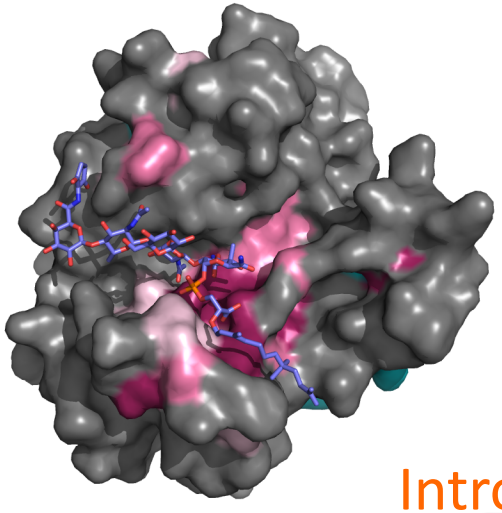
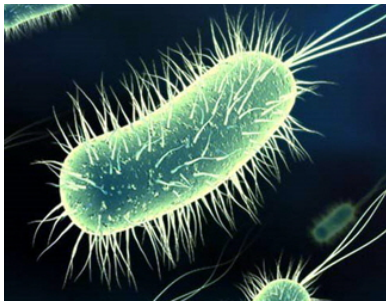
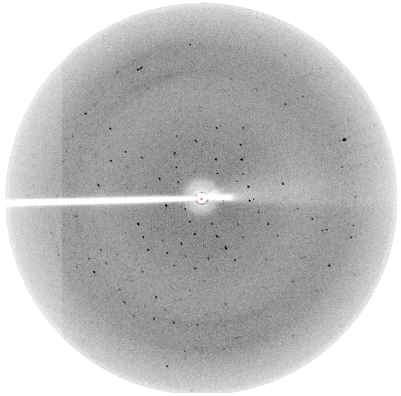
Introduction to Structural Bioinformatics : proteins and DNA

Gwenaëlle André-Leroux

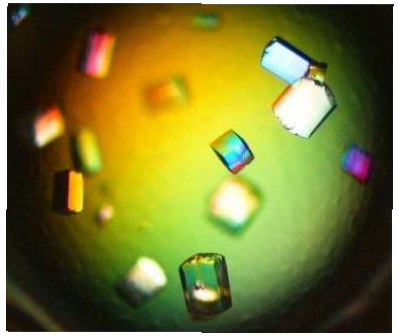




Introduction to Structural Bioinformatics: Proteins and DNA



>3CM3:A|PDBID|CHAIN|SEQUENCE
GPEIRMDKKSLYKYLLLRSTGDMHKAKSP
TIMTRVTNNVYLGNYKNAMDAPSSEVKF
KYVLNLTMDKYTLPNSNINIIHIPLVDDTTT
DISKYFDDVTAFLSKCDQRNEPVLVHSAA
GVNRSGAMILAYLMSKNKESLPM



Gwénaëlle André-Leroux
MaIAGE

Montevideo december 2015

INRA SCIENCE & IMPACT

Mathématiques et Informatique Appliquées du Génome à l'Environnement MaIAGE

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

INRA SCIENCE & IMPACT

CORNELL UNIVERSITY FOUNDED A.D. 1865

INSTITUT PASTEUR

Structural biologist and computational scientist

MaIAGE Mathématiques et Informatique Appliquées du Génome à l'Environnement

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

Coffee breaks

3. Molecular mechanics and docking

3.1. Binding siteidentification and characterization

3.2. Analysis of ligands

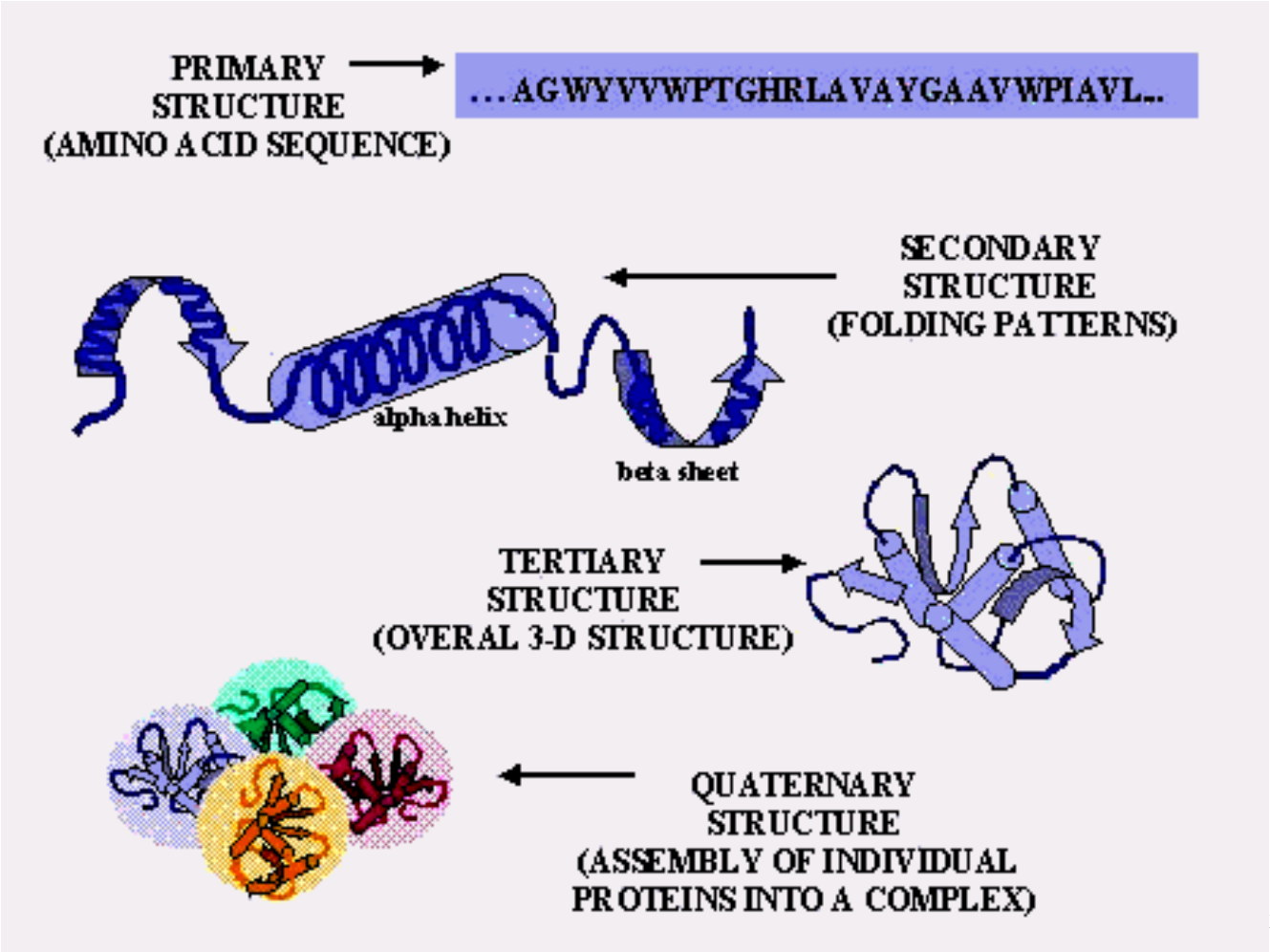
3.3. Docking: rigide, flexible

Conclusions, questions, comments

Mathématiques et Informatique Appliquées du Génome à l'Environnement
MaIAGE

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

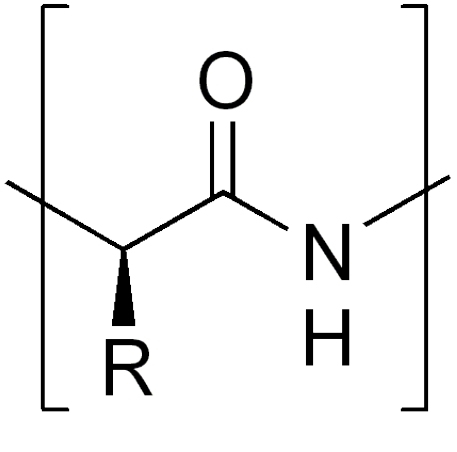
It refers to the sequence of amino acids present in the polypeptide chain

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

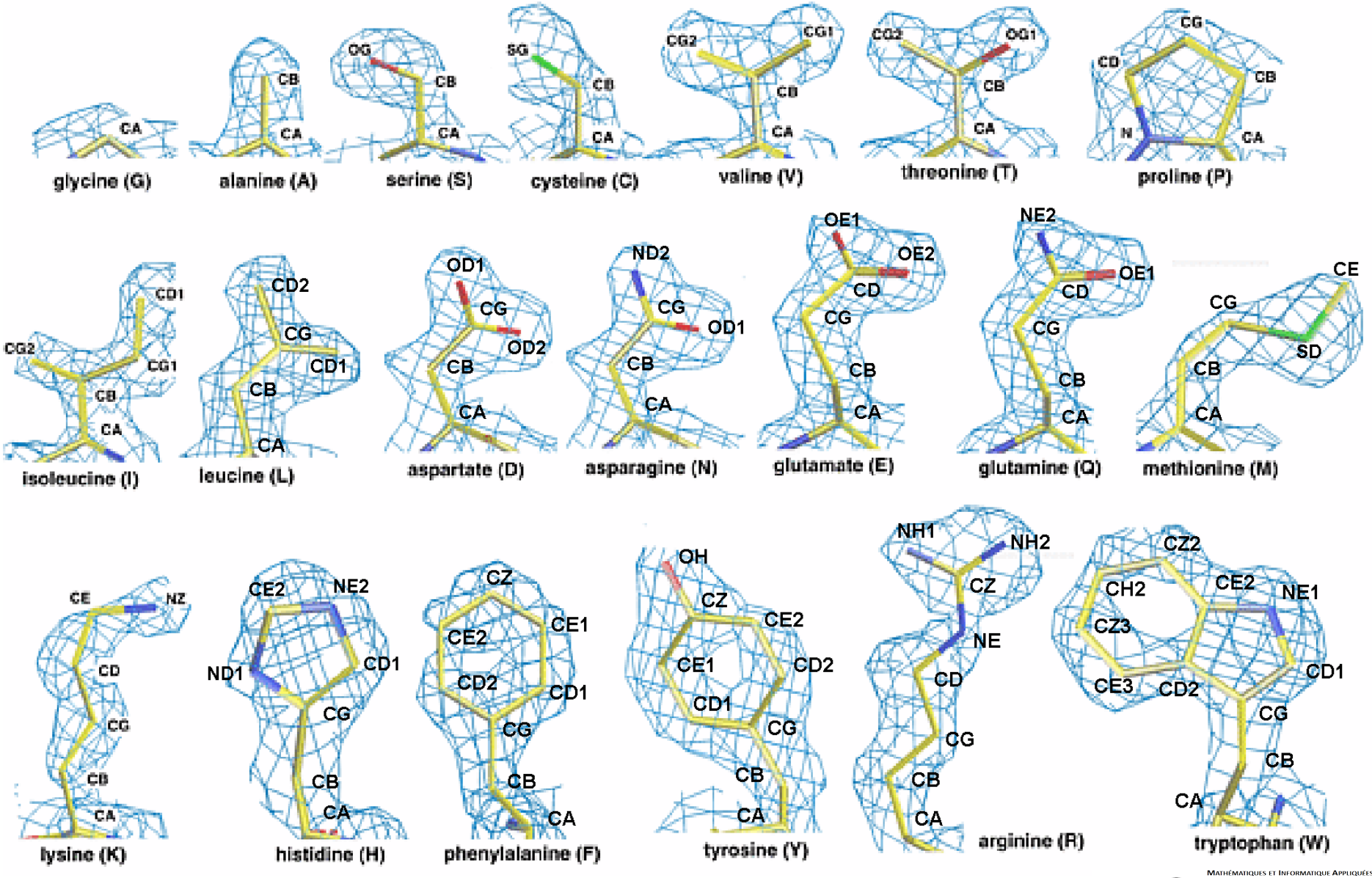


Amino acid residue in a peptidic chain

Each letter refers for an aa residue:
A for Alanine
C for Cysteine ...

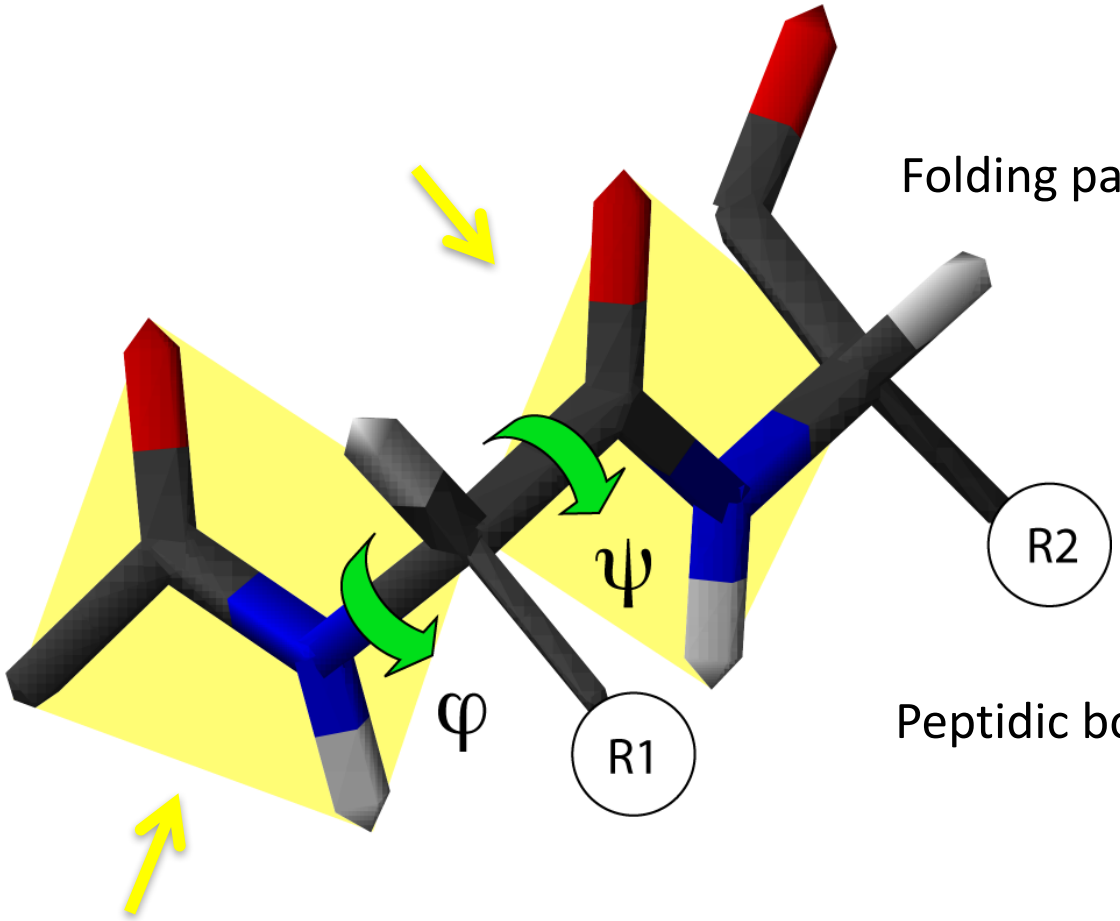
R refers to the side chain of the amino acid

Residues: small to large, polar to apolar



Proteins: secondary structure

The peptidic chain shows dihedral angles



Folding patterns into sheets, helix, coils etc …

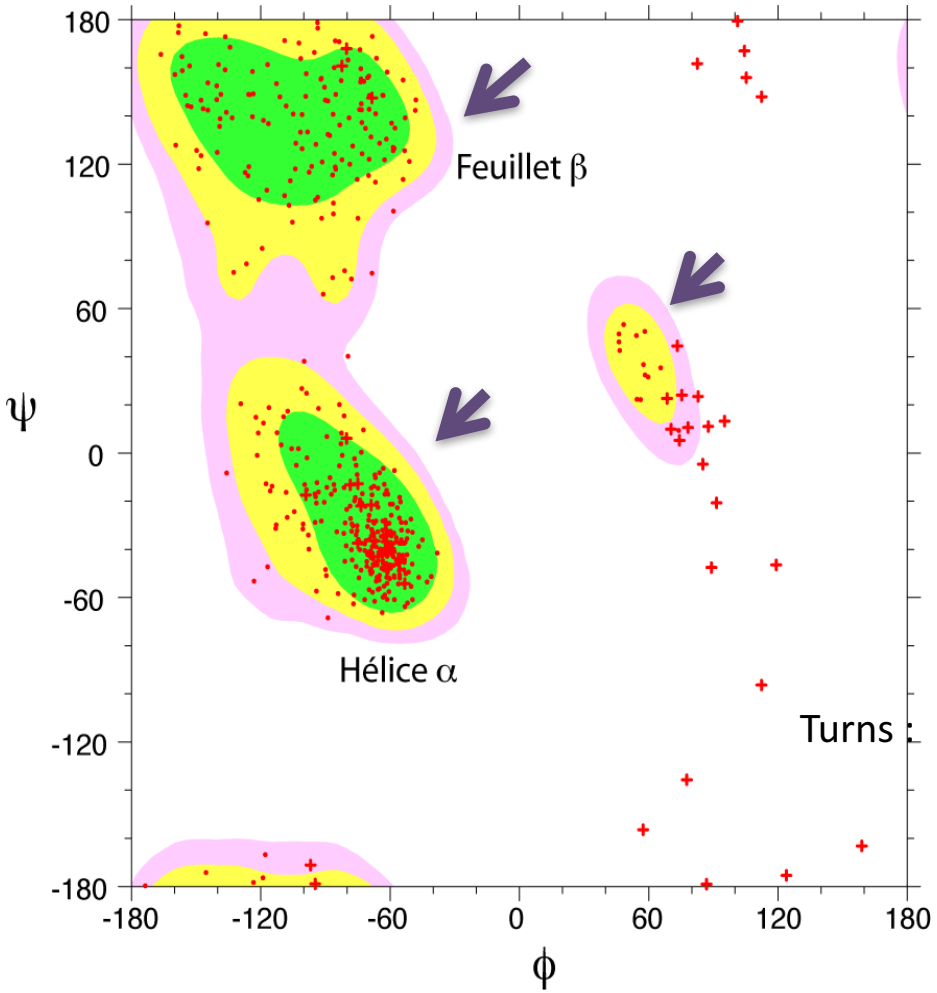
Peptidic bond is planar

Consequence: successive plans linked by dihedrals (φ,ψ)

