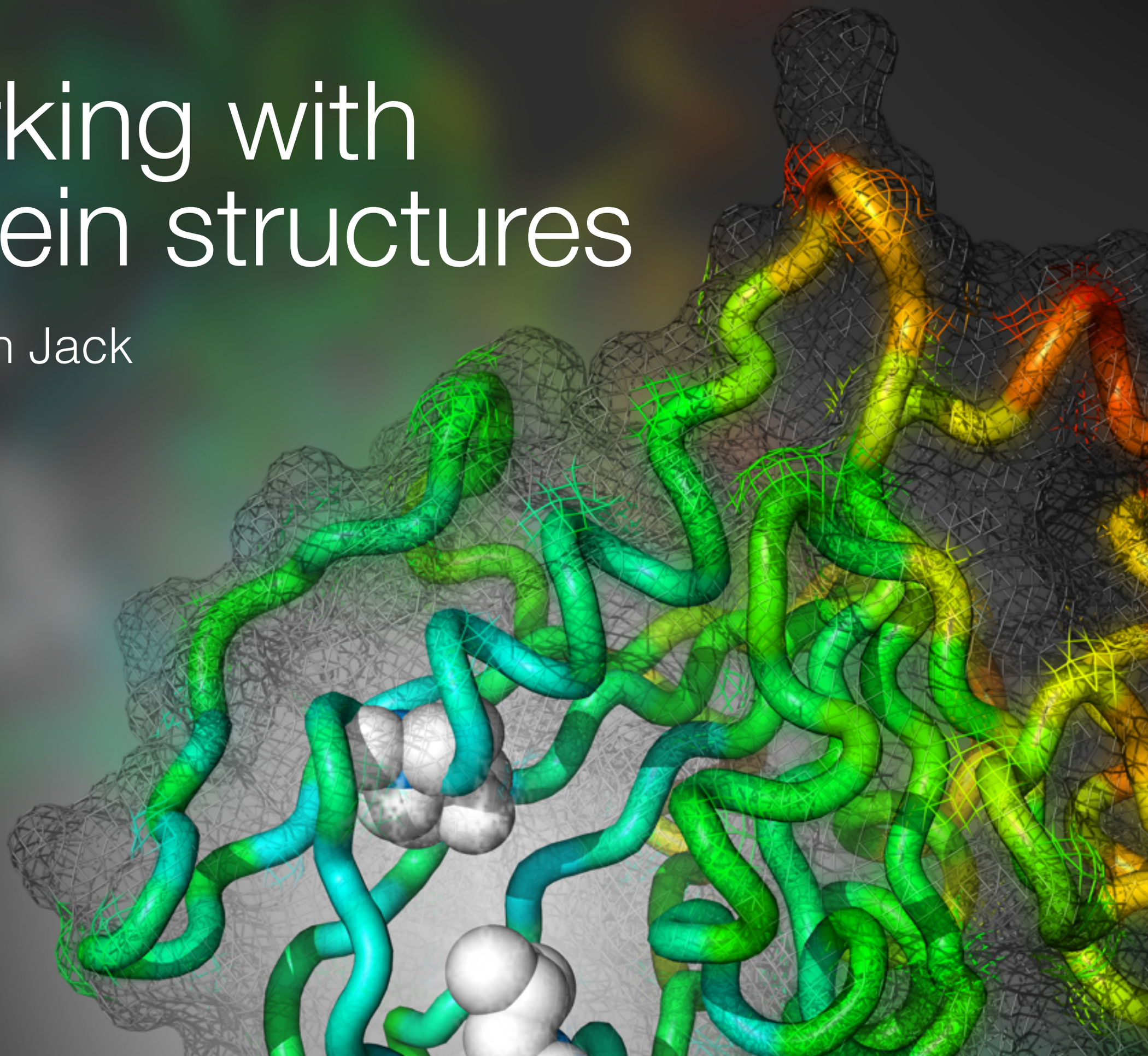
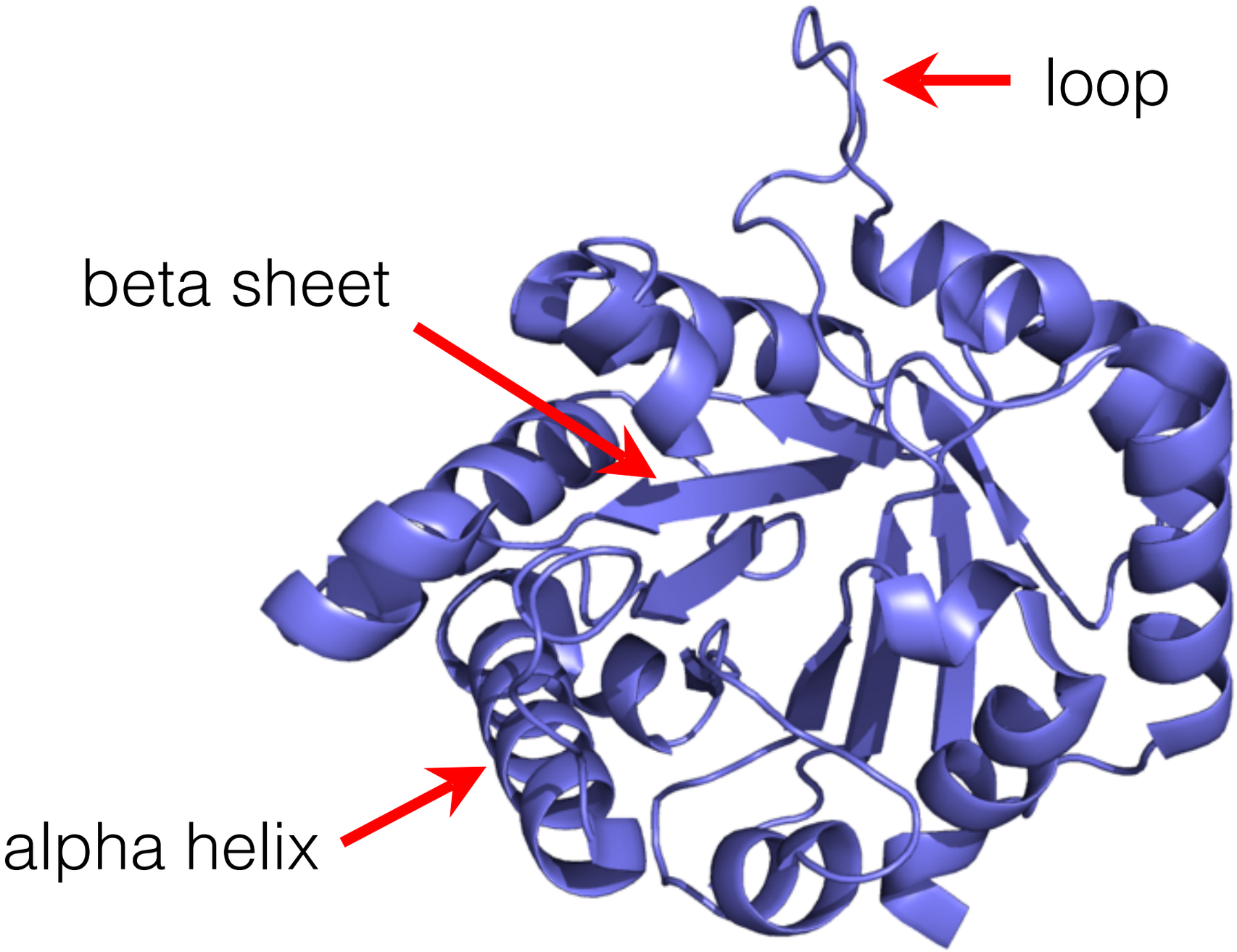
Working with protein structures

