



Introduction to Structural Bioinformatics : proteins and DNA

Gwenaëlle André-Leroux

► To cite this version:

Gwenaëlle André-Leroux. Introduction to Structural Bioinformatics : proteins and DNA. 3rd cycle. Workshop in Structural bioinformatics (Introduction to Structural bioinformatics), 2015, 60 diapos. hal-02795917

HAL Id: hal-02795917

<https://hal.inrae.fr/hal-02795917>

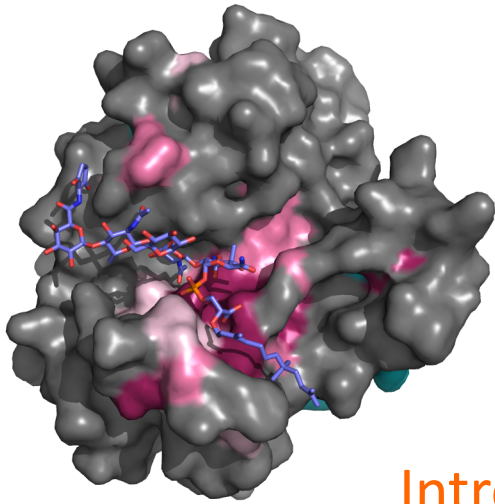
Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



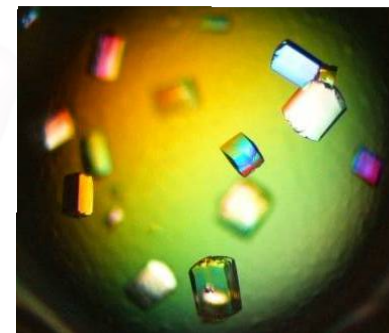
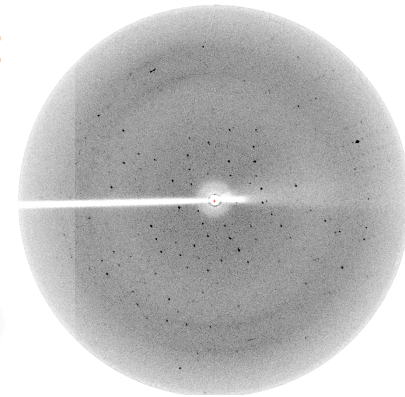
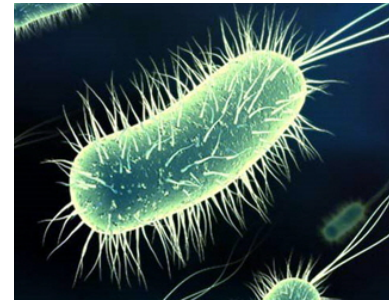
Introduction to Structural Bioinformatics: Proteins and DNA

>3CM3:A|PDBID|CHAIN|SEQUENCE
 GPEIRMDKKSLYKYLLLRSTGDMHKAKSP
 TIMTRVTNNVYLGNYKNAMDAPSSEVKF
 KYVLNLTMDKYTL PNSNIIHIPLVDDTTT
 DISKYFDDVTAFLSKCDQRNEPVLVHSAA
 GVNRS GAMILAYLSKNKESLPM



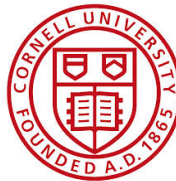
Gwénaëlle André-Leroux
 MaIAGE

Montevideo december 2015



Curriculum Vitae

- ★ 1994 – 1998: PhD in molecular modeling, Inra Nantes, France
- ★ 1999 – 2000: Visiting Assistant Professor, Cornell University, NY – USA
- ★ 1998 – 2002: Research assistant, Inra Nantes, BIA
- ★ 2002 – 2004: Senior researcher, Inra Nantes, BIA
- ★ 2004 – 2013: Senior researcher at Institut Pasteur, Microbiologie Structurale
- ★ 2013 – : Senior researcher, Inra Jouy-en-Josas, MaIAGE



Structural biologist and computational scientist

Structural bioinformatics

Molecular modeling is dedicated to

- ★ Visualize
- ★ Understand
- ★ Predict

Structural bioinformatics

- ★ Integrate metaOmics data
- ★ Develop bioinformatic tools dedicated to analyze 3D fold in metaOmics



1. Visualize 2. Understand 3. Predict 4. Integrate MetaOmics

« The theoretical in silico prediction of protein structures and dynamics is essential for understanding the molecular basis of drug action, metabolic and signaling pathways in living cells, and designing new technologies in the life science and material sciences. » A. Kolinski

Molecular modeling

Introduction, study, expecting in this course

Introduction to proteins

Tool box: softwares, websites, tricks to study proteins, DNA, RNA and ligands.

Hands on tutorials

1. Structures

1.1. Analyse of proteins. Get familiar with PyMOL

1.2. Homology modeling: strategy and use of I-tasser, Phyre2, modeler

1.3. Mutations *in silico*

2. Oligomerisation and motion analysis

2.1. Analyse des oligomers with PyMol

2.2. Modeling of oligomers

2.3. Motion and normal mode

3. Molecular mechanics and docking

3.1. Binding site identification and characterization

3.2. Analysis of ligands

3.3. Docking: rigide, flexible

Coffee breaks



Conclusions, questions, comments

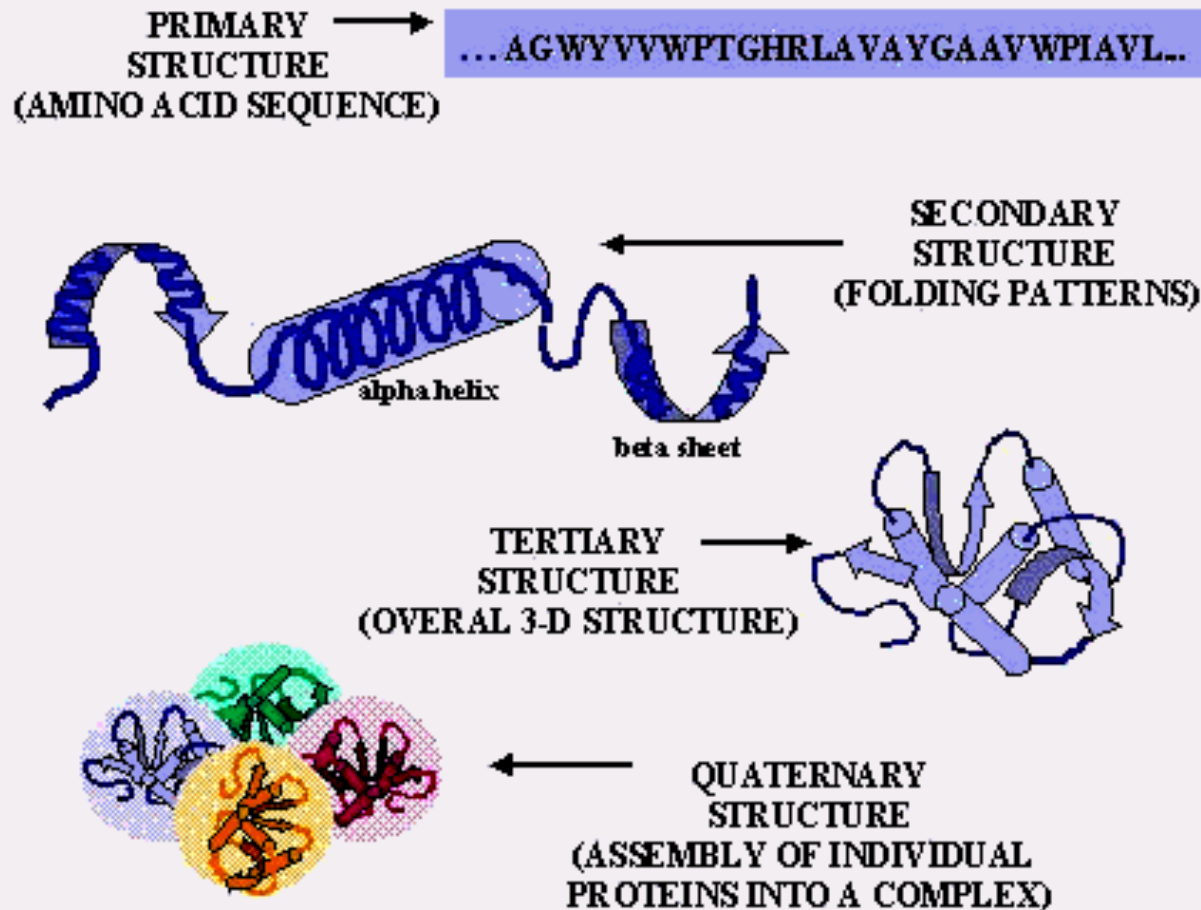
Introduction : proteins

Macromolecule in cell, responsible for metabolism, energy cycle TCA, signaling...

linear repetition of atomic bricks: amino acid residues linked by peptidic bond

Fold (3D space arrangement) is specific and guarantees its function.

From sequence to quaternary arrangement, 4 steps in protein folding



Proteins: primary structure

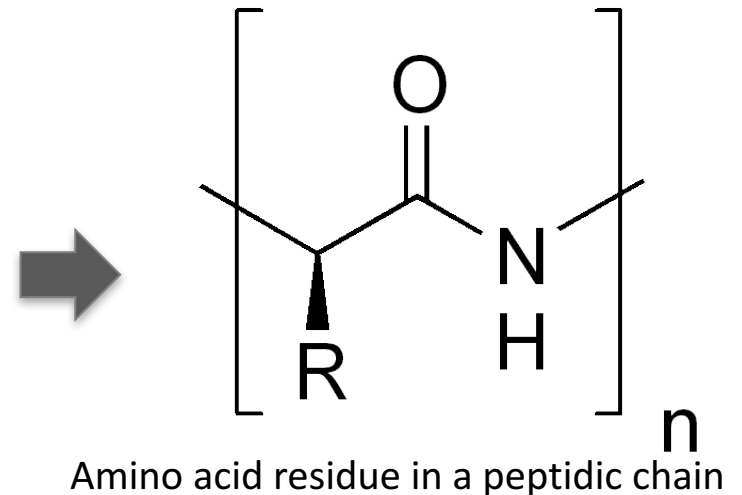
It refers to the sequence of amino acids present in the polypeptide chain

Amino acids -called residues- are covalently linked by a peptide bonds

The first is called amino terminal end or N-terminus

The last is called amino carboxy terminal or C-terminus

>3CM3:A|PDBID|CHAIN|SEQUENCE
GPEIRMDKKSLYKYLLLRSTGDMHKAKSPTIMTRV
TNNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYT
LPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQ
RNEPVLVHSAAGVNRSGAMILAYLMSKNKESLPM



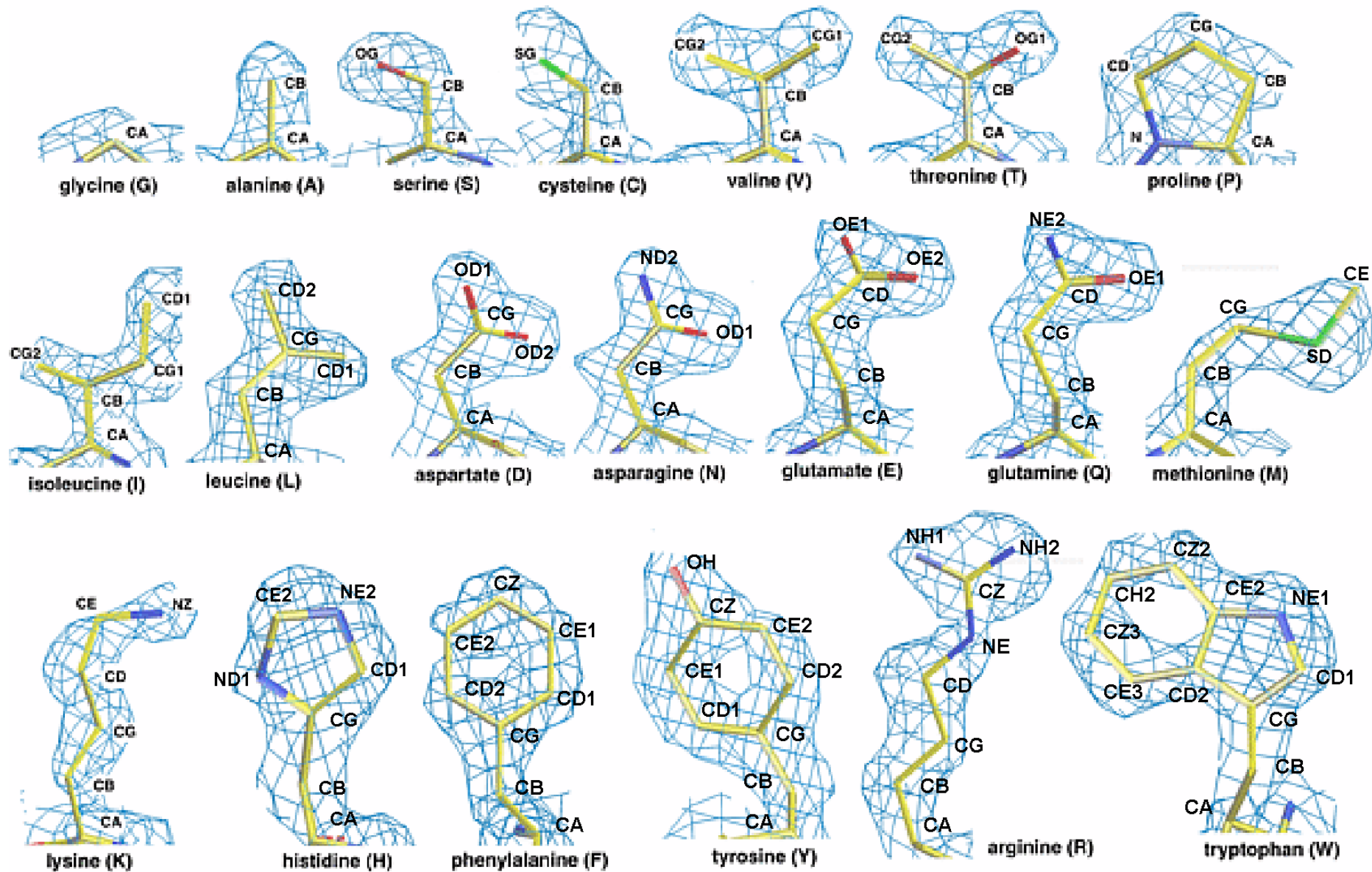
Each letter refers for an aa residue:

A for Alanine

C for Cysteine ...

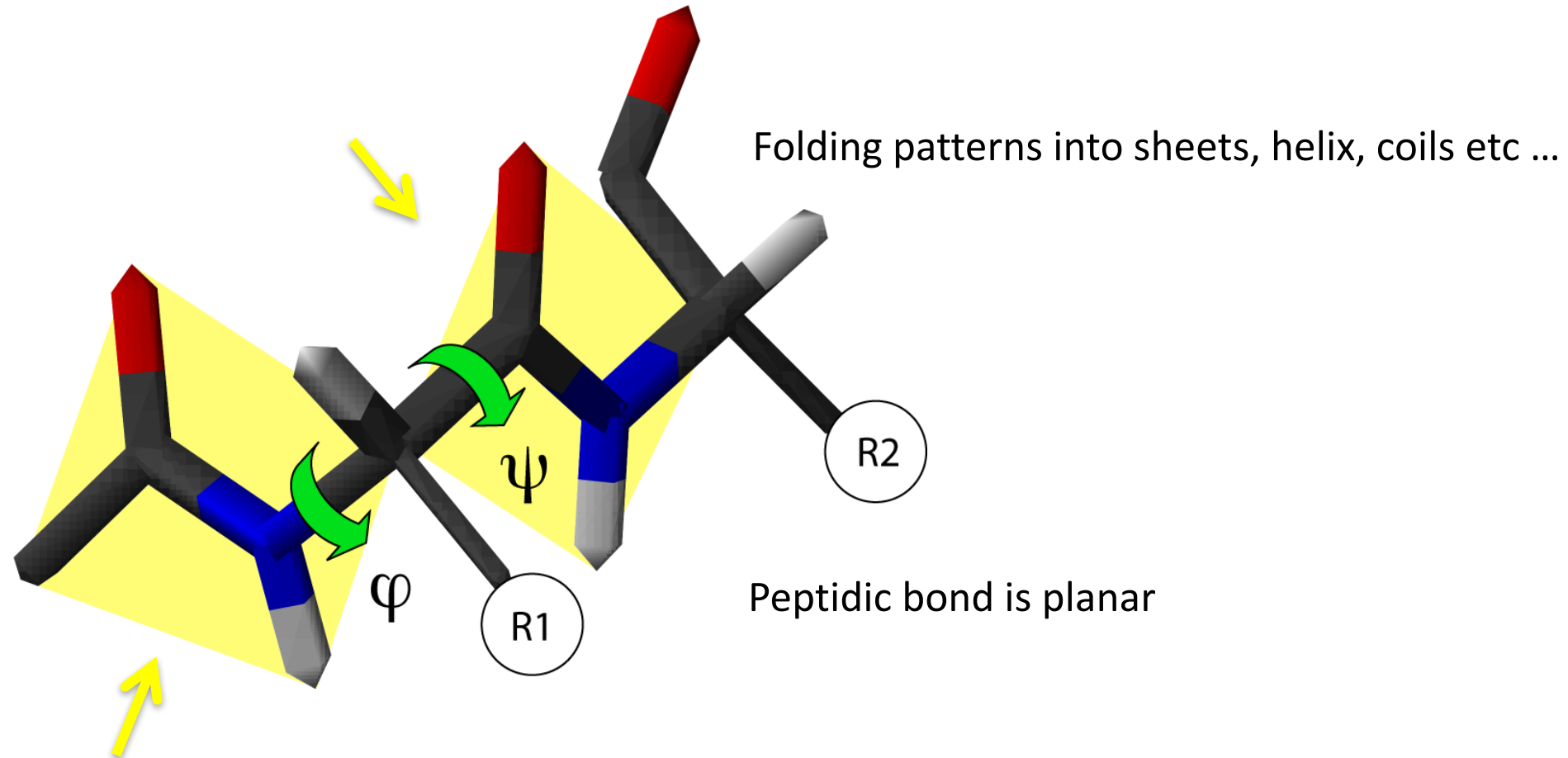
R refers to the side chain of the amino acid

Residues: small to large, polar to apolar



Proteins: secondary structure

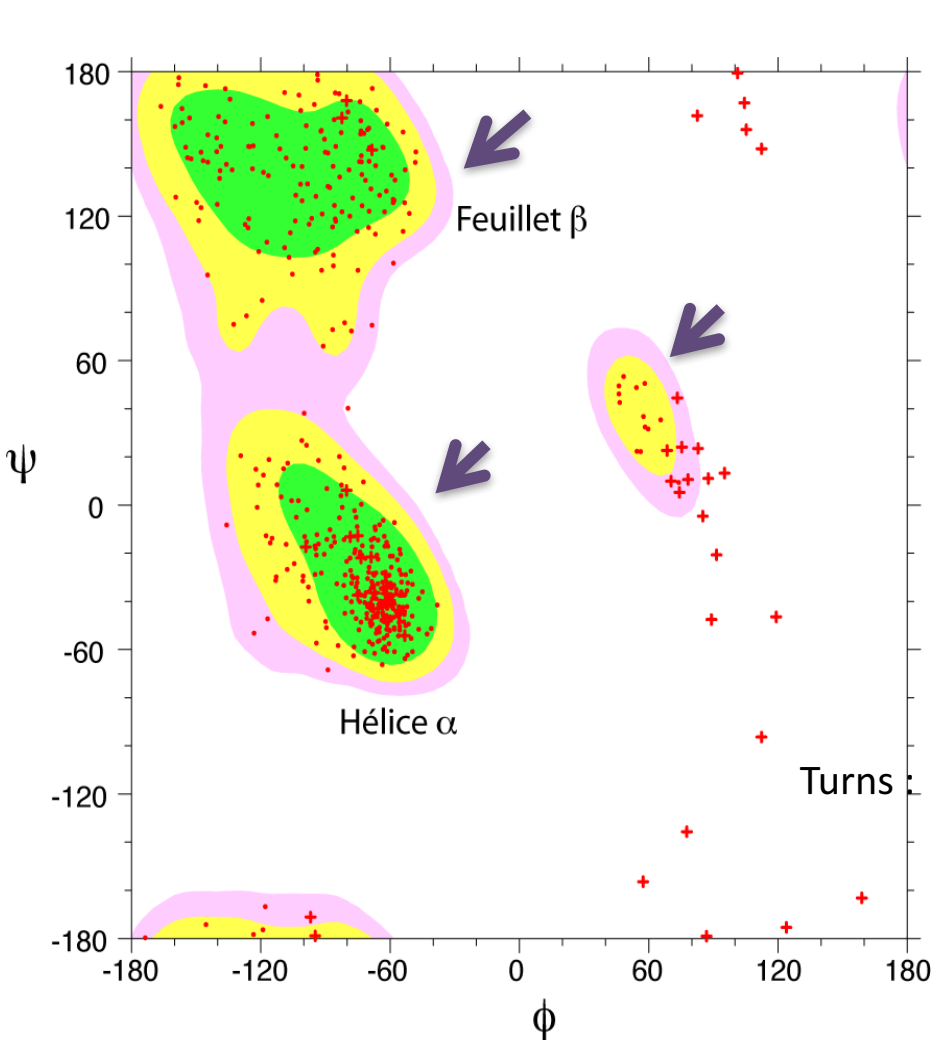
The peptidic chain shows dihedral angles



Proteins: secondary structure

Folding patterns into sheets, helix, coils etc ...