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

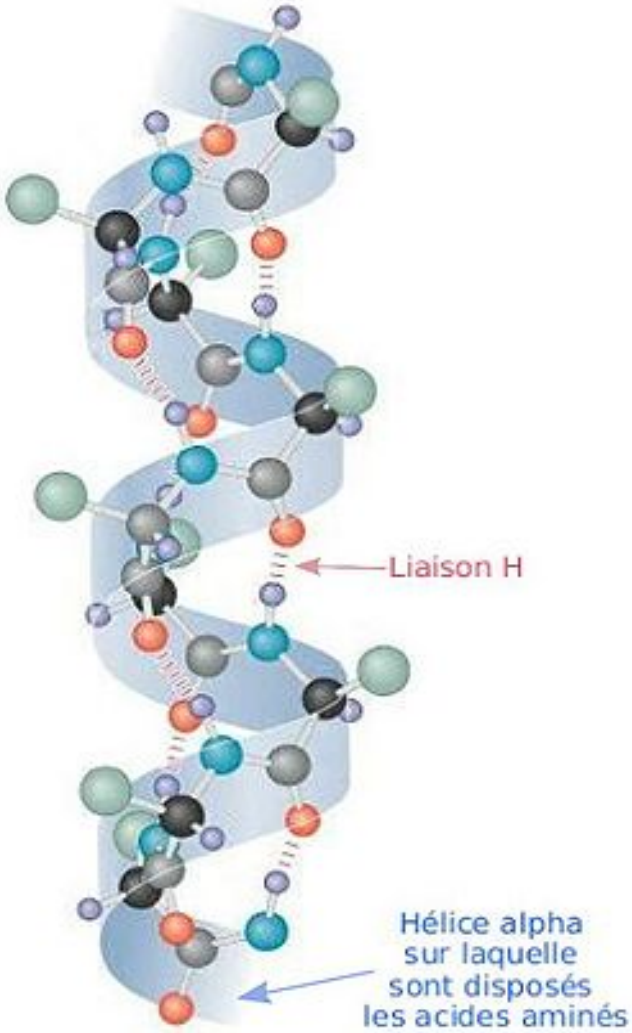
Hélice α : helical periodic structure.

Strand β : flat periodic structure flat. Assemble into sheets

Turns :

Proteins: secondary structure

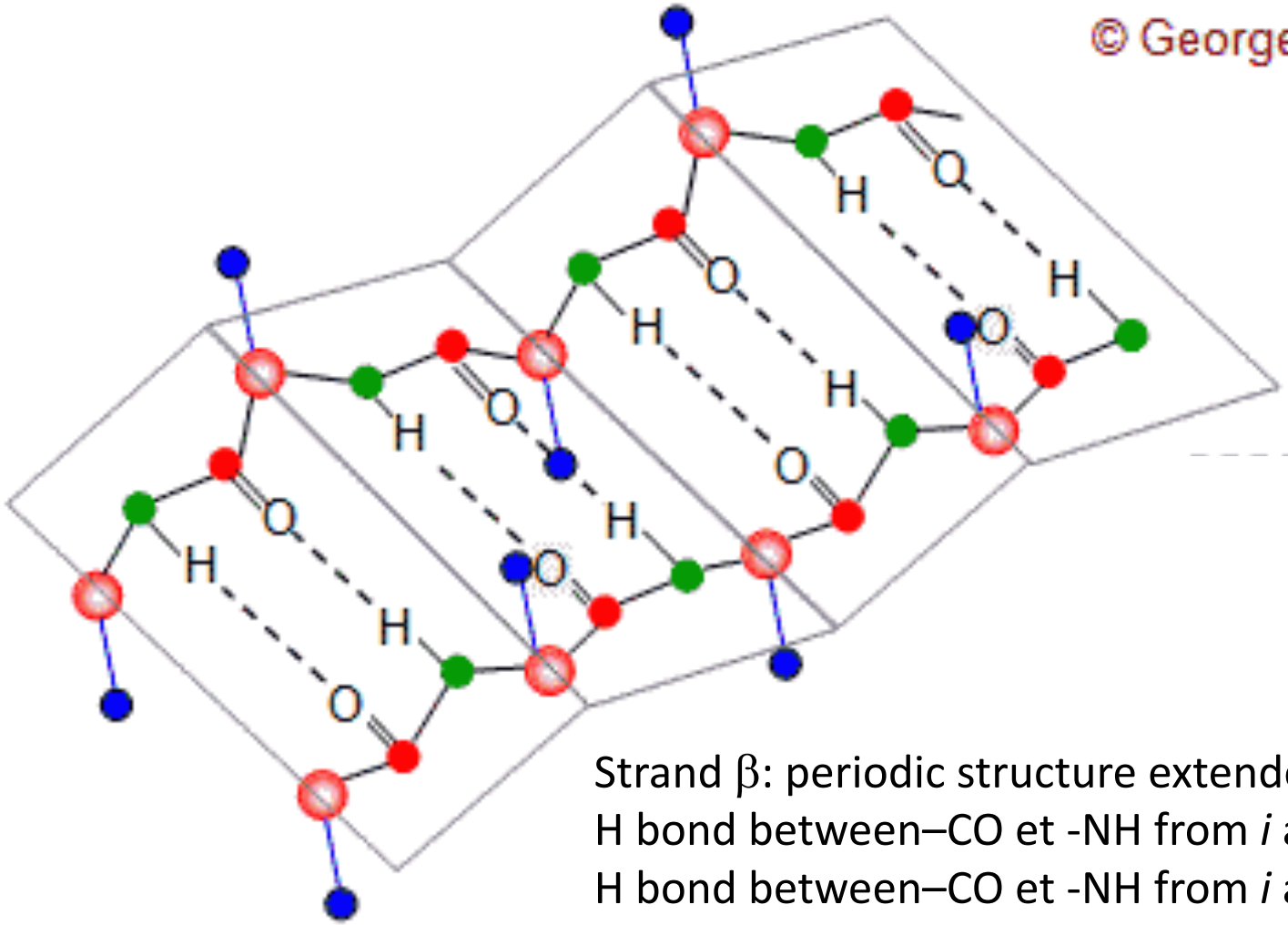
Folding patterns into sheets, helix, coils etc …



α Helix: periodic structure.
H bond between –CO from *i* and-NH from *i+4*

Proteins: secondary structure

Folding patterns into sheets, helix, coils etc …

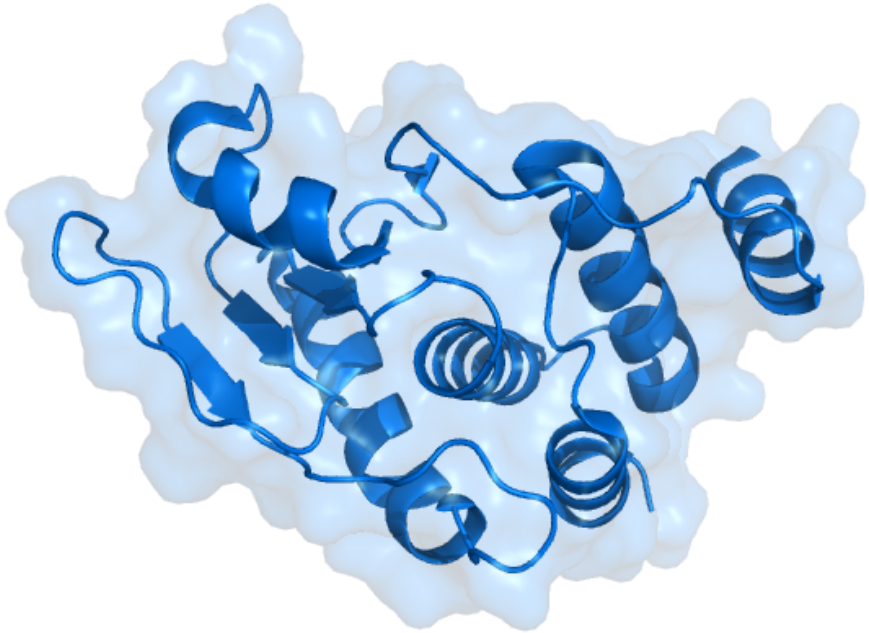


Strand β: periodic structure extended. Forms β sheets
H bond between–CO et -NH from *i* and *j* -> antiparallel
H bond between–CO et -NH from *i* and *j+2* -> parallel

Proteins: tertiary structure

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

3D structure correlates to function.



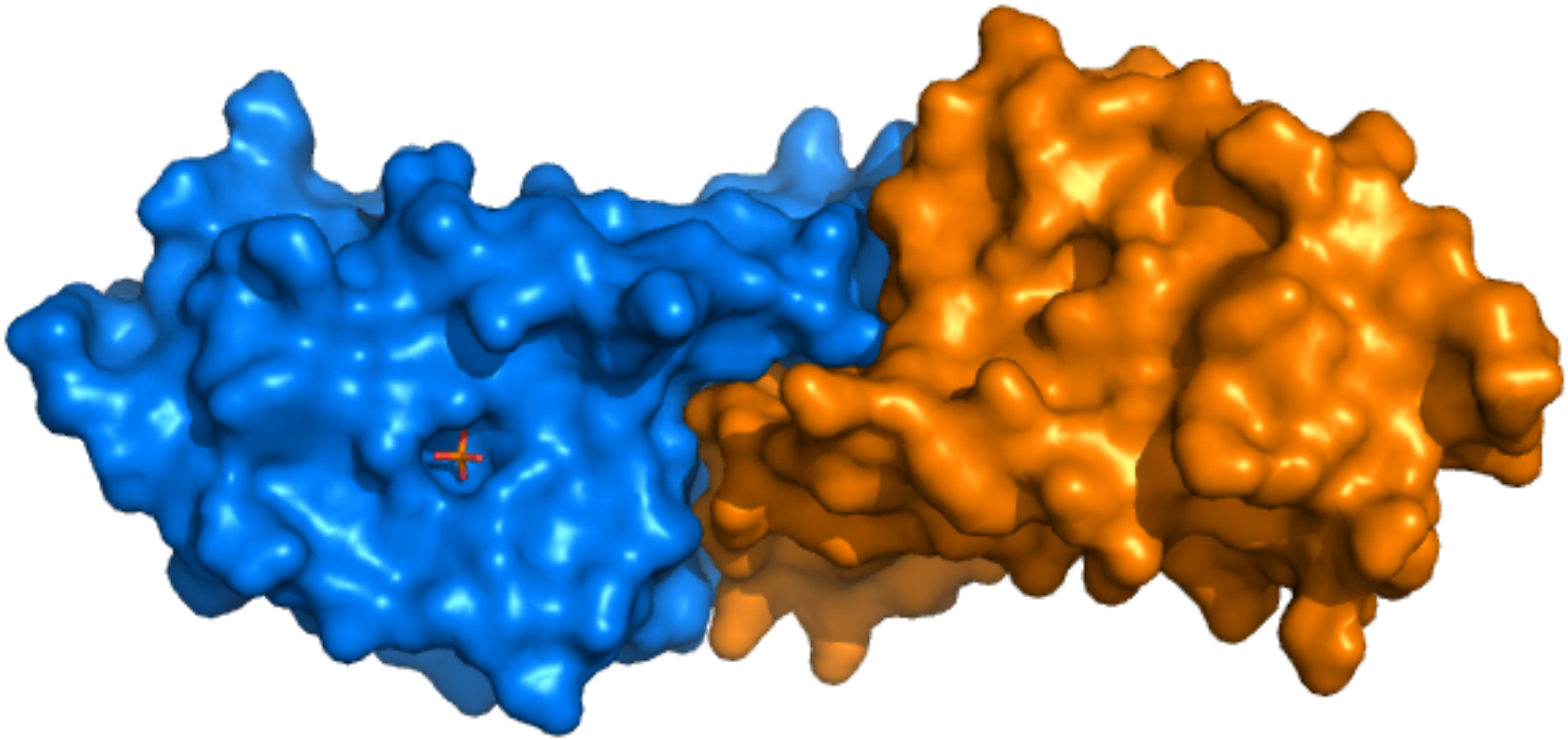
Several interactions garantee its fold and its stability

- Covalent bonds (disulfide bridges between cysteine residues)
- Electrostatic bonds (ioninc 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

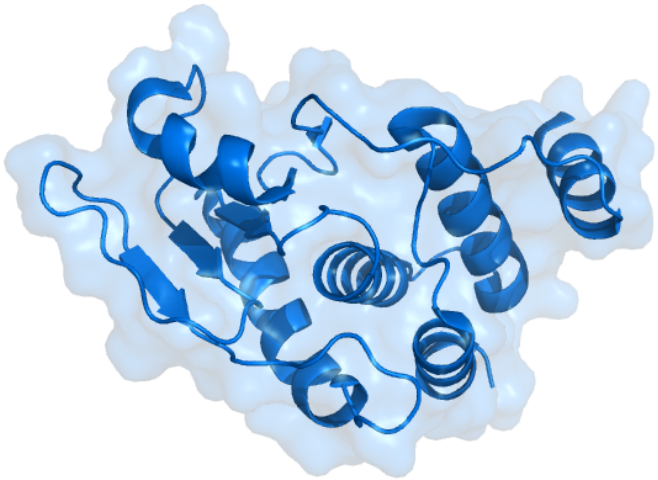


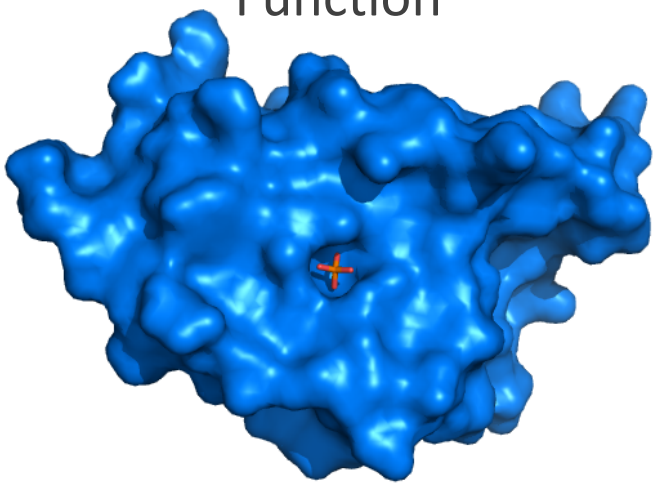
Protein Structure

1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE
GPEIRMDKKSLYKYLLLRSTGDMHKAKSPTIMTRVT
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL
PNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQR
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY
FLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDKN
```

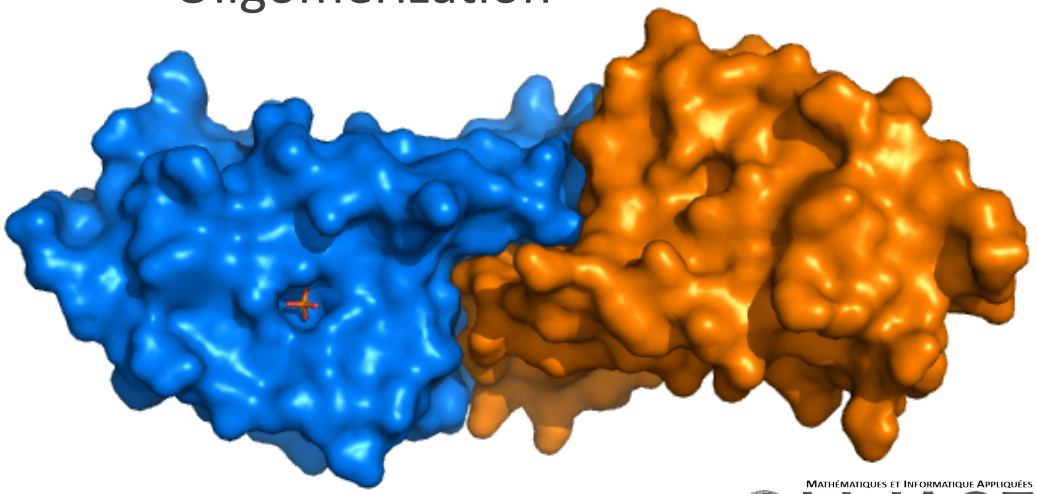
3D Fold





Function

Oligomerization

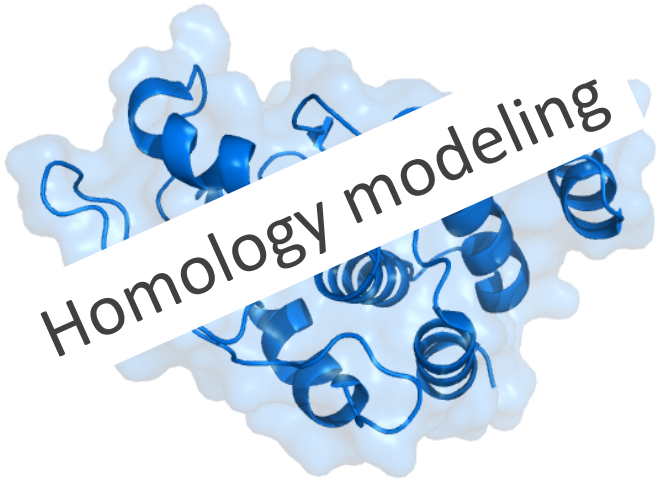


Protein Structure

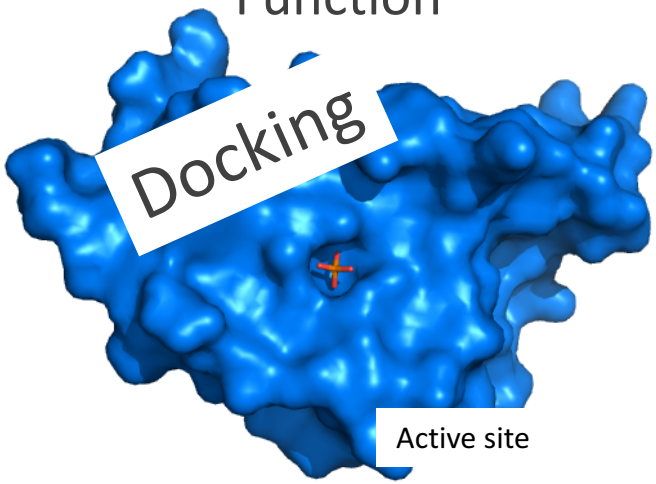
1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE
GPEIRMDKKSLYKYLLLRSTGDMHKAKSPTIMTRVT
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL
PNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQR
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY
FLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDKN
```