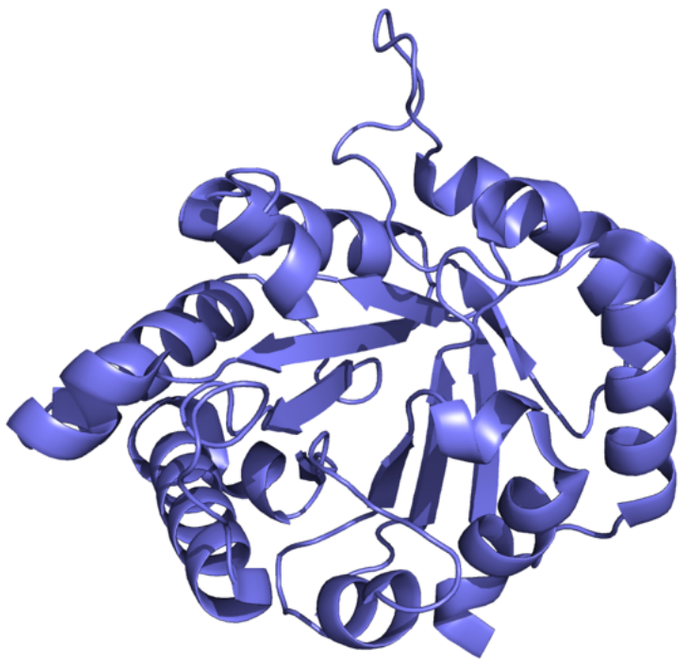
Benjamin Jack

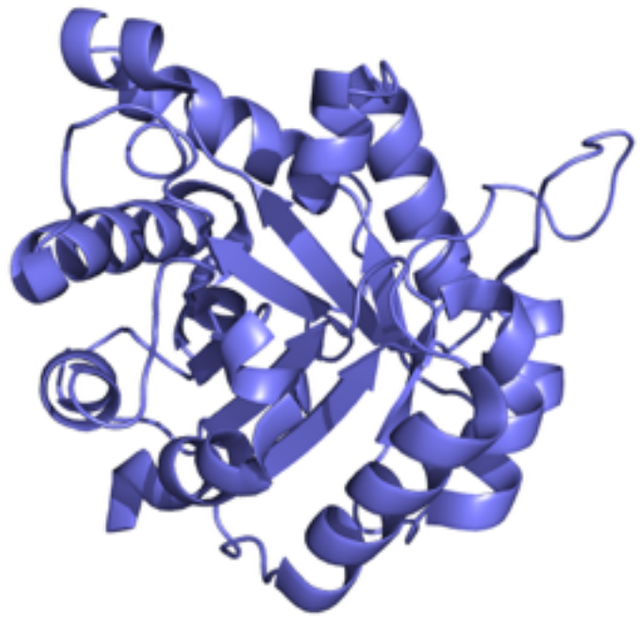


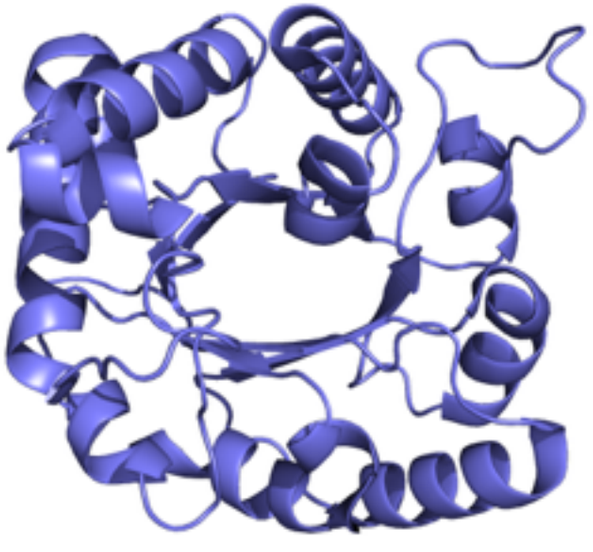
Structure of Triosephosphate Isomerase
PDB ID: 1HTI