Ramachandran diagram calculates the possible combinaisons of dihedrals (ϕ, ψ)



3 Regions energetically allowed

75% is energetically not favorable

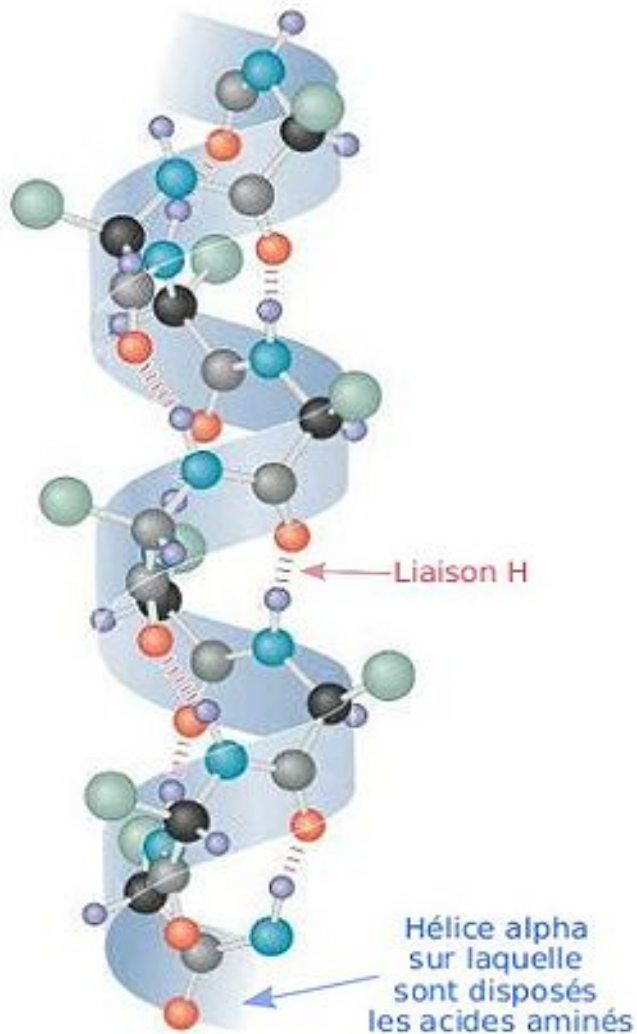
Hélice α : helical periodic structure.

Strand β : flat periodic structure flat.
Assemble into sheets

Turns :

Proteins: secondary structure

Folding patterns into sheets, helix, coils etc ...

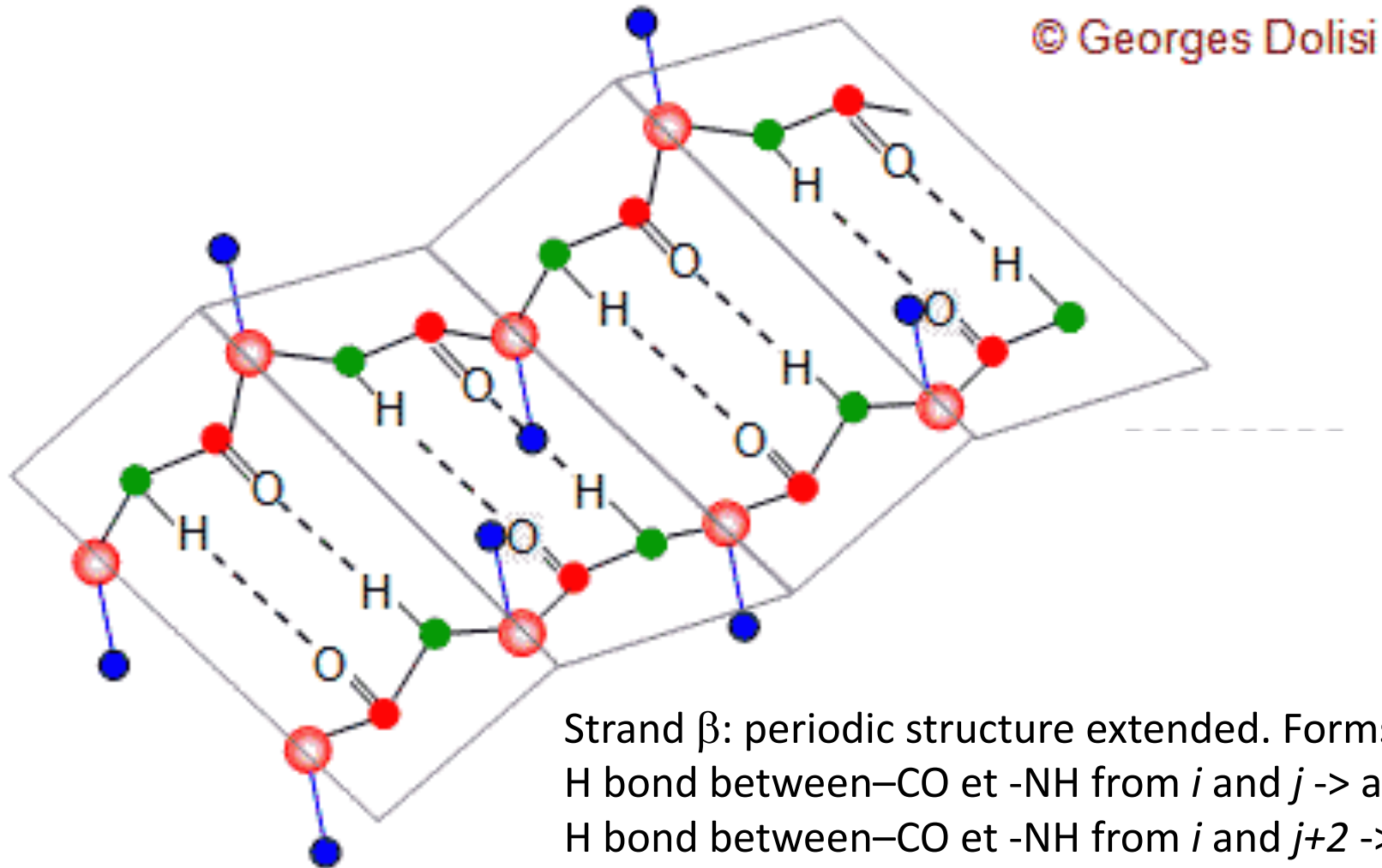


α Helix: periodic structure.

H bond between $-\text{CO}$ from i and $-\text{NH}$ from $i+4$

Proteins: secondary structure

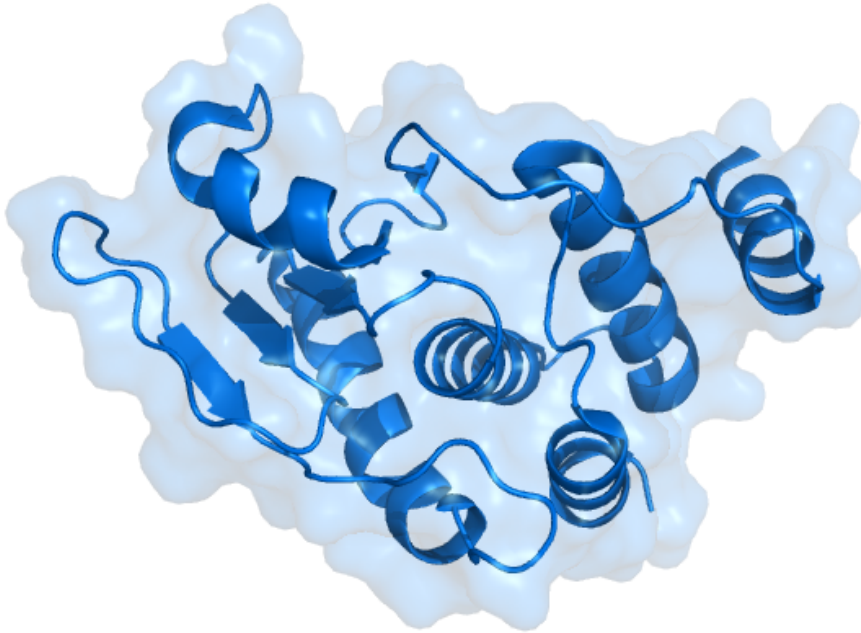
Folding patterns into sheets, helix, coils etc ...



Proteins: tertiary structure

Overall 3D structure with a complete folding in 3D space.

3D structure correlates to function.

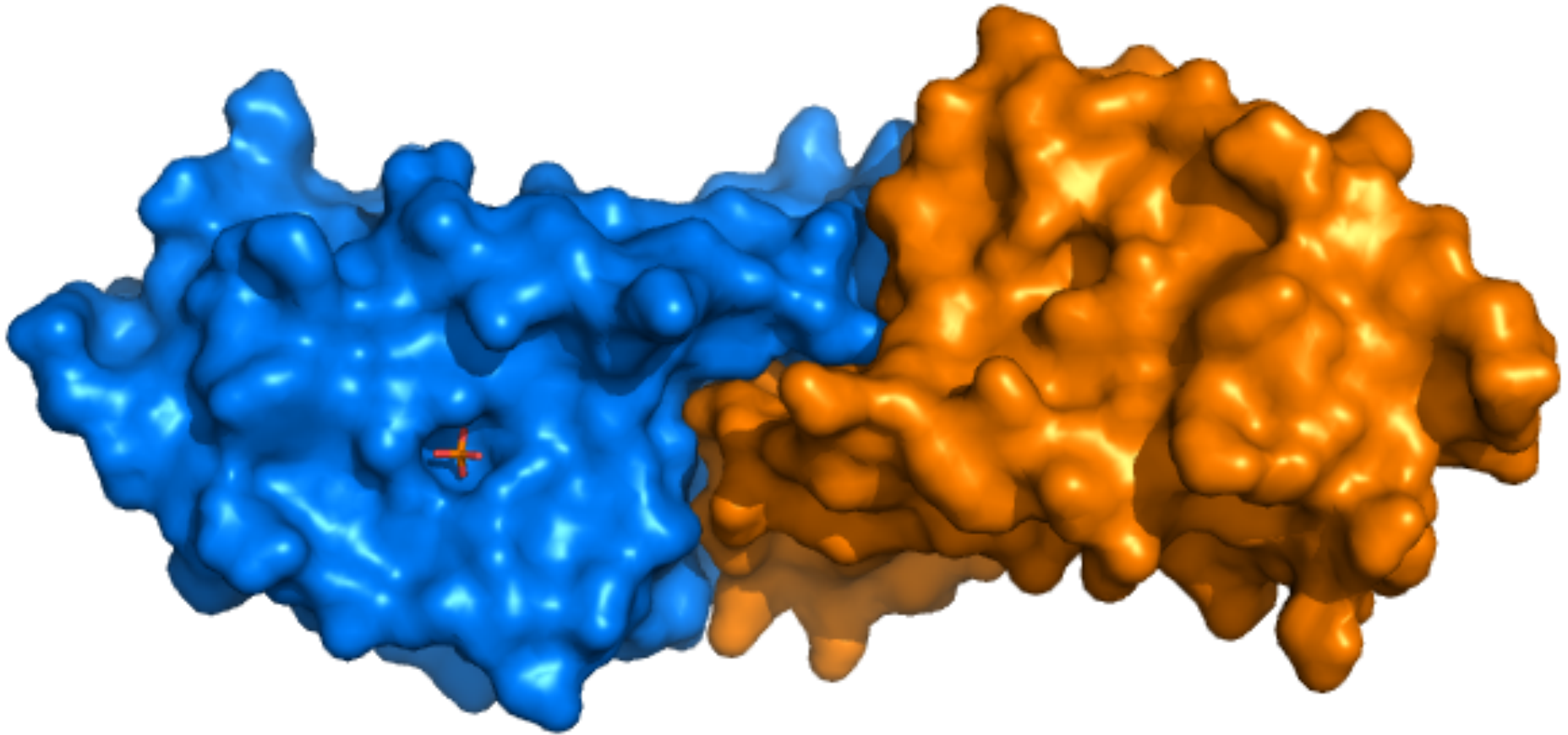


Several interactions guarantee its fold and its stability

- Covalent bonds (disulfide bridges between cysteine residues)
- Electrostatic bonds (ionic bonds, H-bonds)
- van der Waals interactions
- solvent, ions, lipids interactions

Proteins: quaternary structure

Assembly of individual proteins (protomers) into a functional complex:
Either homo-oligomers or hetero-oligomers

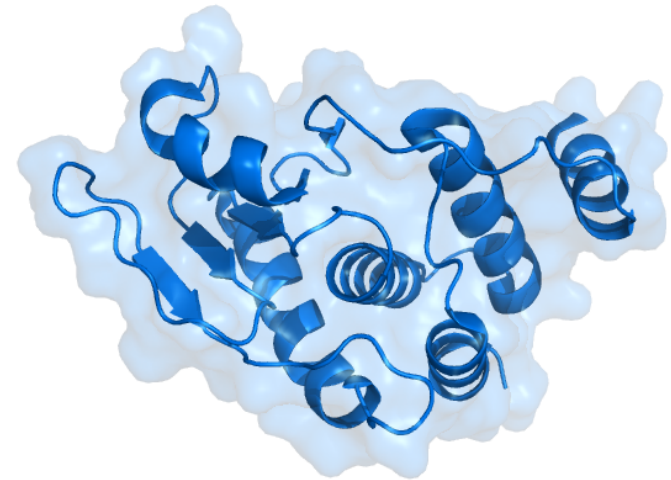


Protein Structure

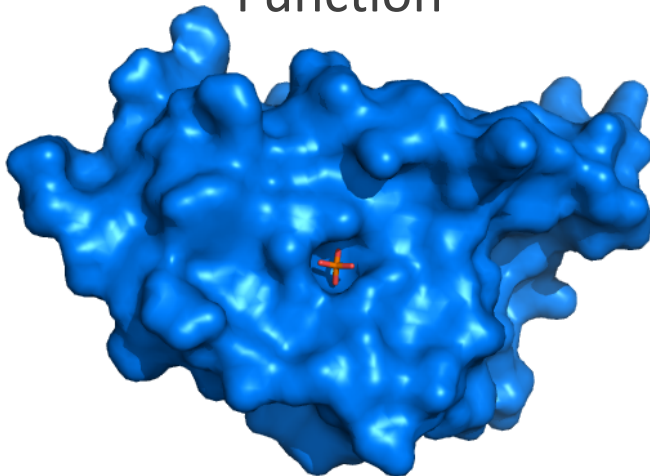
1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE  
GPEIRMDKKSPLYKYLRLRSTGDMHKAKSPTIMTRVT  
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL  
PNSNINIIHIPLVDDTTTDDISKYFDDVTAFLSKCDQR  
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY  
FLYVYHSMRDLRGAFVENPSFKRQIEKYVIDKN
```

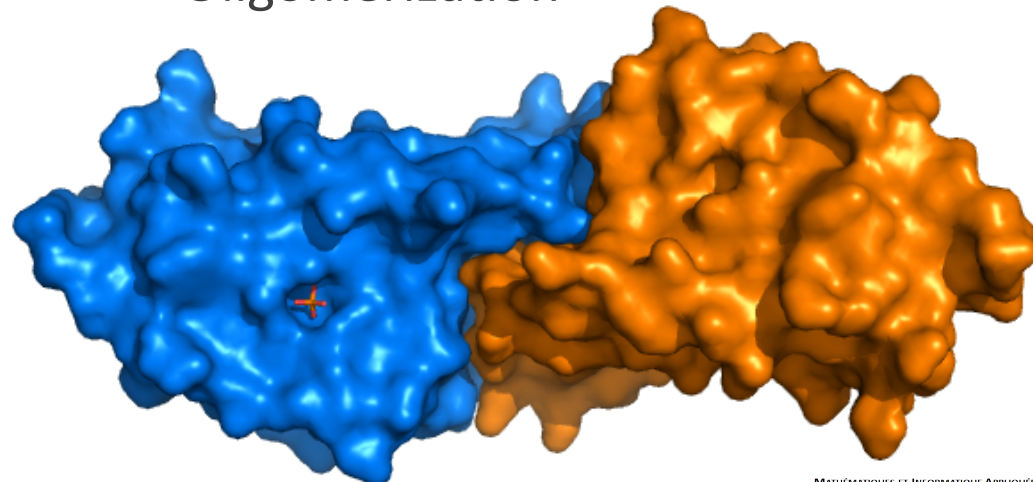
3D Fold



Function



Oligomerization

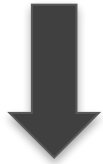
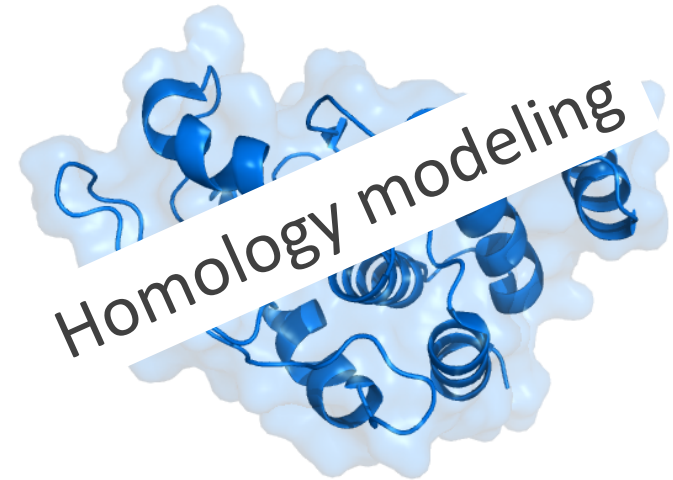


Protein Structure

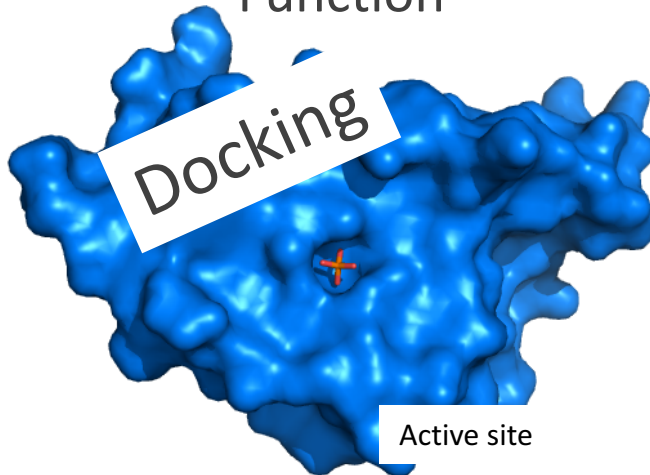
1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE  
GPEIRMDKKSLEYKYLRLSTGDMHKAKSPTIMTRVT  
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL  
PNSNINIIHIPLVDDTTTDDISKYFDDVTAFLSKCDQR  
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY  
FLYVYHSMRDLRGAFVENPSFKRQIEKYVIDKN
```

