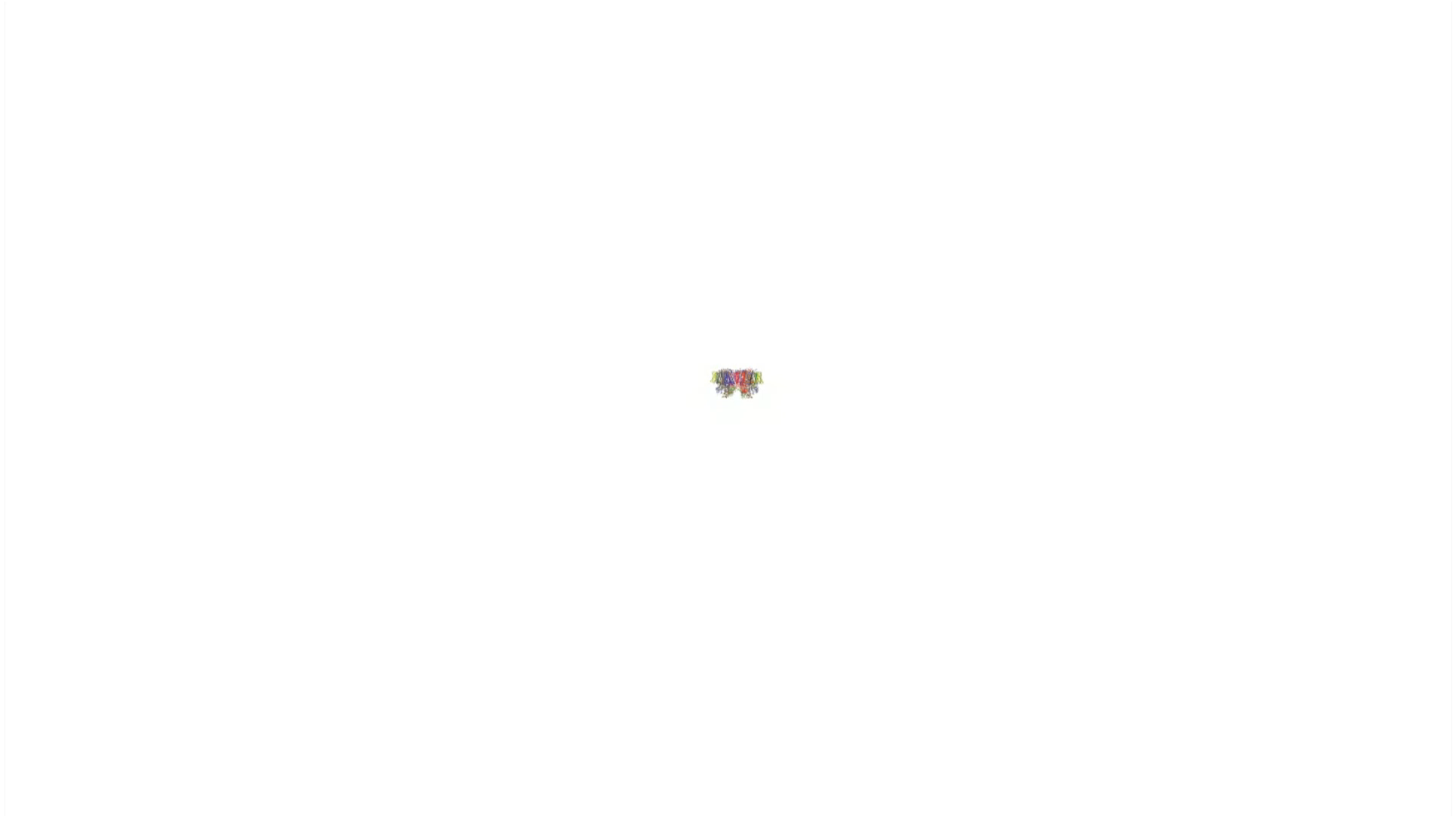
Complex simulations

Most tutorials are about ubiquitine (also today's)

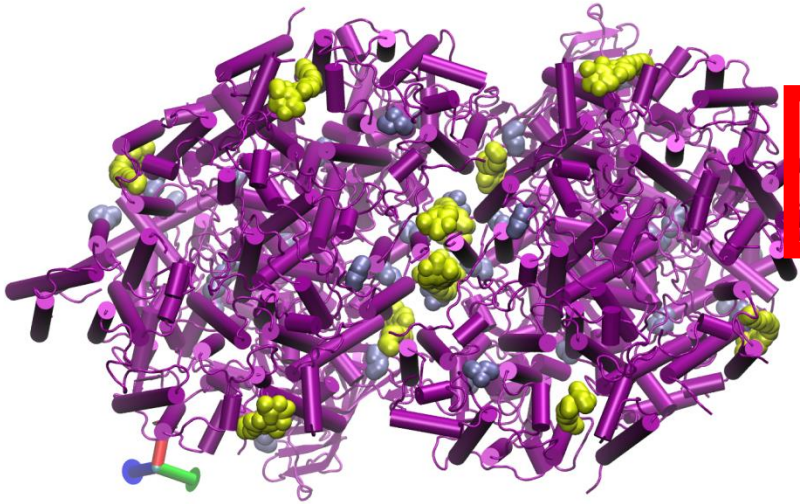
- Single protein chain
- No cofactors
- Easy to set up

What if your system looks more like this?



Preparing your structure for simulating

- Dimer!