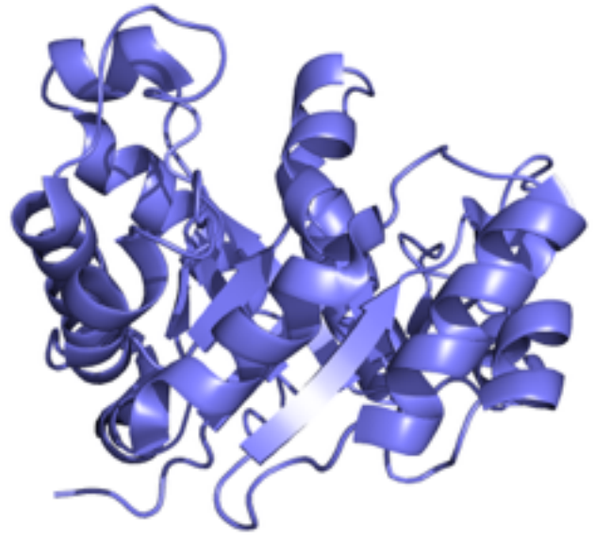


Different perspectives of the same structure

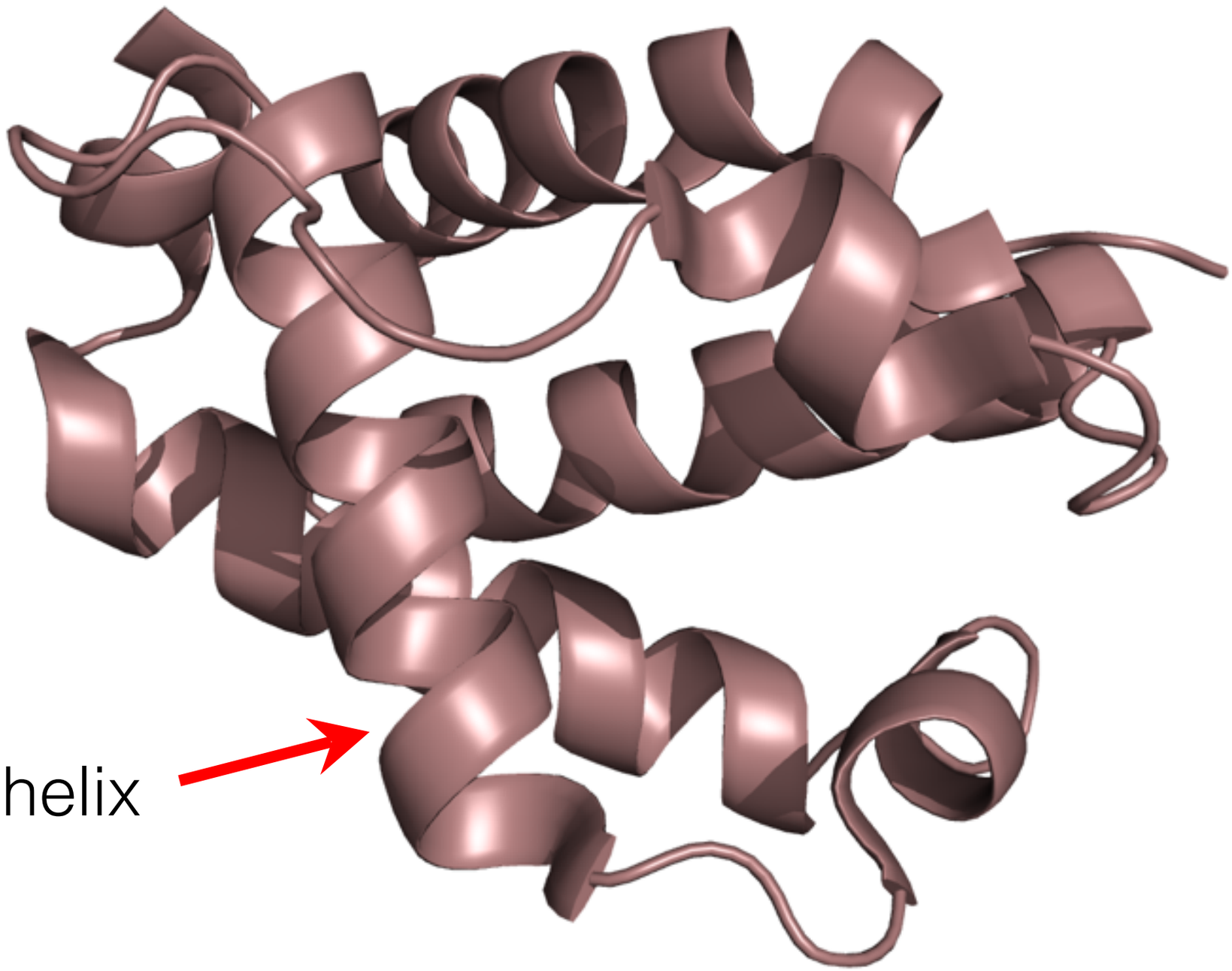