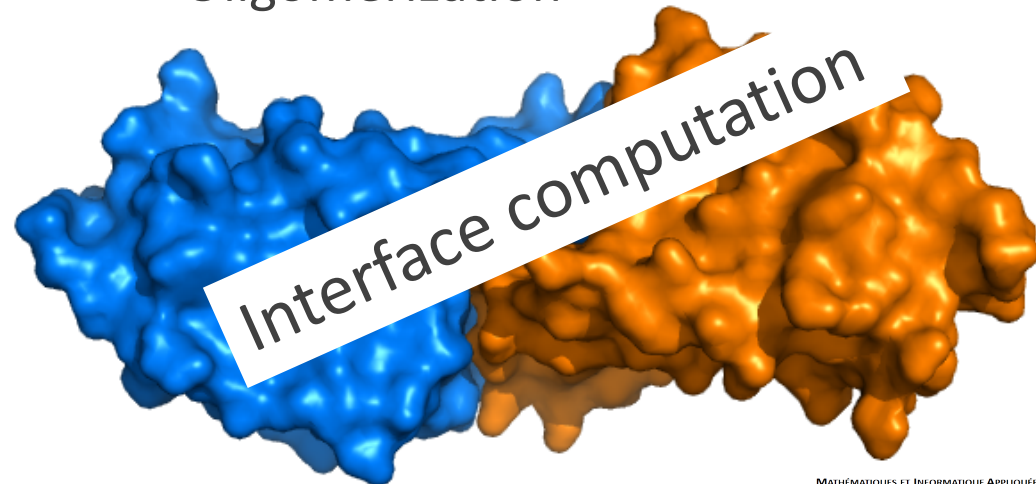
3D Fold



Function



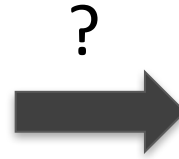
Oligomerization



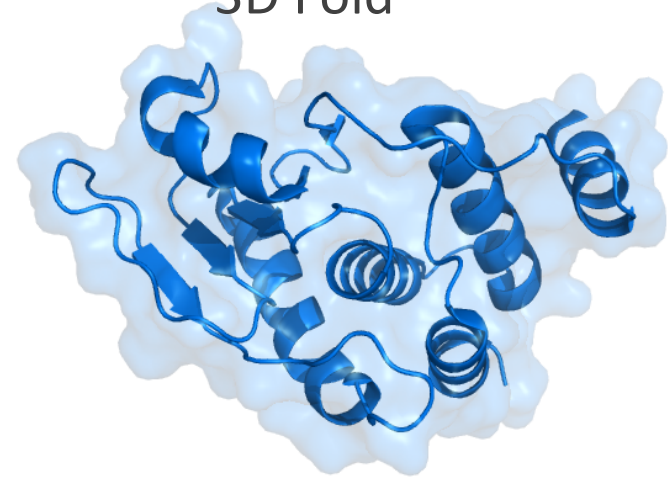
Protein Structure

1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE  
GPEIRMDKKSLYKYLLLRSTGDMHKAKSPTIMTRVT  
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL  
PNSNINIIHIPLVDDTTTDDISKYFDDVTAFLSKCDQR  
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY  
FLYVYHSMRDLRGAFVENPSFKRQIEKYVIDKN
```



3D Fold



Why?

How?

Biological consistency?

Protein Structure

double click on PyMol

http://www.pymolwiki.org/index.php/Main_Page

.....



Show

Label

A complex network graph visualization on a black background. The graph consists of numerous interconnected nodes and edges. The edges are primarily green, with some blue and red segments interspersed. The nodes are represented by small red plus signs (+) scattered throughout the network. The overall structure is dense and irregular, with many branches and loops. The green lines form the main backbone of the network, while the blue and red lines highlight specific paths or clusters. The red plus signs are distributed across the entire area, often marking the vertices of the network.

```

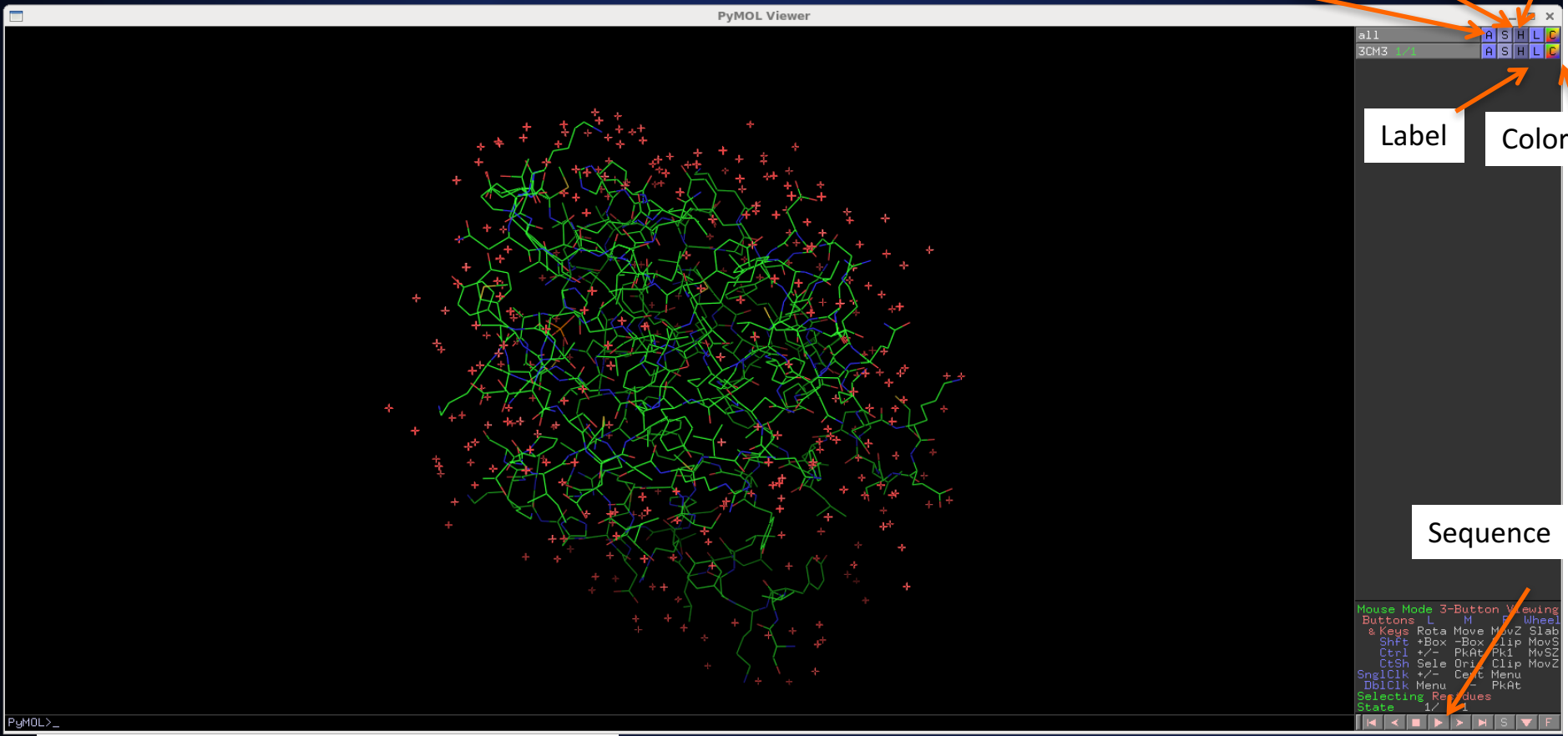
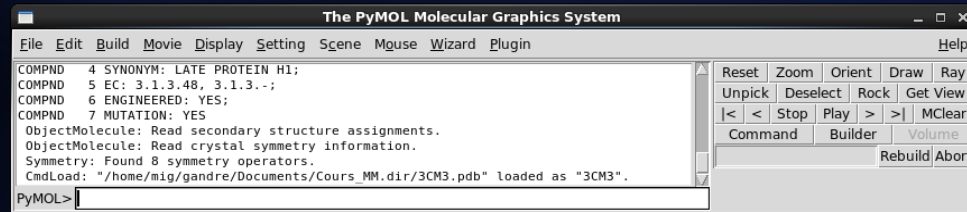
Mouse Mode 3-Button Viewing
Buttons      L      M      R      Wheel
+ Keys
  Shift      +Box -Box Clip MovS Slab
  Ctrl       +/   PKat PK1  MovS
  CtSh       Sele Orig Clip MovZ
SnglClk     +/_   Cent  Menu
DblClk      Menu  -      PKat

Selecting   Residues
State       1/      1

```

Analysis of 3CM3.pdb

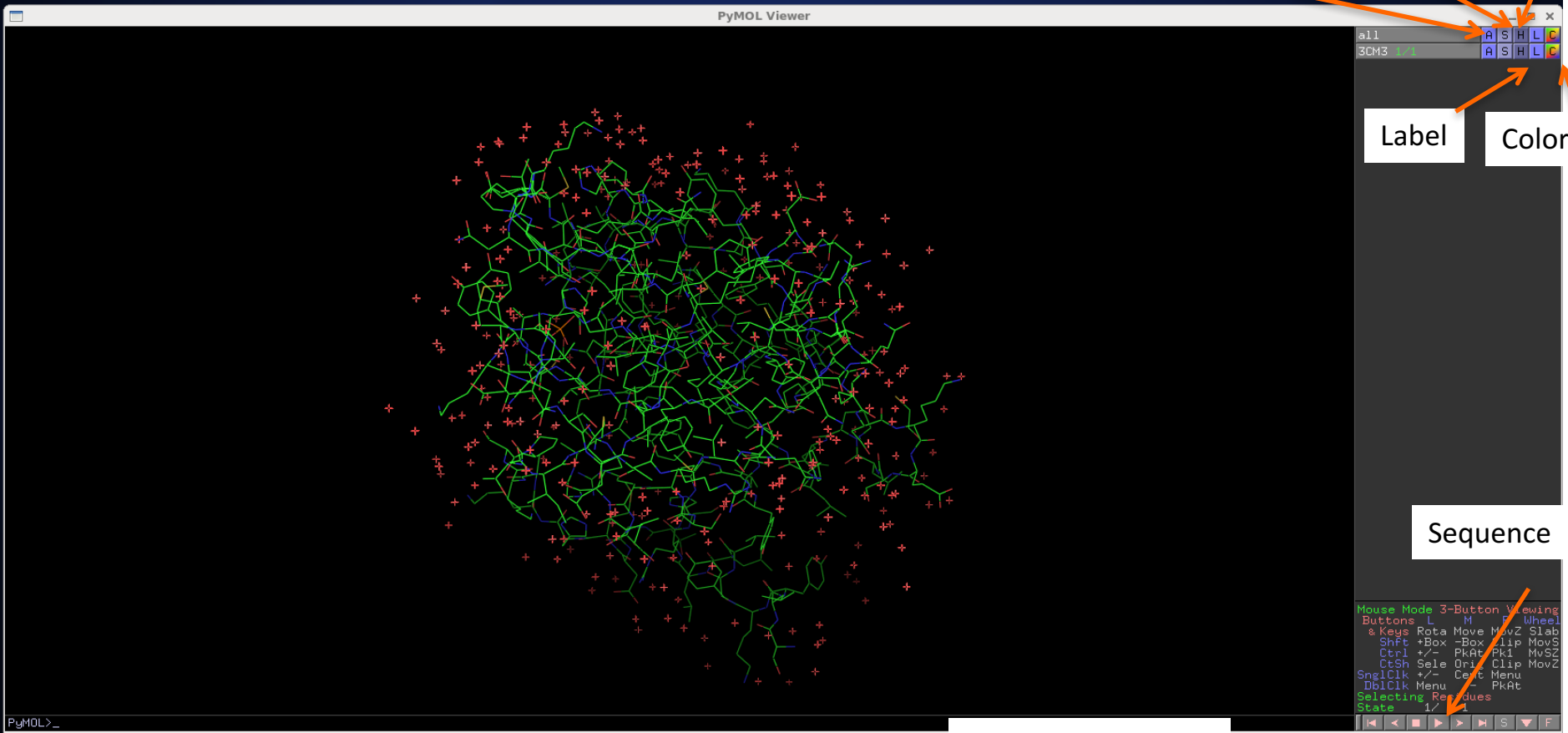
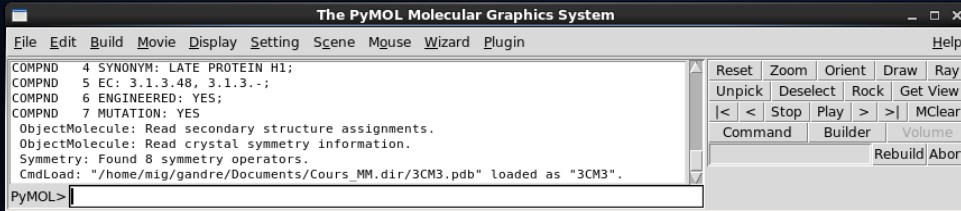
File /open/3CM3.pdb



Keywords of 3CM3? Starts at?

Analysis of 3CM3.pdb

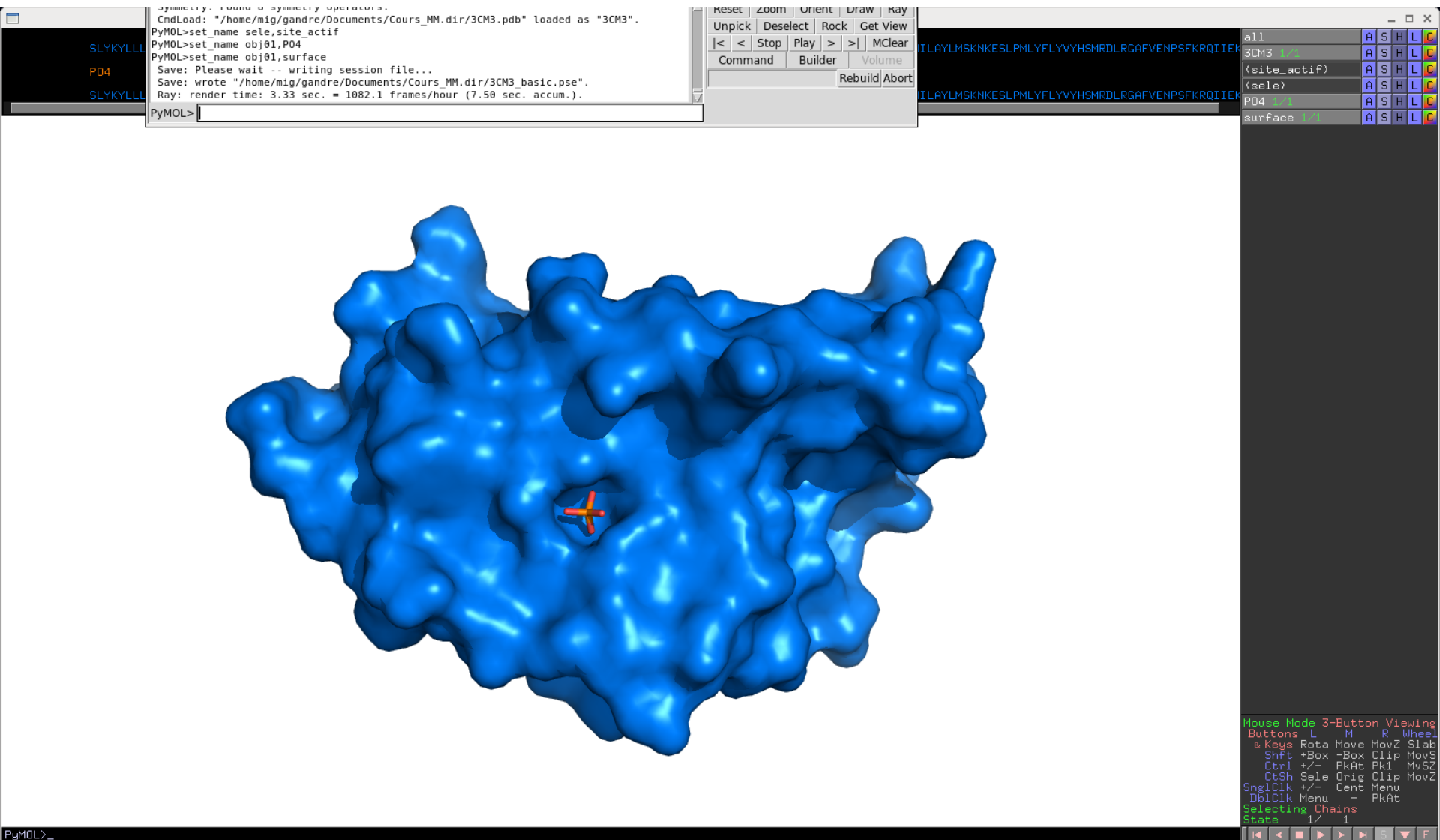
File /open/3CM3.pdb



Starts at SLYK

Analysis of 3CM3.pdb

File /open/3CM3-basic.pse go to Ile 1170 show stick



Show stick

What do we see?

Analysis of 3CM3.pdb

<http://www.rcsb.org/pdb/home/home.do>

3CM3 download fasta & pdb. Piece of informations?

RCSB Protein Data Bank - RCSB PDB - 3CM3 Structure Summary - Mozilla Firefox

www.rcsb.org/pdb/explore/explore.do?structureId=3CM3

RCSB PDB Deposit Search Visualize Analyze Download Learn More

MyPDB Login

RCSB PDB An Information Portal to 108607 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations

PDB-101 PDB EMDBank Structural Biology Knowledgebase

Summary 3D View Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Links

High Resolution Crystal Structure of the Vaccinia Virus Dual-Specificity Phosphatase VH1

DOI:10.2210/pdb3cm3/pdb

Primary Citation

Dimeric quaternary structure of the prototypical dual specificity phosphatase VH1.

Koksal, A.C., Nardozzi, J.D., Cingolani, G.

Journal: (2009) J.Biol.Chem. 284: 10129-10137

PubMed: 19211553

PubMedCentral: PMC2665067

DOI: 10.1074/jbc.M808362200

Search Related Articles in PubMed

PubMed Abstract:

The Vaccinia virus H1 gene product, VH1, is a dual specificity phosphatase that down-regulates the cellular antiviral response by dephosphorylating STAT1. The crystal structure of VH1, determined at 1.32 Å resolution, reveals a novel dimeric quaternary structure, which exposes two active sites spaced approximately 39 Å away from each other. VH1 forms a stable dimer via an extensive domain swap of the N-terminal helix (residues 1-20). In vitro, VH1 can dephosphorylate activated STAT1, in a reaction that is competed by the nuclear transport adapter importin alpha5. Interestingly, VH1 is inactive with respect to STAT1 bound to DNA, suggesting that the viral phosphatase acts predominantly on the cytoplasmic pool of activated STAT1. We propose that the dimeric quaternary structure of VH1 is essential for specific recognition of activated STAT1, which prevents its nuclear translocation, thus blocking interferon-gamma signal transduction and antiviral response.

Keywords:

Active Transport, Cell Nucleus, Catalytic Domain, Circular Dichroism, DNA, Dimerization, Dual Specificity Phosphatase 3, Humans, Interferon-gamma, Models, Molecular, Protein Conformation, Protein Structure, Quaternary, Protein Structure, Tertiary, STAT1 Transcription Factor, Signal Transduction, Vaccinia virus

Organizational Affiliation:

Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, New York 13210, USA.

Click on abstract words and keywords to add them to the search box.

Search PubMed Abstracts

[Hide Abstract]

3CM3 Display Files Download Files Download Citation

Biological Assembly



3D View: JSmol or PV More Images

Symmetry: C2 view

Stoichiometry: Homo 2-mer - A2

Biological assembly 1 assigned by authors and generated by PISA (software)

Downloadable viewers:

Simple Viewer Protein Workshop Kiosk Viewer

MyPDB Personal Annotations Hide

To save personal annotations, please login to your MyPDB account.

Deposition Summary Hide

Authors: Koksal, A.C., Cingolani, G.

Deposition: 2008-03-20

Release: 2009-02-10

Last Modified (REVDAT): 2009-04-28

Revision History Hide

Mouse over text for details

Molecular Description Hide

Classification: Hydrolase

Structure Weight: 20513.83

Molecule: Dual specificity protein phosphatase

Polymer: 1

Type: protein

Length: 176

Phases: A

MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

MaiAGE

DU GÉNOME À L'ENVIRONNEMENT

Analysis of 3CM3.pdb

<http://www.rcsb.org/pdb/home/home.do>

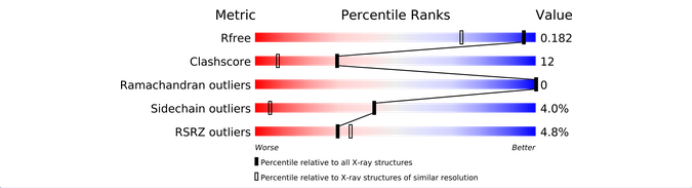
3CM3 download fasta et pdb. Infos?

RCSB Protein Data Bank - RCSB PDB - 3CM3 Structure Summary - Mozilla Firefox

www.rcsb.org/pdb/explore/explore.do?structureid=3CM3

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

View the full validation report



Metric	Percentile Ranks	Value
Rfree		0.182
Clashscore		12
Ramachandran outliers		0
Sidechain outliers		4.0%
RSRR outliers		4.8%

Worse Better
■ Percentile relative to all X-ray structures
■ Percentile relative to X-ray structures of similar resolution

MolProbity Ramachandran Plot
Download Ramachandran Plot PDF (from MolProbity)

Source Hide

Polymer: 1
Scientific Name: Vaccinia virus Taxonomy Expression System: Escherichia coli

Related PDB Entries Hide

Identifier	Details
2RF6	Vaccinia Virus Dual-Specificity Phosphatase VH1
3CEO	Vaccinia Virus Dual-Specificity Phosphatase VH1 Bound to Xenon Gas

Ligand Chemical Component Hide

Identifier	Formula	Name	View Interactions
BME Search Download	C ₂ H ₅ O S	BETA-MERCAPTOETHANOL	Ligand Explorer Jmol
PO4 Search Download	O ₄ P	PHOSPHATE ION	Ligand Explorer Jmol

External Domain Annotations Hide

- CATH Classification v4.0.0: 1 Domain - data from CATH
- Pfam Classification: 2 Domains - data from Pfam
- GO Terms: 16 Terms - data from GO

Structural Biology Knowledgebase Data Hide

Information from the Structural Biology Knowledgebase

- Models from the Protein Model Portal: 6 models
- Related Biological Annotations: >20 annotations
- Related Clones from PSI: Biology Materials Repository: 0 clones
- Related Targets & Protocols from TargetTrack: 0 targets

Data in orange boxes are gathered from external resources (when available). Reset Layout

Quality
Zero outliers

MatIAGE
DU GÉNOME À L'ENVIRONNEMENT

Analysis of 3CM3.pdb

gedit 3CM3.pdb

