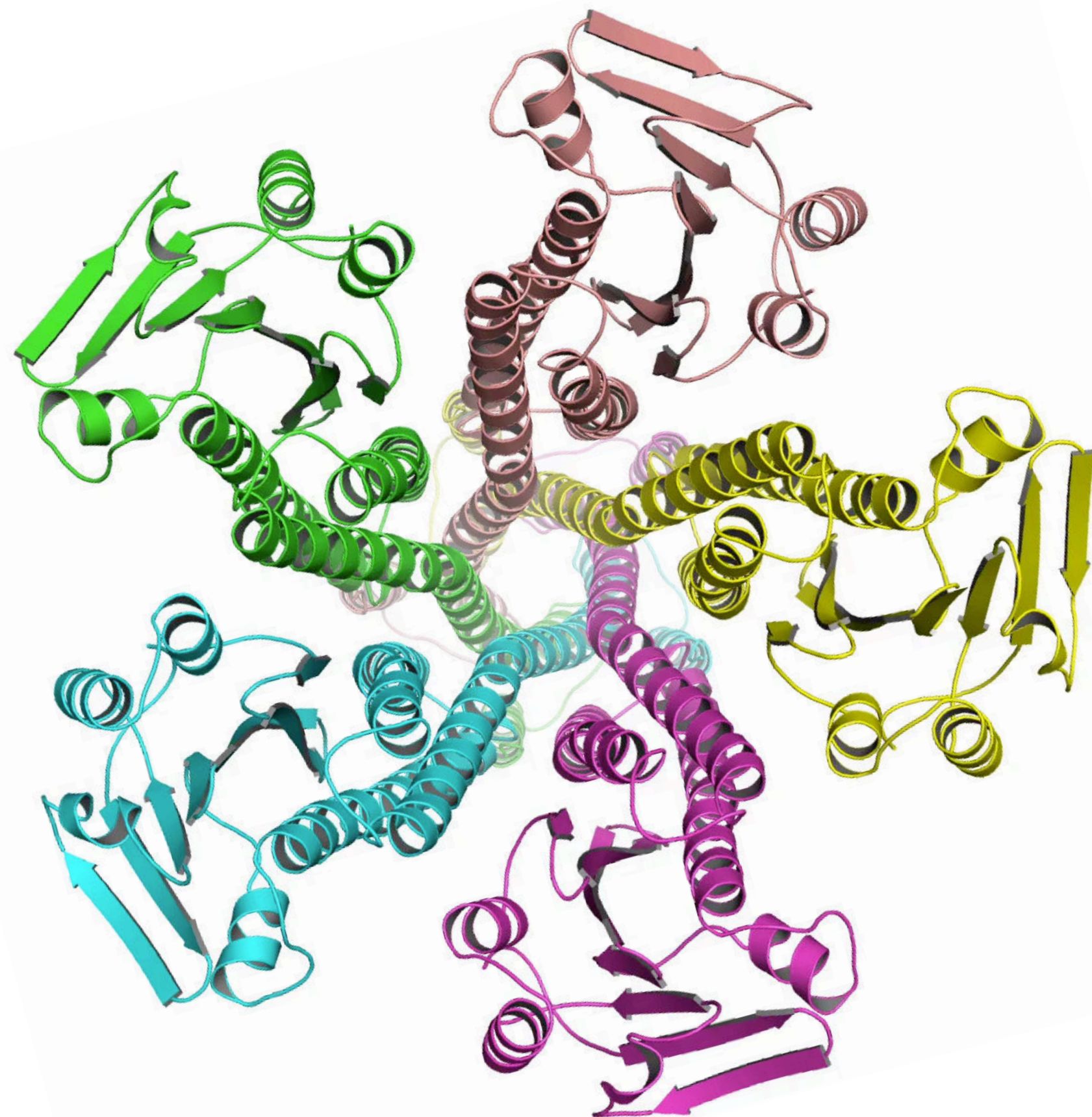


NMA transitions



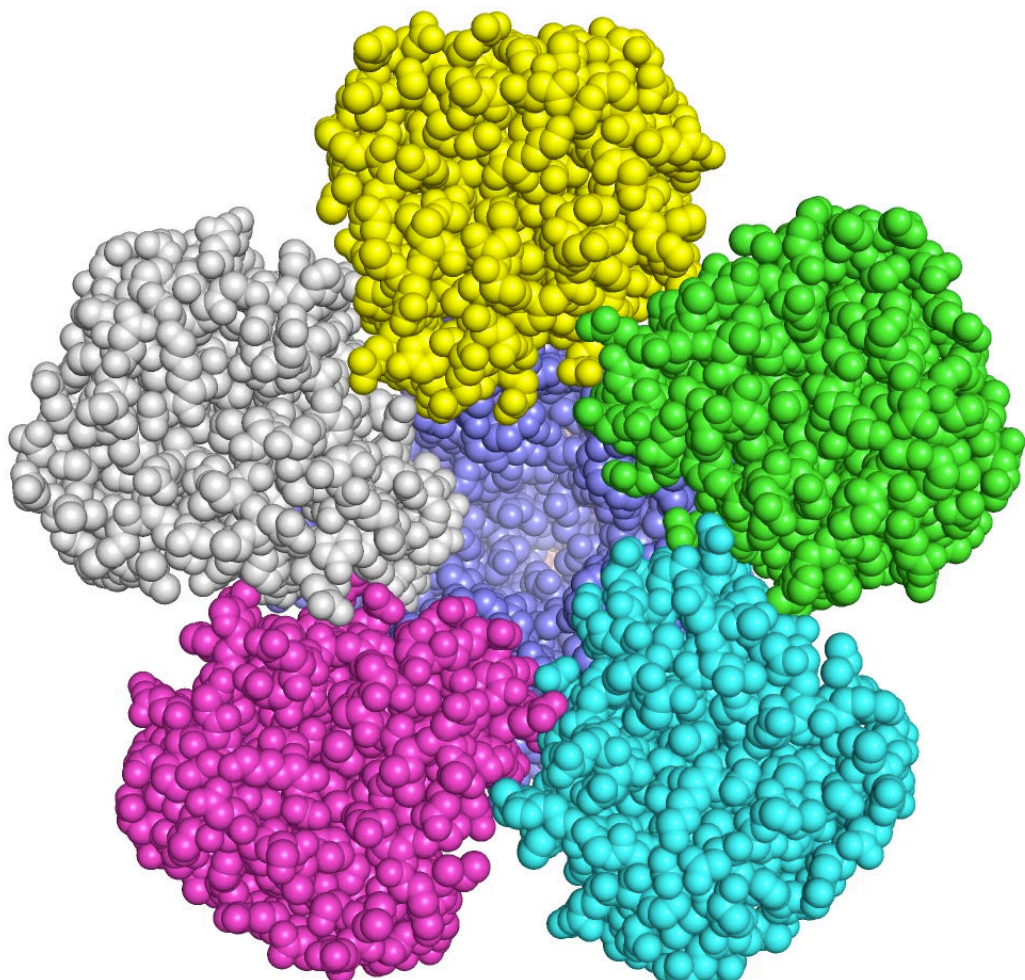
```
*****
*-----*
*-----NOLB : a Non-Linear Rigid Block NMA method-----*
*-----Authors: Alexandre Hoffmann & Sergei Grudinin-----*
*-----Copyright (c): Nano-D team, Inria/CNRS Grenoble, France, 2019.--*
*-----e-mail: sergei.grudinin@inria.fr-----*
*-----http://team.inria.fr/nano-d/software/nolb-normal-modes/-----*
*-----*
*****
Command-line arguments.....:
NOLB 3jcf.pdb 3jcg.pdb
```