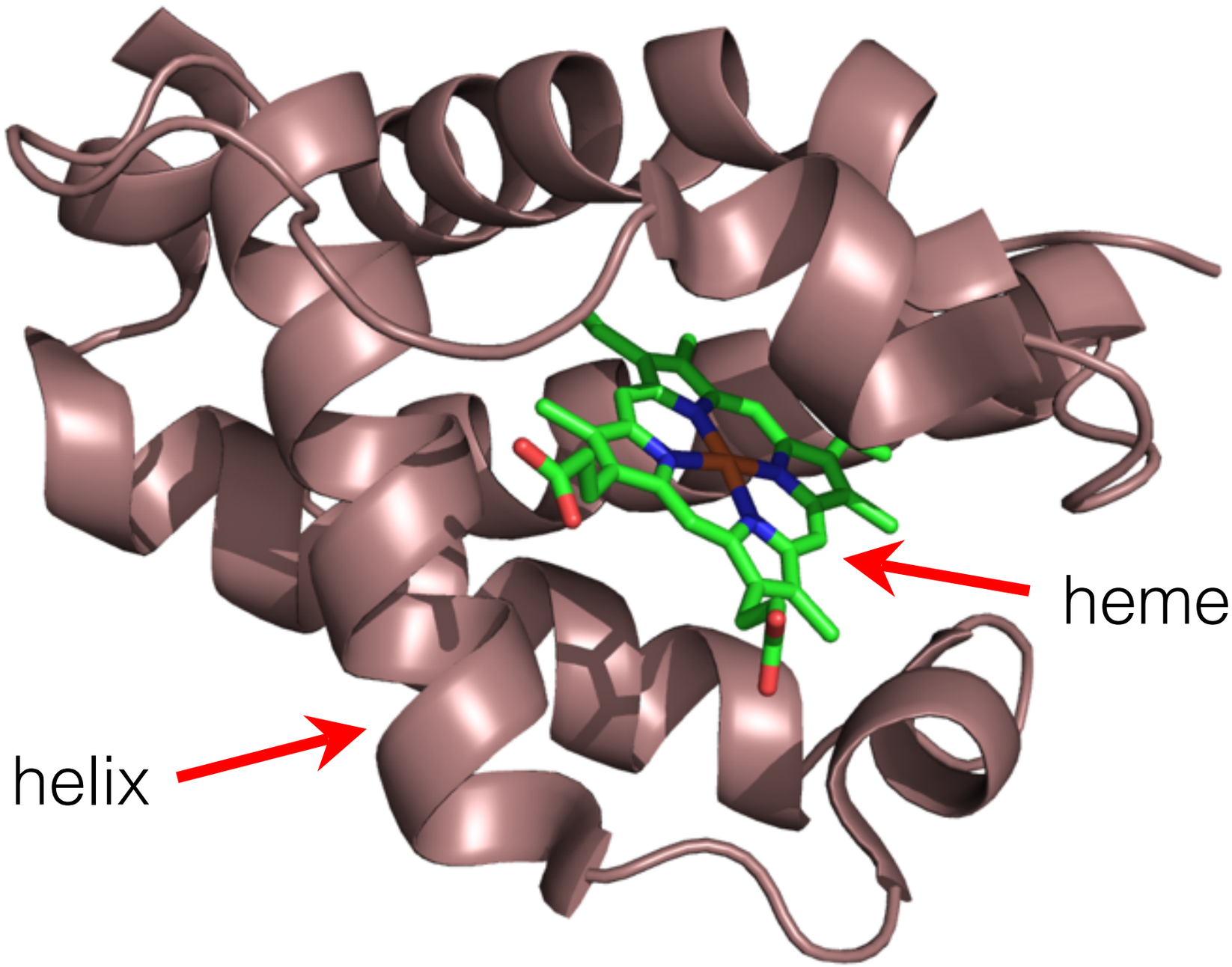




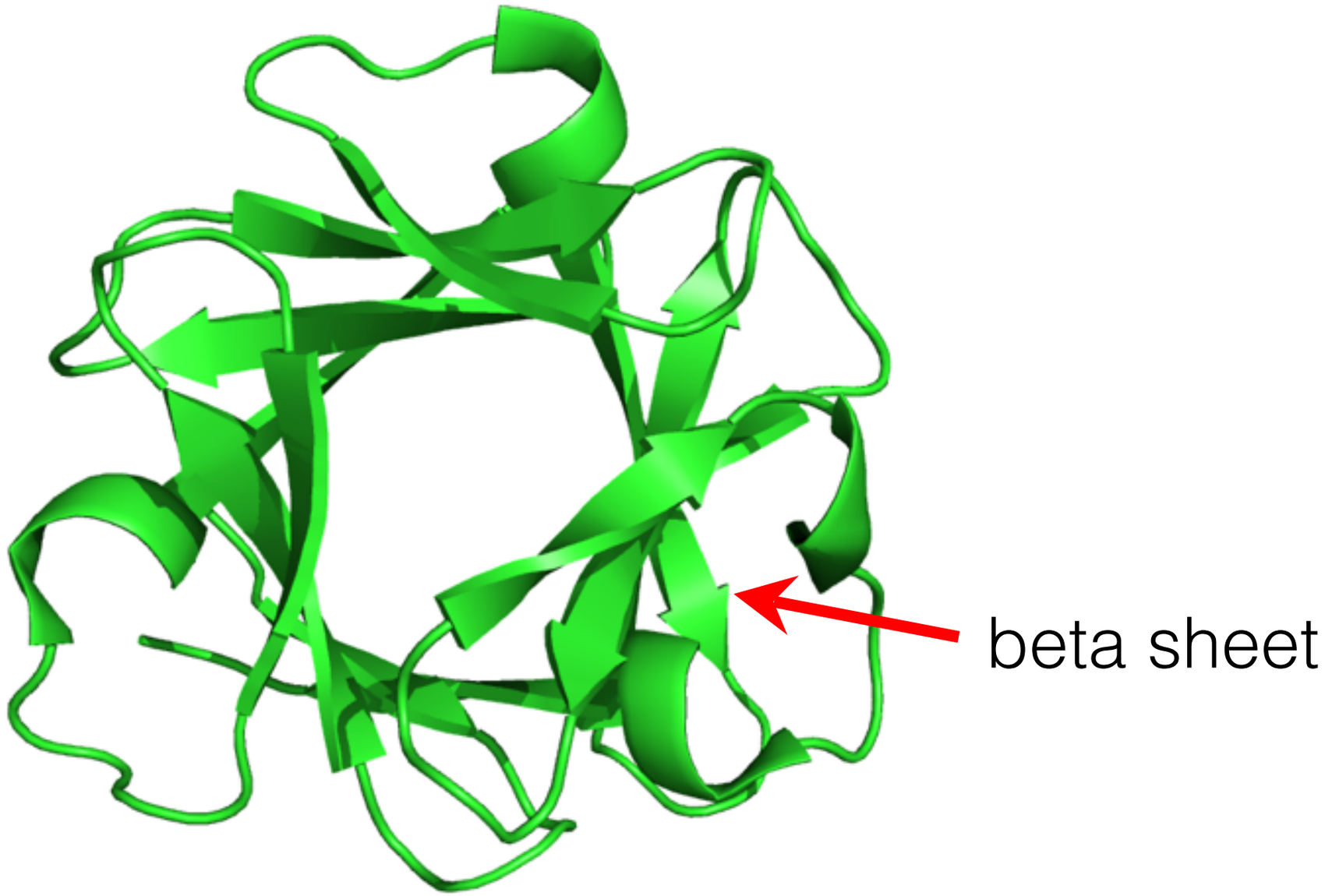
Structure of Truncated Hemoglobin
PDB ID: 1DLW