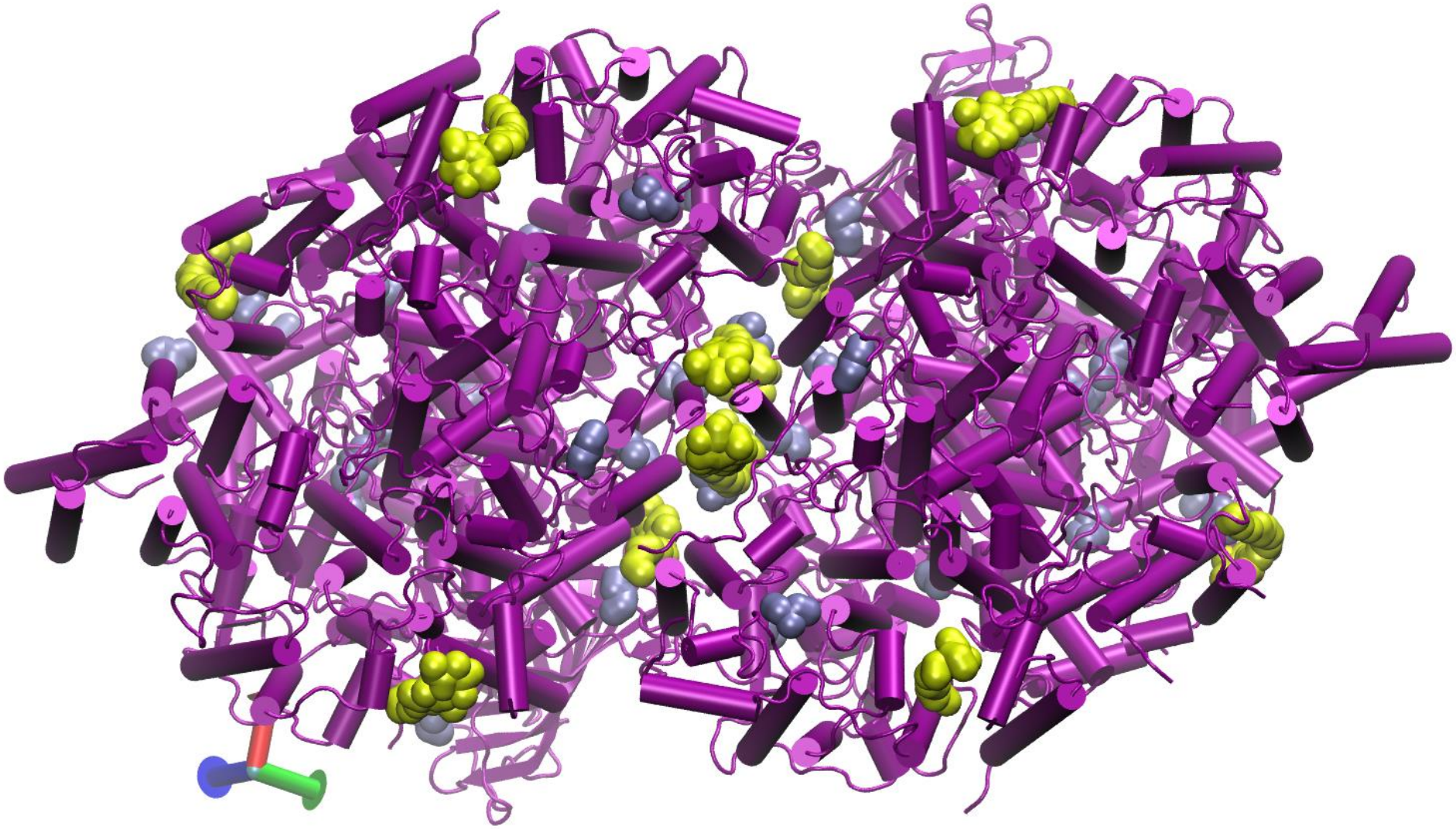


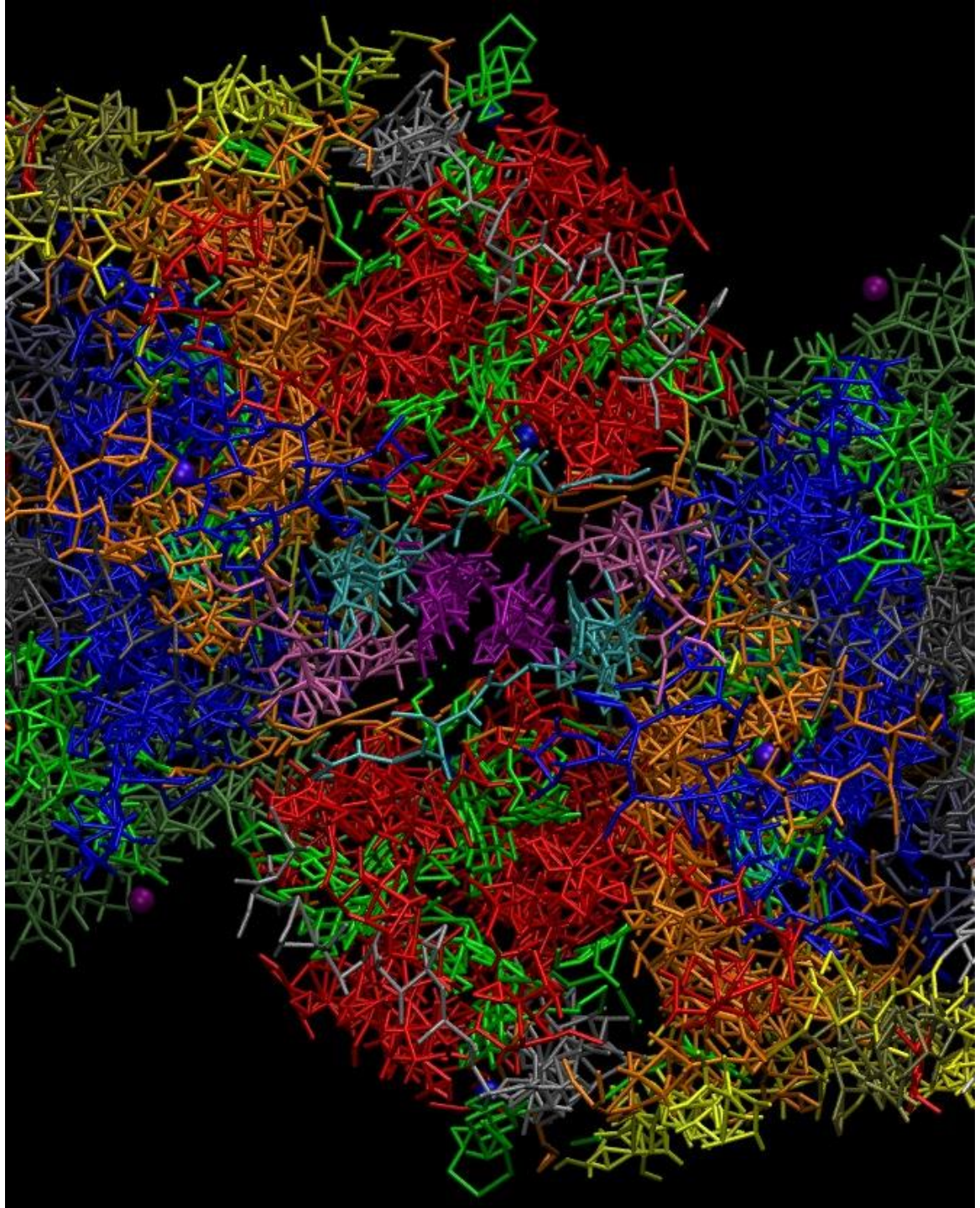
```
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465     MET A     1
REMARK 465     THR A     2
REMARK 465     THR A     3
REMARK 465     THR A     4
REMARK 465     LEU A     5
REMARK 465     GLN A     6
REMARK 465     ARG A     7
REMARK 465     ARG A     8
REMARK 465     GLU A     9
REMARK 465     SER A    10
REMARK 465     ASN C    19
REMARK 465     SER C    20
REMARK 465     ILE C    21
3arc.pdb
REMARK 610 MISSING HETEROATOM
REMARK 610 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 610 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 610 I=INSERTION CODE):
REMARK 610   M RES C SSEQI
REMARK 610     LMG B  669
REMARK 610     DGD C  657
REMARK 610     DGD C  660
REMARK 610     DGD C  661
REMARK 610     LMG C  729
REMARK 610     LMG C  776
REMARK 610     DGD D  755
REMARK 610     SQD D  768
REMARK 610     LMG D  692
REMARK 610     HTG D  726
REMARK 610     LHG E  772
REMARK 610     DGD H  663
REMARK 610     LMG Z  784
REMARK 610     LMG a  751
REMARK 610     LMG b  669
REMARK 610     LMT b  791
REMARK 610     DGD c  657
REMARK 610     DGD c  660
REMARK 610     DGD c  661
REMARK 610     LMG c  729
```

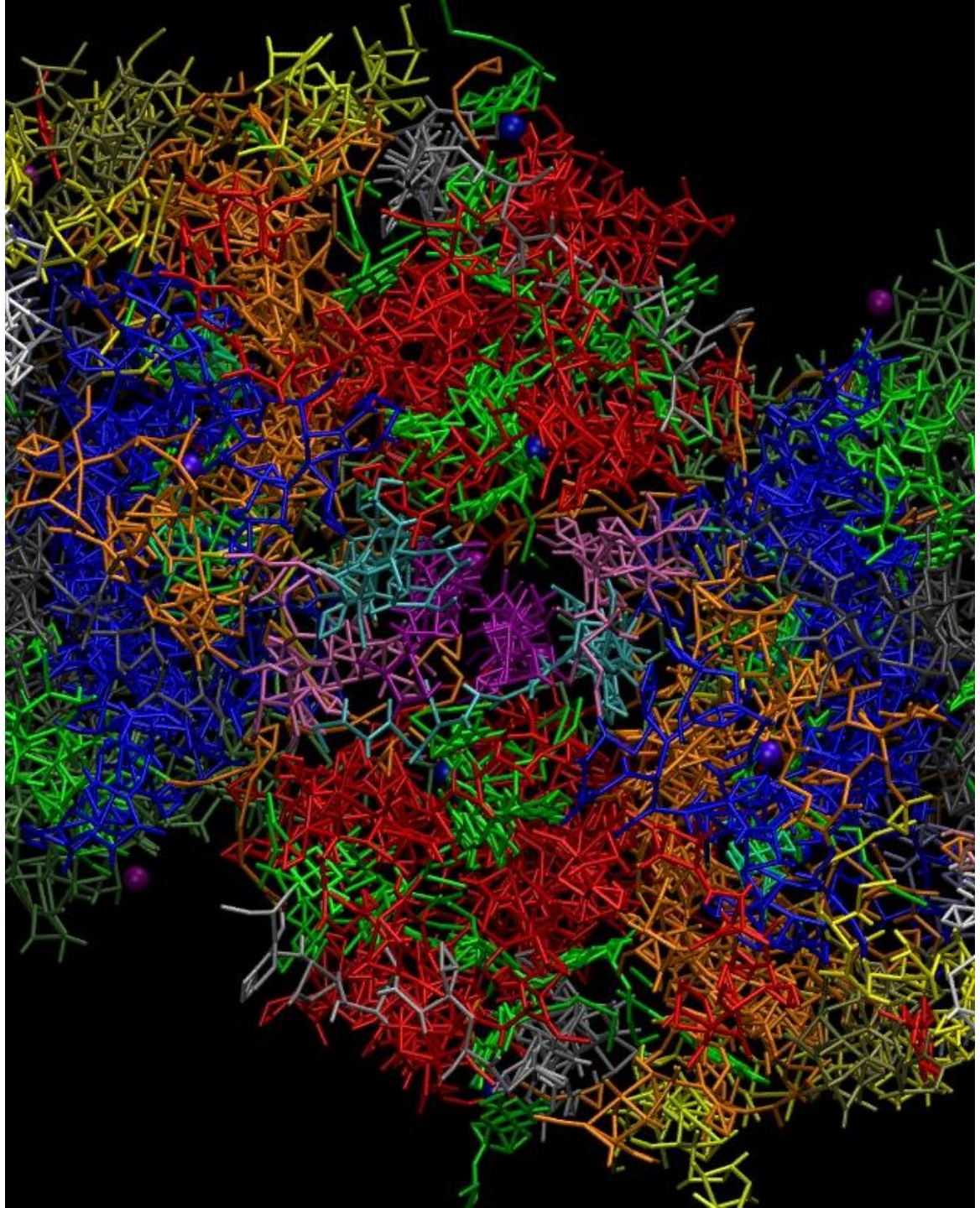
- Add missing residues
- Add missing heteroatoms