```
HEADER          HYDROLASE                      20-MAR-08   3CM3
TITLE           HIGH RESOLUTION CRYSTAL STRUCTURE OF THE VACCINIA VIRUS
TITLE           2 DUAL-SPECIFICITY PHOSPHATASE VH1
COMPND          MOL_ID: 1;
COMPND          2 MOLECULE: DUAL SPECIFICITY PROTEIN PHOSPHATASE;
COMPND          3 CHAIN: A;
COMPND          4 SYNONYM: LATE PROTEIN H1;
COMPND          5 EC: 3.1.3.48, 3.1.3.-;
COMPND          6 ENGINEERED: YES;
COMPND          7 MUTATION: YES;
SOURCE          MOL_ID: 1;
SOURCE          2 ORGANISM_SCIENTIFIC: VACCINIA VIRUS;
SOURCE          3 STRAIN: WESTERN RESERVE / WR;
SOURCE          4 GENE: H1 ORF;
SOURCE          5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE          6 EXPRESSION_SYSTEM_STRAIN: BL21;
SOURCE          7 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE          8 EXPRESSION_SYSTEM_PLASMID: PET14B
KEYWDS          DUAL-SPECIFICITY PHOSPHATASE, VACCINIA VIRUS, VH1,
KEYWDS          2 HYDROLASE, LATE PROTEIN, PROTEIN PHOSPHATASE
EXPDTA          X-RAY DIFFRACTION
AUTHOR          A.C.KOKSAL,G.CINGOLANI
REVSTAT        2 28-APR-09 3CM3 1 JRNL
REVSTAT        1 10-FEB-09 3CM3 0
JRNL            AUTH A.C.KOKSAL,J.D.NARDOZZI,G.CINGOLANI
JRNL            TITL DIMERIC QUATERNARY STRUCTURE OF THE PROTOTYPICAL
JRNL            TITL 2 DUAL SPECIFICITY PHOSPHATASE VH1.
JRNL            REF J.BIOL.CHEM. V. 284 10129 2009
JRNL            REFN ISSN 0021-9258
JRNL            PMID 19211553
JRNL            DOI 10.1074/JBC.M808362200
REMARK          1
REMARK          2
REMARK          2 RESOLUTION. 1.32 ANGSTROMS.
REMARK          3
REMARK          3 REFINEMENT.
REMARK          3 PROGRAM : REFMAC 5.2.0019
REMARK          3 AUTHORS : MURSHUDOV,VAGIN,DODSON
REMARK          3
REMARK          3 REFINEMENT TARGET : ENGH & HUBER
REMARK          3
REMARK          3 DATA USED IN REFINEMENT.
REMARK          3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.32
REMARK          3 RESOLUTION RANGE LOW (ANGSTROMS) : 14.11
REMARK          3 DATA CUTOFF (SIGMA(F)) : 0.000
REMARK          3 COMPLETENESS FOR RANGE (%) : 90.8
REMARK          3 NUMBER OF REFLECTIONS : 33998
REMARK          3
REMARK          3 FIT TO DATA USED IN REFINEMENT.
REMARK          3 CROSS-VALIDATION METHOD : THROUGHOUT
REMARK          3 FREE R VALUE TEST SET SELECTION : RANDOM
REMARK          3 R VALUE (WORKING + TEST SET) : 0.172
REMARK          3 R VALUE (WORKING SET) : 0.171
REMARK          3 FREE R VALUE : 0.185
REMARK          3 FREE R VALUE TEST SET SIZE (%) : 5.000
REMARK          3 FREE R VALUE TEST SET COUNT : 1800
REMARK          3
REMARK          3 FIT IN THE HIGHEST RESOLUTION BIN.
REMARK          3 TOTAL NUMBER OF BINS USED : 20
REMARK          3 BIN RESOLUTION RANGE HIGH (A) : 1.32
```

pdb code

Authors

Resolution

X-ray data quality criteria

Analysis of 3CM3.pdb

gedit 3CM3.pdb

```
3CM3.pdb (~/Documents/Cours_MM.dir) - gedit
File Edit View Search Tools Documents Help
Open Save Undo Cut Copy Paste
3CM3.pdb
SEQADV 3CM3 ARG A 1002 UNP P07239 EXPRESSION TAG
SEQADV 3CM3 SER A 1112 UNP P07239 CYS 110 ENGINEERED
SEQRES 1 A 176 GLY PRO GLU ILE ARG MET ASP LYS LYS SER LEU TYR LYS
SEQRES 2 A 176 TYR LEU LEU LEU ARG SER THR GLY ASP MET HIS LYS ALA
SEQRES 3 A 176 LYS SER PRO THR ILE MET THR ARG VAL THR ASN ASN VAL
SEQRES 4 A 176 TYR LEU GLY ASN TYR LYS ASN ALA MET ASP ALA PRO SER
SEQRES 5 A 176 SER GLU VAL LYS PHE LYS TYR VAL LEU ASN LEU THR MET
SEQRES 6 A 176 ASP LYS TYR THR LEU PRO ASN SER ASN ILE ASN ILE ILE
SEQRES 7 A 176 HIS ILE PRO LEU VAL ASP ASP THR THR THR ASP ILE SER
SEQRES 8 A 176 LYS TYR PHE ASP ASP VAL THR ALA PHE LEU SER LYS CYS
SEQRES 9 A 176 ASP GLN ARG ASN GLU PRO VAL LEU VAL HIS SER ALA ALA
SEQRES 10 A 176 GLY VAL ASN ARG SER GLY ALA MET ILE LEU ALA TYR LEU
SEQRES 11 A 176 MET SER LYS ASN LYS GLU SER LEU PRO MET LEU TYR PHE
SEQRES 12 A 176 LEU TYR VAL TYR HIS SER MET ARG ASP LEU ARG GLY ALA
SEQRES 13 A 176 PHE VAL GLU ASN PRO SER PHE LYS ARG GLN ILE ILE GLU
SEQRES 14 A 176 LYS TYR VAL ILE ASP LYS ASN
HET P04 A 1 5
HET BME A1174 4
HET BME A 2 4
HETNAM P04 PHOSPHATE ION
HETNAM BME BETA-MERCAPTOETHANOL
FORMUL 2 P04 O4 P 3-
FORMUL 3 BME 2(C2 H6 O S)
FORMUL 5 HOH *300(H2 O)
HELIX 1 1 SER A 1007 THR A 1017 1 11
HELIX 2 2 ASN A 1040 ASP A 1046 1 7
HELIX 3 3 ALA A 1047 SER A 1050 5 4
HELIX 4 4 ILE A 1087 LYS A 1089 5 3
HELIX 5 5 TYR A 1090 ASN A 1105 1 16
HELIX 6 6 ASN A 1117 ASN A 1131 1 15
HELIX 7 7 LEU A 1135 GLY A 1152 1 18
HELIX 8 8 ASN A 1157 VAL A 1169 1 13
SHEET 1 A 5 THR A1030 ARG A1031 0
SHEET 2 A 5 VAL A1036 GLY A1039 -1 0 LEU A1038 N THR A1030
SHEET 3 A 5 VAL A1108 HIS A1111 1 0 VAL A1110 N TYR A1037
SHEET 4 A 5 TYR A1056 ASN A1059 1 N LEU A1058 0 LEU A1109
SHEET 5 A 5 ASN A1073 HIS A1076 1 0 ILE A1075 N VAL A1057
SITE 1 AC1 8 ASP A1081 SER A1112 ALA A1113 ALA A1114
SITE 2 AC1 8 GLY A1115 ASN A1117 ARG A1118 HOH A1199
SITE 1 AC2 5 ASN A1117 GLU A1156 ASN A1157 HOH A1251
SITE 2 AC2 5 HOH A1391
SITE 1 AC3 3 ASN A1131 LYS A1132 SER A1134
CRYST1 63.816 38.690 134.987 90.00 90.00 90.00 C 2 2 21 8
ORIGX1 1.000000 0.000000 0.000000 0.000000
ORIGX2 0.000000 1.000000 0.000000 0.000000
ORIGX3 0.000000 0.000000 1.000000 0.000000
```

Sequence

Hetero atoms

Space group and unit cell parameters

Crystallization conditions

```
REMARK 200 REMARK: NULL
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS (%): 40.18
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 2.06
REMARK 280 CRYSTALLIZATION CONDITIONS: 62% PEG 400, 0.1M TRIS PH 8 ,
REMARK 280 BATCH, TEMPERATURE 297K
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: C 2 2 21
REMARK 290
REMARK 290 SYMOP SYMMETRY
REMARK 290 NNNMMM OPERATOR
REMARK 290 1555 X,Y,Z
REMARK 290 2555 -X,-Y,Z+1/2
REMARK 290 3555 -X,Y,-Z+1/2
REMARK 290 4555 X,-Y,-Z
REMARK 290 5555 X+1/2,Y+1/2,Z
REMARK 290 6555 -X+1/2,-Y+1/2,Z+1/2
REMARK 290 7555 -X+1/2,Y+1/2,-Z+1/2
REMARK 290 8555 X+1/2,-Y+1/2,-Z
REMARK 290
REMARK 290 WHERE NNN -> OPERATOR NUMBER
```

Analysis of 3CM3.pdb

gedit 3CM3.pdb

3CM3.pdb

SHEET 1

1 AC1 8 ASP A1081 SER A1112 ALA A1113 ALA A1114

SITE 2 AC1 8 GLY A1115 ASN A1117 ARG A1118 HOH A1199

SITE 1 AC2 5 ASN A1117 GLU A1156 ASN A1157 HOH A1251

SITE 2 AC2 5 HOH A1391

SITE 1 AC3 3 ASN A1131 LYS A1132 SER A1134

CRYST1 63.816 38.690 134.987 90.00 90.00 90.00 C 2 2 21 8

ORIGX1 1.000000 0.000000 0.000000 0.000000

ORIGX2 0.000000 1.000000 0.000000 0.000000

ORIGX3 0.000000 0.000000 1.000000 0.000000

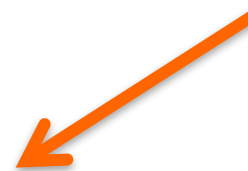
SCALE1 0.015670 0.000000 0.000000 0.000000

SCALE2 0.000000 0.025846 0.000000 0.000000

SCALE3 0.000000 0.000000 0.007408 0.000000

ATOM 1 N SER A1007 23.183 -12.243 7.340 1.00 44.49 N

Positions x y z
Occupancy ($0 < x < 1$)
B factor



4 parameters per atom:

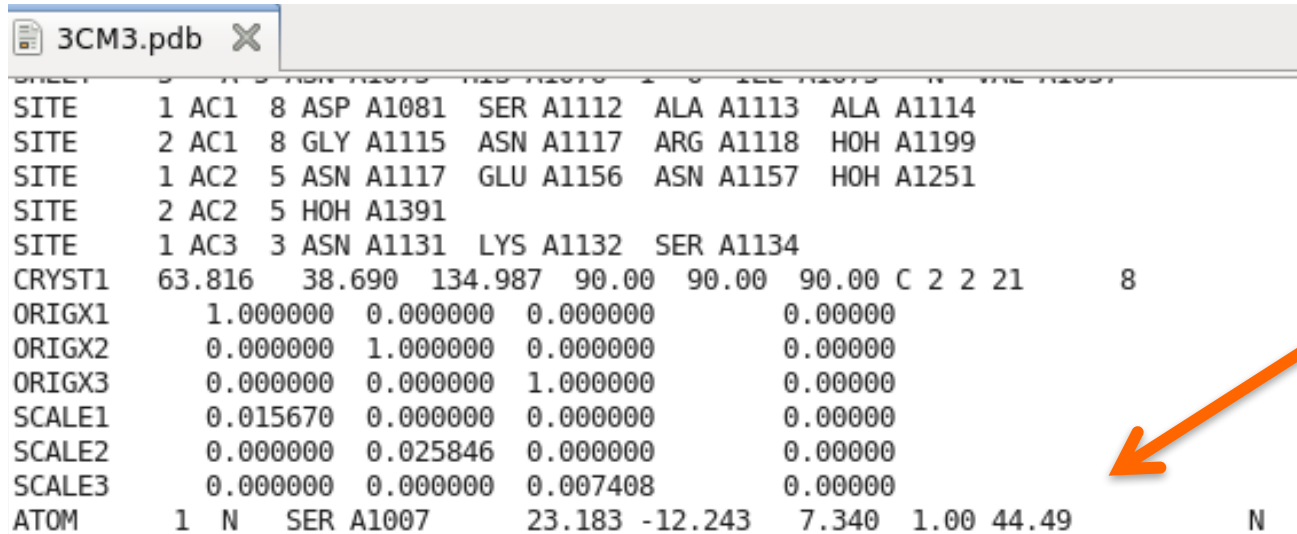
x,y,z location inside the unit cell

Occupancy

Isotropic B or B factor: atomic displacement that accounts for slightly different positions in each unit cell with average position x,y,z.

Analysis of 3CM3.pdb

gedit 3CM3.pdb



SHEET	5	115	ASN A1175	115	1175	1	5	ILE A1175	115	1175	1	115	1175
SITE	1	AC1	8	ASP A1081	SER A1112	ALA A1113	ALA A1114						
SITE	2	AC1	8	GLY A1115	ASN A1117	ARG A1118	HOH A1199						
SITE	1	AC2	5	ASN A1117	GLU A1156	ASN A1157	HOH A1251						
SITE	2	AC2	5	HOH A1391									
SITE	1	AC3	3	ASN A1131	LYS A1132	SER A1134							
CRYST1	63.816	38.690	134.987	90.00	90.00	90.00	C 2 2 21				8		
ORIGX1	1.000000	0.000000	0.000000			0.000000							
ORIGX2	0.000000	1.000000	0.000000			0.000000							
ORIGX3	0.000000	0.000000	1.000000			0.000000							
SCALE1	0.015670	0.000000	0.000000			0.000000							
SCALE2	0.000000	0.025846	0.000000			0.000000							
SCALE3	0.000000	0.000000	0.007408			0.000000							
ATOM	1	N	SER A1007	23.183	-12.243	7.340	1.00	44.49				N	

Positions x y z
Occupancy (0<x<1)
B factor

Go to Ile 1170

Within firefox, go to <http://www.ebi.ac.uk/pdbsum/>
Then 3CM3.pdb.

Within firefox, go to <http://molprobit.biochem.duke.edu/>
Then 3CM3.pdb

Analysis of 3CM3.pdb

gedit 3CM3.pdb