Structure of Truncated Hemoglobin
PDB ID: 1DLW

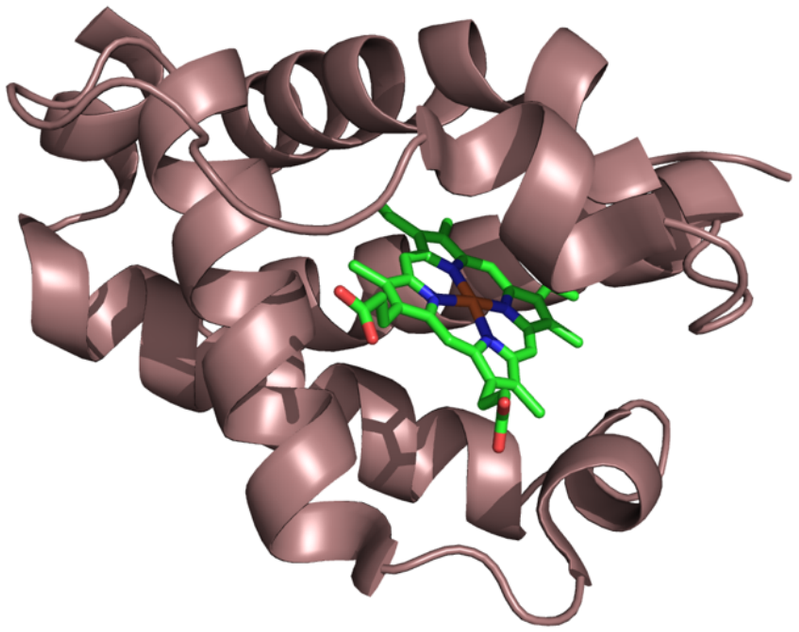


Structure of Basic Fibroblast Growth Factor
PDB ID: 1BFG