- Delete all detergent molecules

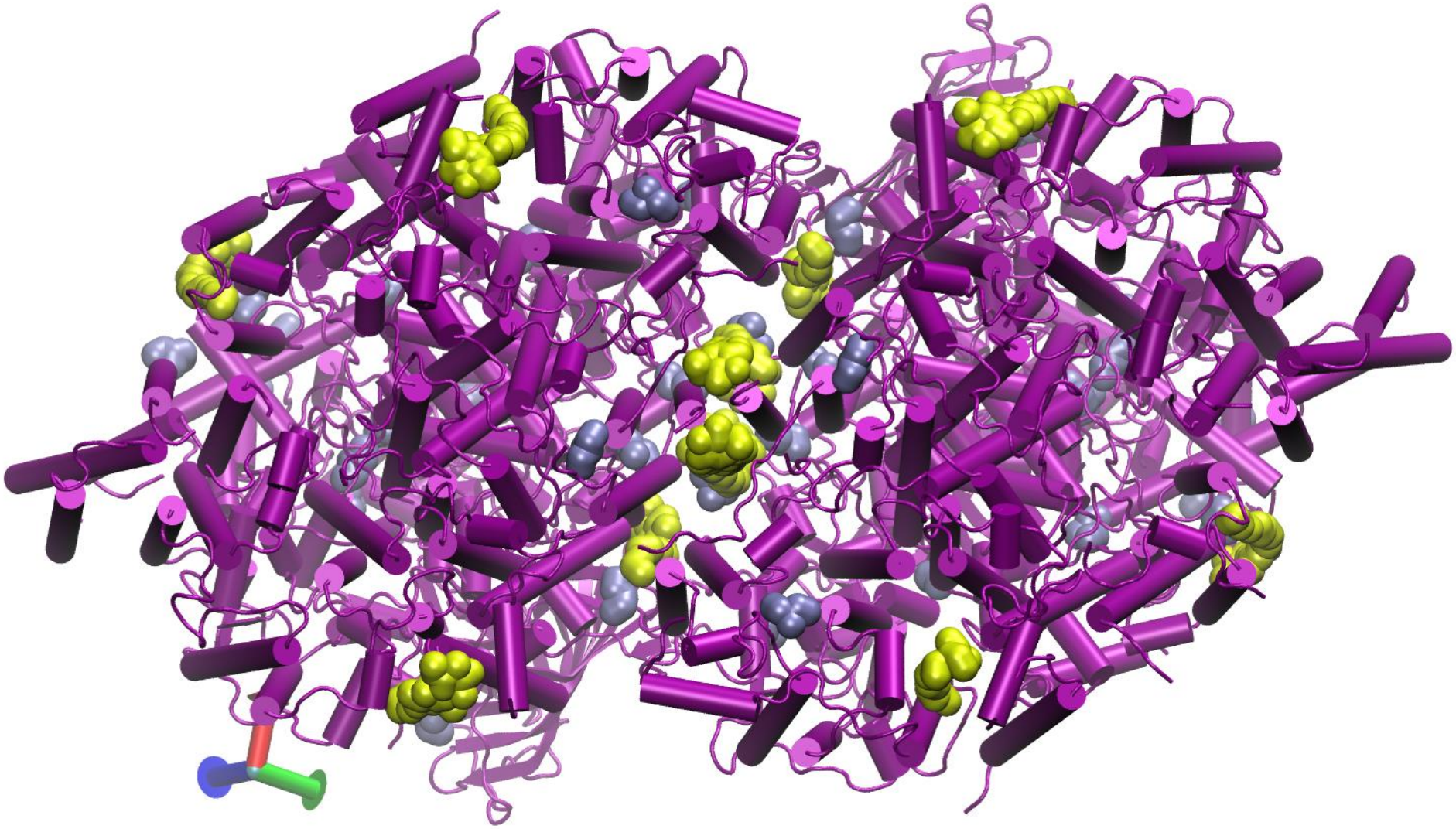
Detergents in crystal structure



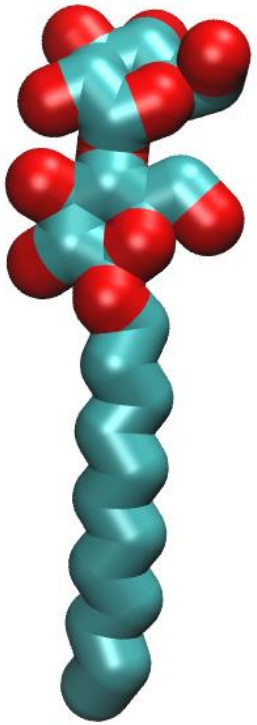




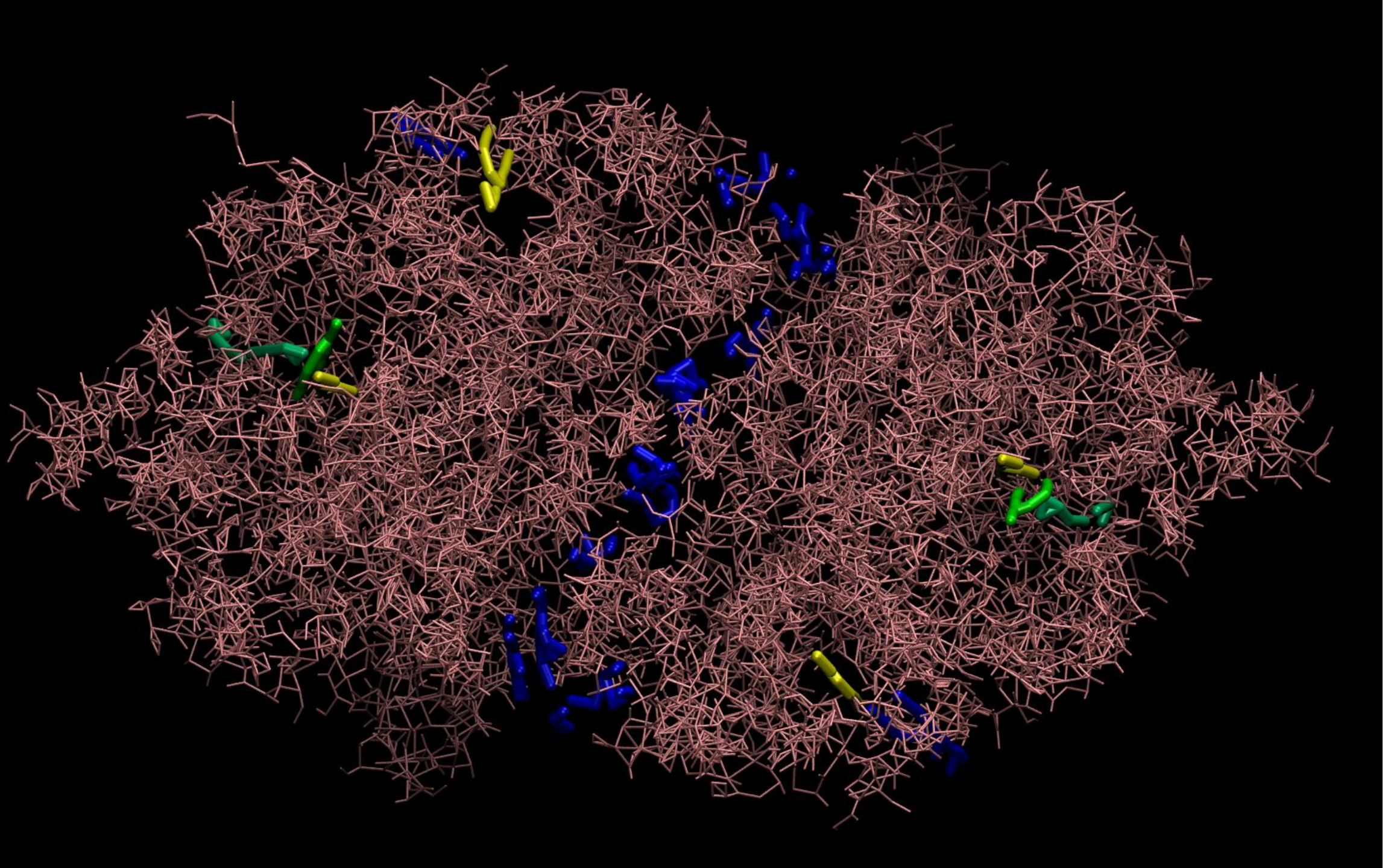
Detergents in crystal structure



Suprise suprise



DODECYL-BETA-D-MALTOSIDE → looks like a lipid



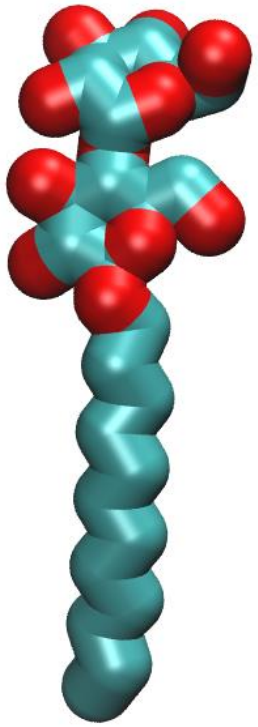
Where should I replace my detergents with?

Lipid that is the most similar?

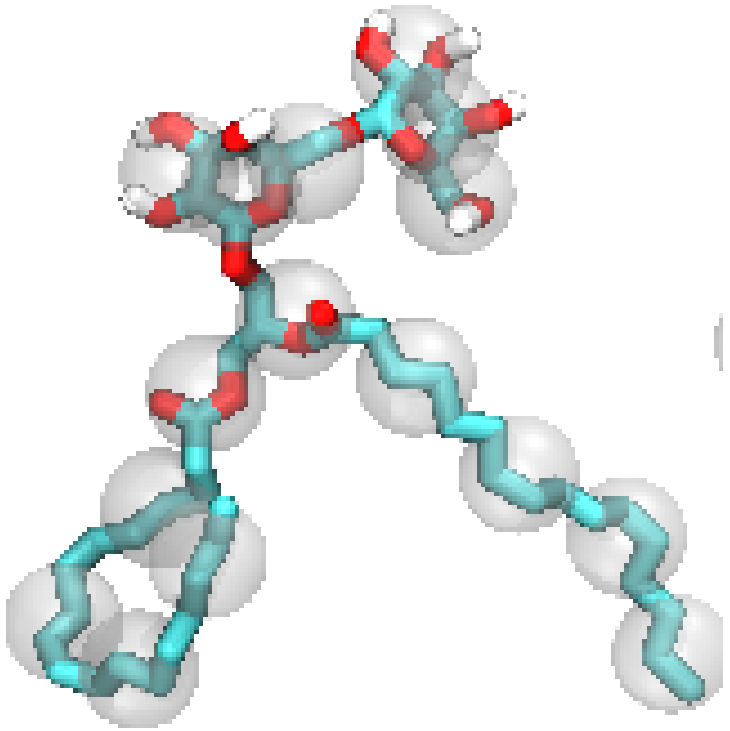
Look in older structures.

Look in literature

→ No real good answer



DGDG



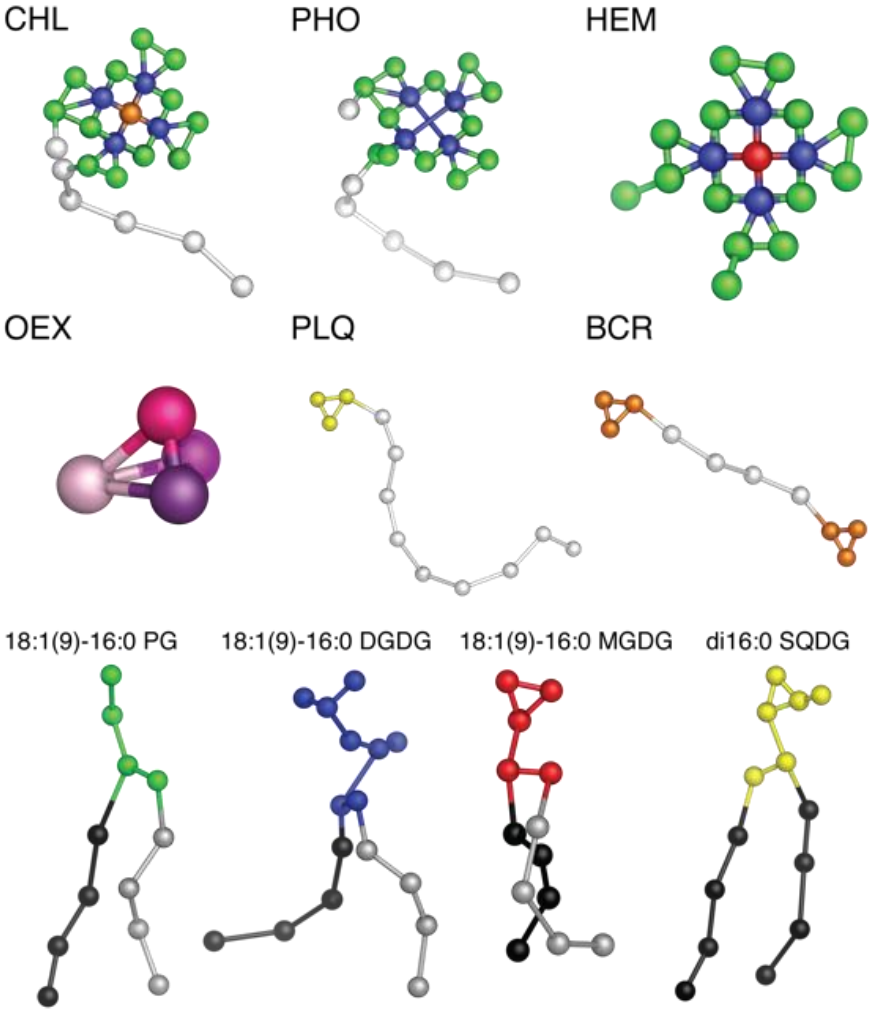
Coarse graining

- Split PDB
  - Protein → Martinize.py
    - Be careful Backward.py can also CG your protein, but it does not make an Elnedyn network, but most importantly the mapping of the backbone bead is different for the Elnedyn version
  - Cofactor → Backward.py