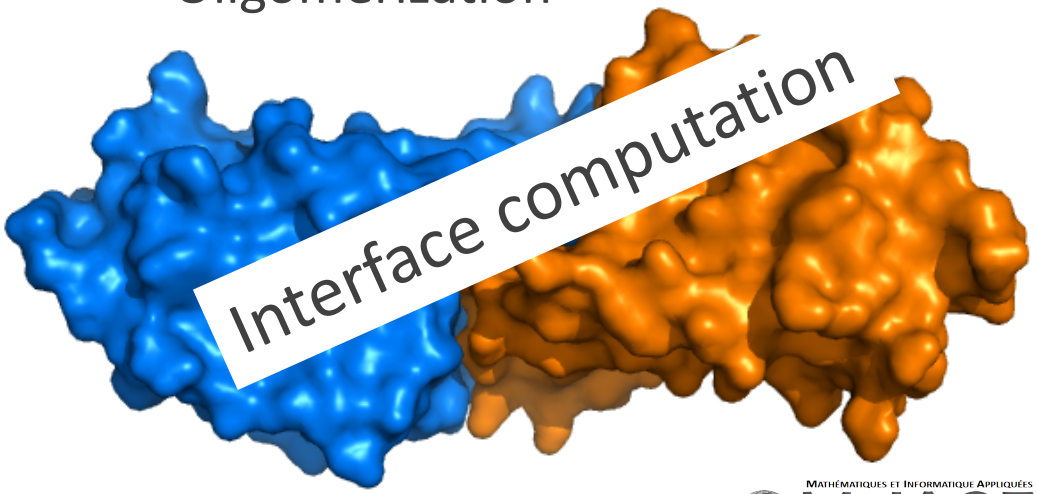
3D Fold



Function



Oligomerization



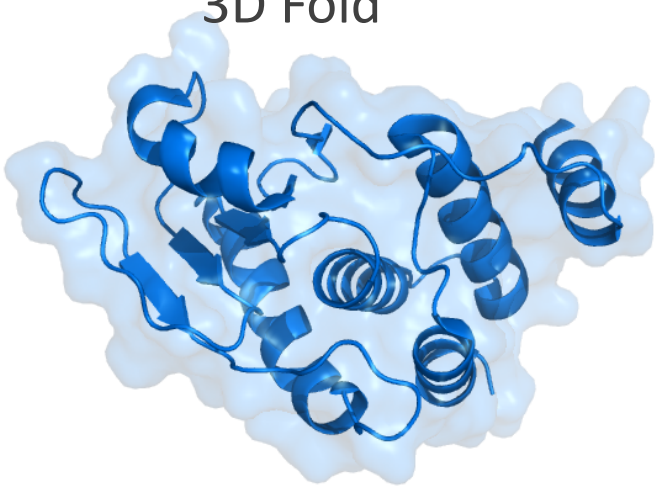
Protein Structure

1D Sequence

```
>3CM3:A|PDBID|CHAIN|SEQUENCE
GPEIRMDKKSLYKYLLLRSTGDMHKAKSPTIMTRVT
NNVYLGNYKNAMDAPSSEVKFKYVLNLTMDKYTL
PNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQR
NEPVLVHSAAGVNRSGAMILAYLMSKNKESLPMLY
FLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDKN
```

?

3D Fold



Why?

How?

Biological consistency?

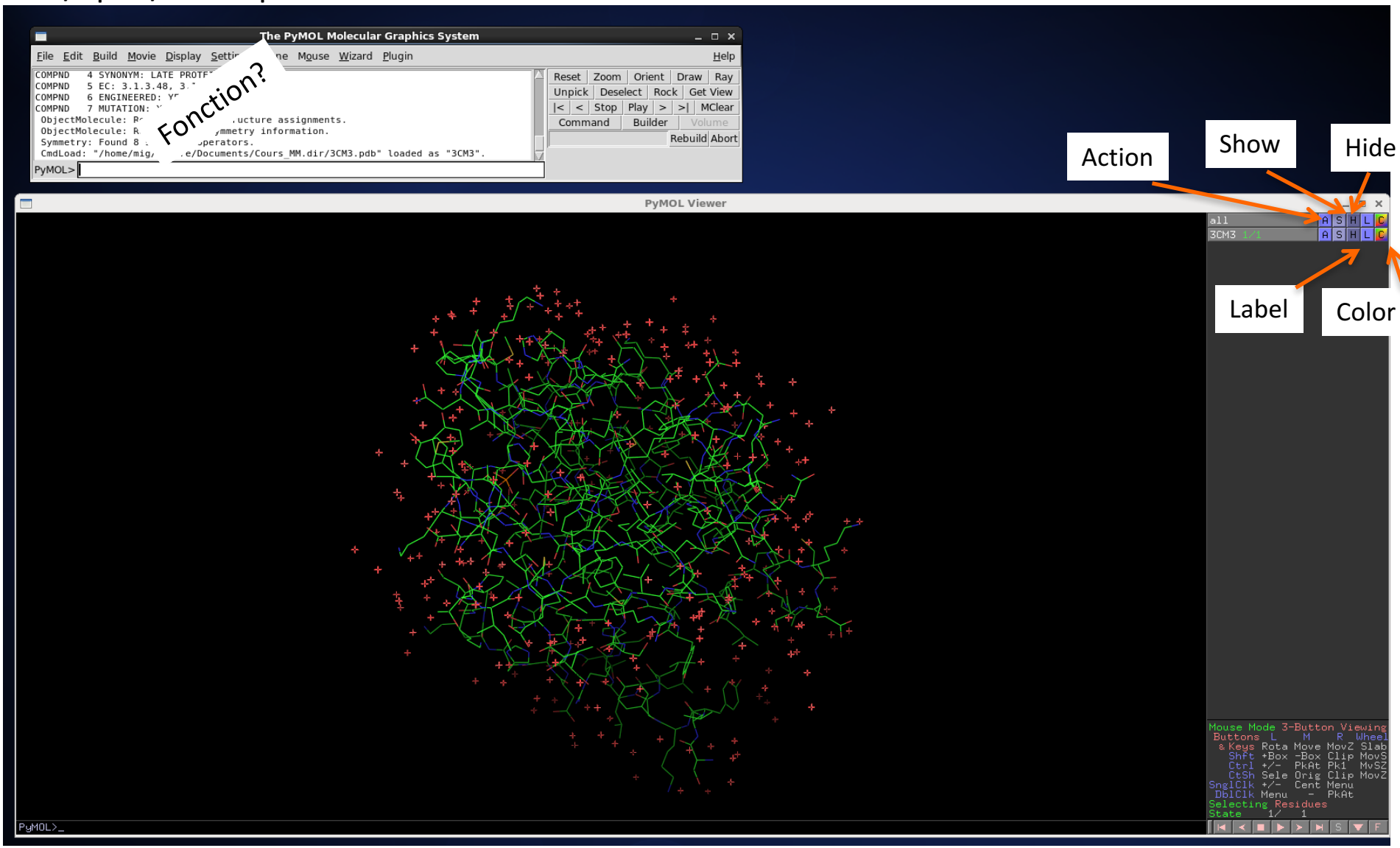
Protein Structure

double click on PyMol

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

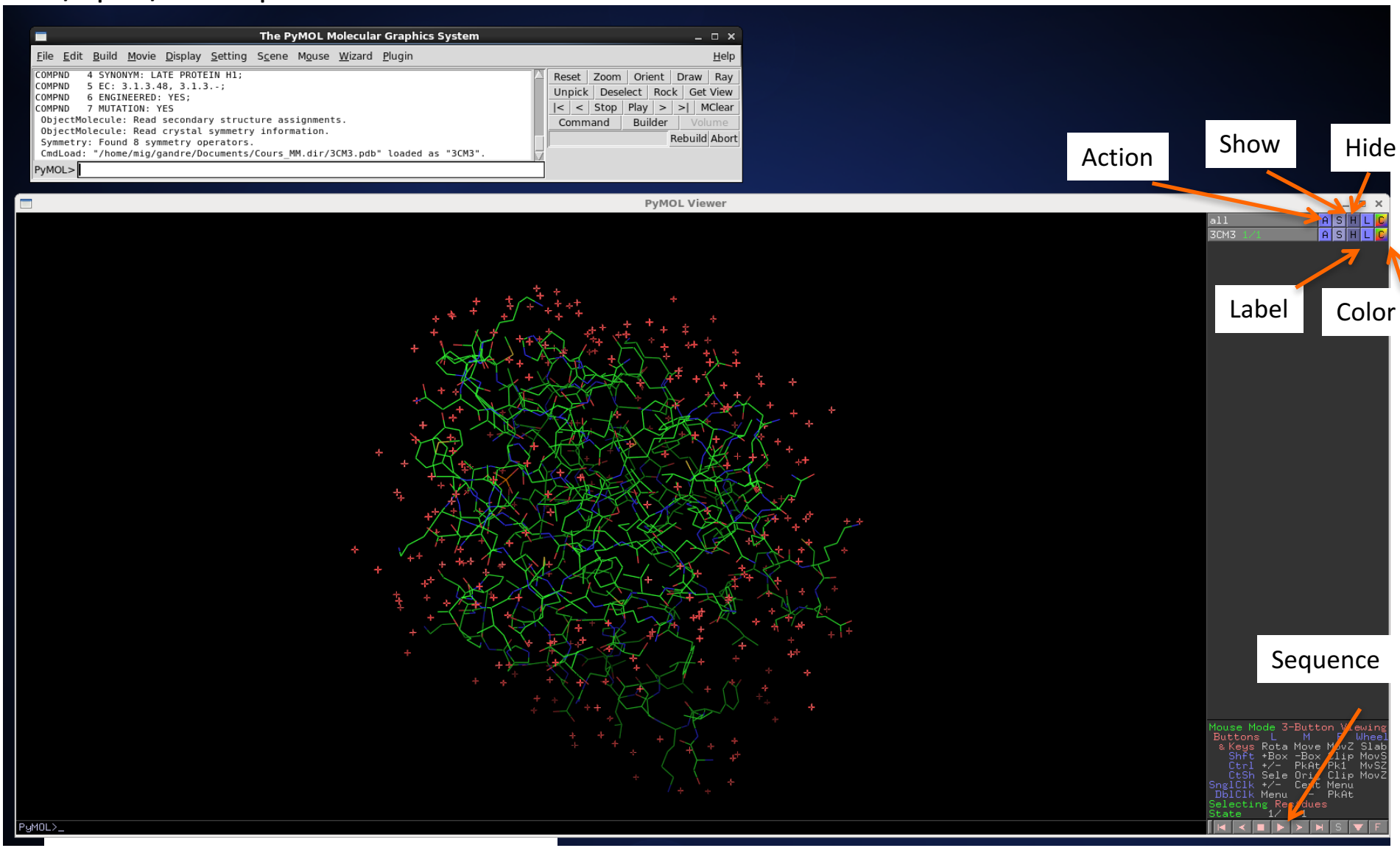
Analysis of 3CM3.pdb

File /open/3CM3.pdb



Analysis of 3CM3.pdb

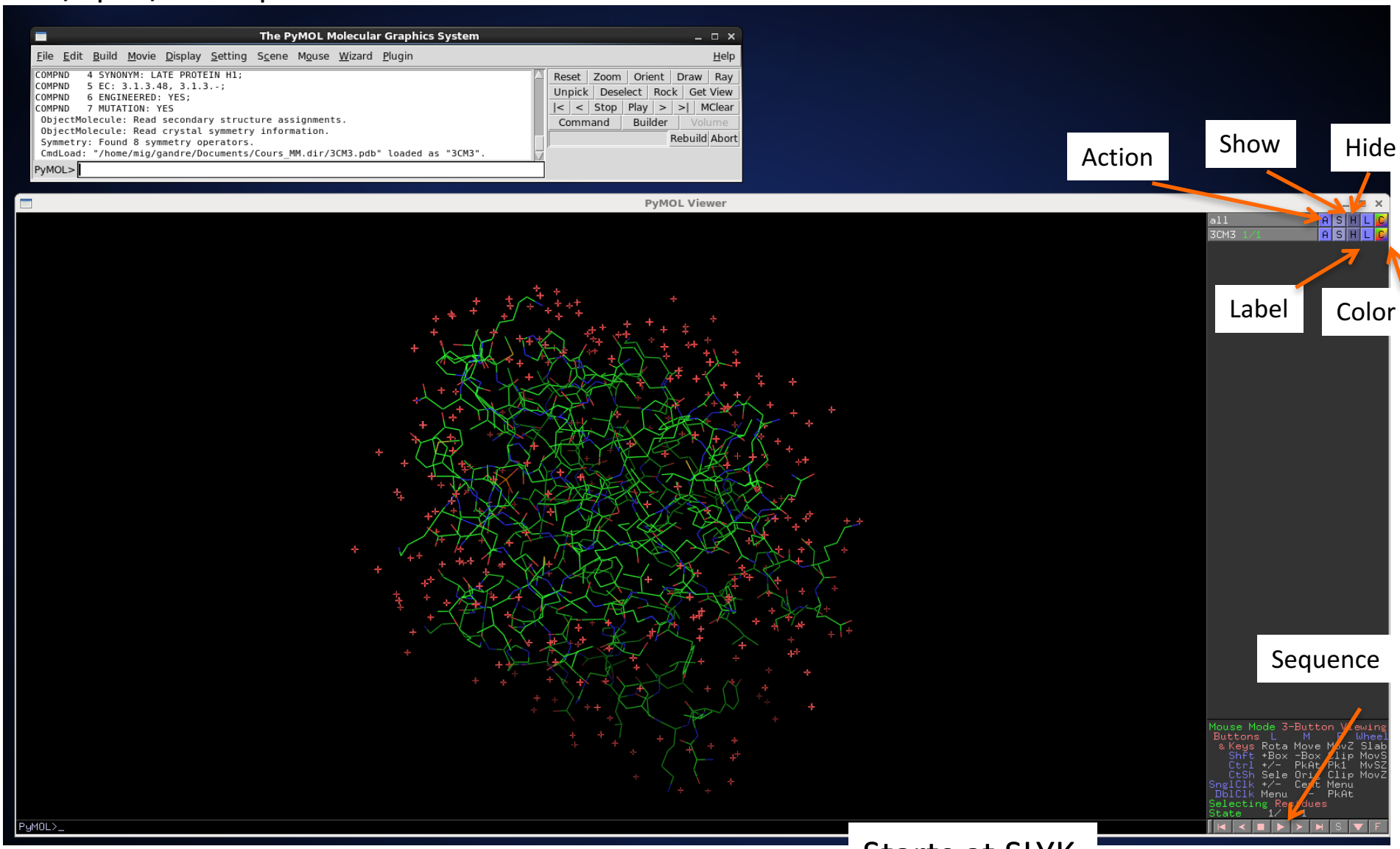
File /open/3CM3.pdb



Keywords of 3CM3? Starts at?