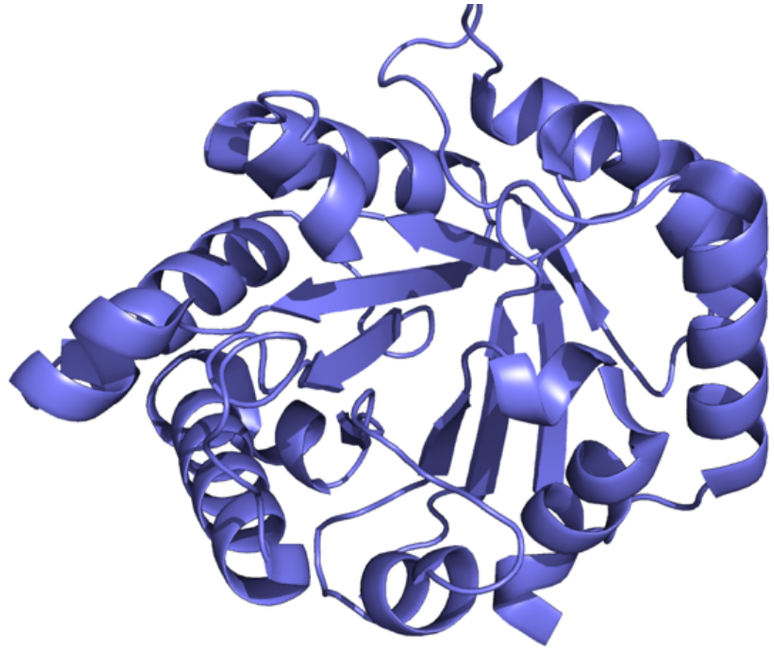


We classify structures by their alpha and beta content

all alpha

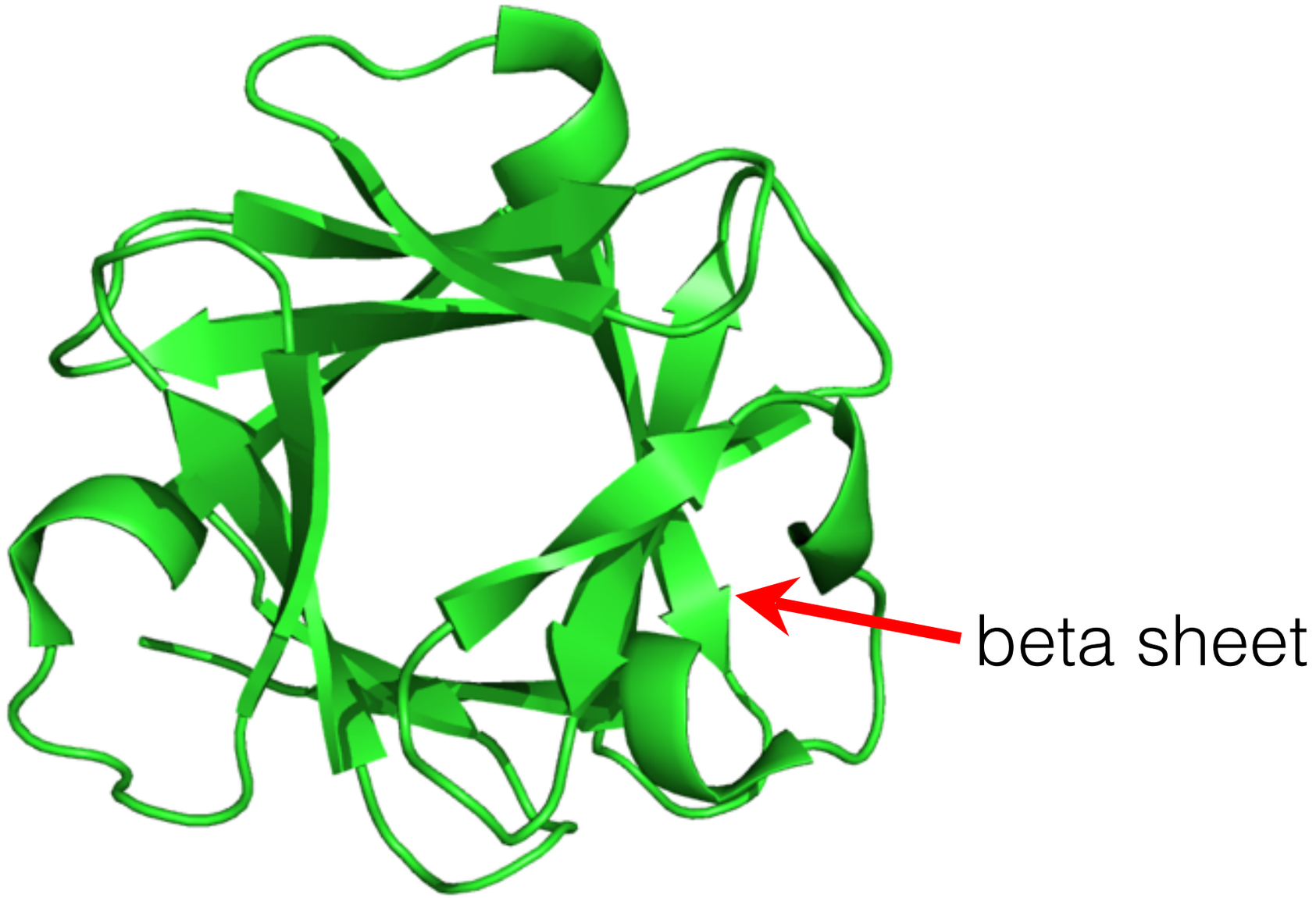


all beta