```
SITE 1 AC1 8 ASP A1081 SER A1112 ALA A1113 ALA A1114
SITE 2 AC1 8 GLY A1115 ASN A1117 ARG A1118 HOH A1199
SITE 1 AC2 5 ASN A1117 GLU A1156 ASN A1157 HOH A1251
SITE 2 AC2 5 HOH A1391
SITE 1 AC3 3 ASN A1131 LYS A1132 SER A1134
CRYST1 63.816 38.690 134.987 90.00 90.00 90.00 C 2 2 21 8
ORIGX1 1.000000 0.000000 0.000000 0.000000
ORIGX2 0.000000 1.000000 0.000000 0.000000
ORIGX3 0.000000 0.000000 1.000000 0.000000
SCALE1 0.015670 0.000000 0.000000 0.000000
SCALE2 0.000000 0.025846 0.000000 0.000000
SCALE3 0.000000 0.000000 0.007408 0.000000
ATOM 1 N SER A1007 23.183 -12.243 7.340 1.00 44.49 N
ANISOU 1 N SER A1007 8216 3592 5095 1023 1280 164 N
ATOM 2 CA SER A1007 21.716 -12.220 7.558 1.00 43.59 C
ANISOU 2 CA SER A1007 8036 3596 4931 834 1225 87 C
ATOM 3 C SER A1007 21.163 -10.798 7.556 1.00 42.26 C
ANISOU 3 C SER A1007 7731 3589 4738 820 900 -142 C
ATOM 4 O SER A1007 20.211 -10.492 6.830 1.00 41.39 O
ANISOU 4 O SER A1007 7764 3474 4487 658 806 -126 O
ATOM 5 CB SER A1007 21.378 -12.924 8.872 1.00 44.08 C
ANISOU 5 CB SER A1007 8223 3643 4881 833 1322 81 C
ATOM 6 OG SER A1007 21.891 -14.249 8.855 1.00 44.41 O
ANISOU 6 OG SER A1007 7896 3876 5102 678 1747 337 O
ATOM 7 N LEU A1008 21.763 -9.925 8.369 1.00 42.33 N
ANISOU 7 N LEU A1008 7350 3720 5012 759 406 -263 N
ATOM 8 CA LEU A1008 21.334 -8.523 8.461 1.00 41.98 C
ANISOU 8 CA LEU A1008 6830 3948 5173 513 240 -274 C
ATOM 9 C LEU A1008 21.441 -7.823 7.107 1.00 39.42 C
ANISOU 9 C LEU A1008 5795 3722 5461 59 181 -5 C
ATOM 10 O LEU A1008 20.696 -6.897 6.781 1.00 38.54 O
ANISOU 10 O LEU A1008 5314 3657 5673 -135 164 153 O
ATOM 11 CB LEU A1008 22.160 -7.770 9.514 1.00 44.48 C
ANISOU 11 CB LEU A1008 7426 4363 5113 635 125 -485 C
ATOM 12 CG LEU A1008 21.858 -8.020 10.999 1.00 50.00 C
ANISOU 12 CG LEU A1008 9524 5374 5140 551 6 477 C
```

Positions x y z, facteur d'occupation ($0 < x < 1$) et facteur B



Le facteur B reflète l'agitation thermique dans le cristal

Aller de nouveau jusqu'à l'le 1170

Depuis firefox, taper <http://www.ebi.ac.uk/pdbsum/>
puis 3CM3.pdb. Voir.

<http://molprobity.biochem.duke.edu/>

puis éventuellement renseigner 3CM3.pdb sous:

Analysis of 3CM3.pdb

gedit 3CM3.pdb



```
3CM3.pdb (~/.Documents/Cours_MM.dir) - gedit
File Edit View Search Tools Documents Help
Open Save Undo Cut Copy Paste Print
3CM3.pdb x
SITE 1 AC1 8 ASP A1081 SER A1112 ALA A1113 ALA A1114
SITE 2 AC1 8 GLY A1115 ASN A1117 ARG A1118 HOH A1199
SITE 1 AC2 5 ASN A1117 GLU A1156 ASN A1157 HOH A1251
SITE 2 AC2 5 HOH A1391
SITE 1 AC3 3 ASN A1131 LYS A1132 SER A1134
CRYST1 63.816 38.690 134.987 90.00 90.00 90.00 C 2 2 21 8
ORIGX1 1.000000 0.000000 0.000000 0.000000
ORIGX2 0.000000 1.000000 0.000000 0.000000
ORIGX3 0.000000 0.000000 1.000000 0.000000
SCALE1 0.015670 0.000000 0.000000 0.000000
SCALE2 0.000000 0.025846 0.000000 0.000000
SCALE3 0.000000 0.000000 0.007408 0.000000
ATOM 1 N SER A1007 23.183 -12.243 7.340 1.00 44.49 N
ANISOU 1 N SER A1007 8216 3592 5095 1023 1280 164 N
ATOM 2 CA SER A1007 21.716 -12.220 7.558 1.00 43.59 C
ANISOU 2 CA SER A1007 8036 3596 4931 834 1225 87 C
ATOM 3 C SER A1007 21.163 -10.798 7.556 1.00 42.26 C
ANISOU 3 C SER A1007 7731 3589 4738 820 900 -142 C
ATOM 4 O SER A1007 20.211 -10.492 6.830 1.00 41.39 O
ANISOU 4 O SER A1007 7764 3474 4487 658 806 -126 O
ATOM 5 CB SER A1007 21.378 -12.924 8.872 1.00 44.08 C
ANISOU 5 CB SER A1007 8223 3643 4881 833 1322 81 C
ATOM 6 OG SER A1007 21.891 -14.249 8.855 1.00 44.41 O
ANISOU 6 OG SER A1007 7896 3876 5102 678 1747 337 O
ATOM 7 N LEU A1008 21.763 -9.925 8.369 1.00 42.33 N
ANISOU 7 N LEU A1008 7350 3720 5012 759 406 -263 N
ATOM 8 CA LEU A1008 21.334 -8.523 8.461 1.00 41.98 C
ANISOU 8 CA LEU A1008 6830 3948 5173 513 240 -274 C
ATOM 9 C LEU A1008 21.441 -7.823 7.107 1.00 39.42 C
ANISOU 9 C LEU A1008 5795 3722 5461 59 181 -5 C
ATOM 10 O LEU A1008 20.696 -6.897 6.781 1.00 38.54 O
ANISOU 10 O LEU A1008 5314 3657 5673 -135 164 153 O
ATOM 11 CB LEU A1008 22.160 -7.770 9.514 1.00 44.48 C
ANISOU 11 CB LEU A1008 7426 4363 5113 635 125 -485 C
ATOM 12 CG LEU A1008 21.858 -8.020 10.999 1.00 50.00 C
ANISOU 12 CG LEU A1008 9524 5374 5140 551 6 477 C
```

Positions x y z, facteur d'occupation ($0 < x < 1$) et facteur B



Le facteur B reflète l'agitation thermique dans le cristal

Critical to assess the quality of template protein +++

Homology modeling

Homology modeling, also known as **comparative modeling** of protein, refers to constructing an atomic-resolution model of the "query" protein from its amino acid sequence from an experimental three-dimensional structure of an homologous protein (the "template"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that align residues in the query sequence with residues in the template sequence.

Protein structures are more conserved than protein sequences amongst homologues.

Sequences falling below a 20% sequence identity can have very different structure. (wiki source)

- 1) Identification de la/des protéines « templates » de référence
- 2) Modélisation de la protéine requête « query »

Qualité de l'alignement +++

<http://robeta.bakerlab.org>

<http://toolkit.tuebingen.mpg.de/hhpred>

<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>



Homology modeling

<http://toolkit.tuebingen.mpg.de/hhpred>

The screenshot shows the Bioinformatics Toolkit interface in a Mozilla Firefox browser. The page title is "Bioinformatics Toolkit" and the subtitle is "Max-Planck Institute for Developmental Biology". The main navigation bar includes "Search", "Alignment", "Sequence Analysis", "2ary Structure", "3ary Structure", "Classification", and "Utils". The "Search" tab is active, showing a list of tools: CS-BLAST, HHblits, HHpred, HHsenser, HMMER3, PatternSearch, ProtBLAST, PSI-BLAST, and SimShiftDB. The "HHpred" tool is selected, and the results page is displayed. The job ID is 2458690, and the date is 17:08 on May 06 2015. The "Results" tab is active, showing a progress bar and a list of sequences: 3cn3_A, 2q05_A, 2hxp_A, 3enu_A, 3s4e_A, 2hcn_A, 2nt2_A, and 1zzw_A. A red arrow points to the "Resubmit section" button. The text "Cliquez pour voir l'alignement 1D /2D" is overlaid on the right side of the image.

by HMM-HMM comparison - Mozilla Firefox

toolkit.tuebingen.mpg.de/hhpred/results/2458690

HOME Login PDBeart Personal Databases Contact Tool Versions Imprint Disclaimer Help

Bioinformatics Toolkit
Max-Planck Institute for Developmental Biology

Search Alignment Sequence Analysis 2ary Structure 3ary Structure Classification Utils

CS-BLAST HHblits HHpred HHsenser HMMER3 PatternSearch ProtBLAST PSI-BLAST SimShiftDB
New job Resubmit Resubmit_HMM Resubmit/HHsenser Realign Log Input-params Delete

HHpred - Results Job-ID: 2458690 Date: 17:08 on May 06 2015

Results Histogram Reduced alignment Representative alignment Full alignment

Create model Merge Q/T alignments Forward to PCoils Forward MSA Save Export

Color alignments
☐ color only SS ☒ color alignments ☐ color alignments

View alignment Activate Jalview

3 100

Resubmit section

3cn3_A
2q05_A
2hxp_A
3enu_A
3s4e_A
2hcn_A
2nt2_A
1zzw_A

Cliquez pour voir l'alignement 1D /2D

Modeller <https://salilab.org/modeller/>

https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html

Direct link to <http://esprict.ibcp.fr/ESPrict/cgi-bin/ESPrict.cgi>

http://www.ebi.ac.uk/Tools/psa/emboss_stretcher/

http://www.ebi.ac.uk/Tools/psa/emboss_water/

Homology modeling

<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>

PHYRE2 Protein Fold Recognition Server - Mozilla Firefox

HHpred - Homology d... x PBIL-IBCP Lyon Gerland x NPS@ : CLUSTALW ALI... x ESPrnt 3.x / ENDscript... x 0-0-1433760249-esp... x PHYRE2 Protein Fold R... x

www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

Standard Mode | [Login](#) for job manager, batch processing, Phyre alarm and other advanced options

Retrieve Phyre Job Id

Phyre²

Protein Homology/analogY Recognition Engine V 2.0

Subscribe to Phyre at Google Groups
Email:

[Visit Phyre at Google Groups](#)
[Follow @Phyre2server](#)

[New Phyre2 paper](#) | [NEW](#) Fast structural search with [PhyreStorm](#) (beta-testing)
[Edinburgh workshop](#) | [Oxford workshop](#)

E-mail Address	
Optional Job description	
Amino Acid Sequence	
Modelling Mode	Normal ☒ Intensive ☐
	Phyre Search Reset

Or try the [sequence finder \(NEW!\)](#)

1401165 submissions since Feb 14 2011

Phyre is for **non-commercial** use only

Homology modeling

https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html

Direct link to <http://espript.ibcp.fr/ESPrIPT/cgi-bin/ESPrIPT.cgi>

NPS@ : CLUSTALW multiple alignment - Mozilla Firefox

https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html

Pôle BioInformatique Lyonnais
Network Protein Sequence Analysis
NPS@ is the IBCP contribution to PBIL in Lyon, France

[HOME] [NPS@] [HELP] [REFERENCES] [NEWS] [MPSA] [ANTHEPROT] [Geno3D] [SuMo] [Positions] [PBIL]

Monday, March 9th 2015 : NPS@ is online again ([see news](#)).
Monday, March 2nd 2015 : NPS@ is offline after disk array hardware failure ([see news](#)).
NPS@ is up and running at new URL <https://npsa-prabi.ibcp.fr/>

CLUSTALW

[Abstract] [NPS@ help] [Original server]

Paste a protein sequence databank in Pearson/Fa

UNLTHDKYTLPSNINIIHI
PLVDDITTDISKYFDDVIAFLSKCDORNEPVLVHSAAGVNRSGAMTLAYLMSKIKESLPW
LYELVYHSDMDLACAFVEN
PSFKDQIIIEKVIDKN
>PTP_Orf_wt
MGDKSEMYARLLLRCTRAGPPLALPSGNTRLTDHVLGSAEDARAVLRGDSGVDFKCLVN
MTMSKYSTAGITATYHPLRDDKTHIASIMPALVKLLARLEAEQKPTLVHCVAGVNRSG
AAANGVYHHRLEAENFMTQFARFVFLKTYEIDRLGGAFLNANFRYQLIKDFVCDSP

Renseigner cette partie avec 3CM3_2Q05_PTP_Orf.fasta

All sequence names must be different !

SUBMIT CLEAR

Output width : 60

CLUSTALW Parameters

Output format : Clustalw

Output order : aligned

Pairwise alignment type : SLOW

Fast pairwise alignment parameters	Slow pairwise alignment parameters
K-tuple (word) size : 1	Protein weight matrix : GONNET
Number of top diagonals : 5	Gap opening penalty : 10.0
Window size : 5	Gap extension penalty : 0.1
Gap penalty : 3	
Scoring method : Percentage	

Multiple Alignment Parameters :

Weight matrix : GONNET

Gap opening penalty : 10.0

MaiAGE
MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES
DU GÉNOME À L'ENVIRONNEMENT

Homology modeling

https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html

Direct link to <http://escript.ibcp.fr/ESPrpt/cgi-bin/ESPrpt.cgi>

ESPrpt 3.0

SUBMIT DISPLAY MODE LAYERS SESSION TIME **EXIT**
RESULTS DOC BEG ADV EXP -1 +1 LOAD SAVE

ENDScript / ESPrpt uses popup windows to display results - please be sure to disable popup blockers before submitting a job
When publishing data resulting from usage of this server, please cite this reference article

Aligned Sequences

Main alignment file Selection

ALN file <https://npsa-prabi.i...>
☐ RESET
Example file • Tutorial

☐ Hide sequences
☐ Number sequences
☒ More info (for PP/NPS@)
☐ Sec. struct. info from PP/NPS@/PDB

Range: all
Start:
Chain ID (for PDB files):

Secondary structure depiction

Top & bottom secondary structures

TOP secondary structures

Input file: **3CM3_A.pdb** (indicated by an orange arrow)
☐ RESET

☐ Chain ID ☐ Relative accessibility
☒ Depict all known structures

BOTTOM secondary structures

Input file: Upload a file below OR click here: **PDB**

Sec. structure labels: ☒ $\alpha 1, \beta 1, \alpha 2, \beta 2, \dots$
☐ $\alpha A, \beta A, \alpha B, \beta B, \dots$
☐ $\alpha 1, \beta A, \alpha 2, \beta B, \dots$

☐ Hide labels
☐ Hide turns
☐ Hide names
☐ Hide alternates
☐ Hide disordered regions

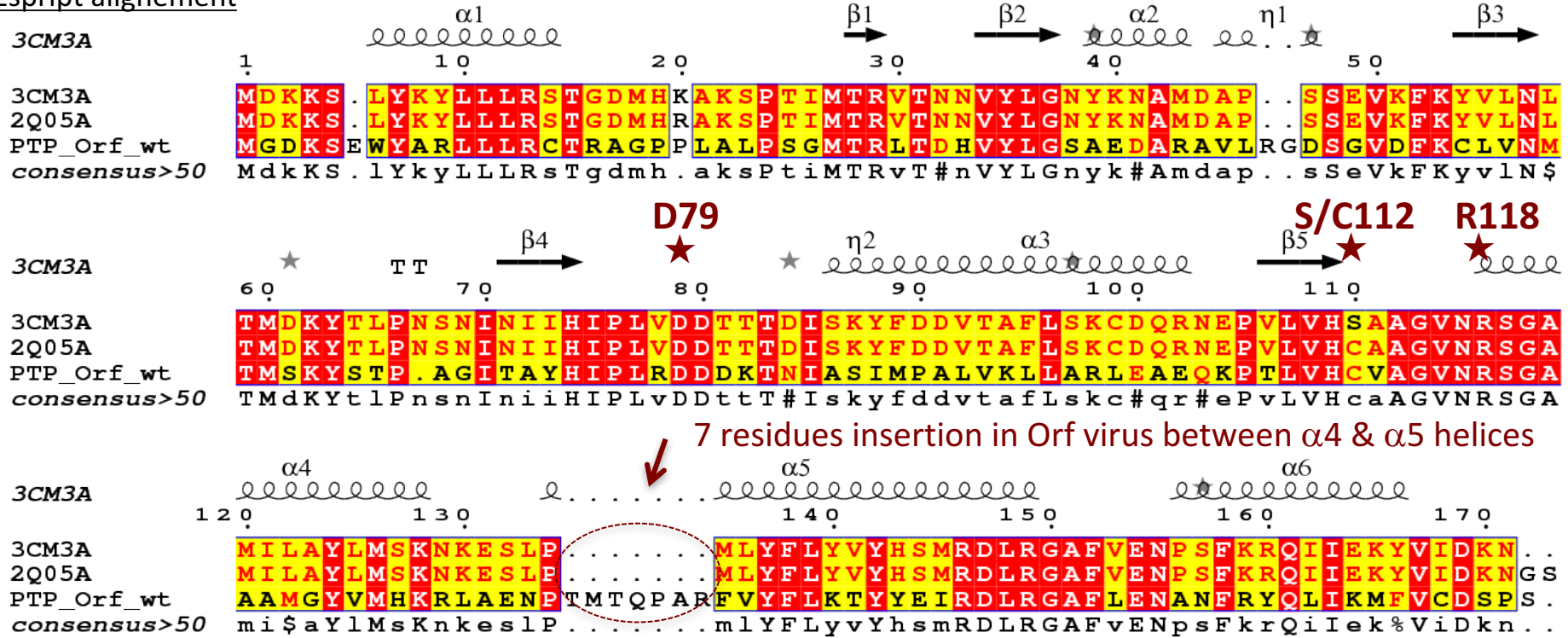
Renseigner cette partie avec 3CM3.pdb

- Aligned sequences
- Secondary structure
- Sequence similarities
- Special commands
- Defining groups
- Output layout
- Output files

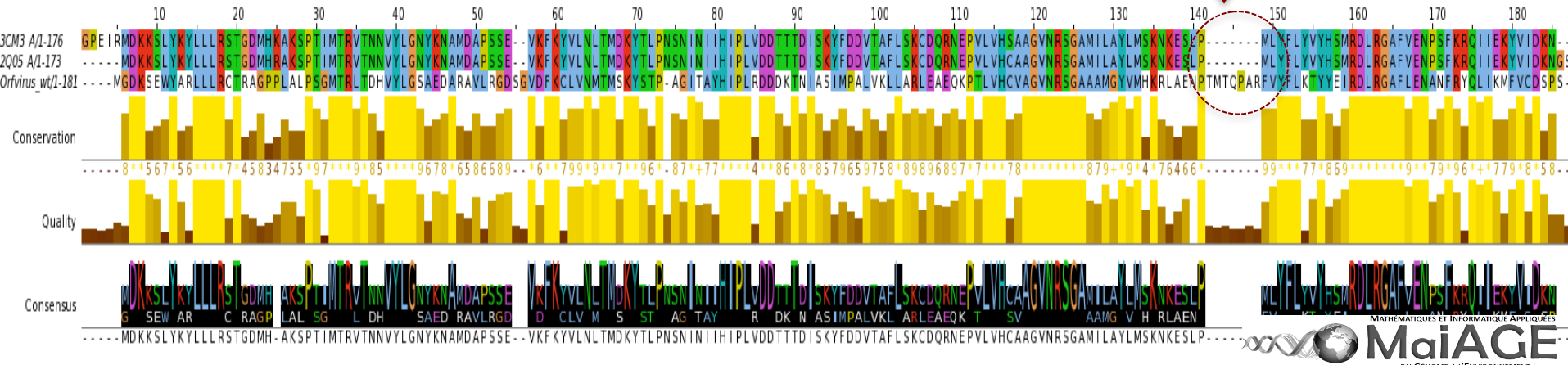
MaiAGE
MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES
DU GÉNOME À L'ENVIRONNEMENT

2D alignment of PTP_ORF wt with VH1 from Vaccinia (3CM3) & Vaccinia WR (2Q05)

Espright alignment



Jalview visualisation



Homology modeling

Need to convert into .pir alignment. Go back to nps Lyon then ask for a .pir output
Copy/paste/save using gedit/nedit.

Go from 3CM3_2Q05_PTPOrf.pir to aln_modeler_PTP_Orf.pir

NB: PyMol is useful to borne out with residues numbering and chain names



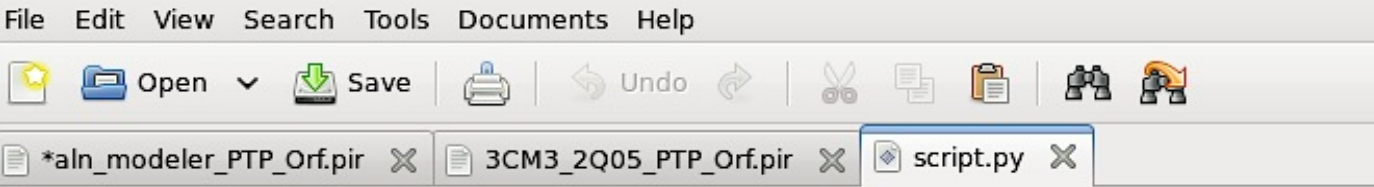
```
aln_modeler_PTP_Orf.pir 3CM3_2Q05_PTP_Orf.pir
>P1;3CM3Ax0
MDKKS-LYKYLLLRSTGDMHAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFKYVLN
LTMDKYTLPSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHSAAGVNR
GAMILAYLMSKNKESLP-----MLYFLYVYHSMRDLRGAFVENPSFKRQIEKYVIDK
N--
*
>P1;2Q05Ax1
MDKKS-LYKYLLLRSTGDMHRAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFKYVLN
LTMDKYTLPSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHCAAGVNR
GAMILAYLMSKNKESLP-----MLYFLYVYHSMRDLRGAFVENPSFKRQIEKYVIDK
NGS
*
>P1;PTP_Orf_wt
MGDKSEWYARLLLRCTRAGPPLALPSGMTRLTDHVYLGSAEDARAVLRGDSGVDFKCLVN
MTMSKYSTP-AGITAYHIPLRDDDKTNIASIMPALVKLLARLEAEQKPTLVHCVAGVNR
GAAAMGYVMHKRLAENPTMTQPARFVYFLKTYEIRDLRGAFLNANFRYQLIKMFVCD
PS-
*
aln_modeler_PTP_Orf.pir 3CM3_2Q05_PTP_Orf.pir
>P1;3CM3_A
structureX:3CM3_A:1007 :A:1170 :A:unknown:unknown:-1.00:-1.00
-----SLYKYLLLRSTGDMHAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFK
YVLNLTMDKYTLPSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHSAAG
VNRSGAMILAYLMSKNKESLP-----MLYFLYVYHSMRDLRGAFVENPSFKRQIEKY
VI-----
*
>P1;2Q05_B
structureX:2Q05_B:1 :B:171 :B:unknown:unknown:-1.00:-1.00
----MDKKSLEYKYLLLRSTGDMHRAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFK
YVLNLTMDKYTLPSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHCAAG
VNRSGAMILAYLMSKNKESLP-----MLYFLYVYHSMRDLRGAFVENPSFKRQIEKY
VIDKN
*
>P1;PTP_ORF_WT
sequence:PTP_ORF_WT:6 :A:181 :A:unknown:unknown:-1.00:-1.0
-----EWYARLLLRCTRAGPPLALPSGMTRLTDHVYLGSAEDARAVLRGDSGVDFK
CLVNMTMSKYSTP-AGITAYHIPLRDDDKTNIASIMPALVKLLARLEAEQKPTLVHCVAG
VNRSGAAAMGYVMHKRLAENPTMTQPARFVYFLKTYEIRDLRGAFLNANFRYQLIKMF
VCDSPS-
*|
```


Homology modeling

Needs to convert the.pir alignment
Copy/paste/save /with gedit.
Go from 3CM3_2Q05_PTPOrf.pir to aln_modeler_PTP_Orf.pir

Modify now python program
Add alignment, templates, nb modèles etc ...

Launch modeler
Mod9.13 script.py