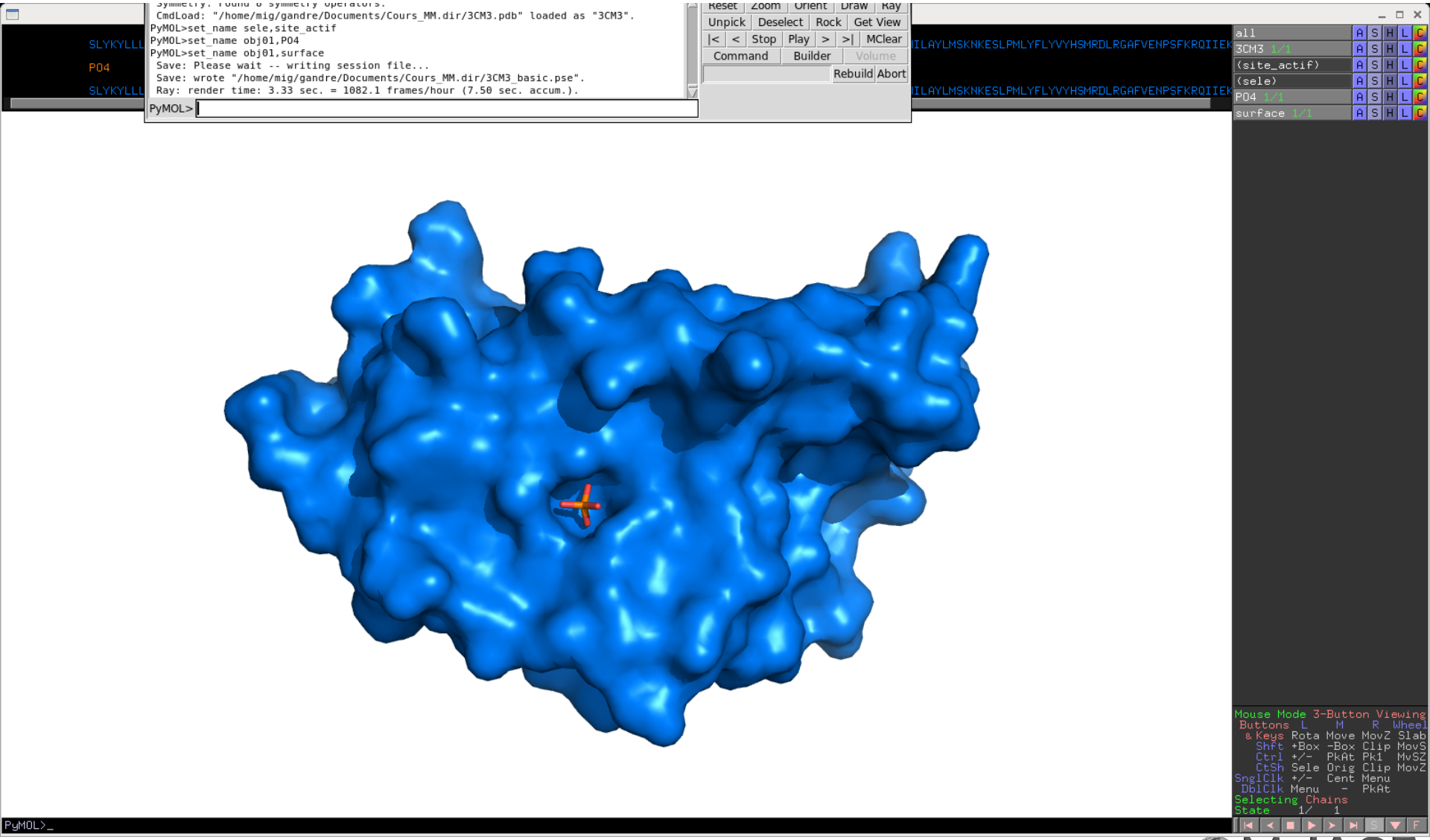
Analysis of 3CM3.pdb

File /open/3CM3.pdb



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

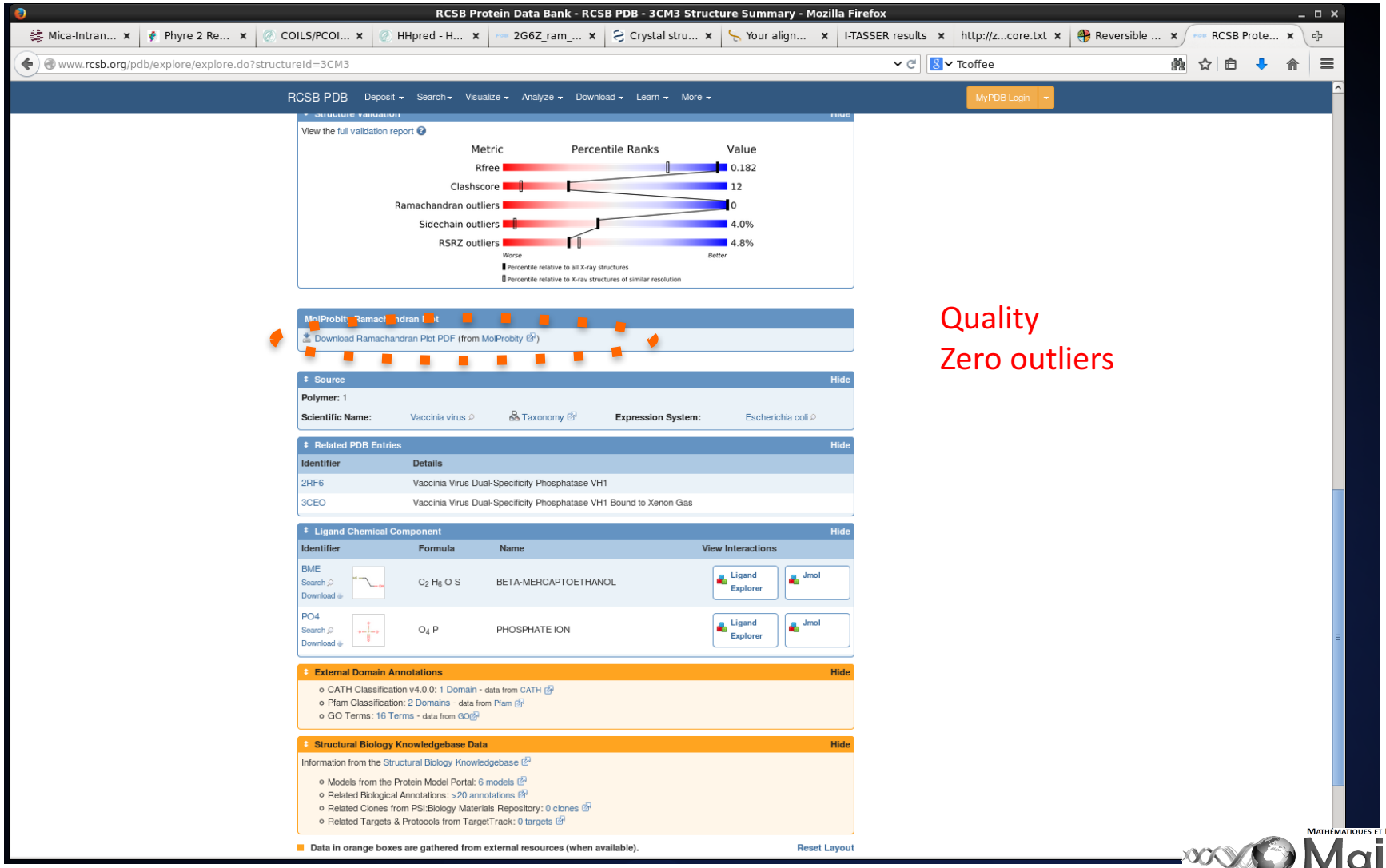
http://www.rcsb.org/pdb/home/home.do

3CM3 download fasta &pdb. Piece of informations?

Analysis of 3CM3.pdb

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

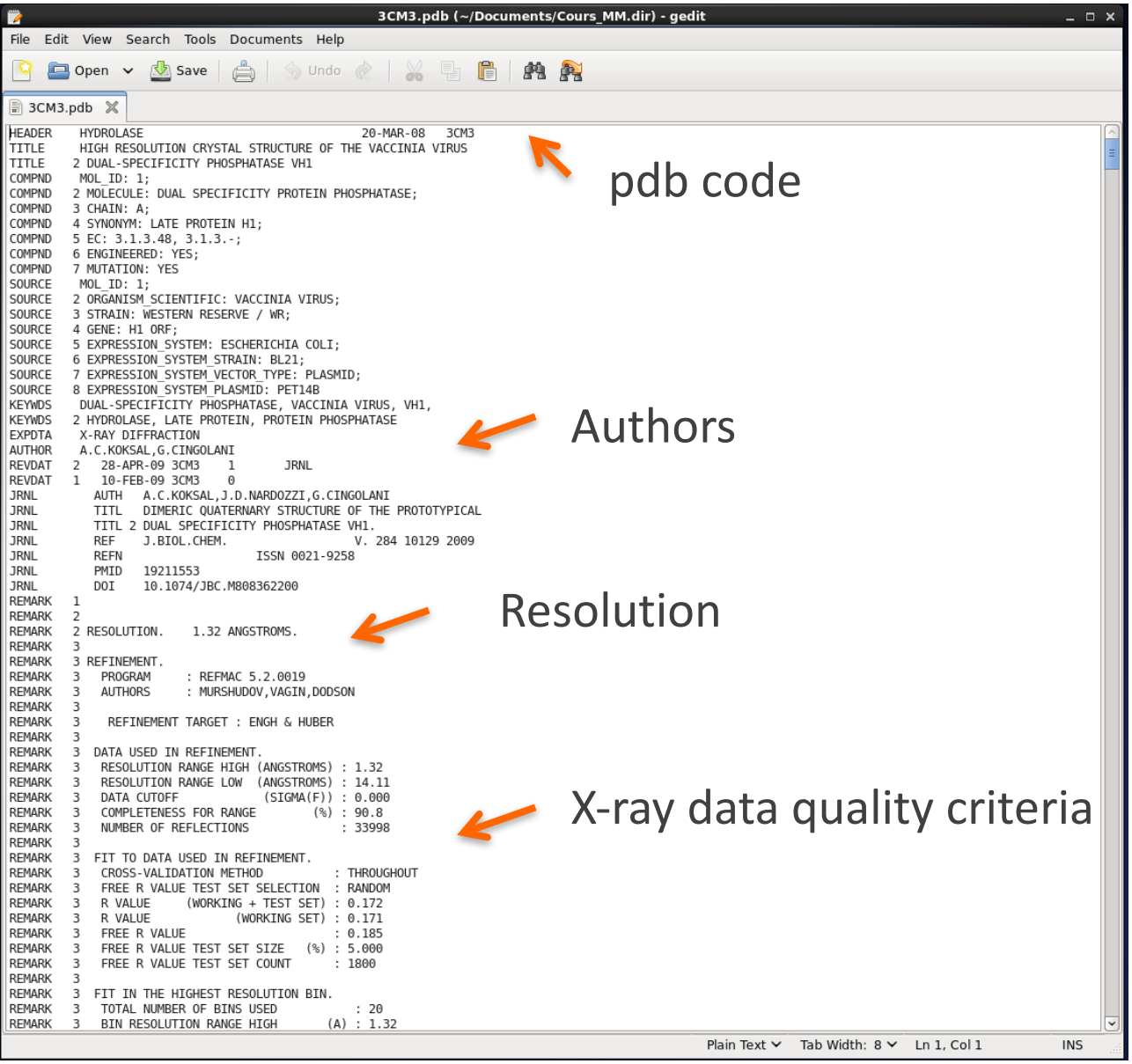
3CM3 download fasta et pdb. Infos?