Embed in a bilayer

- Insane
  - OPM database → contains calculated position of the bilayer

Parameterization

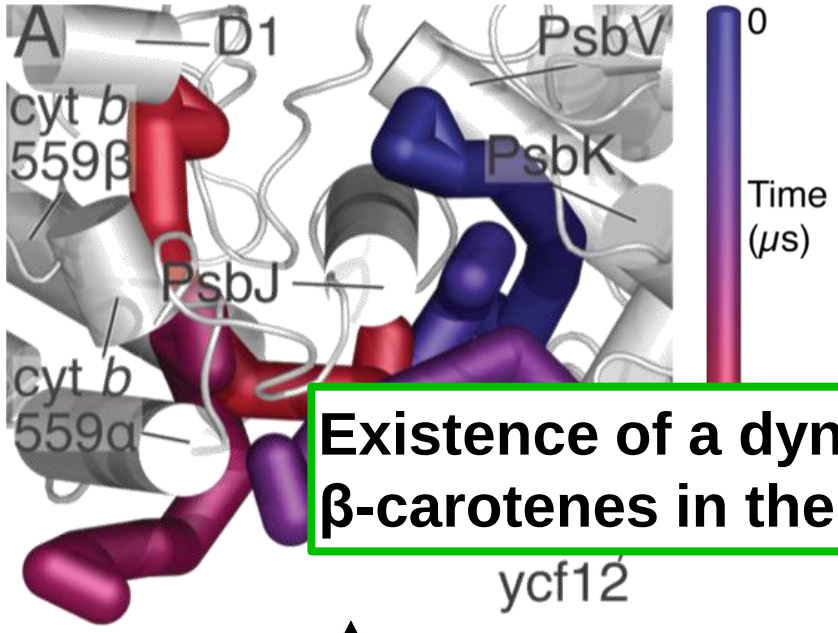


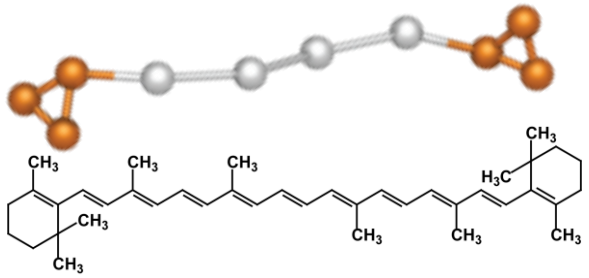
PSII dimer (cyanobacteria)[#]
2x19 protein subunits

cofactors per monomer:
chlorophyll A (35)
pheophytin (2)
heme (2)
β-carotene (12)
plastoquinone (2)
ox. evolving compl. (1)

lipid composition:
18:1-16:0 PG (10%)
18:1-16:0 DGDG (25%)
18:1-16:0 MGDG (40%)
18:1-16:0 SQDG (15%)
di16:0 SQDG (10%)
+ co-crystall. lipids (> 40 total)

β-carotenes are very mobile



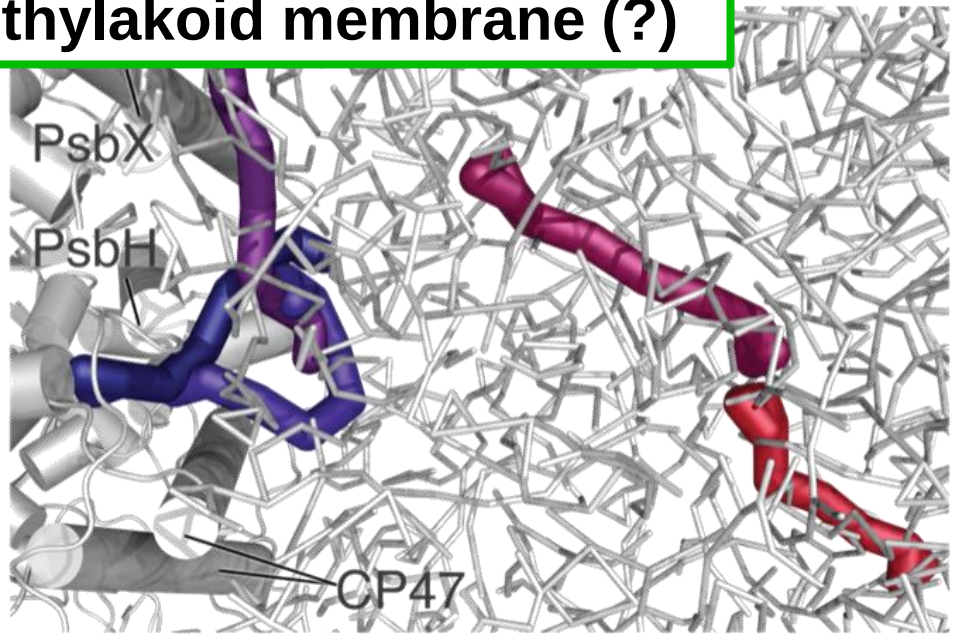


β-carotenes protect against photodamage

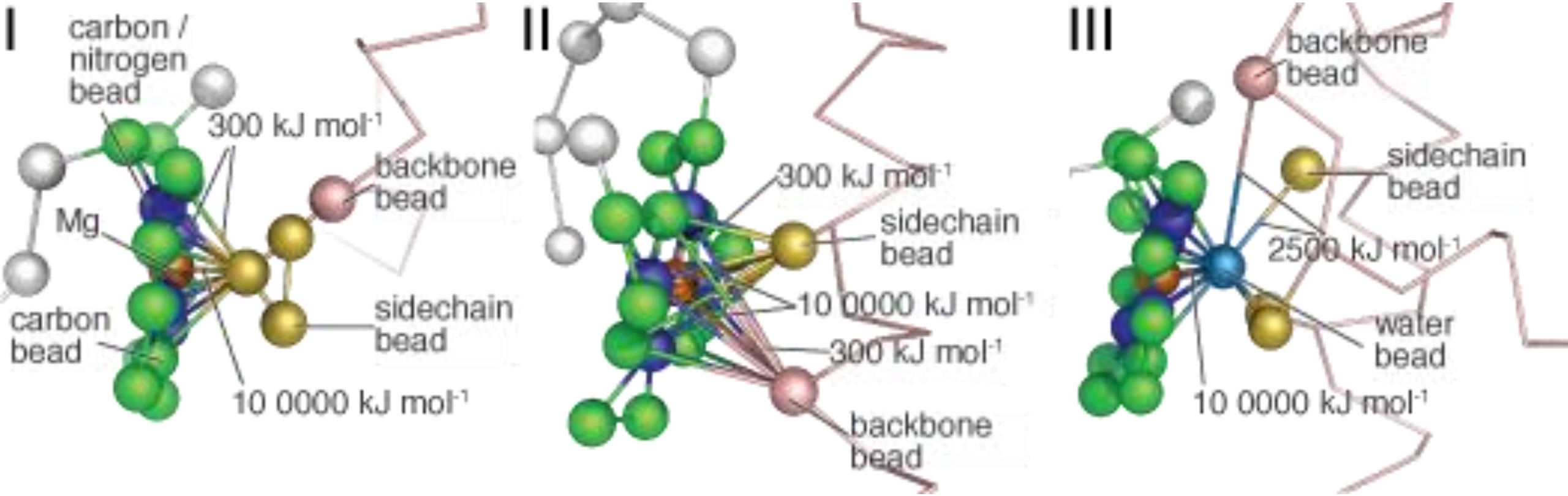
Existence of a dynamic pool of β-carotenes in the thylakoid membrane (?)