Flexible fitting

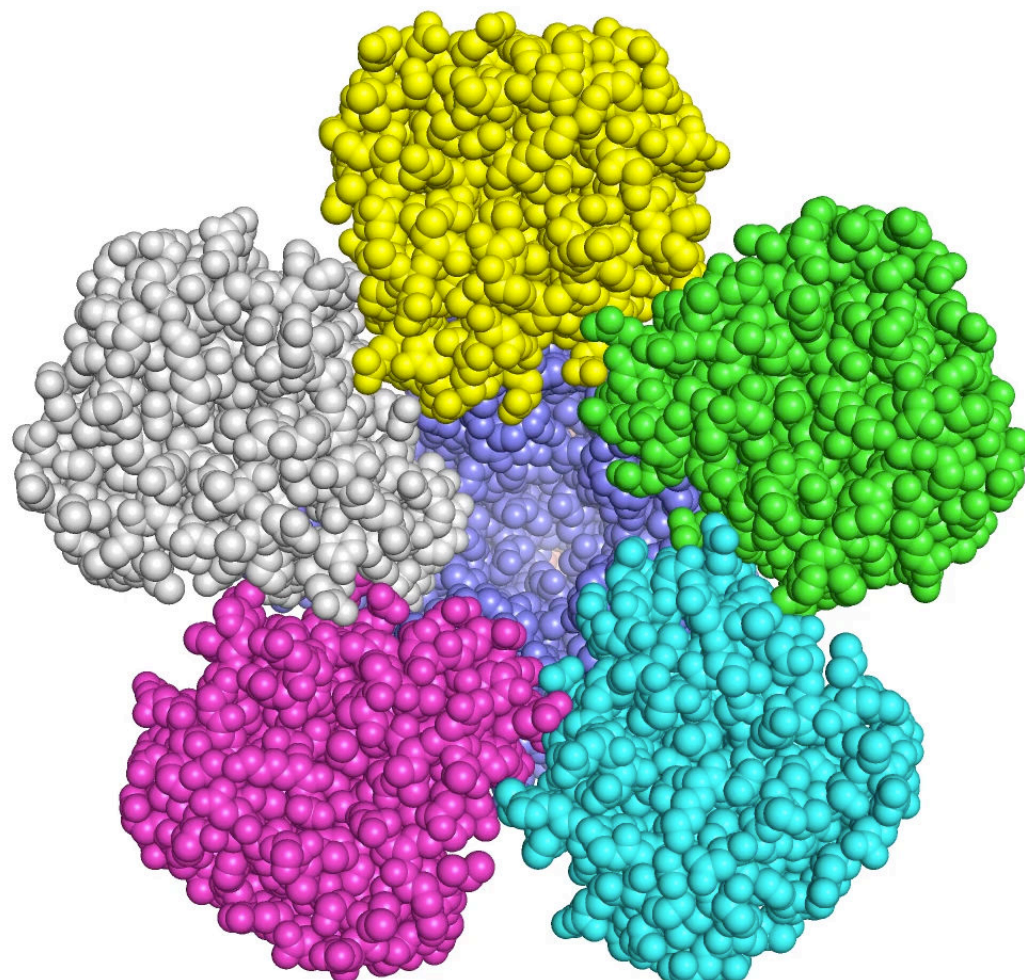


Pepsi-SAXS 3jcf.pdb 