Quality
Zero outliers

Analysis of 3CM3.pdb

gedit 3CM3.pdb



pdb code

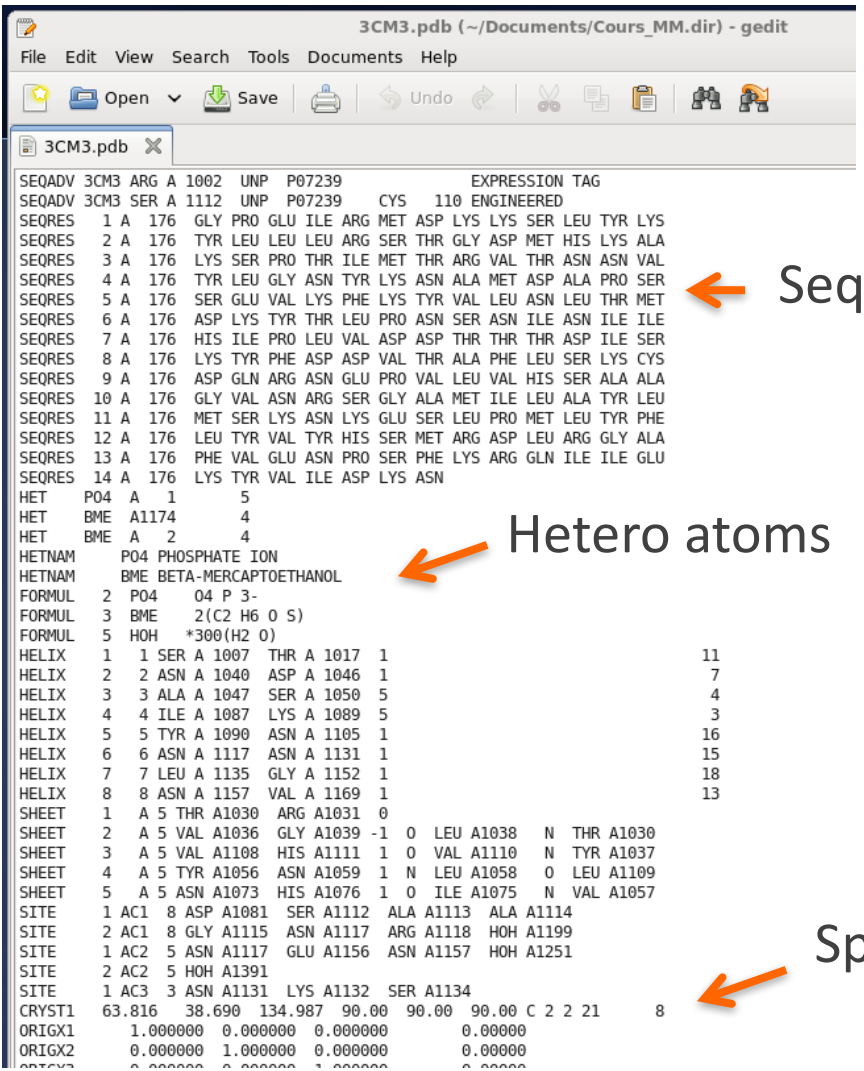
Authors

Resolution

X-ray data quality criteria

Analysis of 3CM3.pdb

gedit 3CM3.pdb



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
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

```
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        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.015670  0.000000  0.000000        0.00000
SCALE2      0.000000  0.025846  0.000000        0.00000
SCALE3      0.000000  0.000000  0.007408        0.00000
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

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        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.015670  0.000000  0.000000        0.00000
SCALE2      0.000000  0.025846  0.000000        0.00000
SCALE3      0.000000  0.000000  0.007408        0.00000
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

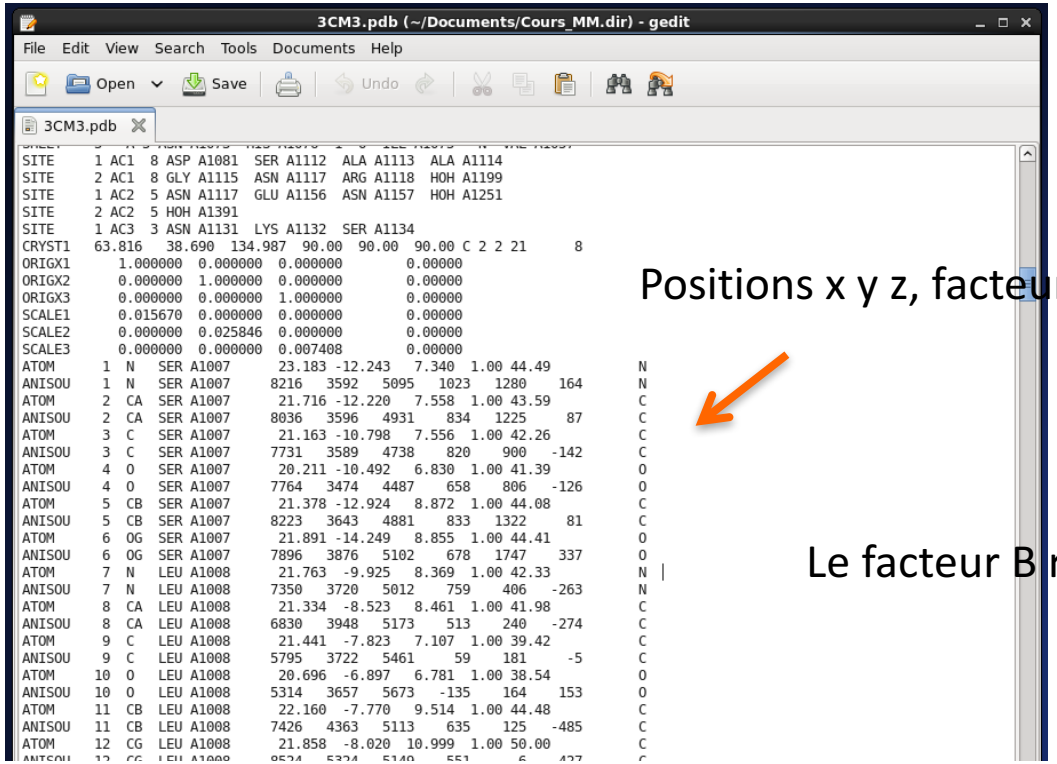
Go to Ile 1170

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

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

Analysis of 3CM3.pdb

gedit 3CM3.pdb



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

Le facteur B reflète l'agitation thermique dans le cristal

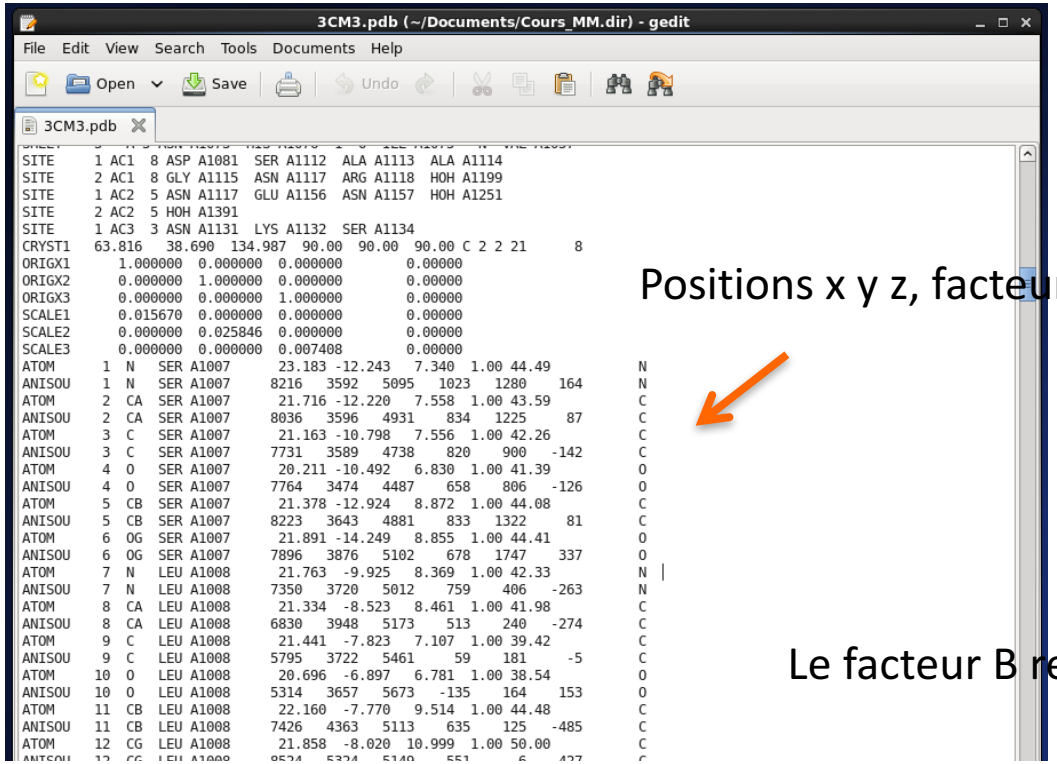
Aller de nouveau jusqu'à Ile 1170

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



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)

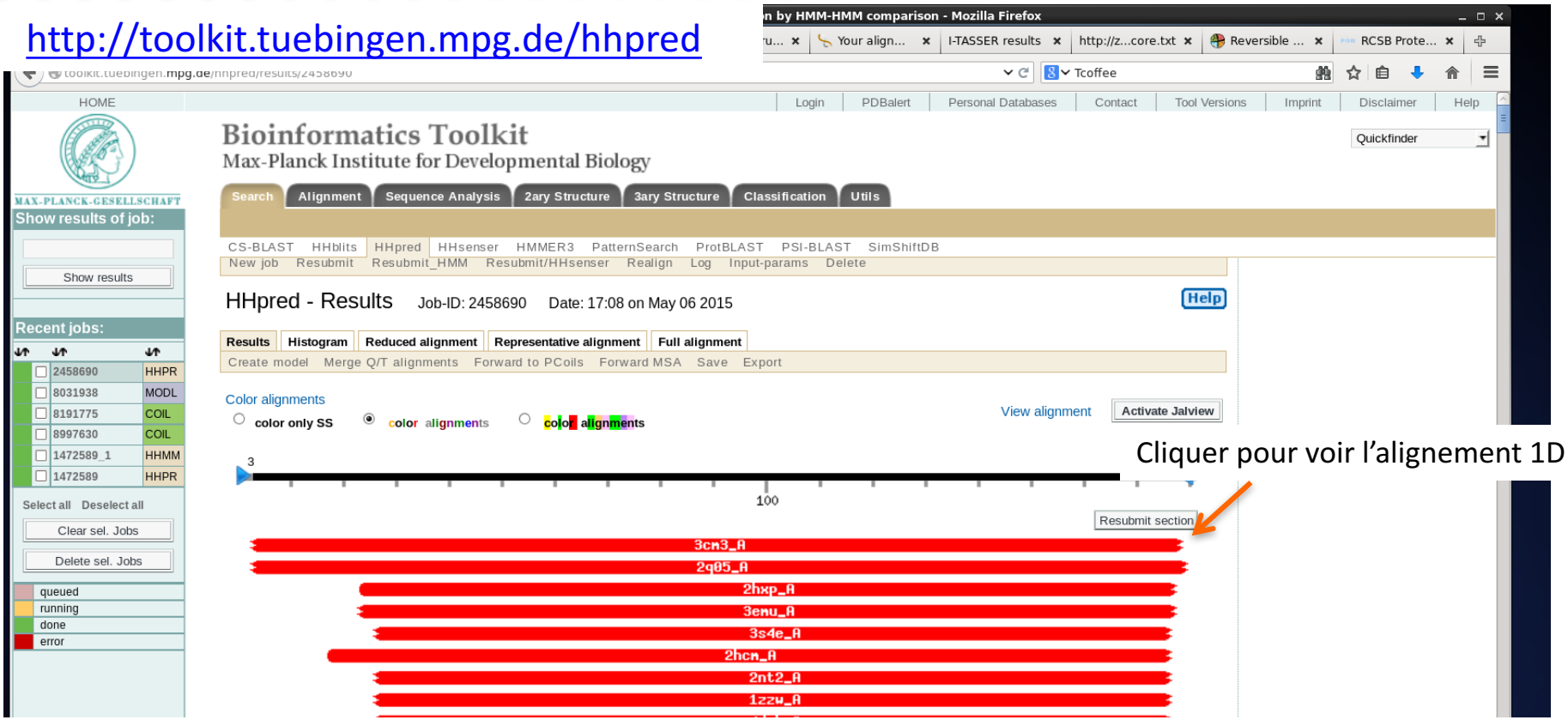
1) Identification de la/des protéines « templates » de référence
2) Modélisation de la protéine requête « query »

Qualité de l'alignement +++

http://robetta.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



Cliquer 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://espript.ibcp.fr/ESPript/cgi-bin/ESPript.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

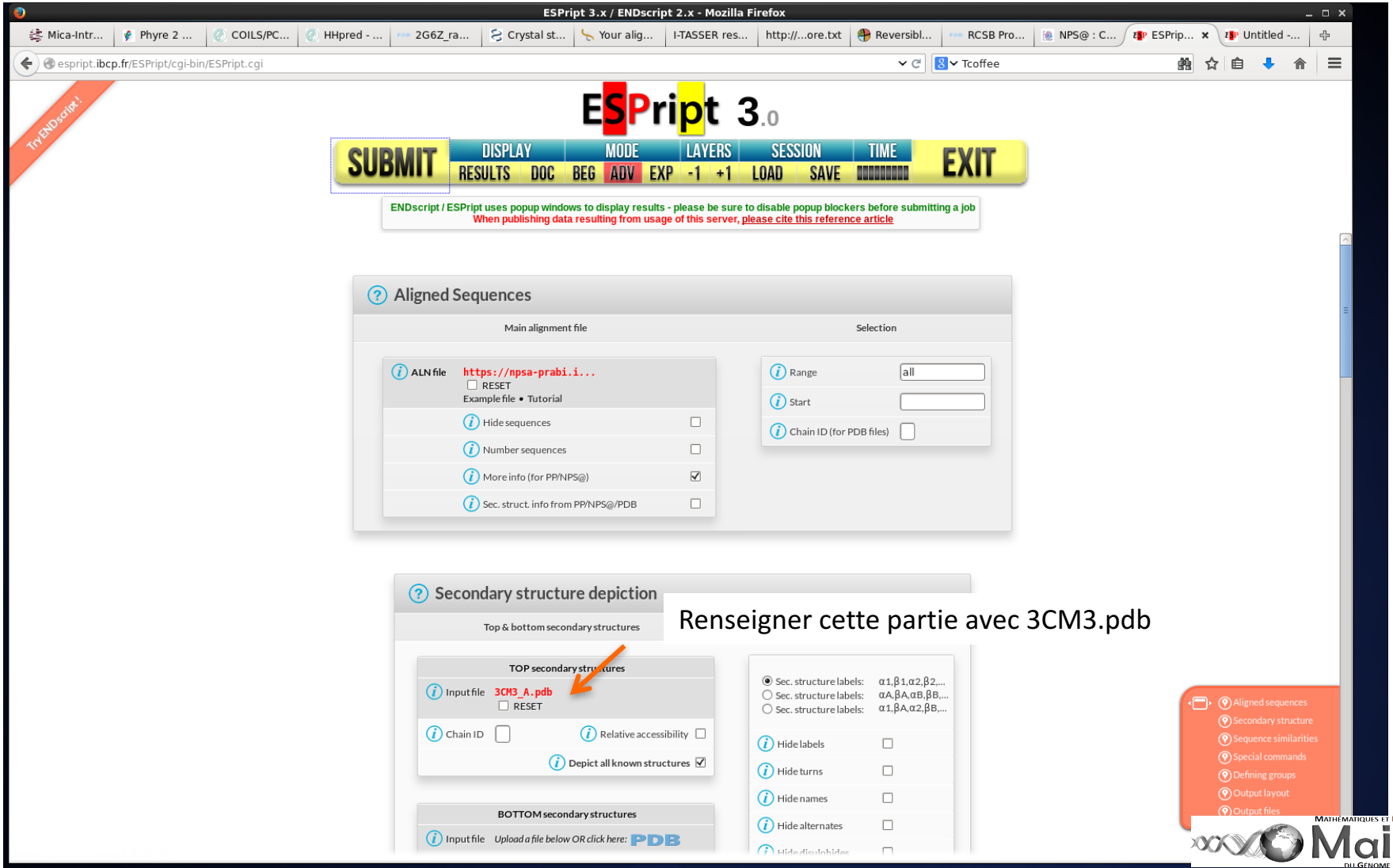
Homology modeling

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

Renseigner cette partie avec 3CM3_2Q05_PTP_Orf.fasta

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

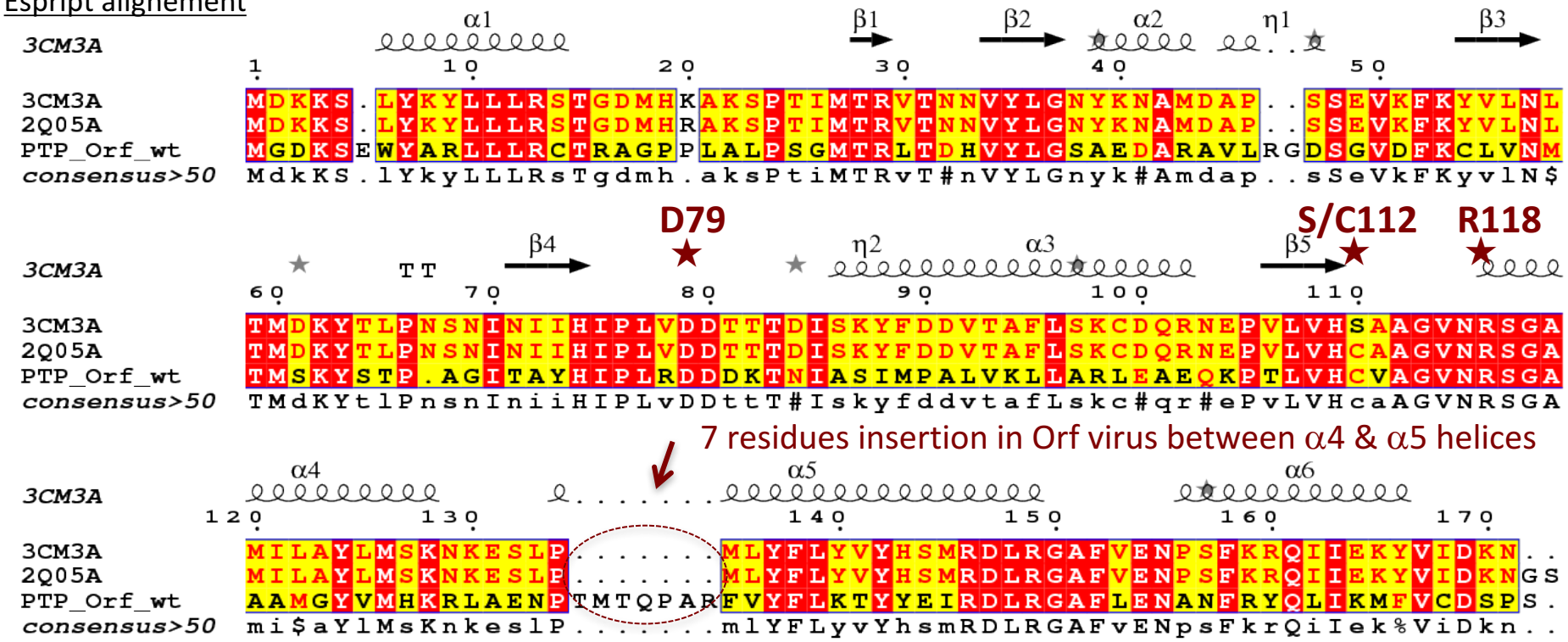


Renseigner cette partie avec 3CM3.pdb

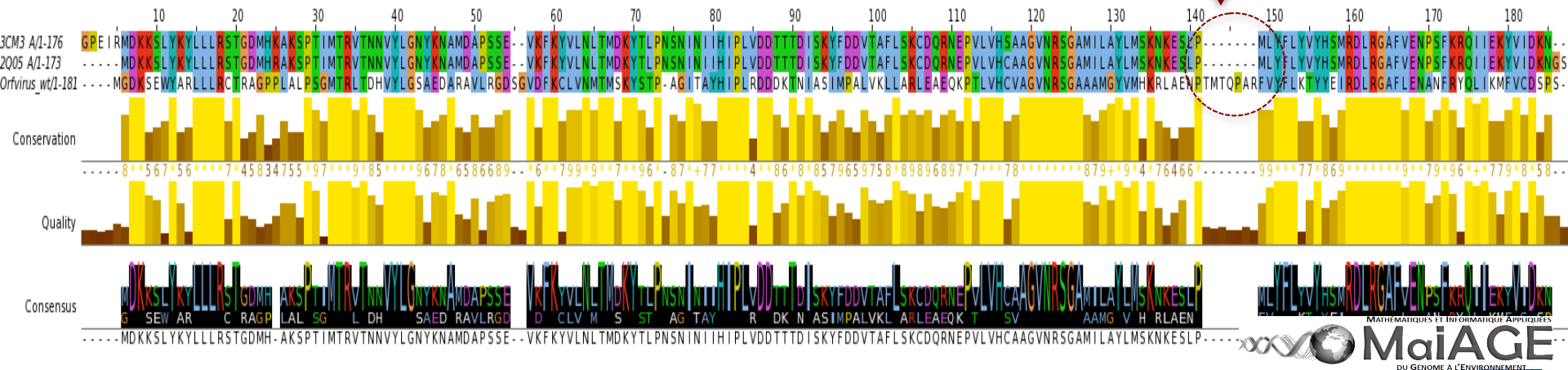
MaIAGE
Mathématiques et Informatique Appliquées du Génome à l'Environnement

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

Espript alignement


3CM3A
α1 β1 β2 α2 η1 β3
1 10 20 30 40 50
3CM3A MDKKS.LYKYLLLRSTGDMHKAKSPTIMTRVTNNVYLGNYKNAMDAP..SSEVKFKYVLNL
2Q05A MDKKS.LYKYLLLRSTGDMHRAKSPTIMTRVTNNVYLGNYKNAMDAP..SSEVKFKYVLNL
PTP_Orf_wt MGDKSEWYARLLLRCTRAGPPLALPSGMTRLTDHVYLGSAEDARAVLRGDSGVDFKCLVNM
consensus>50 MdkKS.lYkyLLLRsTgdmh.aksPtiMTRvT#nVYLGnyk#Amdap..sSeVkFKyvlN$
D79
S/C112
R118
3CM3A
β4 η2 α3 β5
60 70 80 90 100 110
3CM3A TMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHSAAGVNRSGA
2Q05A TMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHCAAGVNRSGA
PTP_Orf_wt TMSKYSTP.AGITAYHIPLRDDDKTNIASIMPALVKLLARLEAEQKPTLVHCVAGVNRSGA
consensus>50 TMdKYtlPnsnInIiiHIPLvDDttT#Iskyfddvtafkskc#qr#ePvLVHcaAGVNRSGA
7 residues insertion in Orf virus between α4 & α5 helices
3CM3A
α4 α5 α6
120 130 140 150 160 170
3CM3A MILAYLMSKNKESLP.....MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDKN..
2Q05A MILAYLMSKNKESLP.....MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDKNGS
PTP_Orf_wt AAMGYVMHKRLAENPTMTQPARFVYFLKTYYEIRDLRGAFLENANFRYQLIKMFVCDSPS.
consensus>50 mi$aYlMsKnkeslP.....mlYFLyvYhsmRDLRGAFvENpsFkrQiIek%ViDkn..
Triade catalytique
7 residues insertion @137 and 143 (PTP orf nb)

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-LYKYLLLRSTGDMHKAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFKYVLN
LTMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHSAAGVNRS
GAMILAYLMSKNKESLP-------MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDK
N--
*
>P1;2Q05Ax1

MDKKS-LYKYLLLRSTGDMHRAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFKYVLN
LTMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHCAAGVNRS
GAMILAYLMSKNKESLP-------MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKYVIDK
NGS
*
>P1;PTP_Orf_wt

MGDKSEWYARLLLRCTRAGPPLALPSGMTRLTDHVYLGSAEDARAVLRGDSGVDFKCLVN
MTMSKYSTP-AGITAYHIPLRDDDKTNIASIMPALVKLLARLEAEQKPTLVHCVAGVNRS
GAAAMGYVMHKRLAENPTMTQPARFVYFLKTYYEIRDLRGAFLENANFRYQLIKMFVCDS
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
---------SLYKYLLLRSTGDMHKAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFK
YVLNLTMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHSAAG
VNRSGAMILAYLMSKNKESLP-------MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKY
VI-----
*
>P1;2Q05_B
structureX:2Q05_B:1    :B:171  :B:unknown:unknown:-1.00:-1.00
-----MDKKSLYKYLLLRSTGDMHRAKSPTIMTRVTNNVYLGNYKNAMDAP--SSEVKFK
YVLNLTMDKYTLPNSNINIIHIPLVDDTTTDISKYFDDVTAFLSKCDQRNEPVLVHCAAG
VNRSGAMILAYLMSKNKESLP-------MLYFLYVYHSMRDLRGAFVENPSFKRQIIEKY
VIDKN
*
>P1;PTP_ORF_WT
sequence:PTP_ORF_WT:6    :A:181  :A:unknown:unknown:-1.00:-1.0
---------EWYARLLLRCTRAGPPLALPSGMTRLTDHVYLGSAEDARAVLRGDSGVDFK
CLVNMTMSKYSTP-AGITAYHIPLRDDDKTNIASIMPALVKLLARLEAEQKPTLVHCVAG
VNRSGAAAMGYVMHKRLAENPTMTQPARFVYFLKTYYEIRDLRGAFLENANFRYQLIKMF
VCDSPS-
*
```