internal rearrangements

exchange with thylakoid membrane

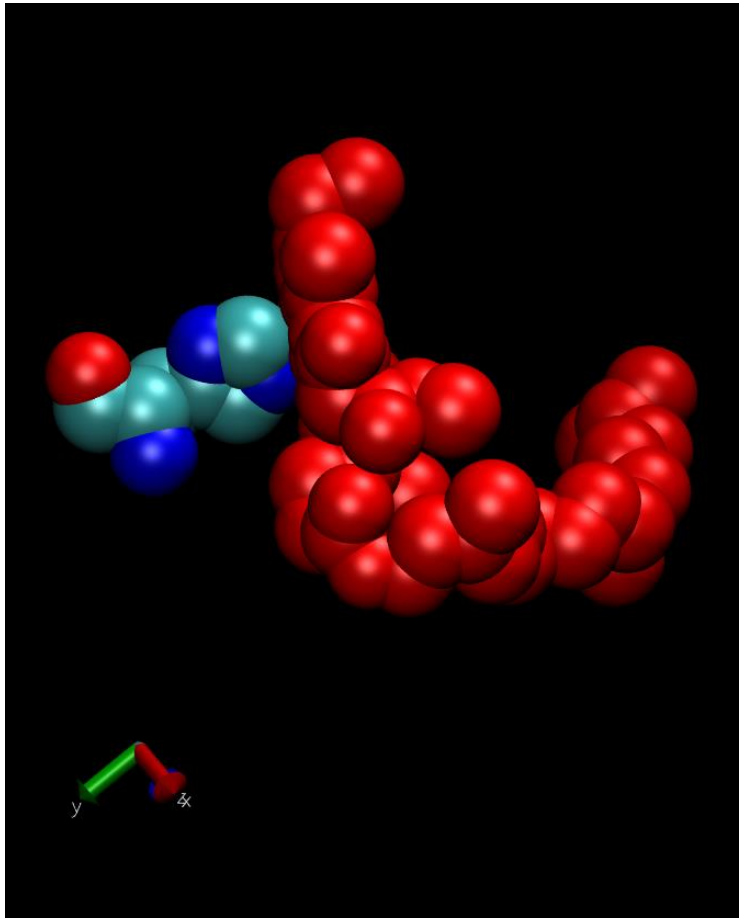


Chlorophylls are bound in 3 different ways

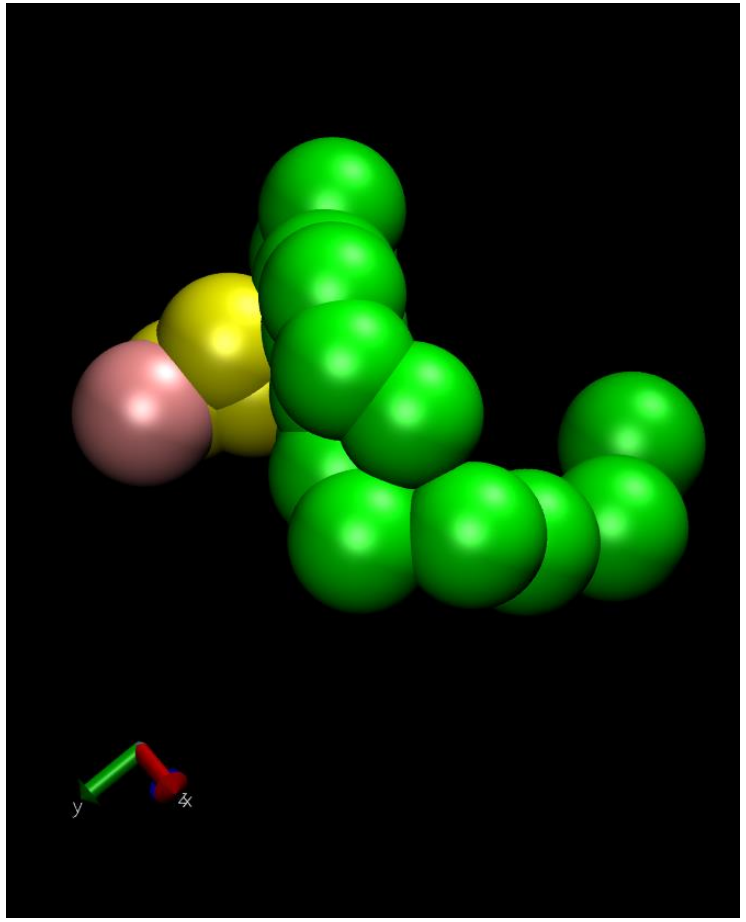


CG a bit too fat…

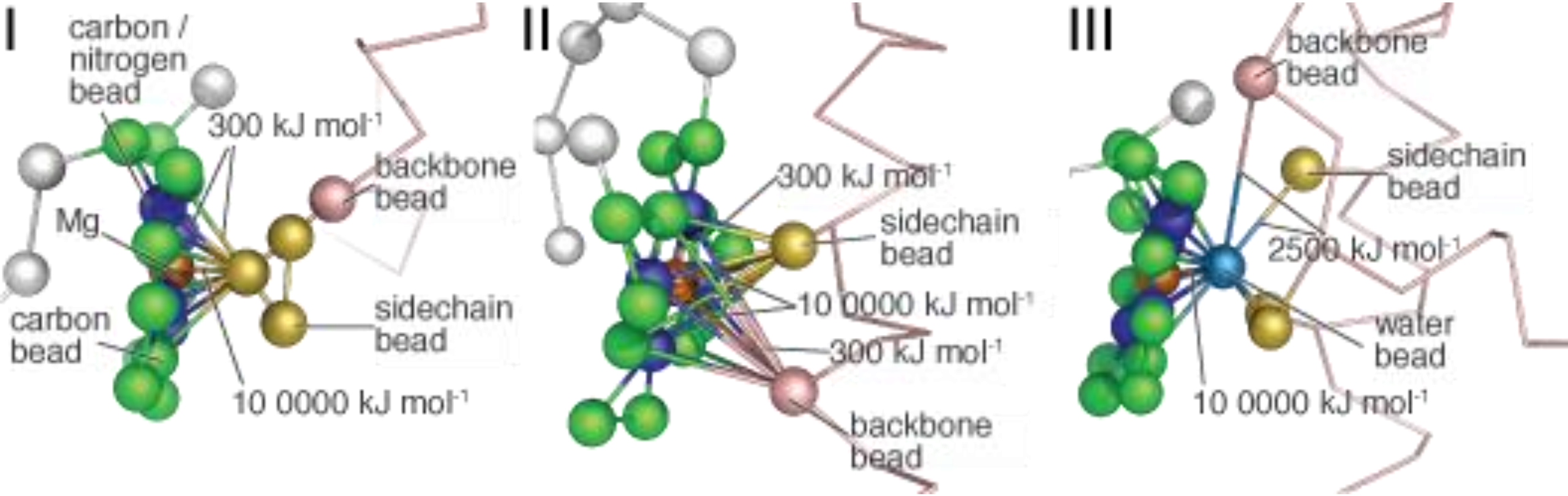
Fine grained



Coarse grained



Chlorophylls are bound in 3 different ways



Martinize.py → a itp file for every chain and cofactor

```
./Protein_T.itp
./Protein_Y.itp
./Protein_I.itp
./Protein_A.itp
./Protein_M.itp
./Protein_C.itp
./Protein_J.itp
./Protein_F.itp
./Protein_D.itp
./Protein_U.itp
./Protein_E.itp
./Protein_O.itp
./Protein_K.itp
./Protein_H.itp
./Protein_X.itp
./Protein_V.itp
./Protein_Z.itp
./Protein_L.itp
./Protein_B.itp
11:13:18@~/parameter/PS
```

```
; MARTINI (elnedyn22) Coarse Grained topology file for "Prote
; Created by py version 2.2
; Using the following options:  -f monomer_no_COfactor.pdb -c
; Sequence:
; MTTTLQRRESANLWERFCNWVTSTDNRLYVGWFGVIMIPTLLAATICFVIAFIAAPPVD
QGSFSDGMPLGISGTFNFMIVFQAEHNILMHPFHQLGVAGVFGGALFCAMHGSLVTSSLIR
MHERNAHNFPLDLA
; Secondary Structure:
; CCCCSCCCSSCC1111H2222TCTTSSSCC1111HHHHHHHHHHHHHHHHH2222CCCCC
2TCCCCSCCEE1111HHHHHHH2222CCCCC1111HHHHHHHHHHHHHHHHHH2222TCCC
2TTTTTCCSSCCCC

[ moleculetype ]
; Name            Exclusions
Protein_A           1

[ atoms ]
    1    Qd     1   MET    BB     1  1.0000 ; C
    2    C5     1   MET    SC1    2  0.0000 ; C
    3    P5     2   THR    BB     3  0.0000 ; C
    4    P1     2   THR    SC1    4  0.0000 ; C
    5    P5     3   THR    BB     5  0.0000 ; C
    6    P1     3   THR    SC1    6  0.0000 ; C
    7    P5     4   THR    BB     7  0.0000 ; C
    8    P1     4   THR    SC1    8  0.0000 ; C
    9    P5     5   LEU    BB     9  0.0000 ; S
   10    C1     5   LEU    SC1   10  0.0000 ; S
   11    P5     6   GLN    BB    11  0.0000 ; C
   12    P4     6   GLN    SC1   12  0.0000 ; C
   13    P5     7   ARG    BB    13  0.0000 ; C
   14    N0     7   ARG    SC1   14  0.0000 ; C
   15    Qd     7   ARG    SC2   15  1.0000 ; C
   16    P5     8   ARG    BB    16  0.0000 ; C
   17    N0     8   ARG    SC1   17  0.0000 ; C
   18    Qd     8   ARG    SC2   18  1.0000 ; C
   19    P5     9   GLU    BB    19  0.0000 ; S
   20    Qa     9   GLU    SC1   20 -1.0000 ; S
   21    P5    10   SER    BB    21  0.0000 ; S
   22    P1    10   SER    SC1   22  0.0000 ; S
   23    P4    11   ALA    BB    23  0.0000 ; C
   24    P5    12   ASN    BB    24  0.0000 ; C
   25    P5    12   ASN    SC1   25  0.0000 ; C
   26    Nd    13   LEU    BB    26  0.0000 ; 1
   27    C1    13   LEU    SC1   27  0.0000 ; 1
   28    Nd    14   TRP    BB    28  0.0000 ; 1
   29   SC4    14   TRP    SC1   29  0.0000 ; 1
   30   SNd    14   TRP    SC2   30  0.0000 ; 1
   31   SC5    14   TRP    SC3   31  0.0000 ; 1
   32   SC5    14   TRP    SC4   32  0.0000 ; 1
```

```
; Chlorophyll-A Coarse-grained
;MODIFIED CLA_CG topology JANUARY 2014, modiefied di
[ moleculetype ]
; molname         nrexcl
CLA               1

[ atoms ]
;id type resnr residu atom cgnr charge
;mapping to keep molecule symmetric and use small be
;Eight beads in the ring preserves symmetry better t
;Eight beads in the ring gives more rigid molecule
;Eight bead ring is to heavy.
   1    SP3       1     CLA    NA    1      -0.25
   2    SC3       1     CLA    CHA   2       0.00
   3    SP3       1     CLA    NB    3      -0.25
   4    SC3       1     CLA    CHB   4       0.00
   5    SP3       1     CLA    NC    5      -0.25
   6    SC3       1     CLA    CHC   6       0.00
   7    SP3       1     CLA    ND    7      -0.25
   8    SC3       1     CLA    CHD   8       0.00
   9    SC3       1     CLA    C1A   9       0.00
  10    SC3       1     CLA    C2A  10       0.00
  11    SC3       1     CLA    C1B  11       0.00
  12    SC3       1     CLA    C2B  12       0.00
  13    SC4       1     CLA    C1C  13       0.00
  14    SNa       1     CLA    C2C  14       0.00
  15    Na        1     CLA    T1   15       0.00
  16    SC3       1     CLA    C1D  16       0.00
  17    SC3       1     CLA    C2D  17       0.00
  18    Na        1     CLA    T2   18       0.00
  19    C3        1     CLA    T3   19       0.00
  20    C1        1     CLA    T4   20       0.00
  21    C1        1     CLA    T5   21       0.00
  22    C1        1     CLA    T6   22       0.00
  23    SQ0       1     CLA    MG   23       1.00

[ bonds ]
; bonds between beads
; The bonds are finetuned against a backmapped gromos
   1       9       1   0.273  30000.0
   9      10       1   0.298  23000.0
   1      10       1   0.318  25000.0
   3      11       1   0.274  25000.0
   3      12       1   0.319  23000.0
  11      12       1   0.29   21000.0
   5      13       1   0.24   25000.0
   5      14       1   0.355  21000.0
   6      14       1   0.28   25000.0
  13      14       1   0.29   20000.0
  14      15       1   0.305  10000.0
```

```
; If using this topology, please site:
; De Jong et al ...

[ moleculetype ]
; molname         nrexcl
PHO               1

[ atoms ]
;id type resnr residu atom cgnr charge
;mapping to keep molecule symmetric and use smal
;Eight beads in the ring preserves symmetry bett
;Eight beads in the ring gives more rigid molecu
;Eight bead ring is to heavy.
   1    SP3       1     PHO    NA    1       0.0
   2    SC3       1     PHO    CHA   2       0.0
   3    SP3       1     PHO    NB    3       0.0
   4    SC3       1     PHO    CHB   4       0.0
   5    SP3       1     PHO    NC    5       0.0
   6    SC3       1     PHO    CHC   6       0.0
   7    SP3       1     PHO    ND    7       0.0
   8    SC3       1     PHO    CHD   8       0.0
   9    SC3       1     PHO    C1A   9       0.0
  10    SC3       1     PHO    C2A  10       0.0
  11    SC3       1     PHO    C1B  11       0.0
  12    SC3       1     PHO    C2B  12       0.0
  13    SC4       1     PHO    C1C  13       0.0
  14    SNa       1     PHO    C2C  14       0.0
  15    Na        1     PHO    T1   15       0.0
  16    SC3       1     PHO    C1D  16       0.0
  17    SC3       1     PHO    C2D  17       0.0
  18    Na        1     PHO    T2   18       0.0
  19    C3        1     PHO    T3   19       0.0
  20    C1        1     PHO    T4   20       0.0
  21    C1        1     PHO    T5   21       0.0
  22    C1        1     PHO    T6   22       0.0

[ bonds ]
; bonds between beads
; The bonds are finetuned against a backmapped g
   1       9       1   0.273  30000.0
   9      10       1   0.298  23000.0
   1      10       1   0.318  25000.0
   3      11       1   0.274  25000.0
   3      12       1   0.319  23000.0
  11      12       1   0.29   21000.0
   5      13       1   0.24   25000.0
   5      14       1   0.355  21000.0
   6      14       1   0.28   25000.0
  13      14       1   0.29   20000.0
  14      15       1   0.305  10000.0
```