```
File Edit View Search Tools Documents Help
[Icons: Open, Save, Undo, Redo, Cut, Copy, Paste, Find, Print]
*aln_modeler_PTP_Orf.pir 3CM3_2Q05_PTP_Orf.pir script.py

from modeller import *
from modeller.automodel import *
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in
env.io.atom_files_directory = './../atom_files'
a = automodel(env,
               alnfile = 'aln_VH1_PTPOrf.pir',
               knowns = ('3CM3_A', '2Q05_B'),
               sequence = 'PTP_ORF_WT',
               assess_methods=(assess.DOPE, assess.GA341))
a.starting_model= 1
a.ending_model = 100
a.make()
```

<https://salilab.org/modeller/>

Homology modeling

Needs to convert the.pir alignment

Copy/paste/save /with gedit.

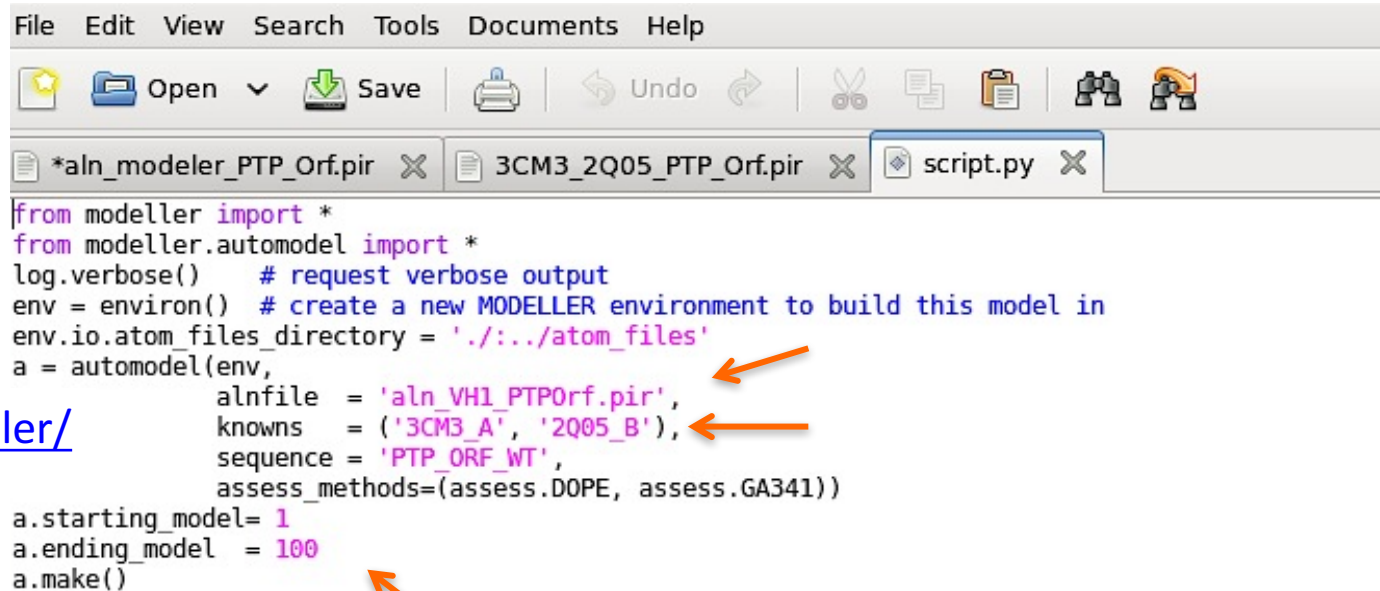
Go from 3CM3_2Q05_PTPOrf.pir to aln_modeler_PTP_Orf.pir

Modify now python program

Add alignment, templates, nb modèles etc ...

Launch modeler

Mod9.13 script.py