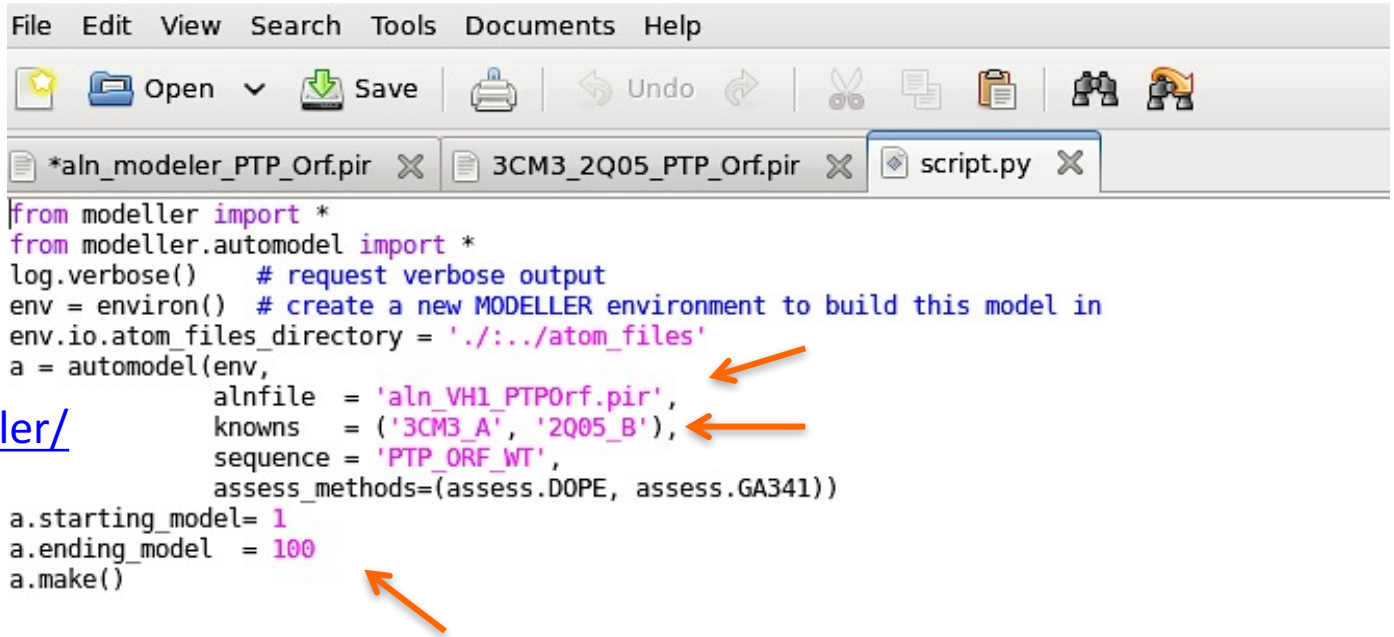
Homology modeling

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

https://salilab.org/modeller/



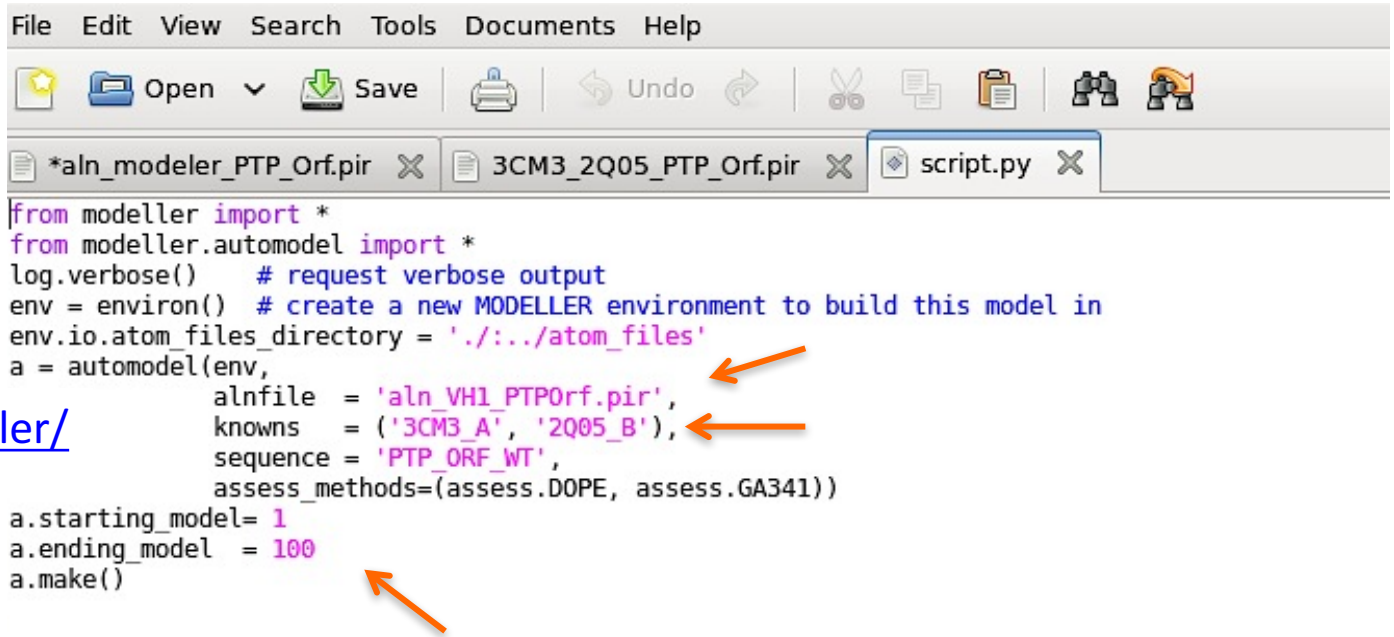
Homology modeling

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

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.

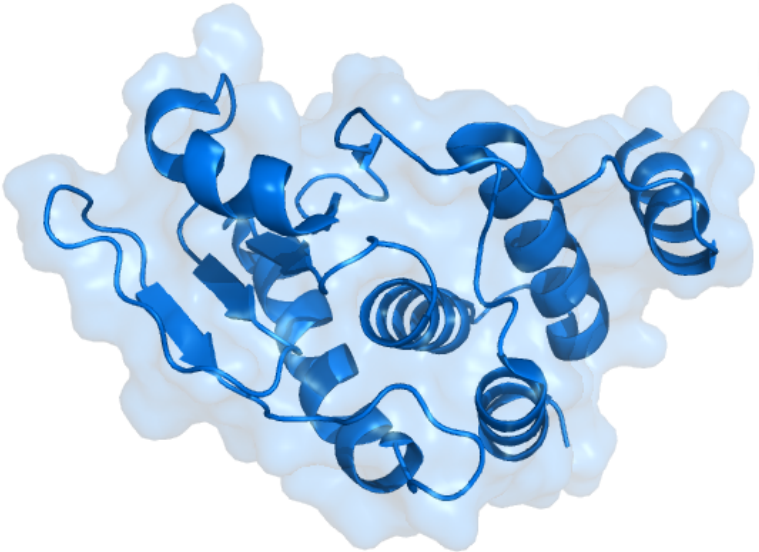
Protein structures are more conserved than protein sequences amongst homologues.

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

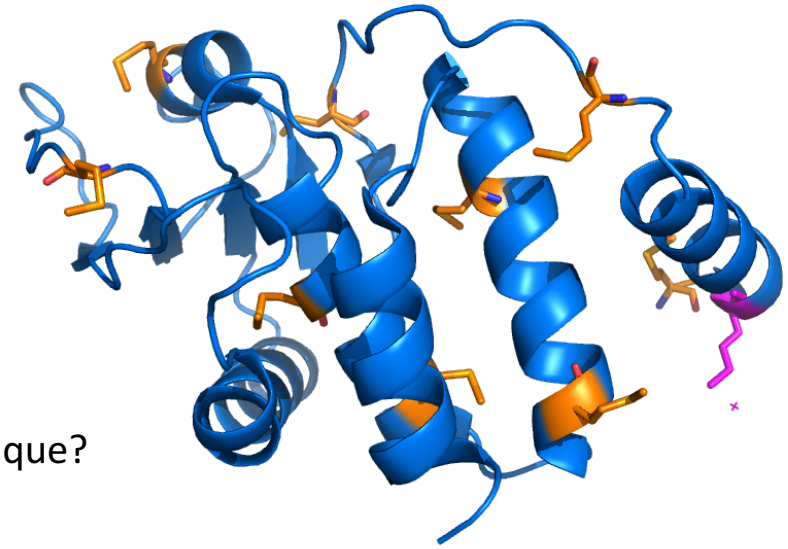
1) Identification of protein(s) as reference « template(s) »
2) Homology modeling of protein « query »

Open PyMol session: compil_Hmodeles.pse

Mutations



?



Pourquoi?
Comment?
Pertinence biologique?

Model the point mutants by substituting residues

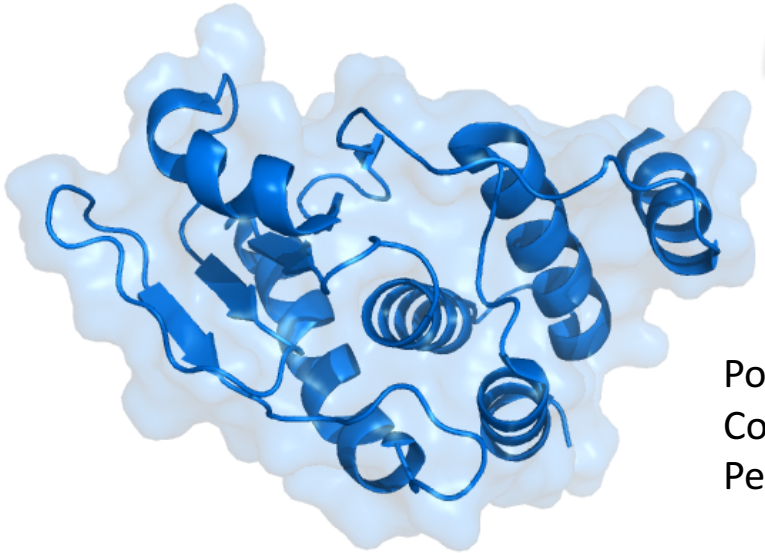
1- Modéliser les mutants, les variants pour lesquels on a des données expérimentales.
2- Modifier le site de fixation
3- Vérifier l'accessibilité
4- Analyser l'interface

Pymol/Plugin/PDB Loader Service/ faire la mutation dans 3CM3 S112C
Ouvrir 2Q05

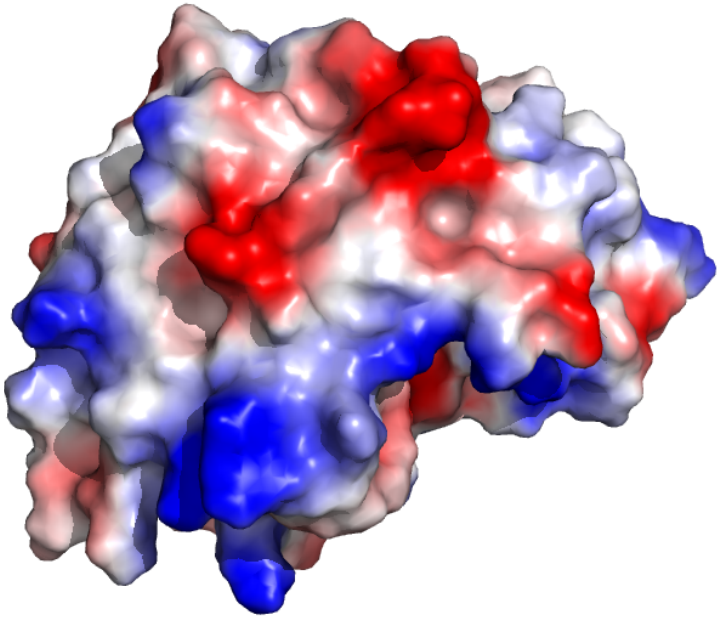
Show sequence. Que voit-on?

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

Electrostatic profile



?

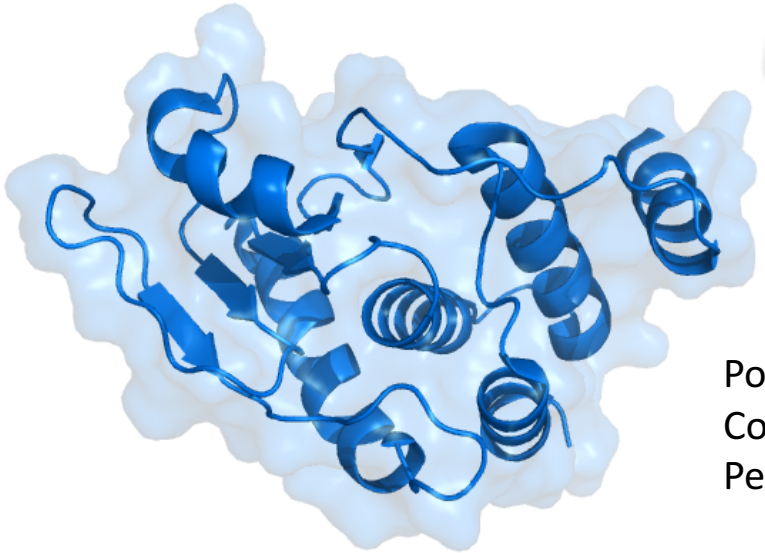


Pourquoi?
Comment?
Pertinence biologique?

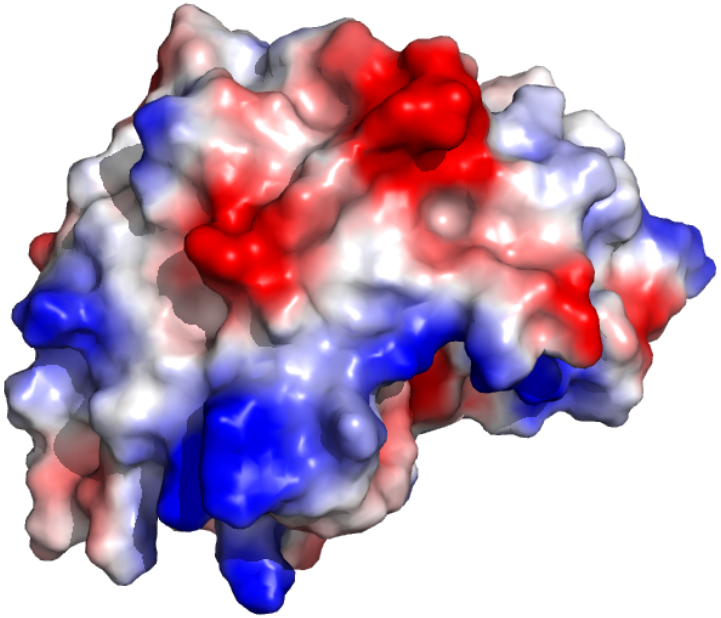
Calcul du profil électrostatique : cartographie des patches hydrophobes/chargés + ou –.
Analyse de la solvatation des systèmes biologiques

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

Electrostatic profile



?



Pourquoi?
Comment?
Pertinence biologique?