3jcg.out --opt --modes 5

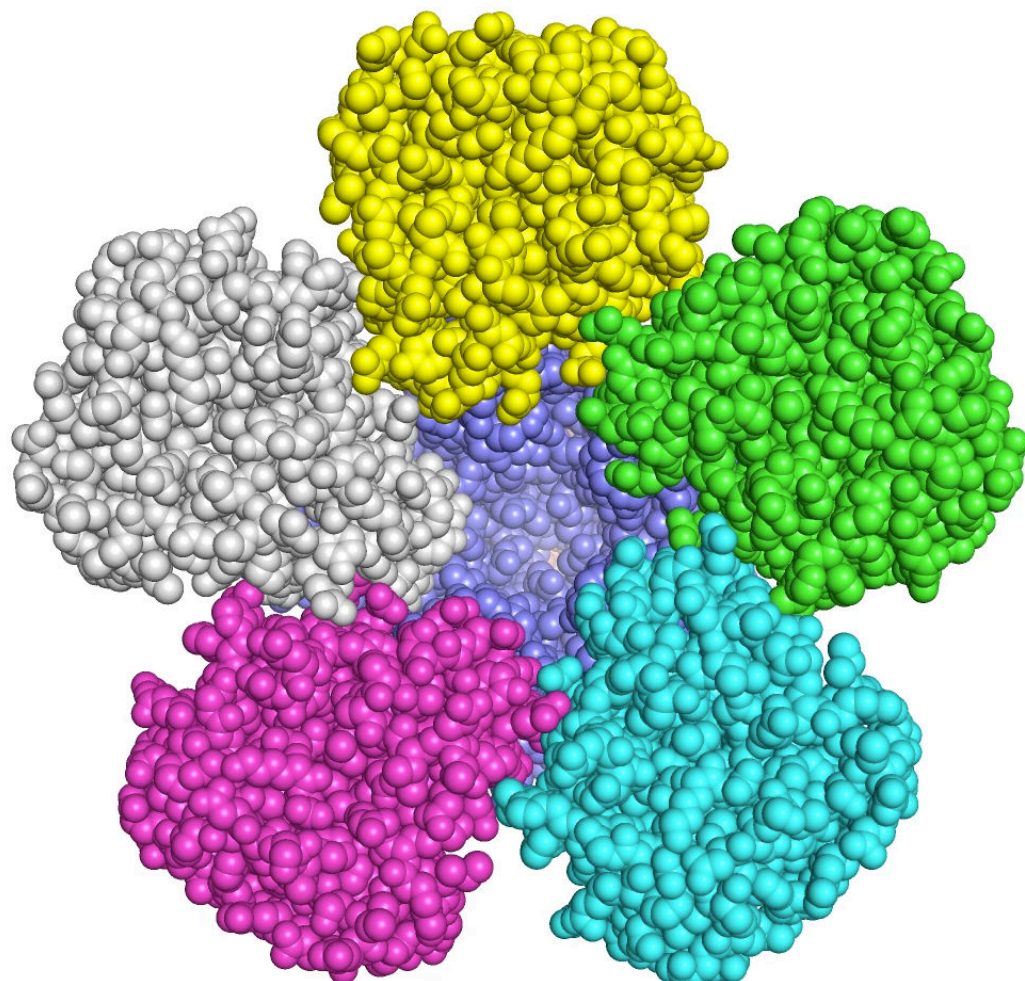