Binding cofactors into your protein? → One ITP needed

```
; Name          Exclusions
PSII_Protein    1

[ atoms ]
;CHAIN A
     1    Qd        1     MET     BB     1          1 ; C
     2    C5        1     MET    SC1     2          0 ; C
     3    P5        2     THR     BB     3          0 ; C
     4    P1        2     THR    SC1     4          0 ; C
     5    P5        3     THR     BB     5          0 ; C
     6    P1        3     THR    SC1     6          0 ; C
     7    P5        4     THR     BB     7          0 ; C
     8    P1        4     THR    SC1     8          0 ; C
     9    P5        5     LEU     BB     9          0 ; S
    10    C1        5     LEU    SC1    10          0 ; S
    11    P5        6     GLN     BB    11          0 ; C
PSII.itp
   770    P5      343     LEU     BB    770          0 ; C
   771    C1      343     LEU    SC1    771          0 ; C
   772    Qa      344     ALA     BB    772         -1 ; C
;CHAIN B
   773    Qd        2     GLY     BB    773          1 ; C
   774    P5        3     LEU     BB    774          0 ; C
   775    C1        3     LEU    SC1    775          0 ; C
   776    P4        4     PRO     BB    776          0 ; C
   777    C3        4     PRO    SC1    777          0 ; C
   778    P5        5     TRP     BB    778          0 ; C
   779   SC4        5     TRP    SC1    779          0 ; C
   780   SNd        5     TRP    SC2    780          0 ; C
   781   SC5        5     TRP    SC3    781          0 ; C
   782   SC5        5     TRP    SC4    782          0 ; C
   783    P5        6     TYR     BB    783          0 ; C
   784   SC4        6     TYR    SC1    784          0 ; C
   785   SC4        6     TYR    SC2    785          0 ; C
PSII.itp
  5775    P5       38     GLN     BB   5775          0 ; C
  5776    P4       38     GLN    SC1   5776          0 ; C
  5777    P5       39     ARG     BB   5777          0 ; C
  5778    N0       39     ARG    SC1   5778          0 ; C
  5779    Qd       39     ARG    SC2   5779          1 ; C
  5780    Qa       40     SER     BB   5780         -1 ; C
  5781    P1       40     SER    SC1   5781          0 ; C
;CHAIN Z
  5782    Qd        1     MET     BB   5782          1 ; C
  5783    C5        1     MET    SC1   5783          0 ; C
  5784    Nd        2     THR     BB   5784          0 ; 1
  5785    P1        2     THR    SC1   5785          0 ; 1
  5786    Nd        3     ILE     BB   5786          0 ; 1
  5787    C1        3     ILE    SC1   5787          0 ; 1
  5788    Nd        4     LEU     BB   5788          0 ; 1
  5789    C1        4     LEU    SC1   5789          0 ; 1
PSII.itp
/CHA
```

```
  5892    C1        52    LEU    SC1   5892          0 ; H
  5893    N0        53    VAL     BB   5893          0 ; H
  5894    C2        53    VAL    SC1   5894          0 ; H
  5895    Na        54    VAL     BB   5895          0 ; 2
  5896    C2        54    VAL    SC1   5896          0 ; 2
  5897    Na        55    GLY     BB   5897          0 ; 2
  5898    Na        56    VAL     BB   5898          0 ; 2
  5899    C2        56    VAL    SC1   5899          0 ; 2
  5900    Na        57    LEU     BB   5900          0 ; 2
  5901    C1        57    LEU    SC1   5901          0 ; 2
  5902    P5        58    ASN     BB   5902          0 ; C
  5903    P5        58    ASN    SC1   5903          0 ; C
  5904    P5        59    PHE     BB   5904          0 ; C
  5905   SC5        59    PHE    SC1   5905          0 ; C
  5906   SC5        59    PHE    SC2   5906          0 ; C
  5907   SC5        59    PHE    SC3   5907          0 ; C
  5908    P5        60    PHE     BB   5908          0 ; C
  5909   SC5        60    PHE    SC1   5909          0 ; C
  5910   SC5        60    PHE    SC2   5910          0 ; C
  5911   SC5        60    PHE    SC3   5911          0 ; C
  5912   Nda        61    VAL     BB   5912          0 ; T
  5913    C2        61    VAL    SC1   5913          0 ; T
  5914    Qa        62    VAL     BB   5914         -1 ; C
  5915    C2        62    VAL    SC1   5915          0 ; C
;============================================================
;CLA
  5916   SP3       604    CLA     NA   5916      -0.25
  5917   SC3       604    CLA    CHA   5917          0
  5918   SP3       604    CLA     NB   5918      -0.25
  5919   SC3       604    CLA    CHB   5919          0
  5920   SP3       604    CLA     NC   5920      -0.25
  5921   SC3       604    CLA    CHC   5921          0
  5922   SP3       604    CLA     ND   5922      -0.25
  5923   SC3       604    CLA    CHD   5923          0
  5924   SC3       604    CLA    C1A   5924          0
  5925   SC3       604    CLA    C2A   5925          0
  5926   SC3       604    CLA    C1B   5926          0
  5927   SC3       604    CLA    C2B   5927          0
  5928   SC4       604    CLA    C1C   5928          0
  5929   SNa       604    CLA    C2C   5929          0
  5930    Na       604    CLA     T1   5930          0
  5931   SC3       604    CLA    C1D   5931          0
  5932   SC3       604    CLA    C2D   5932          0
  5933    Na       604    CLA     T2   5933          0
  5934    C3       604    CLA     T3   5934          0
  5935    C1       604    CLA     T4   5935          0
  5936    C1       604    CLA     T5   5936          0
  5937    C1       604    CLA     T6   5937          0
  5938   SQ0       604    CLA     MG   5938          1
  5939   SP3       607    CLA     NA   5939      -0.25
  5940   SC3       607    CLA    CHA   5940          0
  5941   SP3       607    CLA     NB   5941      -0.25
```

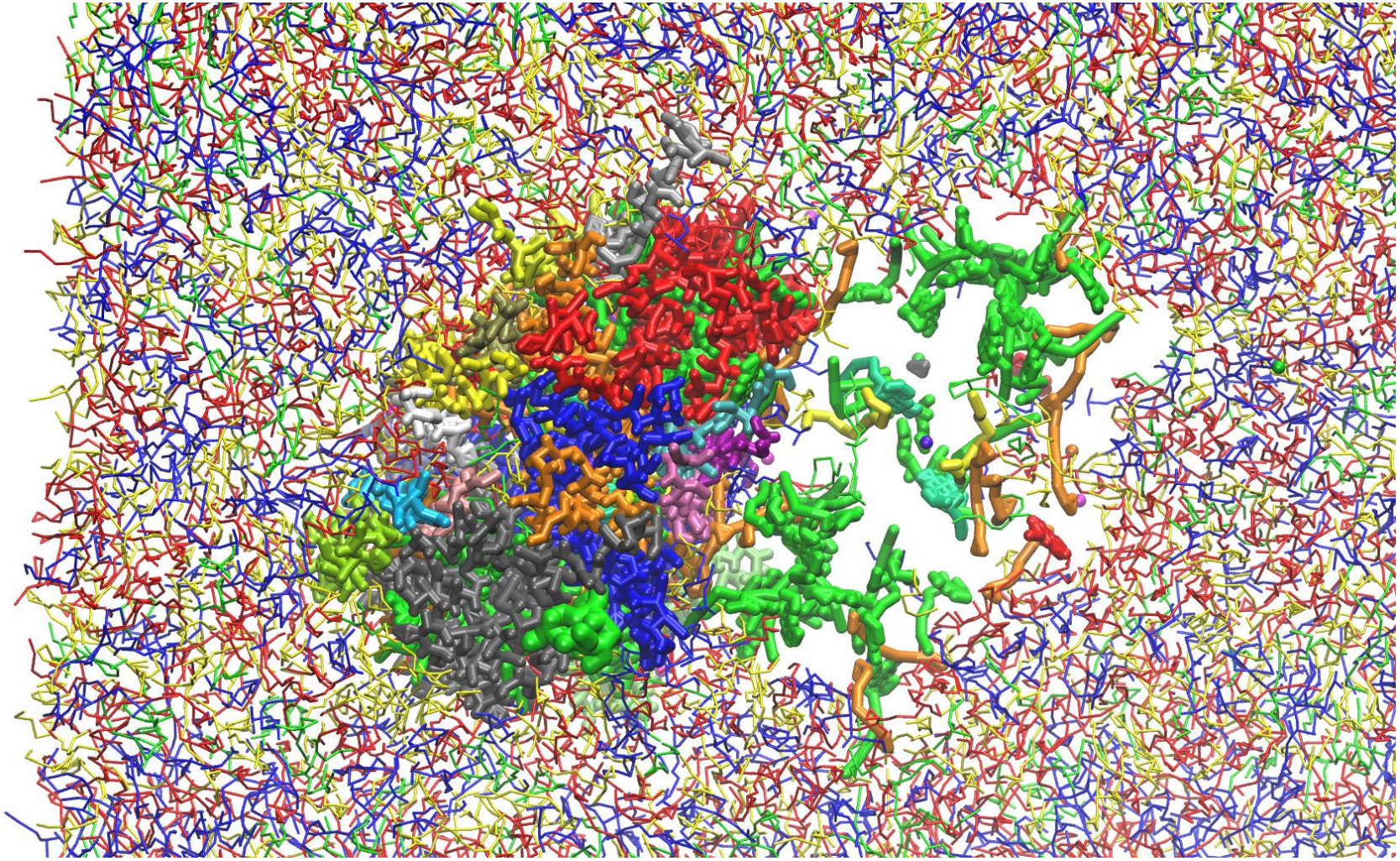
```
  6750   SC3       642    HEM    CHC   6750          0
  6751   SP1       642    HEM     ND   6751       -0.1
  6752   SC3       642    HEM    CHD   6752          0
  6753   SC3       642    HEM    C2A   6753          0
  6754   SC3       642    HEM    C3A   6754          0
  6755   SC3       642    HEM    C2B   6755          0
  6756   SC3       642    HEM    C3B   6756          0
  6757   SC3       642    HEM    C1C   6757          0
  6758   SC3       642    HEM    C2C   6758          0
  6759    Qa       642    HEM    C3C   6759         -1
  6760   SC3       642    HEM    C1D   6760          0
  6761   SC3       642    HEM    C2D   6761          0
  6762    Qa       642    HEM    C3D   6762         -1
  6763   SQ0       642    HEM     FE   6763        0.4
;OEX
  6764    Q0       601    OEX    XCA   6764        1.5 ; the tot
oxyl groups, of which the charges are neutralized in vivo by the OE
  6765    Q0       601    OEX     X1   6765        1.5
  6766    Q0       601    OEX     X2   6766        1.5
  6767    Q0       601    OEX     X3   6767        1.5
;FE
  6768    Q0       603    FE2     FE   6768         +2
;BCT
  6769   Qda       681    BCT    BCT   6769         -1
;CA
  6770    Q0       767     CA     CA   6770         +2
  6771    Q0       796     CA     CA   6771         +2
  6772    Q0       803     CA     CA   6772         +2
;MG
  6773    Q0       771     MG     MG   6773         +2
;CL
  6774    Q0       679     CL    CL-   6774         -1
  6775    Q0       680     CL    CL-   6775         -1
  6776    Q0       808     CL    CL-   6776         -1
;PHEOPHYTINS
  6777   SP3       608    PHO     NA   6777          0
  6778   SC3       608    PHO    CHA   6778          0
  6779   SP3       608    PHO     NB   6779          0
  6780   SC3       608    PHO    CHB   6780          0
  6781   SP3       608    PHO     NC   6781          0
  6782   SC3       608    PHO    CHC   6782          0
  6783   SP3       608    PHO     ND   6783          0
  6784   SC3       608    PHO    CHD   6784          0
  6785   SC3       608    PHO    C1A   6785          0
  6786   SC3       608    PHO    C2A   6786          0
  6787   SC3       608    PHO    C1B   6787          0
  6788   SC3       608    PHO    C2B   6788          0
  6789   SC4       608    PHO    C1C   6789          0
  6790   SNa       608    PHO    C2C   6790          0
  6791    Na       608    PHO     T1   6791          0
  6792   SC3       608    PHO    C1D   6792          0
  6793   SC3       608    PHO    C2D   6793          0
```