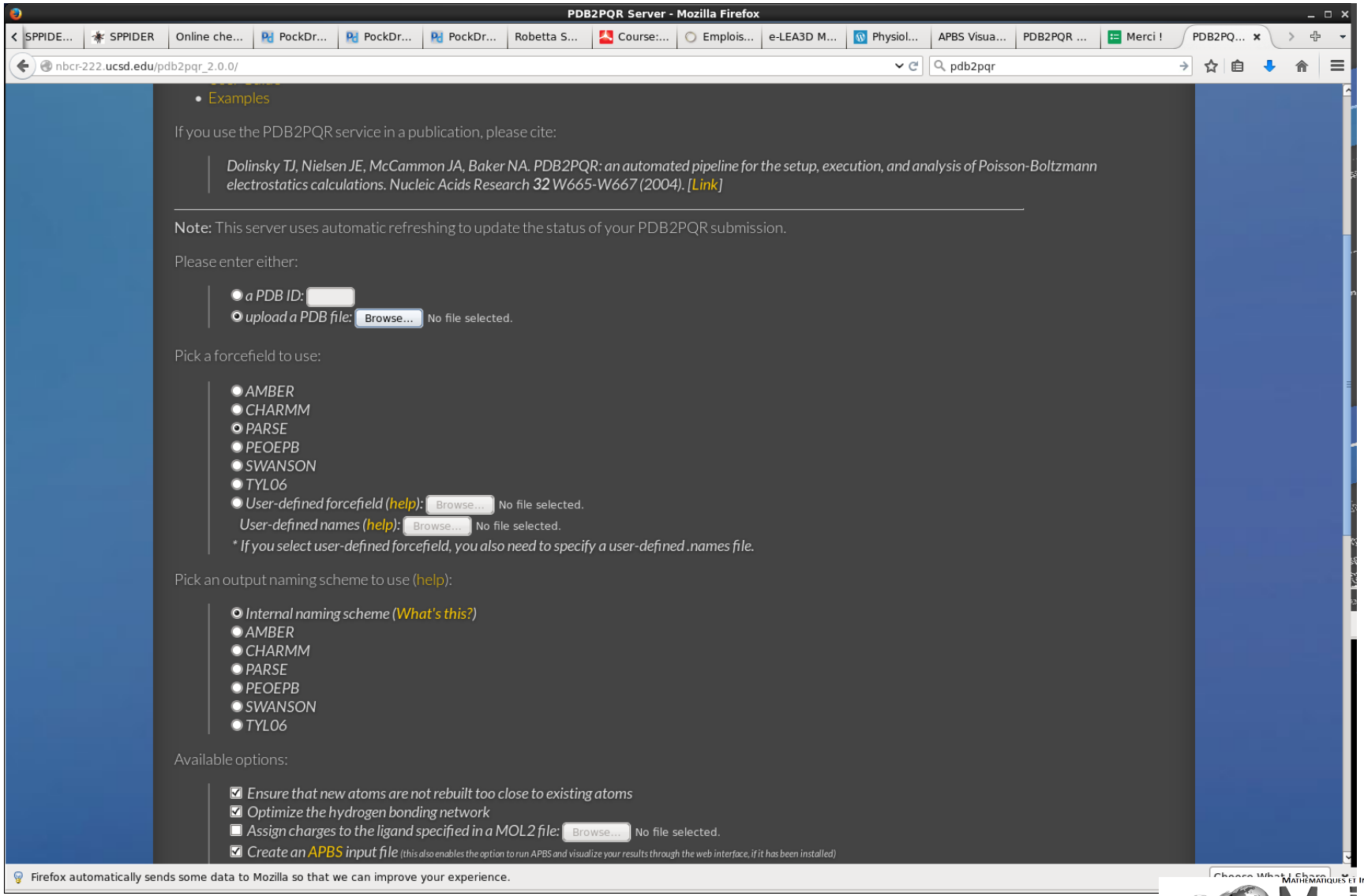
Comment? PDB2PQR et APBS Adaptative Poisson-Boltzmann Solver