```
File Edit View Search Tools Documents Help
[Icons: Open, Save, Undo, Redo, Cut, Copy, Paste, Find, Print]
*aln_modeler_PTP_Orf.pir 3CM3_2Q05_PTP_Orf.pir script.py
from modeller import *
from modeller.automodel import *
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in
env.io.atom_files_directory = './../atom_files'
a = automodel(env,
               alnfile = 'aln_VH1_PTPOrf.pir',
               knowns = ('3CM3_A', '2Q05_B'),
               sequence = 'PTP_ORF_WT',
               assess_methods=(assess.DOPE, assess.GA341))
a.starting_model= 1
a.ending_model = 100
a.make()
```

<https://salilab.org/modeller/>

ouvrir script.log

Comparison of results: gedit *.log

DOPE score

GABE score

Visualization with PyMOL

Homology modeling

Homology modeling, also known as **comparative modeling** of protein, refers to constructing an atomic-resolution model of the "query" [protein](#) from its [amino acid sequence](#) from an experimental three-dimensional structure of an homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an [alignment](#) that align residues in the query sequence with residues in the template sequence.

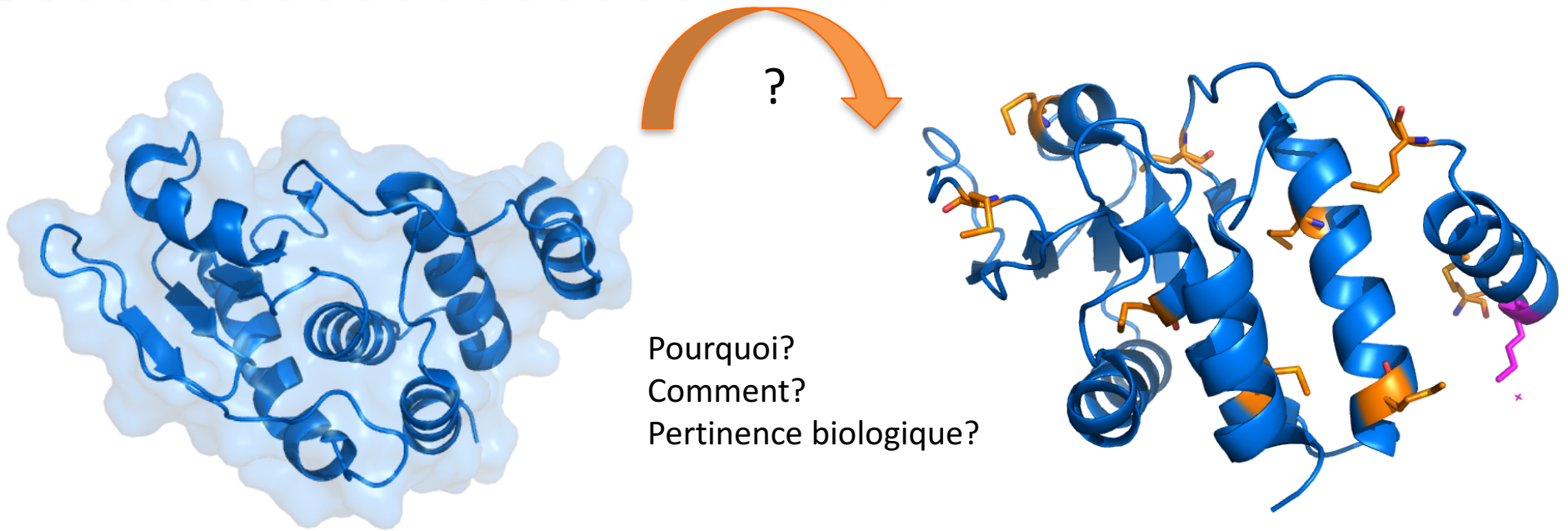
Protein structures are more conserved than protein sequences amongst homologues.

Sequences falling below a 20% sequence identity can have very different structure. ([wiki source](#))

- 1) Identification of protein(s) as reference « template(s) »
- 2) Homology modeling of protein « query »

Open PyMol session: [compil_Hmodeles.pse](#)

Mutations



Model the point mutants by substituting residues

- 1- Modéliser les mutants, les variants pour lesquels on a des données expérimentales.
- 2- Modifier le site de fixation
- 3- Vérifier l'accessibilité
- 4- Analyser l'interface

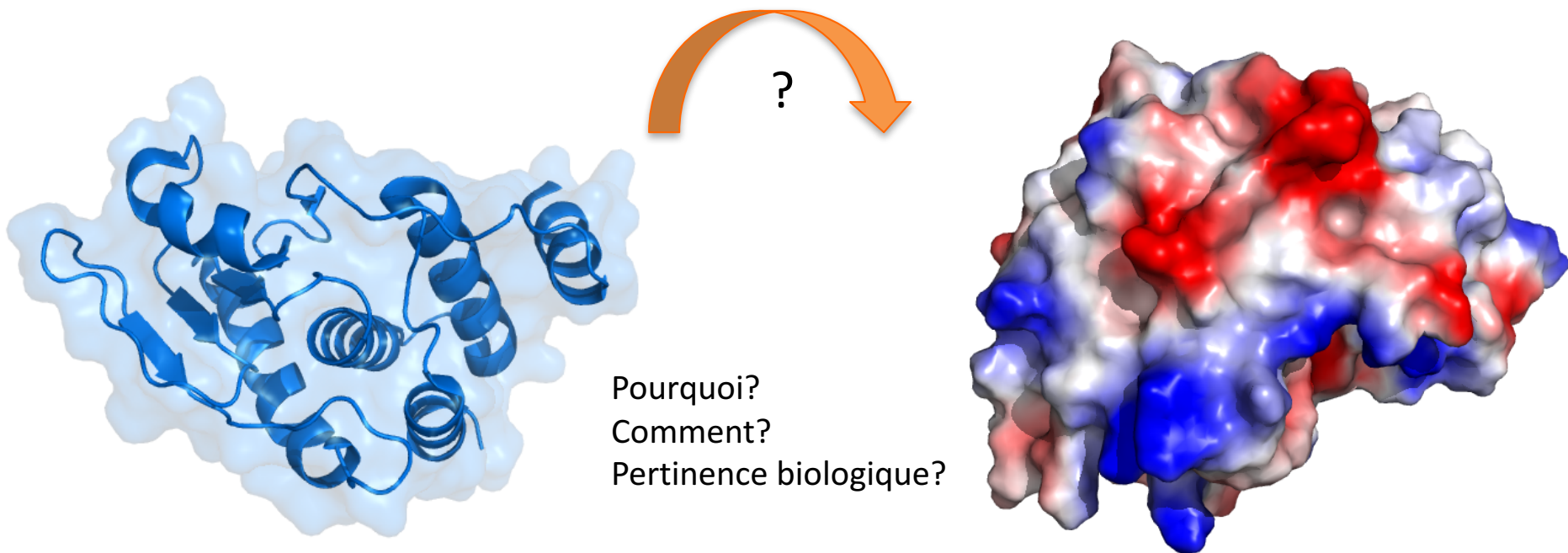
Pymol/Plugin/PDB Loader Service/ faire la mutation dans 3CM3 S112C

Ouvrir 2Q05

Show sequence. Que voit-on?

Show cartoon/color by chain/ select/action copy to objet/ Wizard ...

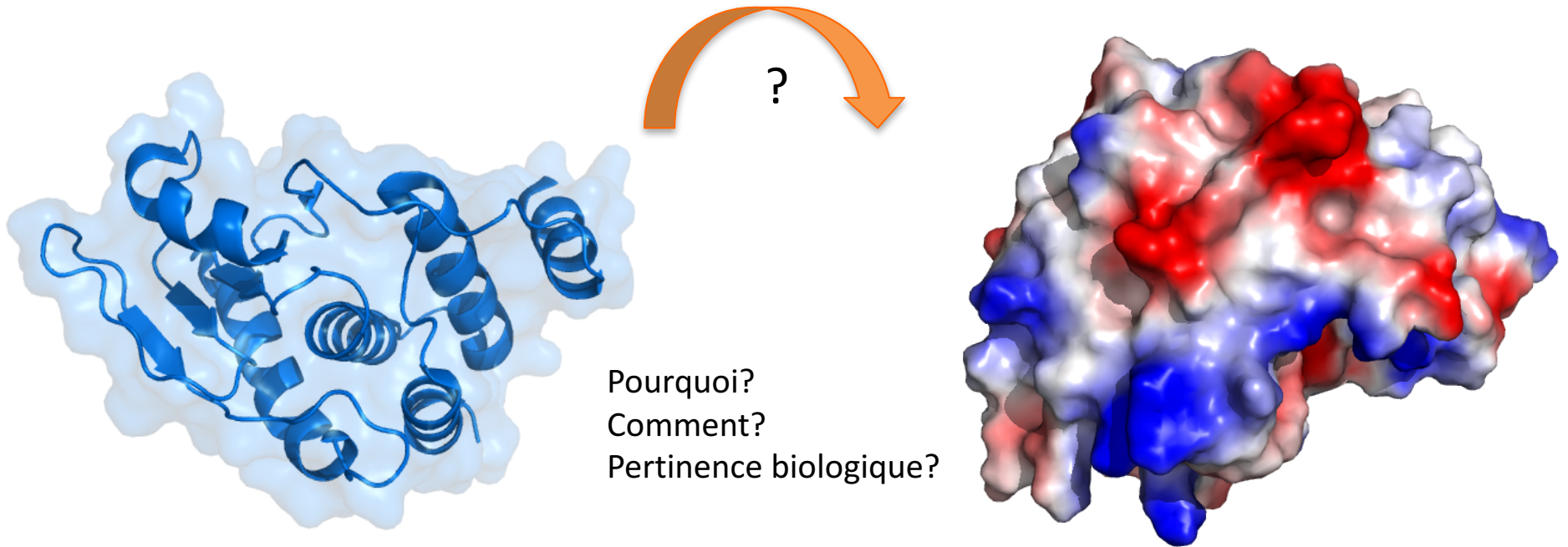
Electrostatic profile



Calcul du profil électrostatique : cartographie des patches hydrophobes/chargés + ou –.
Analyse de la solvation des systèmes biologiques

- 1- Simulation des processus de diffusion -> cinétiques de fixation protéine/ligand; protéine/protéine.
- 2- Modélisation du solvant implicite pour la DM des biomolécules.
- 3- Solvation et calcul d'énergie d'affinité et de constantes d'équilibre. Rational drug-design.
- 4- Etude de titrations des biomolécules.

Electrostatic profile



Comment? PDB2PQR et APBS Adaptive Poisson-Boltzmann Solver

APBS Adaptive Poisson-Boltzmann Solver dans pymol

PDB2PQR aller à http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/

Electrostatic profile

PDB2PQR aller à http://nbc222.ucsd.edu/pdb2pqr_2.0.0/

PDB2PQR Server - Mozilla Firefox

nbcr-222.ucsd.edu/pdb2pqr_2.0.0/

Examples

If you use the PDB2PQR service in a publication, please cite:

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* 32 W665-W667 (2004). [\[Link\]](#)

Note: This server uses automatic refreshing to update the status of your PDB2PQR submission.

Please enter either:

☐ a PDB ID:

☐ upload a PDB file: No file selected.

Pick a forcefield to use:

☐ AMBER

☐ CHARMM

☐ PARSE

☐ PEOEPB

☐ SWANSON

☐ TYLO6

☐ User-defined forcefield (help): No file selected.

User-defined names (help): No file selected.

* If you select user-defined forcefield, you also need to specify a user-defined .names file.

Pick an output naming scheme to use (help):

☐ Internal naming scheme (What's this?)

☐ AMBER

☐ CHARMM

☐ PARSE

☐ PEOEPB

☐ SWANSON

☐ TYLO6

Available options:

☒ Ensure that new atoms are not rebuilt too close to existing atoms

☒ Optimize the hydrogen bonding network

☐ Assign charges to the ligand specified in a MOL2 file: No file selected.

☒ Create an APBS input file (this also enables the option to run APBS and visualize your results through the web interface, if it has been installed)

Firefox automatically sends some data to Mozilla so that we can improve your experience.

Electrostatic profile

PDB2PQR aller à http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/

Electrostatic profile

PDB2PQR aller à http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/

PDB2PQR Job Status Page - Mozilla Firefox

Run time: 0:00:34
Current time: Wed Jun 3 06:08:47 2015

Status: **complete** ✓

Here are the results:

- Input files
 - 2Q05_ChainA.pdb
- Output files
 - 2Q05_ChainA.propka
 - 2Q05_ChainA.in
 - 2Q05_ChainA.pqr
- Runtime and debugging information
 - Program output (stdout)
 - Program errors and warnings (stderr)

[Click here](#) to run APBS with your results.

Electrostatic profile

PDB2PQR aller à http://nbc-222.ucsd.edu/pdb2pqr_2.0.0/

2Q05_A_Met.pdb

2Q05_A_Met.pqr

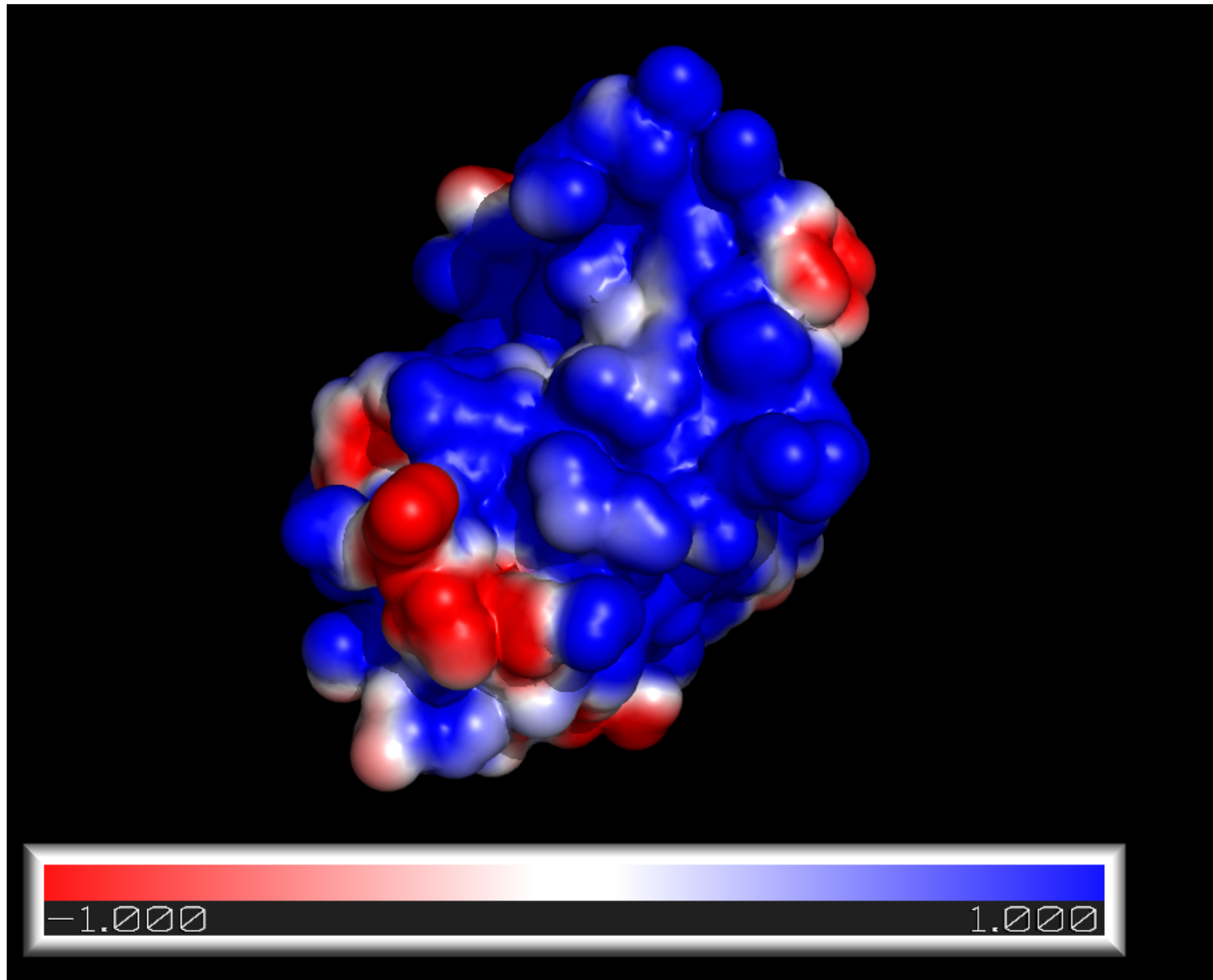
ATOM	1	N	MET	A	1	92.038	61.305	57.368	1.00	0.00	N
ATOM	2	CA	MET	A	1	91.096	61.791	58.372	1.00	0.00	C
ATOM	3	C	MET	A	1	91.815	62.519	59.482	1.00	0.00	C
ATOM	4	O	MET	A	1	93.019	62.442	59.651	1.00	0.00	O
ATOM	5	CB	MET	A	1	90.267	60.604	58.934	1.00	0.00	C
ATOM	6	CG	MET	A	1	89.317	59.911	57.933	1.00	0.00	C
ATOM	7	SD	MET	A	1	88.087	61.085	57.345	1.00	0.00	S
ATOM	8	CE	MET	A	1	87.151	61.276	58.868	1.00	0.00	C
ATOM	9	N	ASP	A	2	91.000	63.233	60.290	1.00	46.15	N
ATOM	10	CA	ASP	A	2	91.465	63.913	61.491	1.00	44.81	C
ATOM	11	C	ASP	A	2	92.278	62.945	62.353	1.00	44.30	C
ATOM	12	O	ASP	A	2	91.788	61.889	62.768	1.00	43.84	O
ATOM	13	CB	ASP	A	2	90.268	64.473	62.272	1.00	44.45	C
ATOM	14	CG	ASP	A	2	90.679	65.206	63.536	1.00	43.48	C
ATOM	15	OD1	ASP	A	2	89.901	65.181	64.515	1.00	42.53	O
ATOM	16	OD2	ASP	A	2	91.777	65.803	63.555	1.00	41.93	O1-
ATOM	17	N	LYS	A	3	93.518	63.317	62.616	1.00	0.00	N
ATOM	18	CA	LYS	A	3	94.423	62.396	63.298	1.00	0.00	C
ATOM	19	C	LYS	A	3	94.117	62.330	64.775	1.00	0.00	C
ATOM	20	O	LYS	A	3	94.090	61.225	65.337	1.00	0.00	O
ATOM	21	CB	LYS	A	3	95.892	62.839	63.054	1.00	0.00	C
ATOM	22	CG	LYS	A	3	96.937	61.855	63.637	1.00	0.00	C
ATOM	23	CD	LYS	A	3	98.394	62.301	63.487	1.00	0.00	C
ATOM	24	CE	LYS	A	3	98.799	62.218	62.010	1.00	0.00	C
ATOM	25	NZ	LYS	A	3	100.230	62.544	61.875	1.00	0.00	N1+