Everything ready?

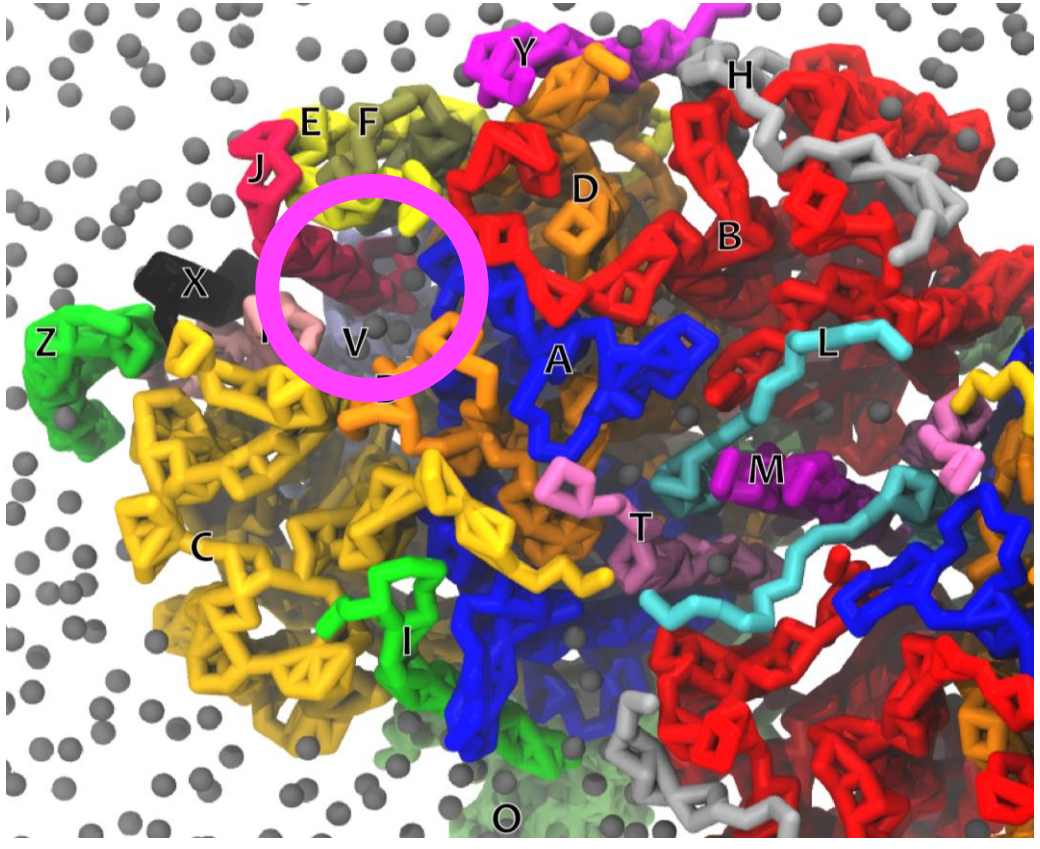
- Good energy minimiaztion
  - Muliple rounds
- Careful equilibration
  - Slowly increasing
    - Time step
    - temperature

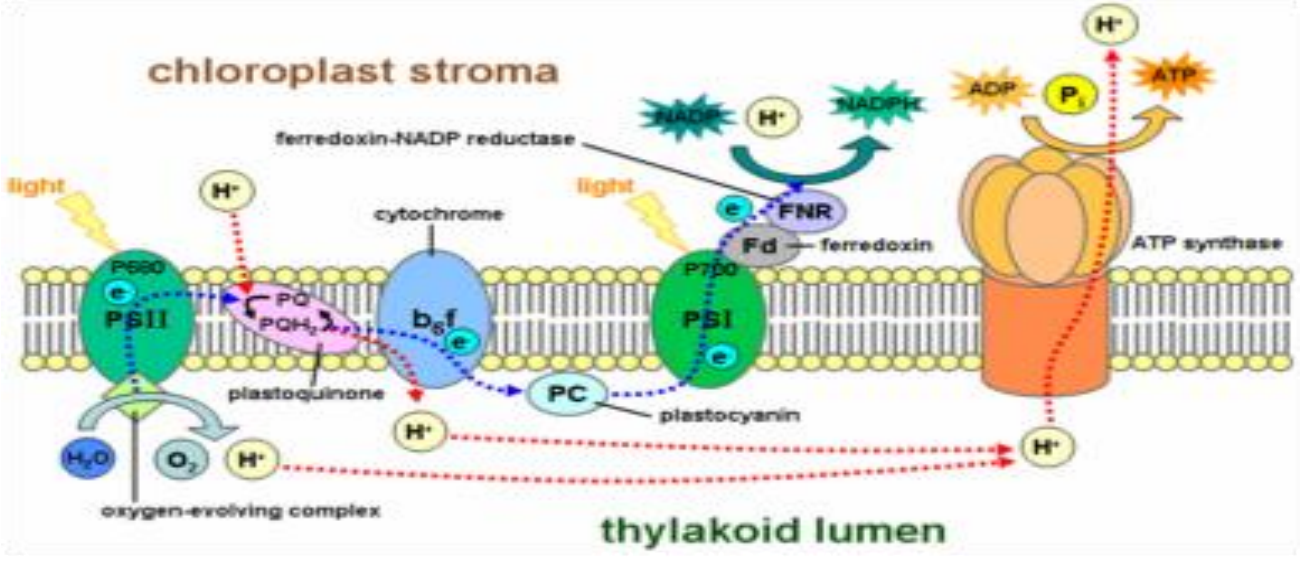
Where should i look at?

- [MOVIE PSII TOP]
- Nothing seems to happen
- /to much to look at.



PLQ exchange cavity





Analysis

The following sequence is a visualisation of an actual molecular dynamics simulation.

Counting PLQs entering/leaving PSII

- PLQs are long molecules
- Entry/exit is slow
- PLQs can enter partly and then go back
- PLQ can be in 2 channels at the same time
- **Channels move over time**

Track the PLQs on a bead level

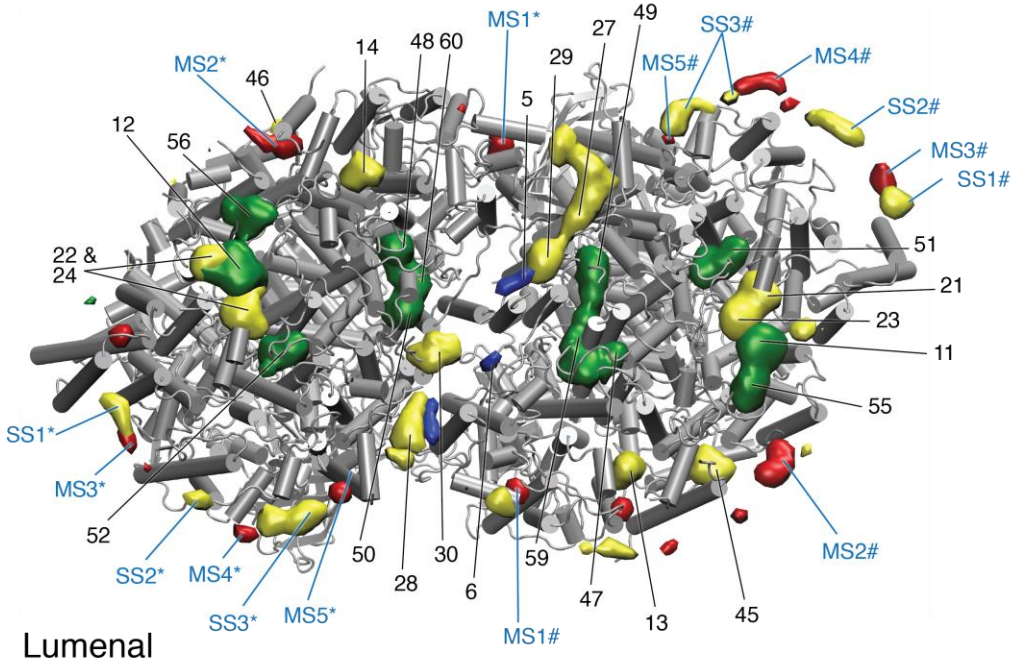
- For every PLQ bead, each frame:
  - Is it in contact with any of the three channels
  - If it is inside or outside the cavity

- For every PLQ
  - If more than 70% of its beads moved through the same channel
  - entry/exit event