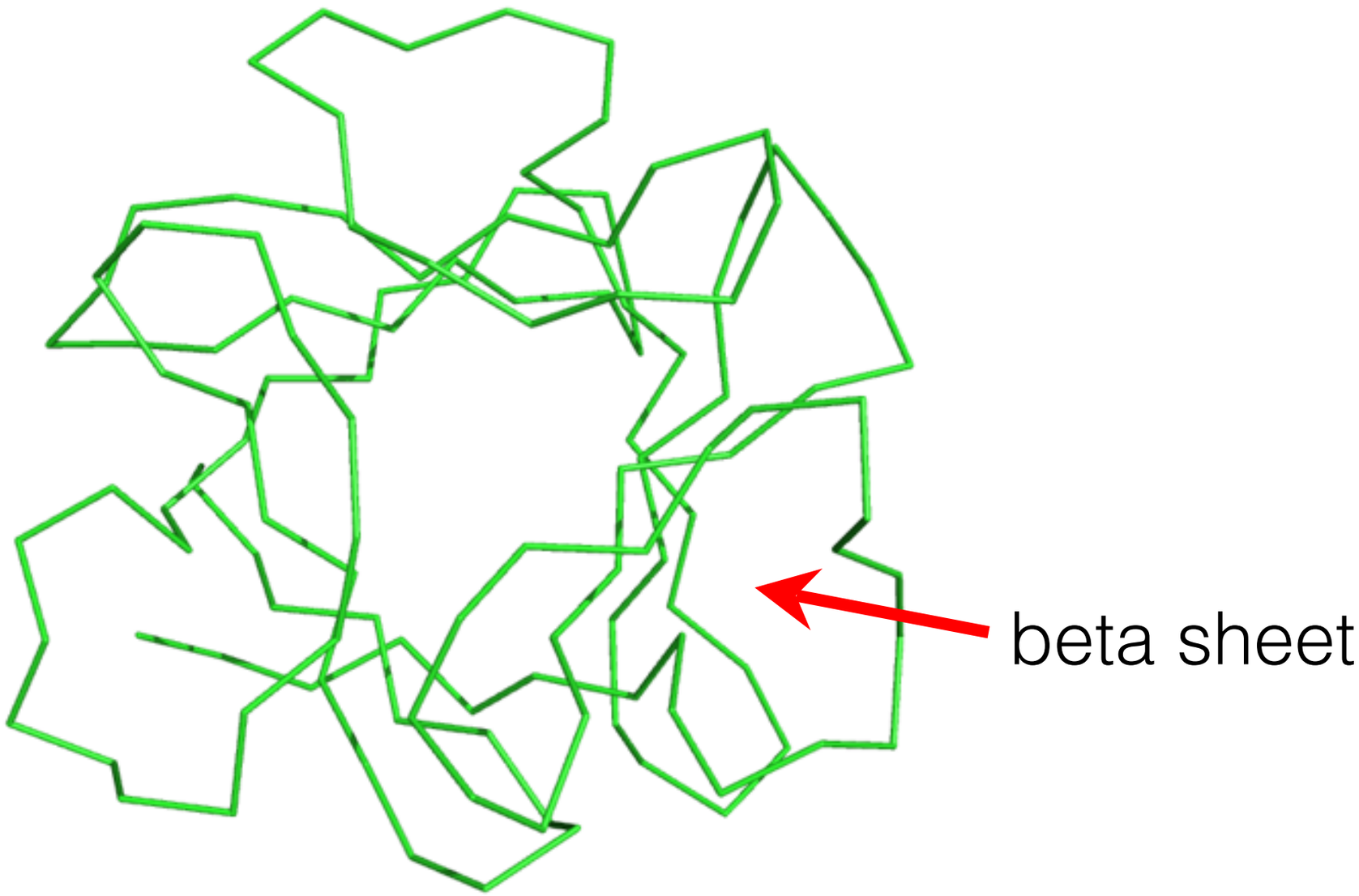


alpha and beta

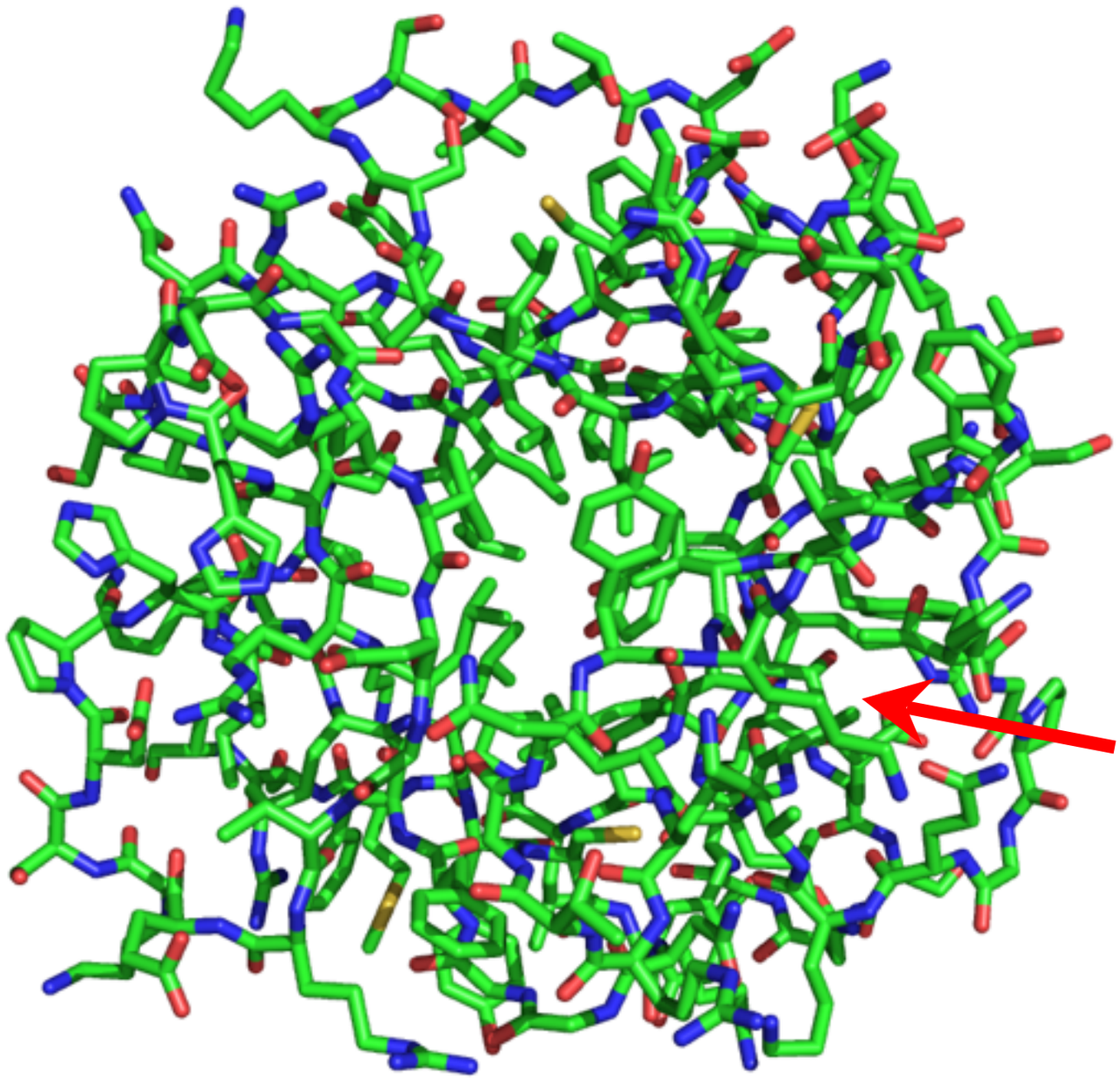
Different visualizations of a structure: Cartoon