Domain analysis



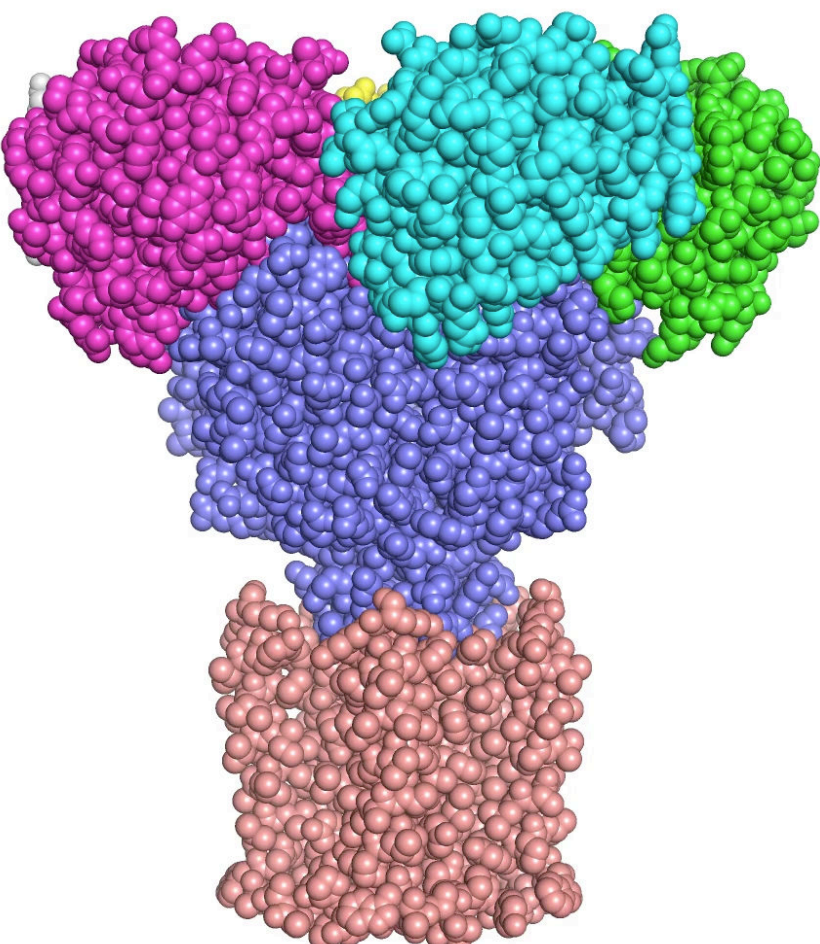
mode 1



mode 2