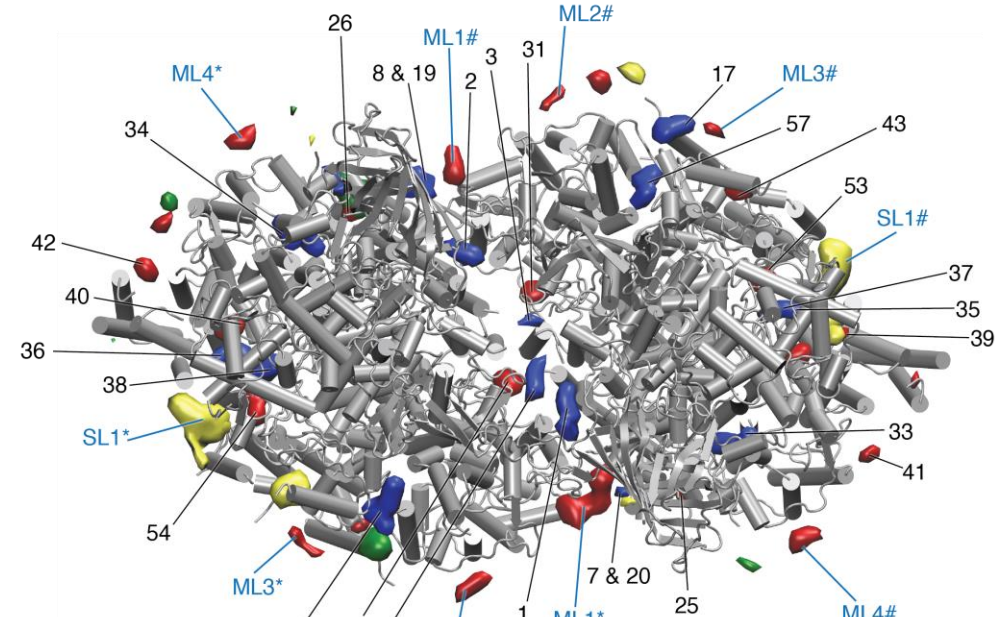
PLQ ENTERING MOVIE

Lipid binding sites

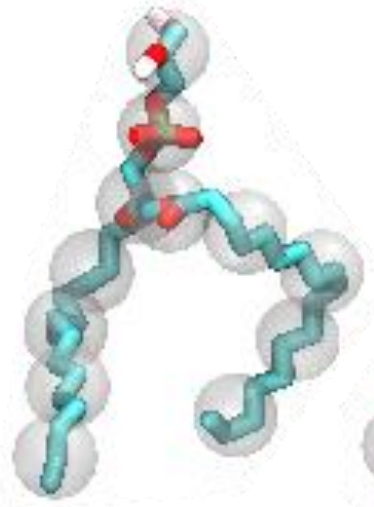
Stromal



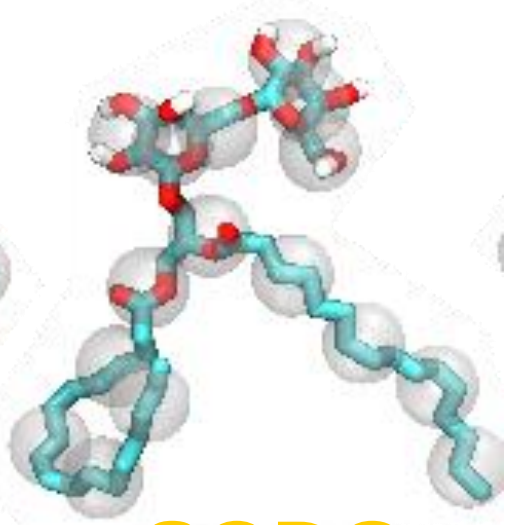
Lumenal



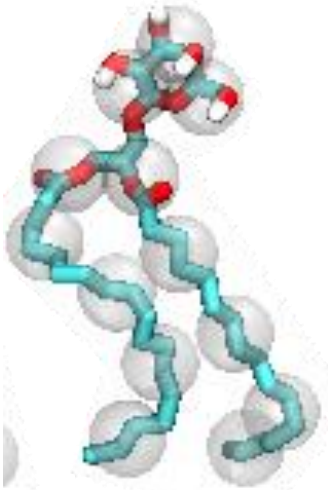
PG



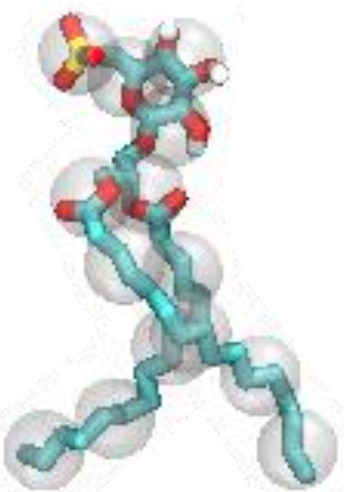
DGDG



MGDG

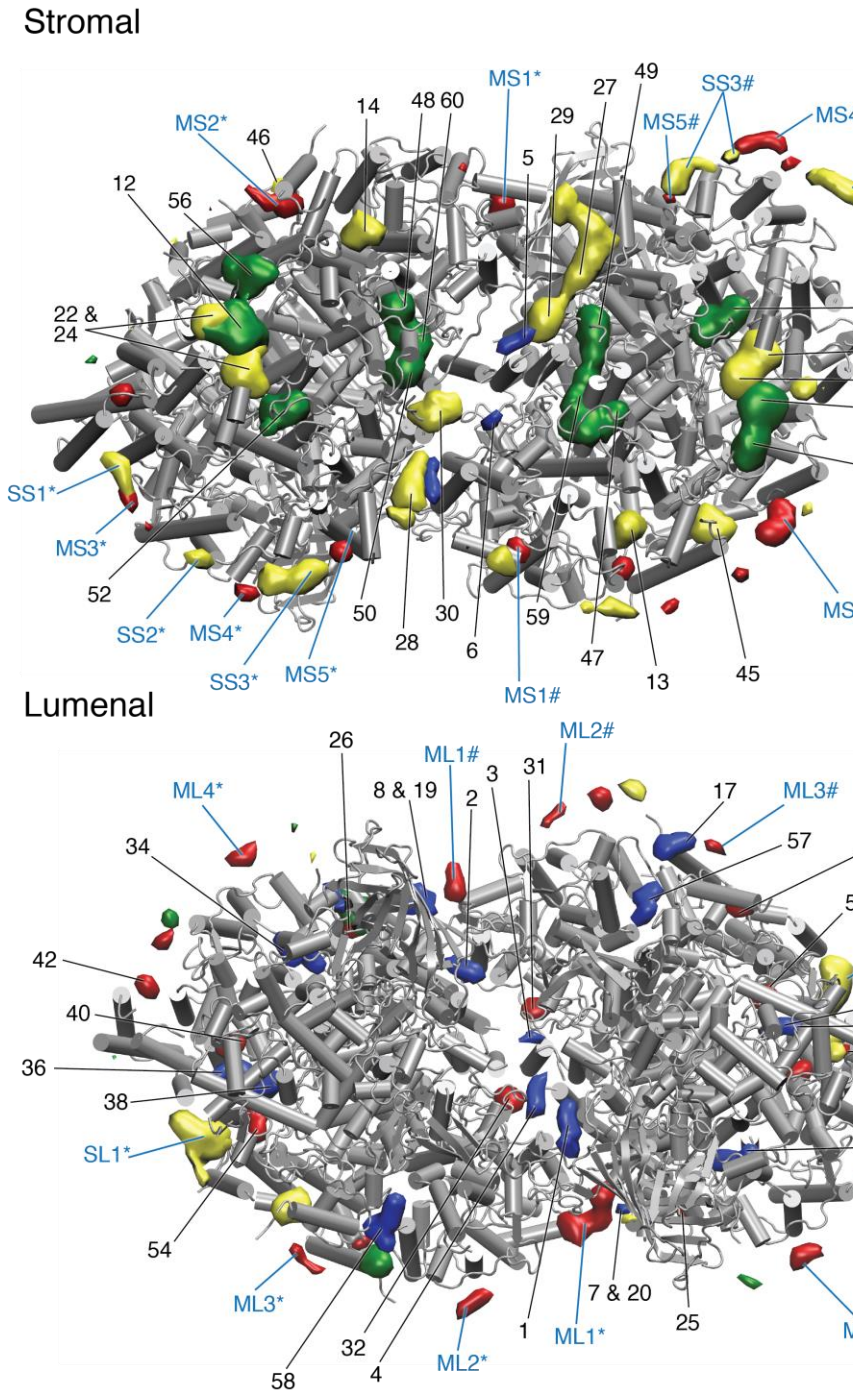


SQDG



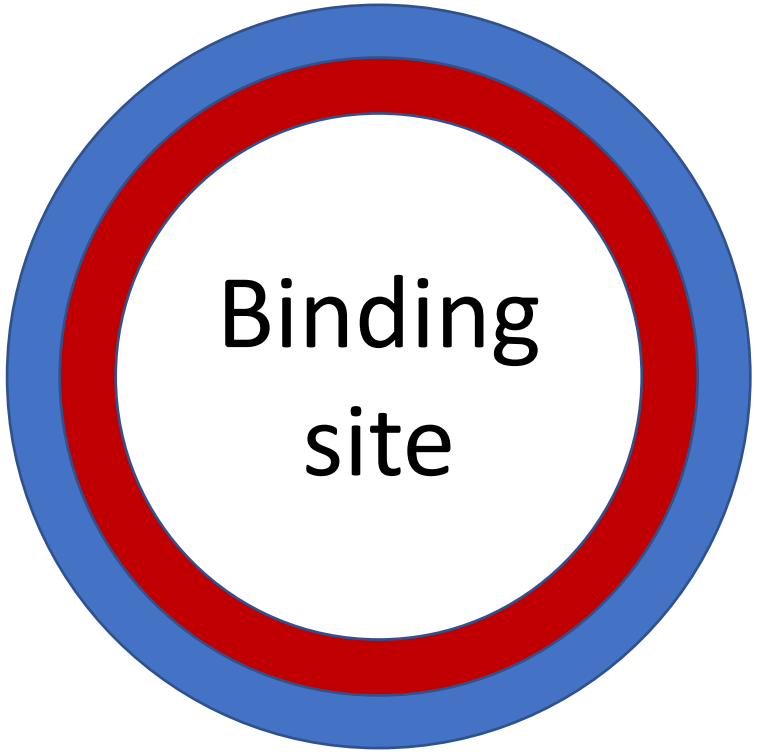
How to define a binding site

- Define ‘phrases’
  - Amino acids a lipid is in contact with at a frame
  - Total number of phrases = number of lipids x frames
- Cluster phrases
- From clusters define binding sites



Residence times

- Double cut off method → removes noise
- Lipid is in contact with binding site
  - Once it enters red ring
- Lipid is not in contact anymore
  - When leaving blue ring



Conclusion

- Carefully prepare your system, it will take a long time!

- Look a lot at your simulations:
  - Only protein
  - Protein with lipids
  - Protein cofactors
  - Cofactors lipids
- Team up with smart colleagues (e.g. writing scripts)
- Make descisions
  - Often there is not a right way to do things