ATOM	26	N	LYS	A	4	93.866	63.458	65.420	1.00	43.16	N
ATOM	27	CA	LYS	A	4	93.415	63.466	66.808	1.00	43.30	C

ATOM	1	N	MET	1	92.038	61.305	57.368	-0.3200	2.0000
ATOM	2	CA	MET	1	91.096	61.791	58.372	0.3300	2.0000
ATOM	3	C	MET	1	91.815	62.519	59.482	0.5500	1.7000
ATOM	4	O	MET	1	93.019	62.442	59.651	-0.5500	1.4000
ATOM	5	CB	MET	1	90.267	60.604	58.934	0.0000	2.0000
ATOM	6	CG	MET	1	89.317	59.911	57.933	0.2650	2.0000
ATOM	7	SD	MET	1	88.087	61.085	57.345	-0.5300	1.8500
ATOM	8	CE	MET	1	87.151	61.276	58.868	0.2650	2.0000
ATOM	9	HE1	MET	1	87.599	61.954	59.451	0.0000	0.0000
ATOM	10	HE2	MET	1	86.223	61.581	58.652	0.0000	0.0000
ATOM	11	HE3	MET	1	87.112	60.399	59.346	0.0000	0.0000
ATOM	12	H2	MET	1	91.895	61.794	56.505	0.3300	0.0000
ATOM	13	H3	MET	1	92.977	61.459	57.683	0.3300	0.0000
ATOM	14	HG2	MET	1	89.872	59.548	57.171	0.0000	0.0000
ATOM	15	HG3	MET	1	88.883	59.127	58.400	0.0000	0.0000
ATOM	16	H	MET	1	91.897	60.324	57.219	0.3300	0.0000
ATOM	17	HA	MET	1	90.469	62.439	57.921	0.0000	0.0000
ATOM	18	HB3	MET	1	89.712	60.939	59.712	0.0000	0.0000
ATOM	19	HB2	MET	1	90.906	59.906	59.293	0.0000	0.0000
ATOM	20	N	ASP	2	91.000	63.233	60.290	-0.4000	1.5000
ATOM	21	CA	ASP	2	91.465	63.913	61.491	-0.0000	2.0000
ATOM	22	C	ASP	2	92.278	62.945	62.353	0.5500	1.7000
ATOM	23	O	ASP	2	91.788	61.889	62.768	-0.5500	1.4000
ATOM	24	CB	ASP	2	90.268	64.473	62.272	0.0000	2.0000
ATOM	25	CG	ASP	2	90.679	65.206	63.536	0.1000	1.7000
ATOM	26	OD1	ASP	2	89.901	65.181	64.515	-0.5500	1.4000
ATOM	27	OD2	ASP	2	91.777	65.803	63.555	-0.5500	1.4000

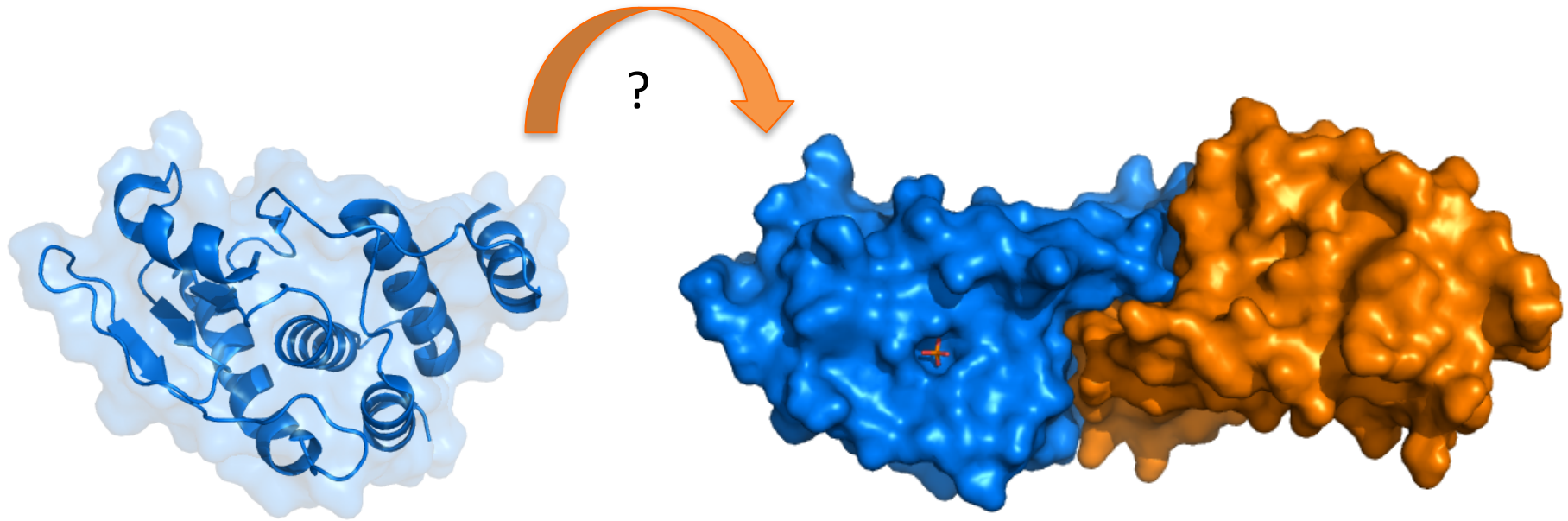
Facteur occupation $0 < x < 1$ Facteur B agitation thermique

Q charge R rayon

Mapping of the electrostatic profile



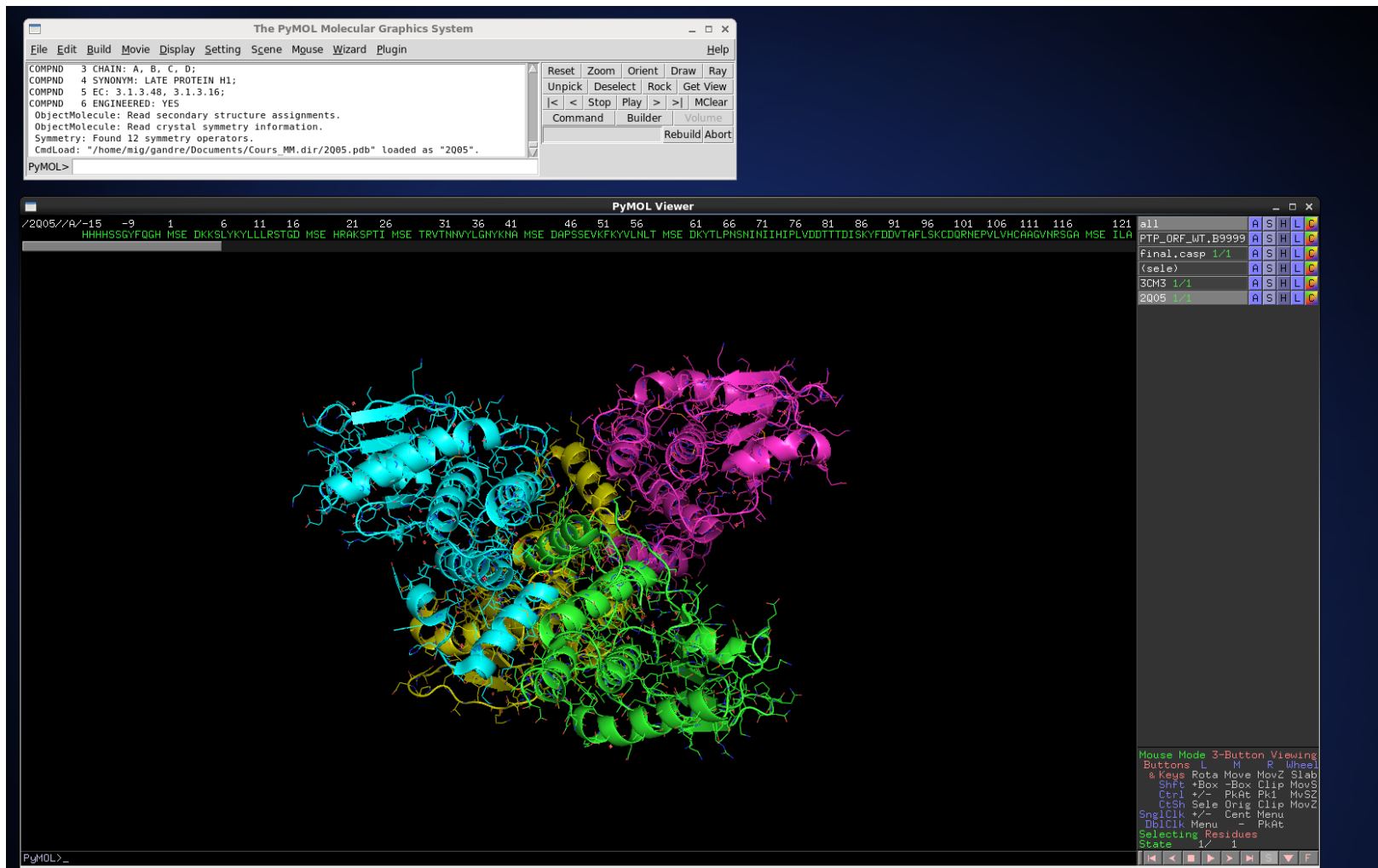
Oligomerization



Pourquoi?
Comment?
Pertinence biologique?

Oligomerization

File /open/2Q05.pdb

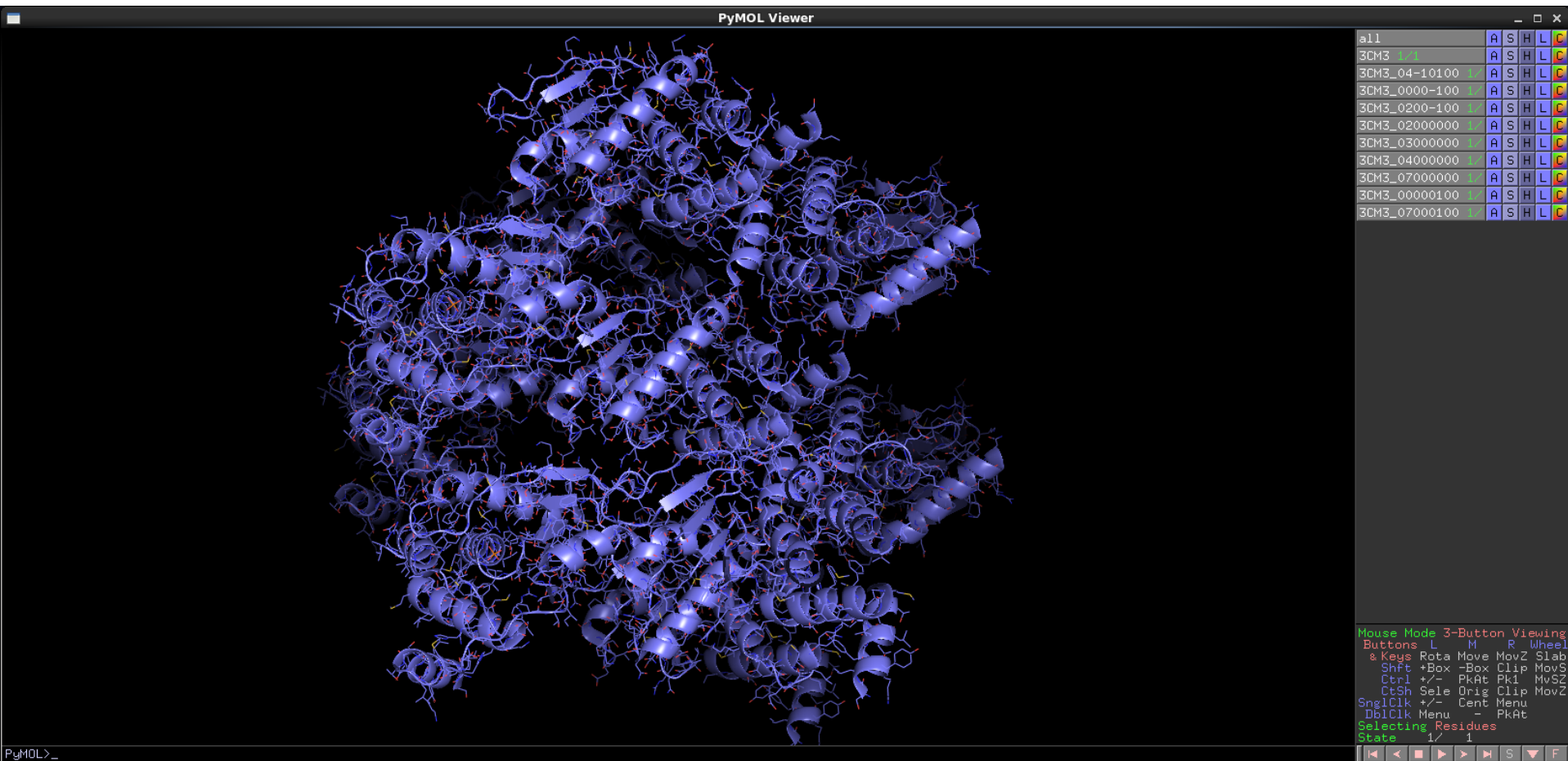


PDBePISA : Proteins, Interfaces, Structures and Assemblies is an interactive tool for the exploration of macromolecular interfaces.

<http://www.ebi.ac.uk/pdbe/pisa/>

Oligomerization

File /open/3CM3.pdb Action/ remove water Show cartoon Action/generate/symetry mates 4A



Informations on symetry related mates in crystal packing. Comparison. Computation. Interface profiling.

PDBePISA : Proteins, Interfaces, Structures and Assemblies is an interactive tool for the exploration of macromolecular interfaces.

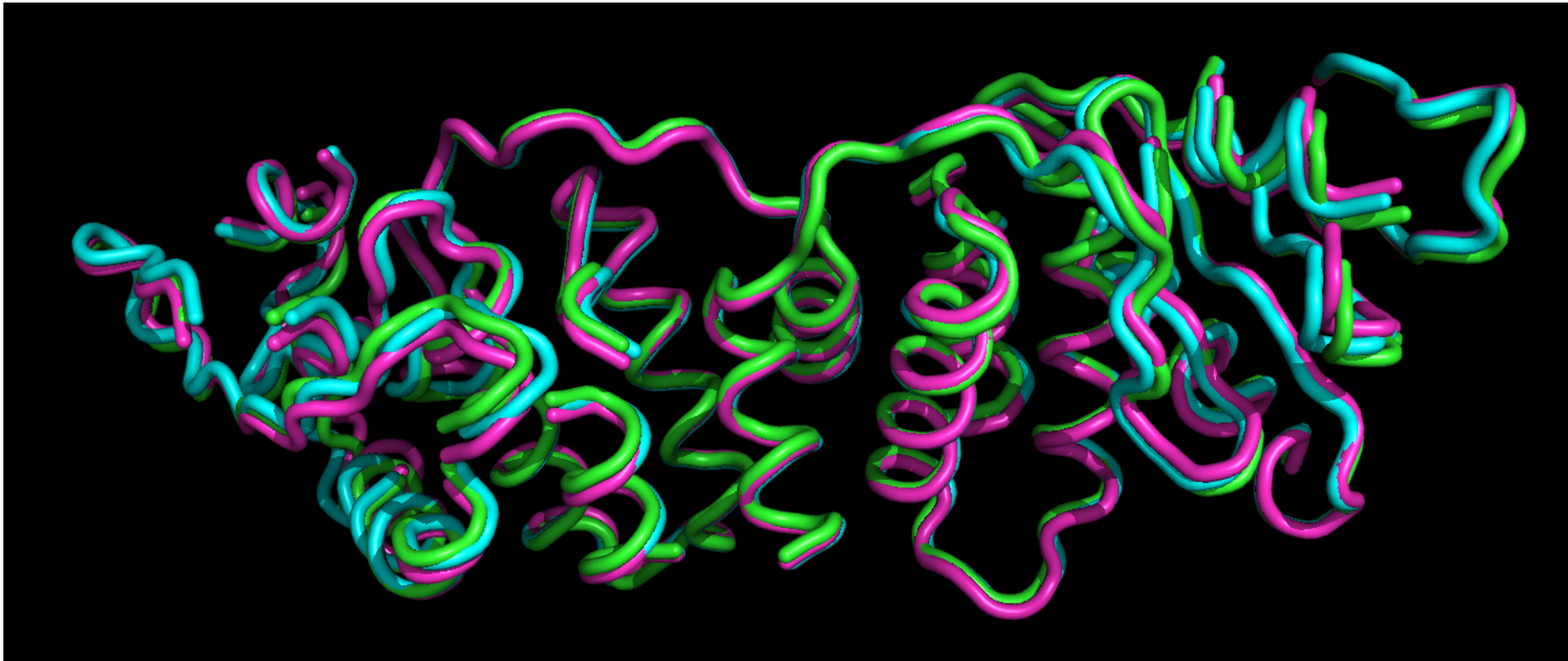
<http://www.ebi.ac.uk/pdbe/pisa/>

Normal mode analysis

<http://lorentz.dynstr.pasteur.fr/nma/submission.php>

Browse 2Q05_AB.pdb keep the selected default parameter then submit

In PyMol: File /open/2Q05_AB.pdb -> arrow below right



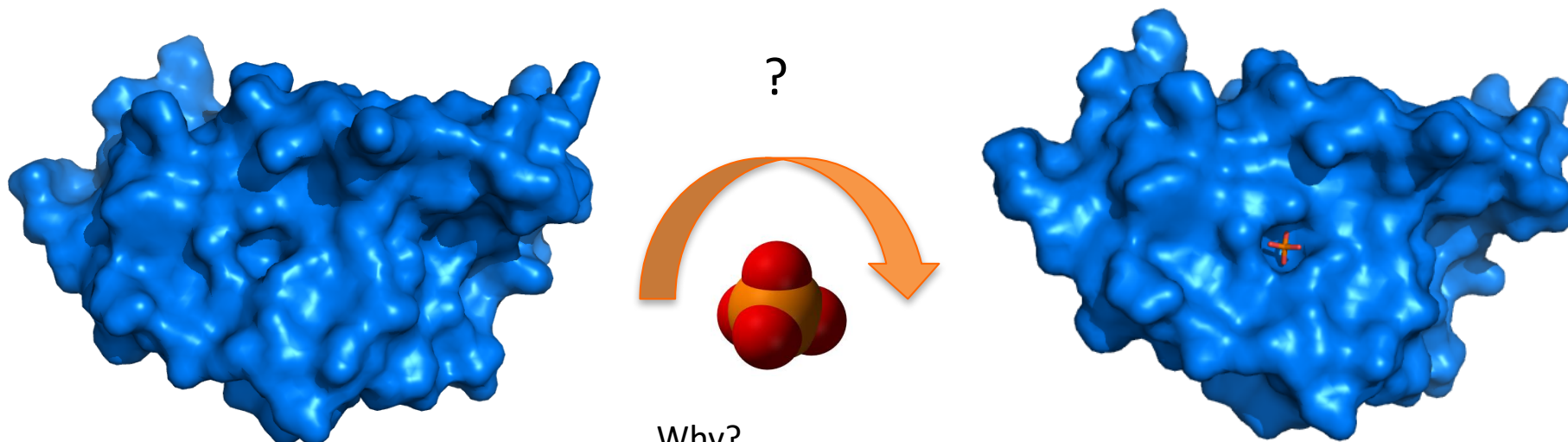
Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex.

Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using (energy as scoring functions).

Docking used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule.

Docking plays an important role in the rational design of drugs. [Wiki site](#)



Why?

How?

Biological relevance?

Binding pocket and ligands

Binding site:

Topology: crevice/groove/funnel/tunnel etc ...: show surface / generate ligand binding site

Electrostatic profile: charged? Polar? Hydrophobic?

Mutants which impede this interaction?

Comparison apo/holo : what are the flexible elements? Color by B-factor?

Accessibility? <http://sts.bioe.uic.edu/castp/calculation.php>

Druggability? <http://pockdrug.rpbs.univ-paris-diderot.fr/cgi-bin/index.py?page=home>

Ligand

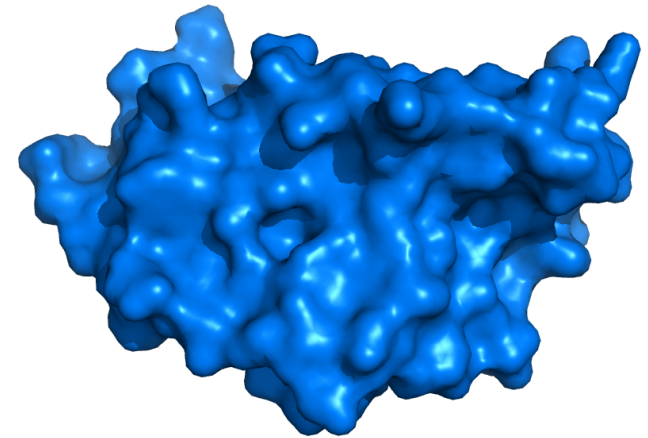
Protein: see webservers

ADN/ARN : voir sites web

Ligands : ADT tools/ DiscoveryStudio biovia©

Construction of ligands in mol2 format using PyMol

Prodrugs token/free registration mandatory



+



?

Scoring :

Total potential energy

Interaction energy

Docking of protein/protein

Rosetta : D. Baker's lab

<http://rosie.rosettacommons.org/docking2/submit>

<http://rosie.rosettacommons.org/>

Zdock : Z. Weng's lab

<http://zdock.umassmed.edu/>

Haddock : A. Bonvin's lab

<http://haddock.science.uu.nl/services/HADDOCK2.2/haddock.php>

Grammx server : I. A. Vakser's lab

<http://vakser.compbio.ku.edu/resources/gramm/grammx/>

Docking of ligands

ADT Auto-dock tools fichier à suivre. Pdf à envoyer

DiscoveryStudio © Biovia ex msi ex Accelrys

Visualisation for seminars, labmeetings

« photo »

Background

Surface transparency/fancy helix

Ray Save as png. Insérer une photo png par ppt

« video »

Open 3CM3.pdb

Movie/Program/Camera/X-roll/8seconds

Play

Voir pour plus d'informations www.pymolwiki.org/index.php/MovieSchool_1

« pse »

Save the PyMol session

Listing non exhaustif

<http://toolkit.tuebingen.mpg.de/hhpred>

<http://molprobit.biochem.duke.edu/>

<https://salilab.org/modeller/>

! Ne pas oublier de les citer dans les publications

<http://www.rcsb.org/pdb/home/home.do>

<http://esript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>

<http://www.ebi.ac.uk/pdbsum/>

<http://www.ebi.ac.uk/services>

(...)

<http://pockdrug.rpbs.univ-paris-diderot.fr/>

Take home messages

modélisation par homologie

Passer du temps sur l'analyse de la séquence et la qualité de l'alignement.

oligomérisation

Compiler données physico-chimiques (UAC, ITC), structurales (RX, RMN, SAXS) et biochimiques (mutations, tests d'activité). La modélisation seule ne suffit pas.

docking

L'affinité n'est pas tout, penser à l'accessibilité côté protéine et à la flexibilité côté ligand.

formation

Sites non exhaustifs. Boîte à outils toujours en évolution.

Gwenaelle.andre-leroux@jouy.inra.fr ou gandre@jouy.inra.fr

Veronique.martin@jouy.inra.fr

Structural Biology and Bioinformatics

A big big thank to Mabel, Andrea and to you

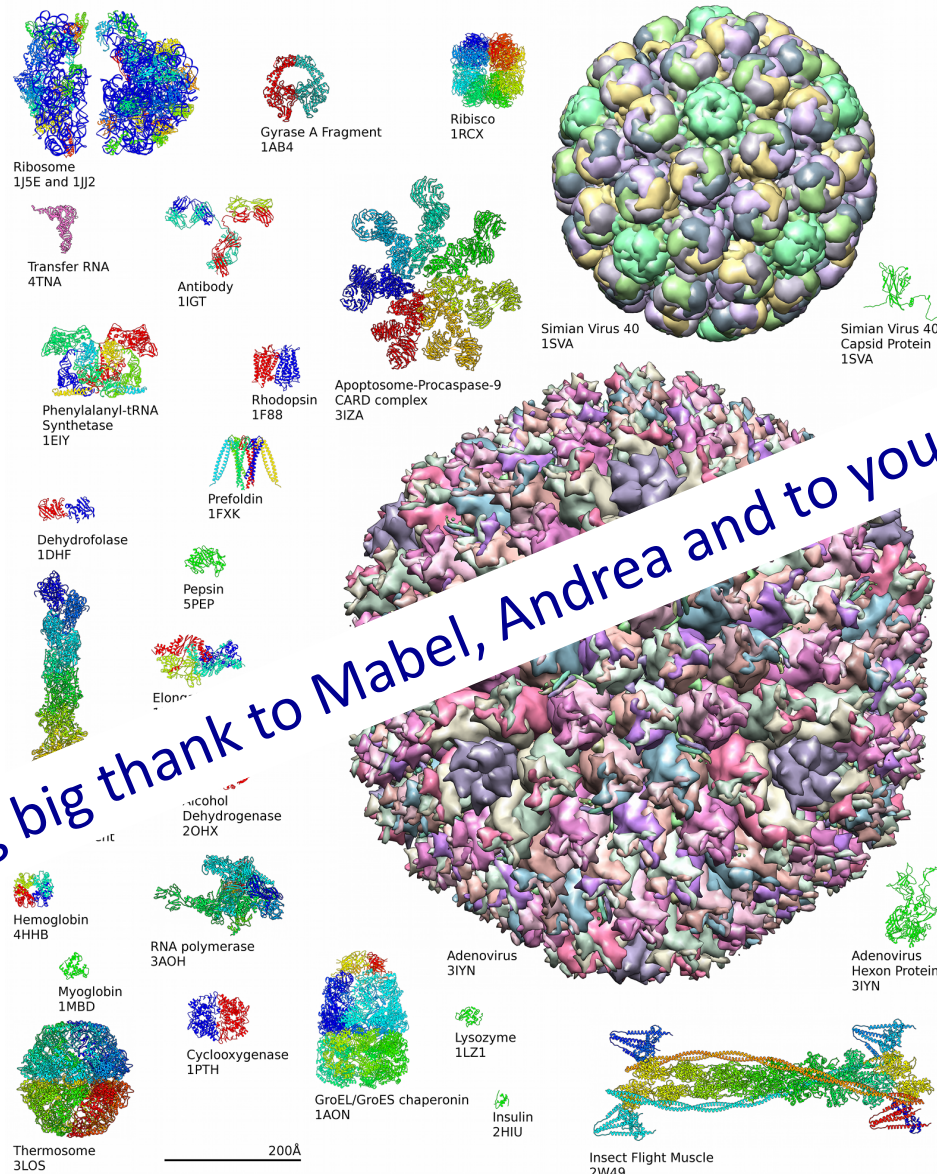


Image d'Axel Griewel



La nuit n'est jamais complète
 Il y a toujours puisque je le dis
 Puisque je l'affirme
 Au bout du chagrin
 une fenêtre ouverte
 une fenêtre éclairée
 Il y a toujours un rêve qui veille
 désir à combler
 faim à satisfaire
 un cœur généreux
 une main tendue
 une main ouverte
 des yeux attentifs
 une vie : la vie à se partager

Paul Eluard



Pablo Picasso



Luc Gwiazdzinski