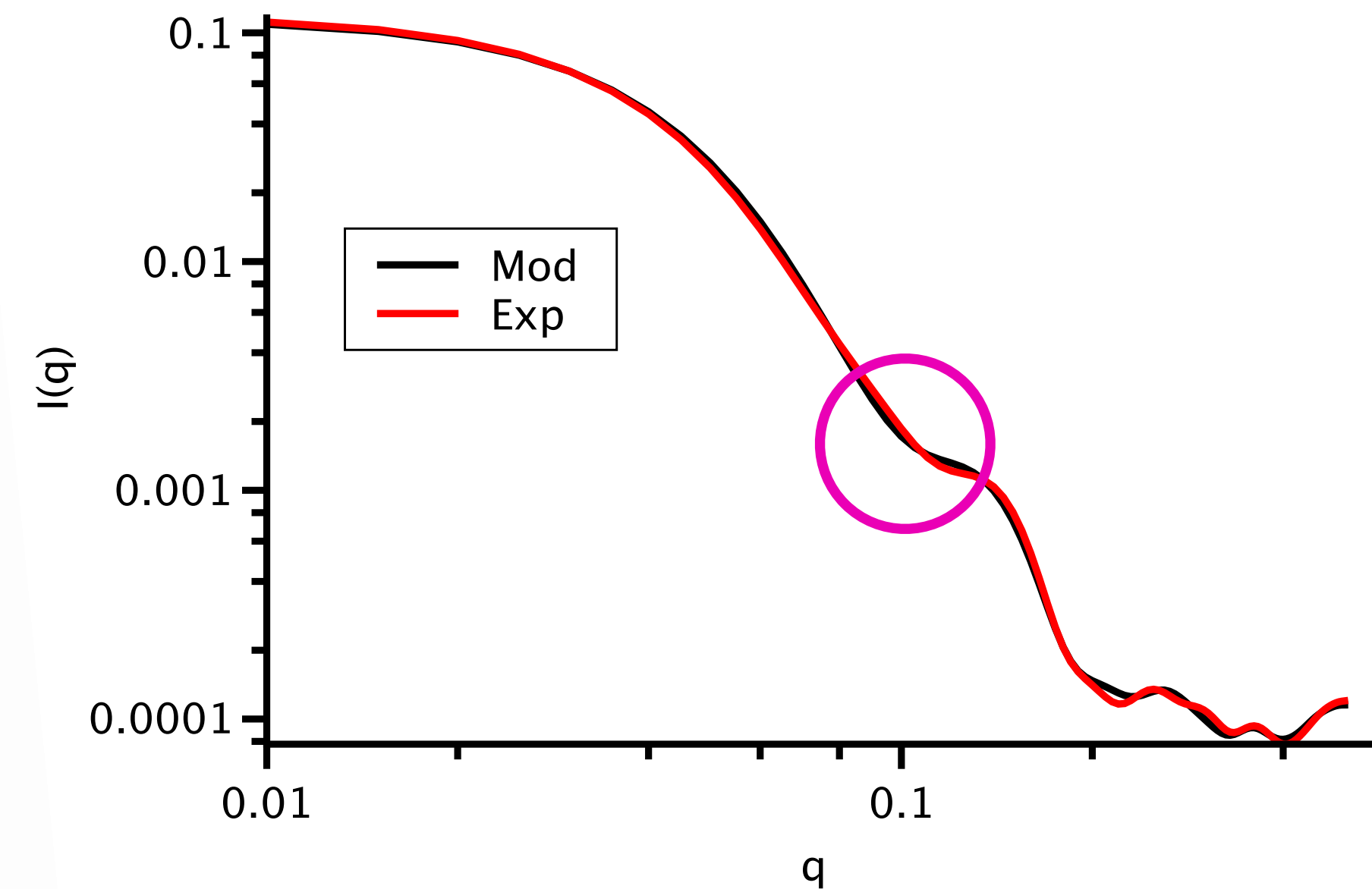
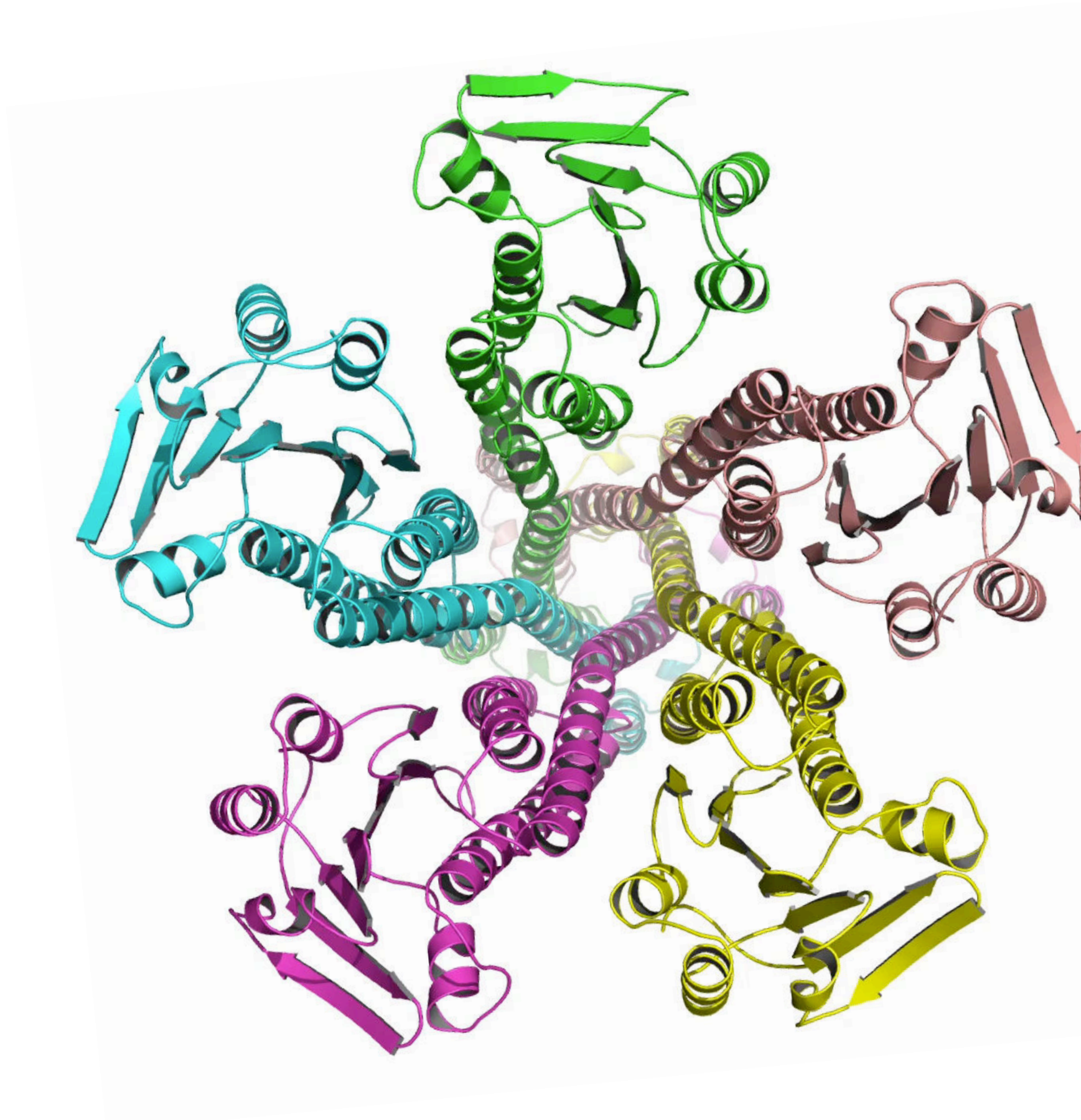