APBS Adaptative Poisson-Boltzmann Solver dans pymol

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

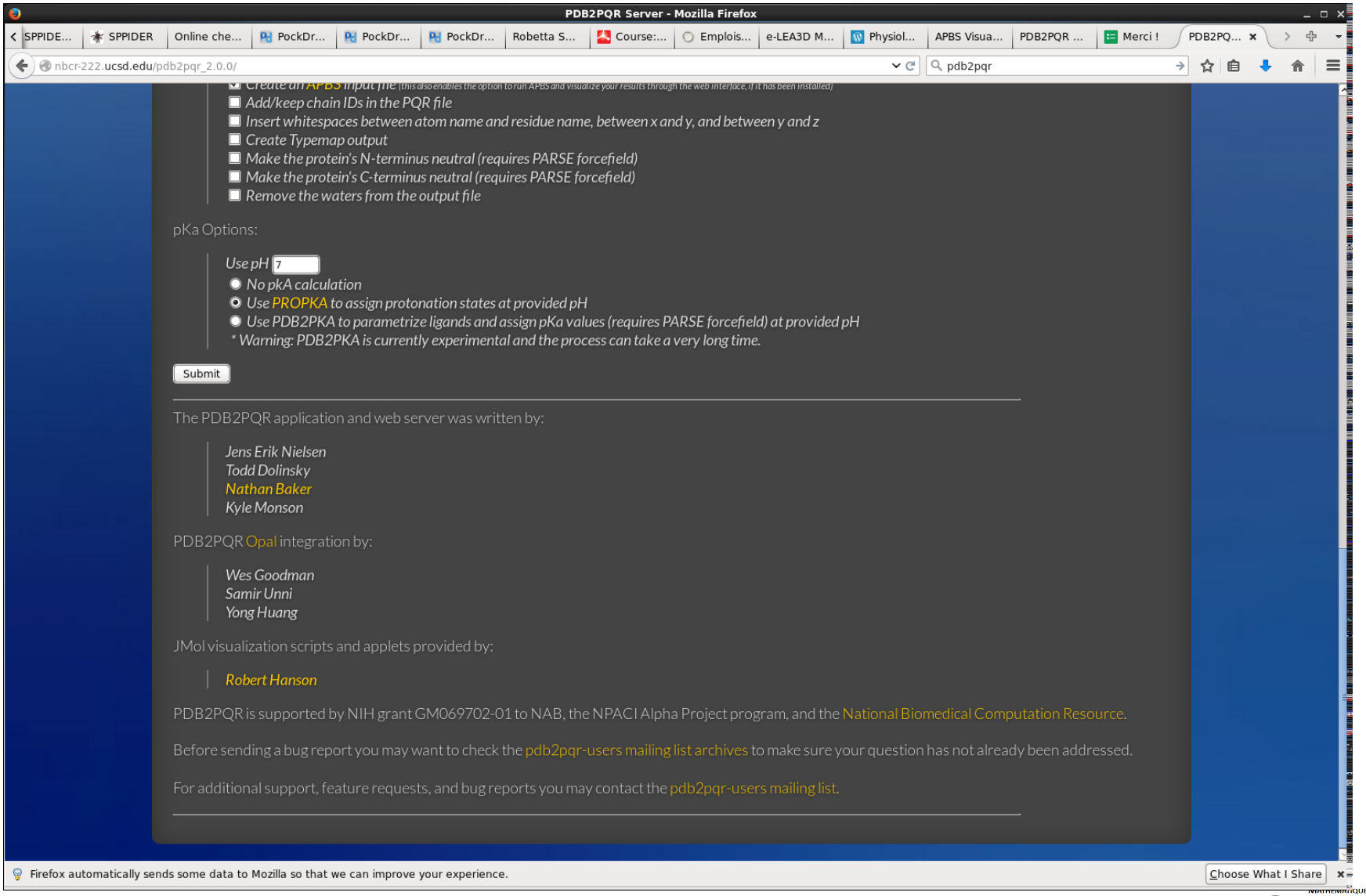
Electrostatic profile

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



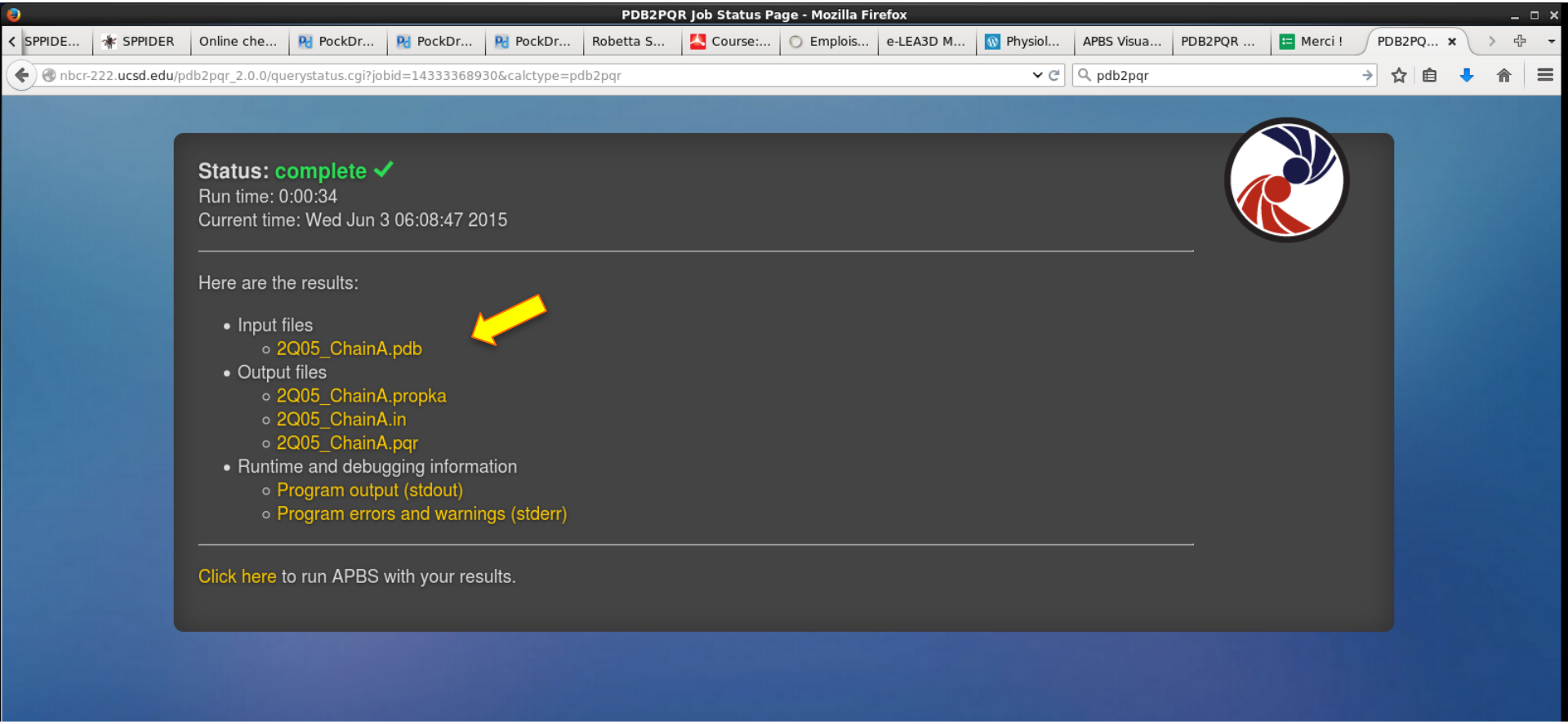
Electrostatic profile

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



Electrostatic profile

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



Electrostatic profile

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

2Q05_A_Met.pdb

```
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
```

2Q05_A_Met.pqr

```
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
```

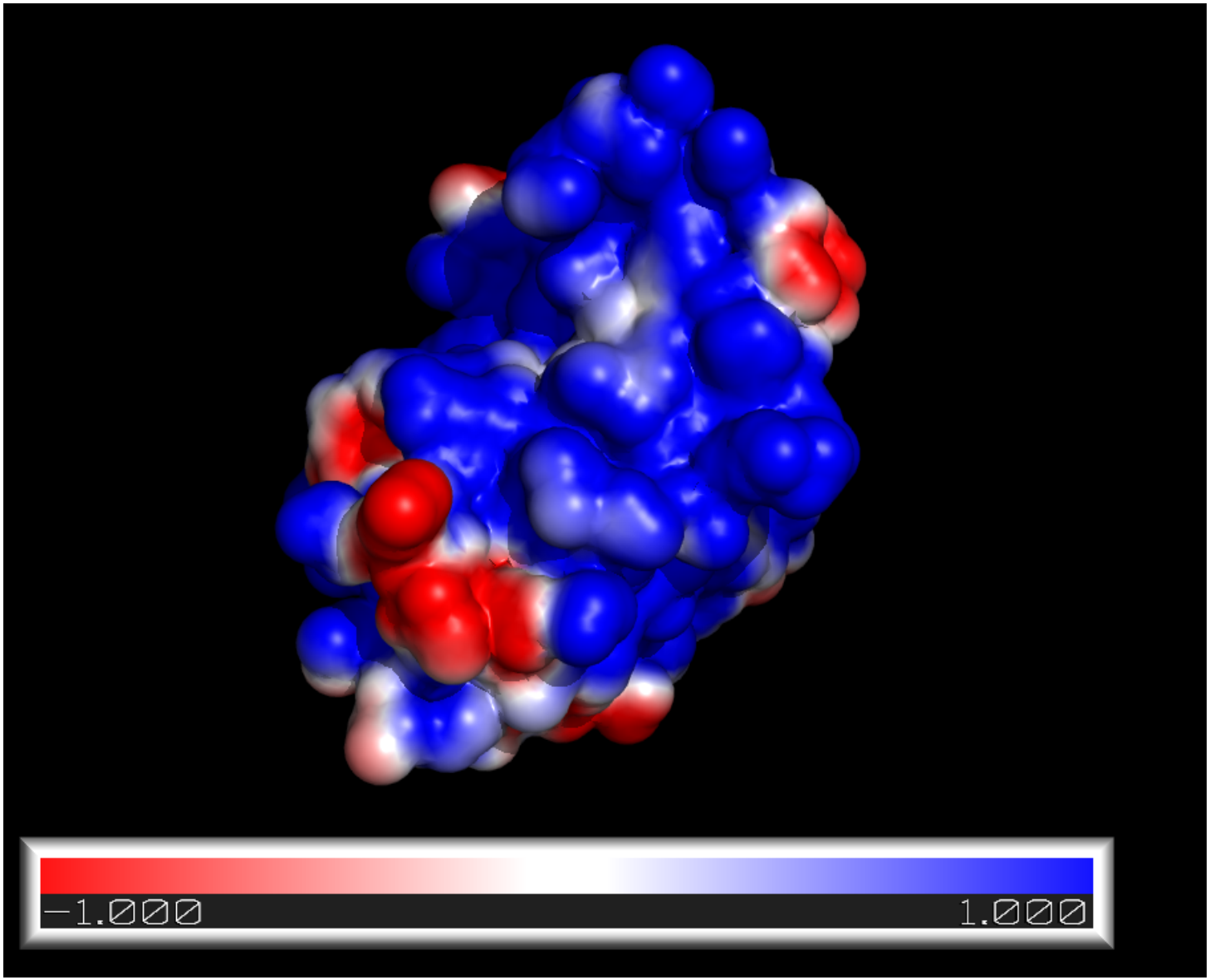
Facteur occupation 0< x< 1

Facteur B agitation thermique

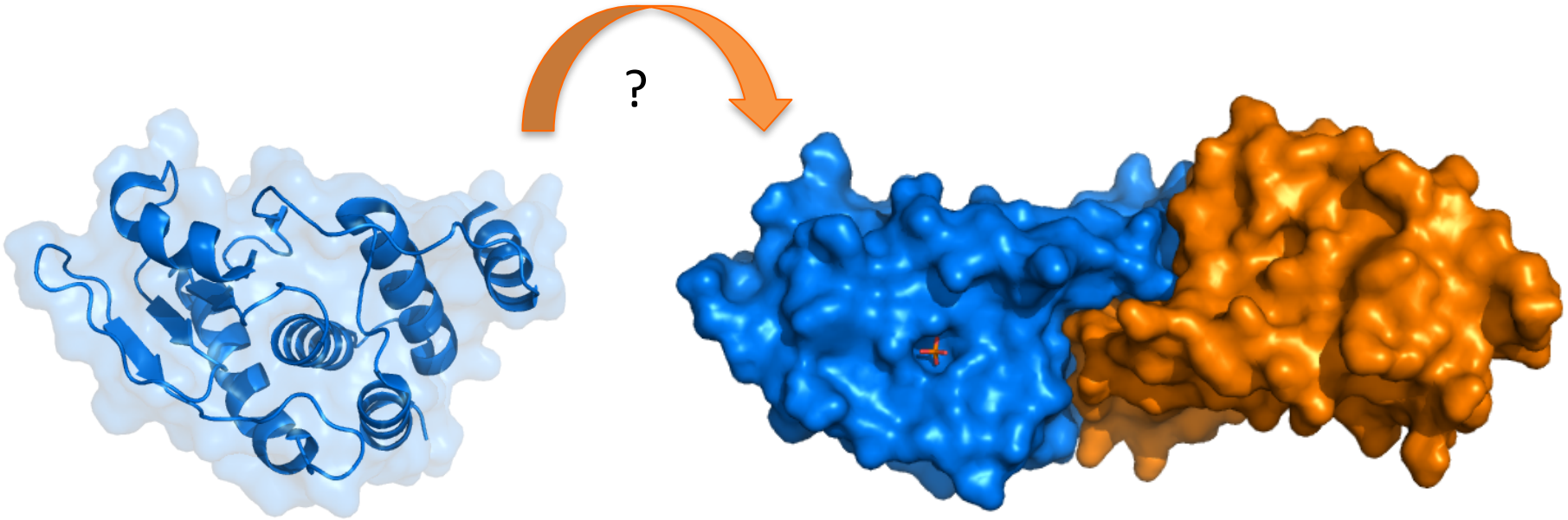
Q charge

R rayon

Mapping of the electrostatic profile



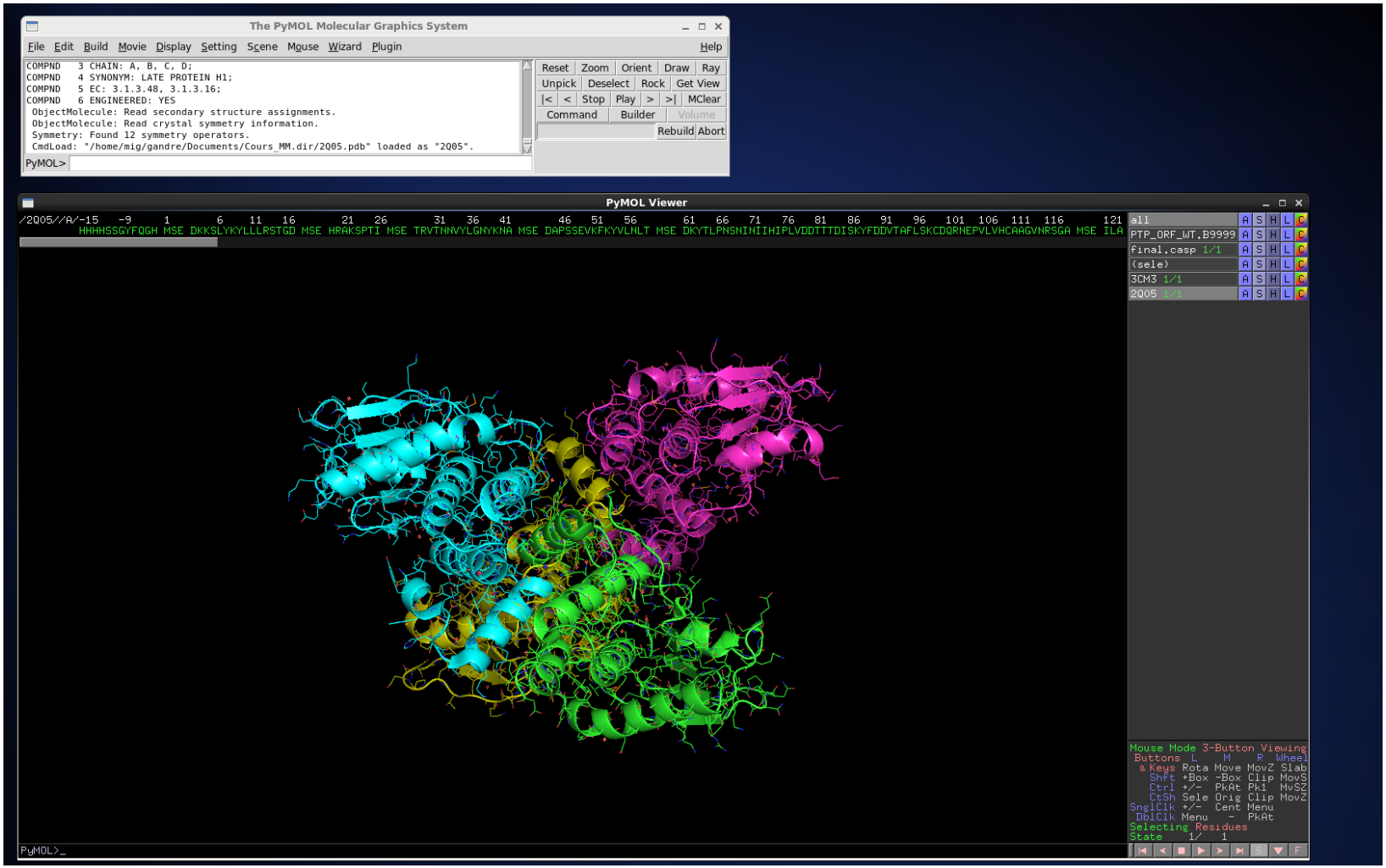
Oligomerization



Pourquoi?
Comment?
Pertinence biologique?

Oligomerization

File /open/2Q05.pdb

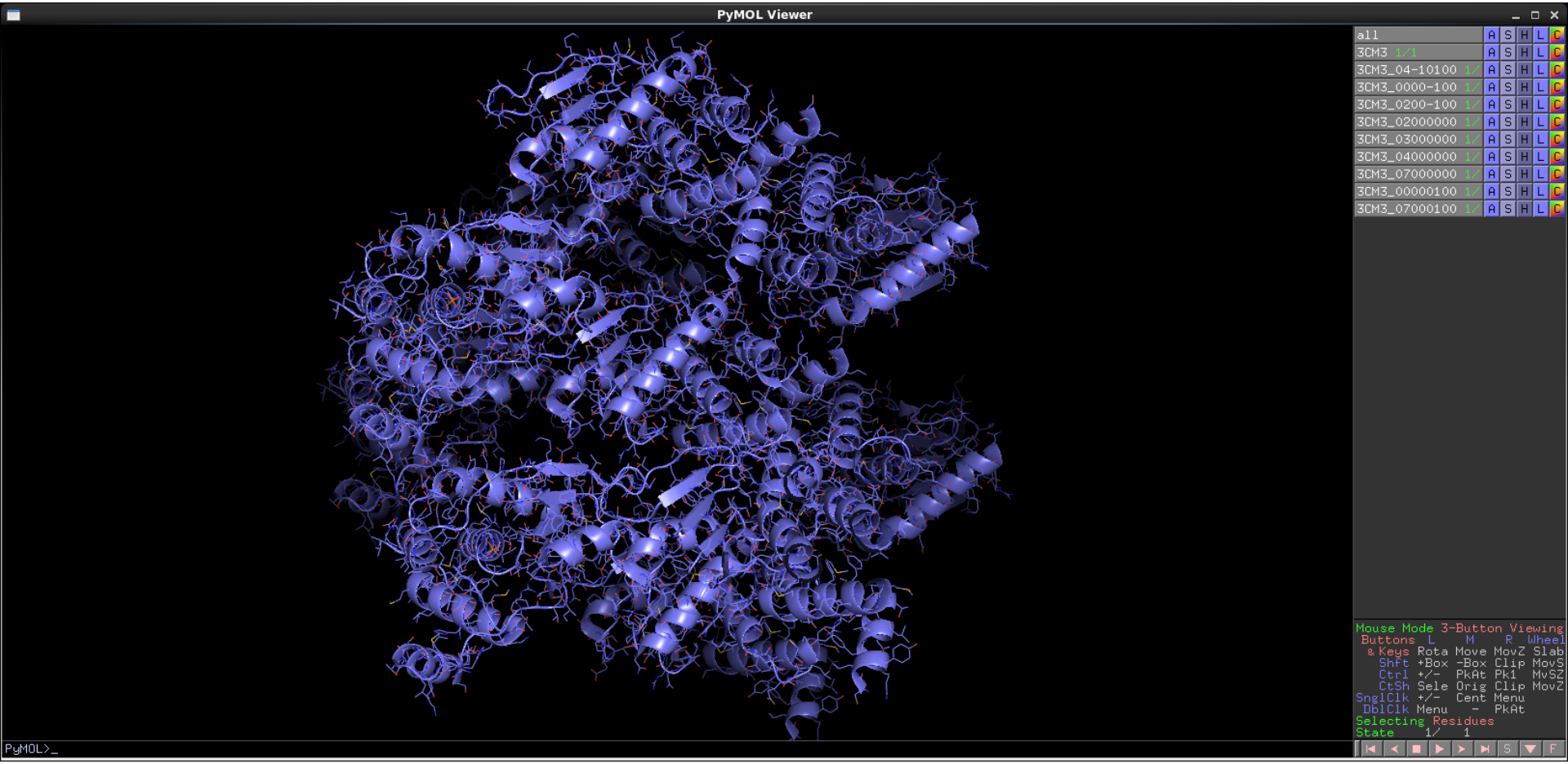


PDBePISA : Proteins, Interfaces, Structures and Assemblies is an interactive tool for the exploration of macromolecular interfaces.

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/

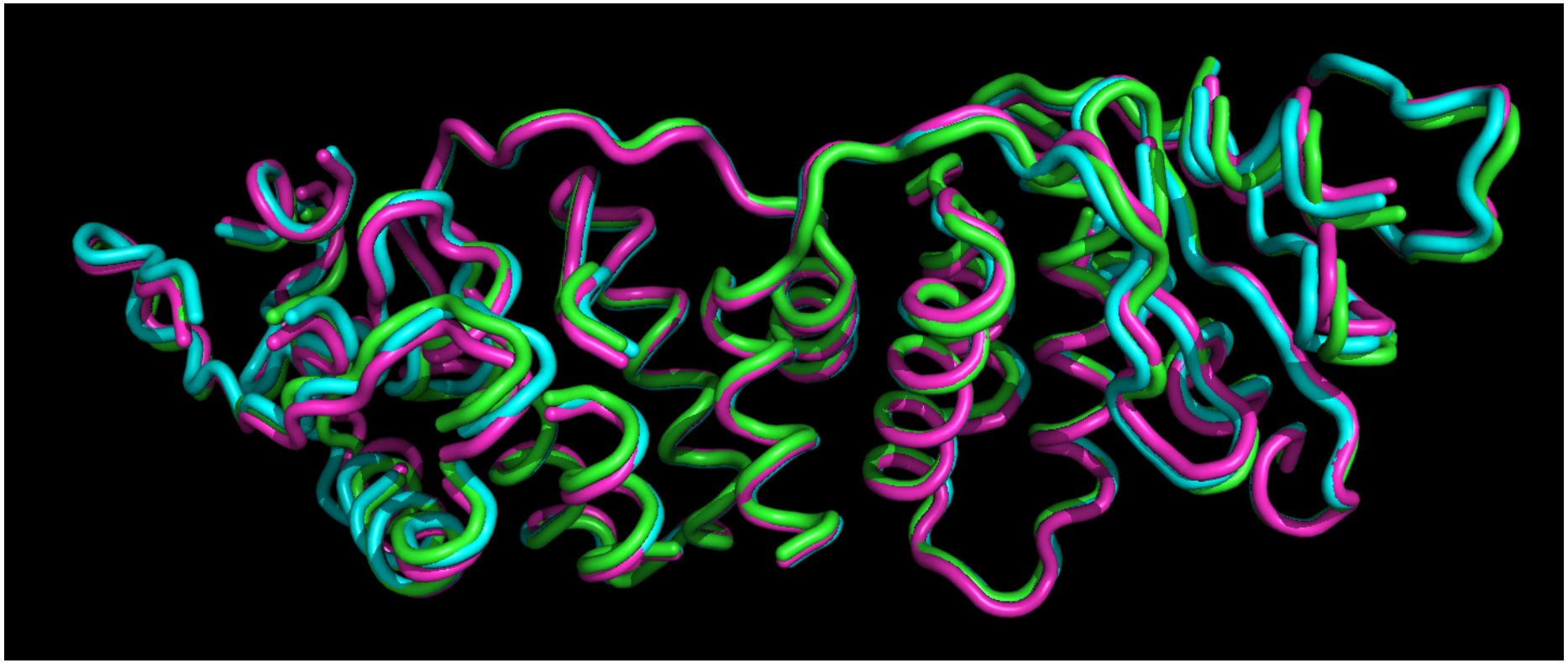
MaiAGE Mathématiques et Informatique Appliquées du Génome à l'Environnement

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



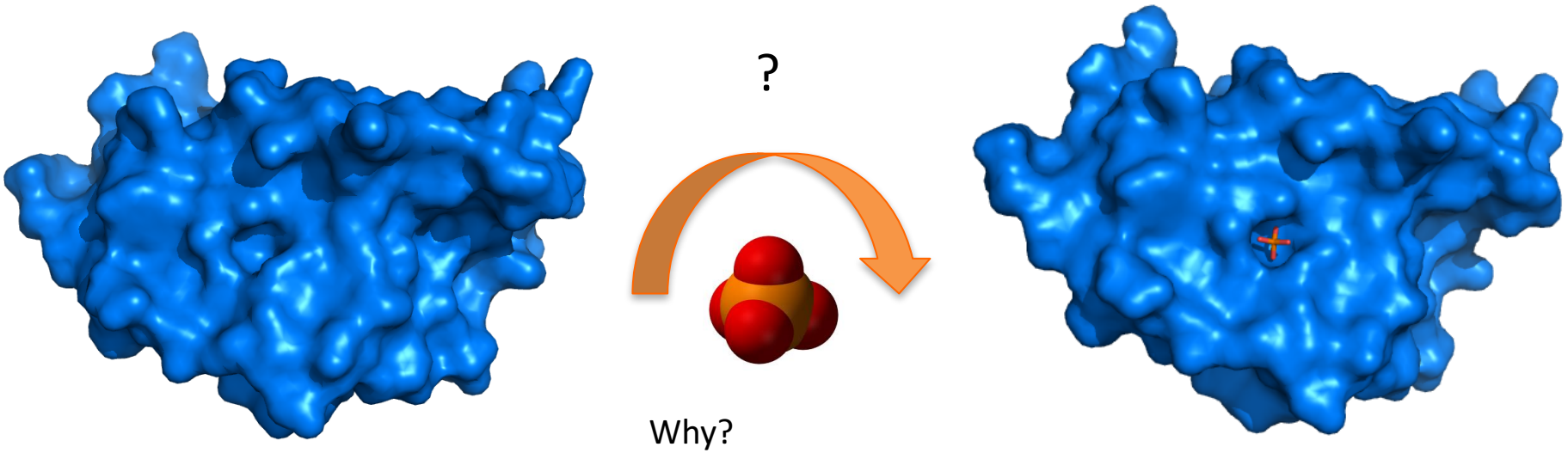
Docking

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?

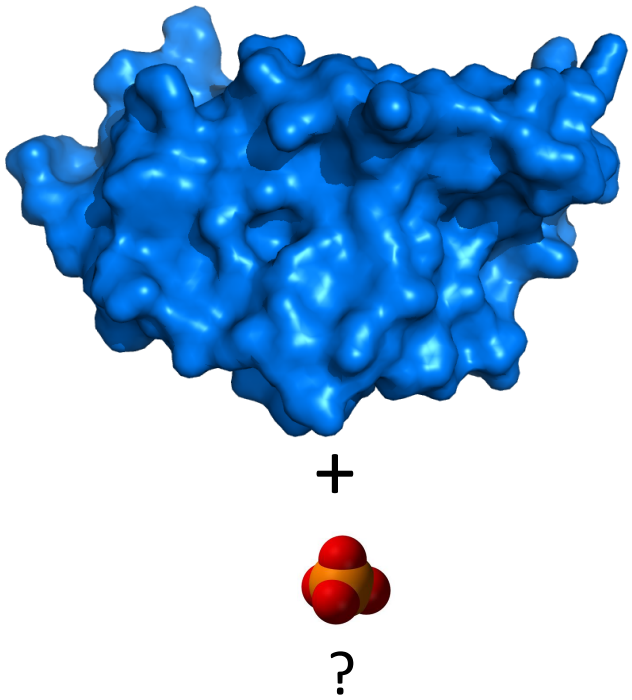
Binding pocket and ligands

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
Prodrg token/free registration mandatory



Scoring :

Total potential energy
Inetraction 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 transparancy/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://molprobity.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://espript.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.

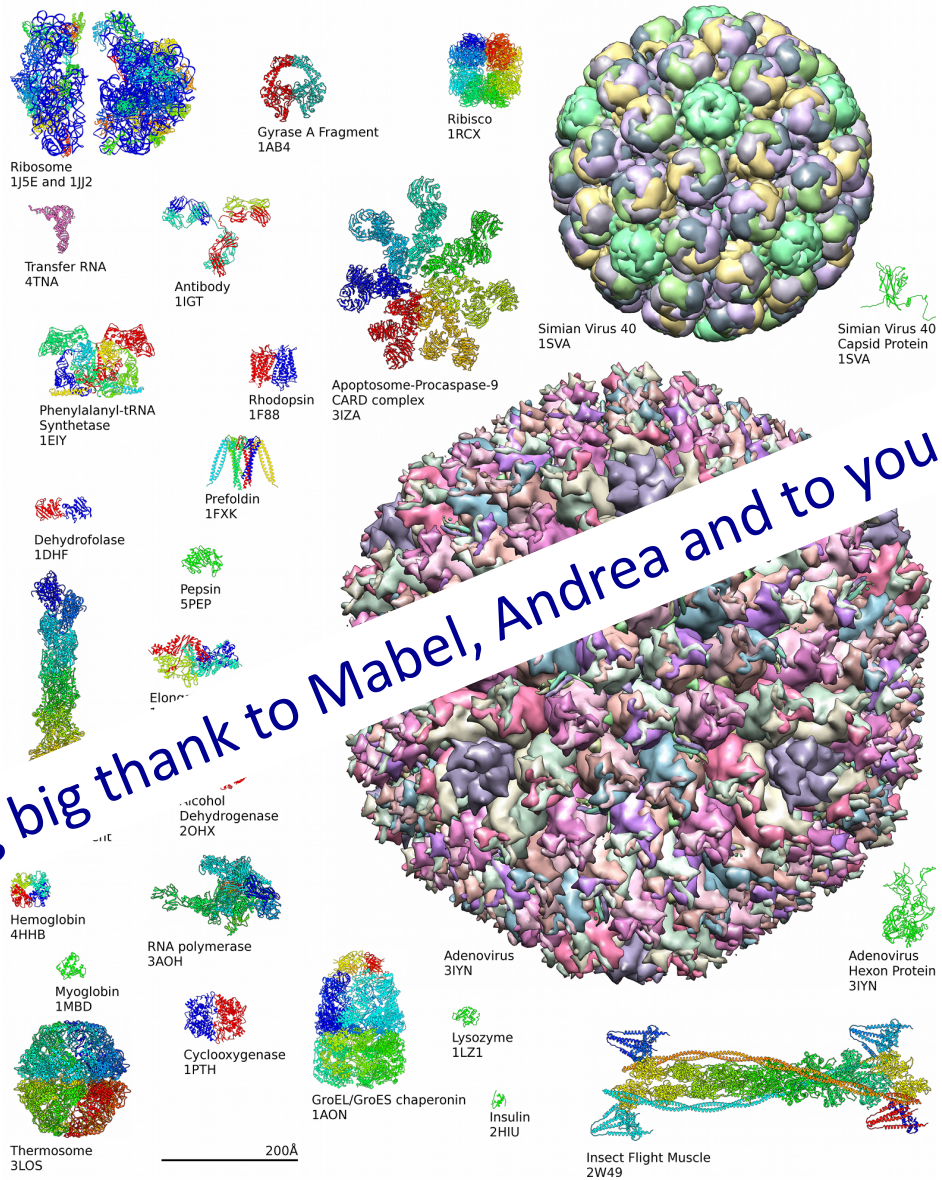
formation

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

MaIAGE
Mathématiques et Informatique Appliquées du Génome à l'Environnement

Pablo Picasso

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