mode 3



```
NOLB 3jcf-noions-aligned.pdb --nDomains 7
*****
*-----*
*-----NOLB : a Non-Linear Rigid Block NMA method-----*
*-----Authors: Alexandre Hoffmann & Sergei Grudinin-----*
*-----Copyright (c): Nano-D team, Inria/CNRS Grenoble, France, 2019.--*
*-----e-mail: sergei.grudinin@inria.fr -----*
*-----http://team.inria.fr/nano-d/software/nolb-normal-modes/ ----*
*-----*
*****
Command-line arguments..... : NOLB
3jcf-noions-aligned.pdb --nDomains 7
=====Writing rigid domains=====
Rigid Block 1..... :
E19:112 E125:177 E230:257 E259
Rigid Block 2..... :
D19:112 D125:177 D230:257 D259:260
Rigid Block 3..... : B260
C19:112 C125:177 C230:257 C259:260
Rigid Block 4..... : E260
A19:112 A125:177 A230:257 A259:260
Rigid Block 5..... :
A295:347 C295:347 B295:347 D295:347 E295:347
Rigid Block 6..... :
B19:112 B125:177 B219:259 B261:269
Rigid Block 7..... :
A113:124 A178:229 A258 A261:294 E113:124 E178:229 E258 E261:294 D113:124
D178:229 D258 D261:294 C113:124 C178:229 C258 C261:294 B113:124 B178:218
B270:294
```


Rigid body fitting



Pepsi-SAXS 3jcf.pdb 3jcg.out --opt --modes 5 --blocks domainFile.txt

Thank you!

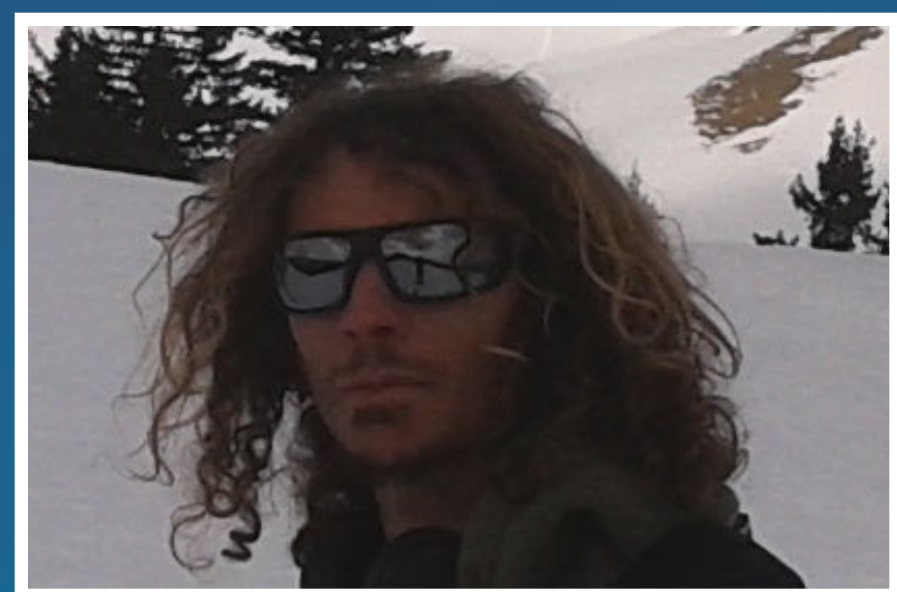
Guillaume Pages



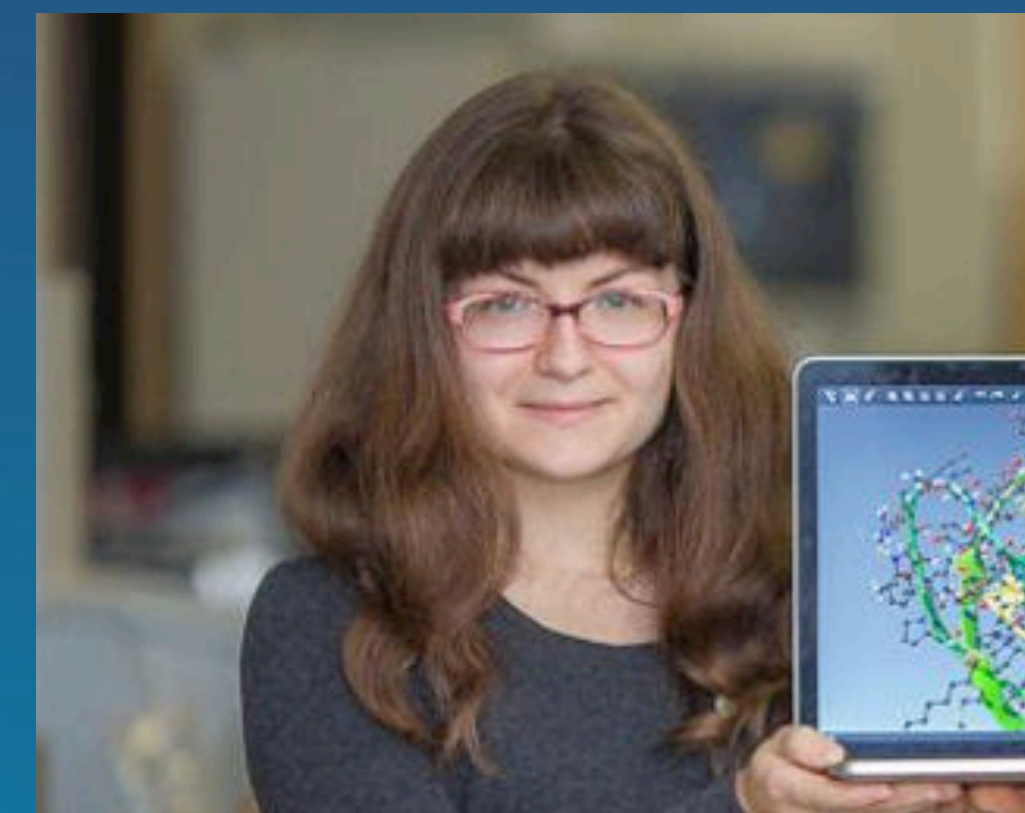
Mikhail Karasikov



Alexandre Hoffmann



Maria Kadukova



<https://team.inria.fr/nano-d>

Agence Nationale de la Recherche