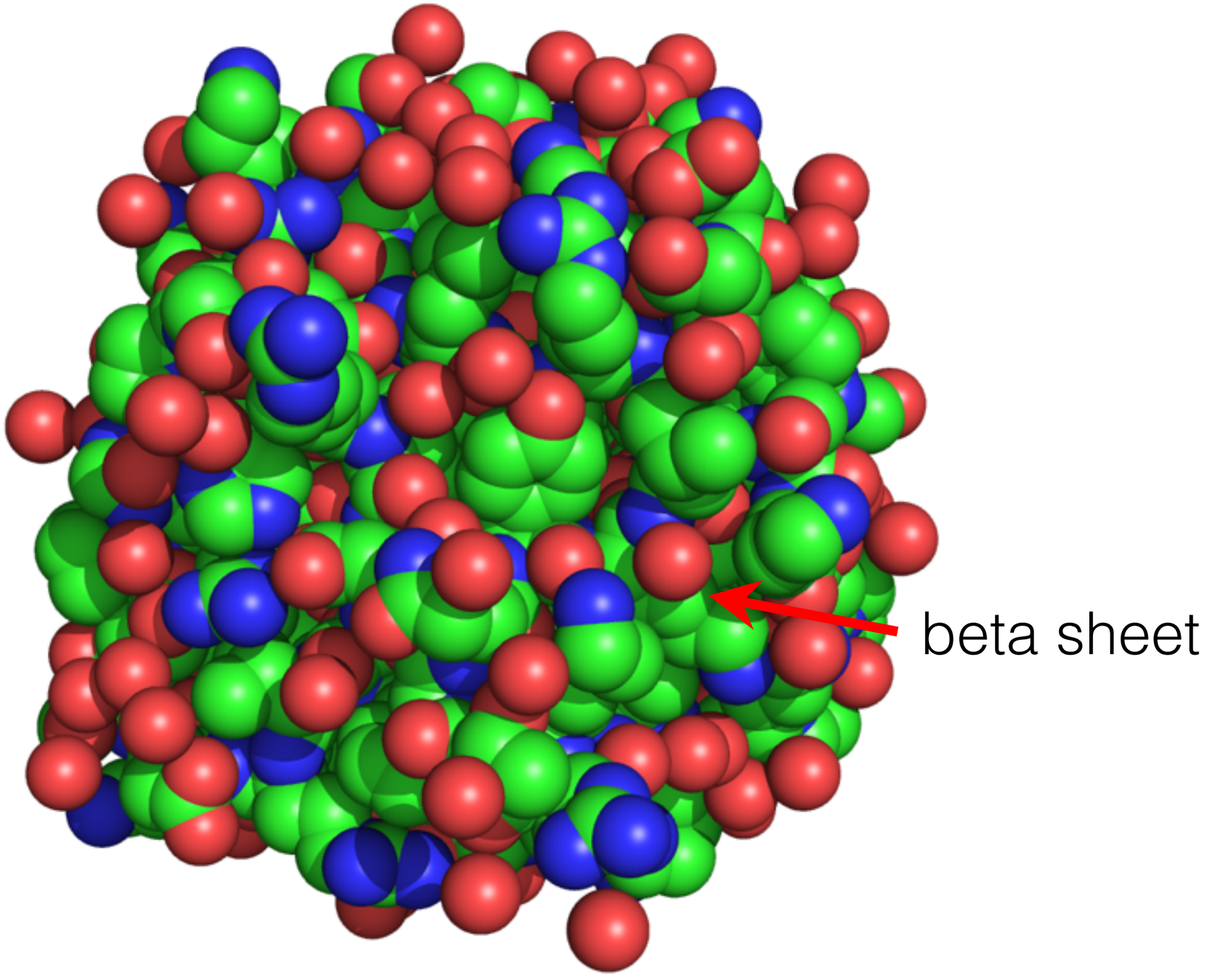
Different visualizations of a structure: Ribbon



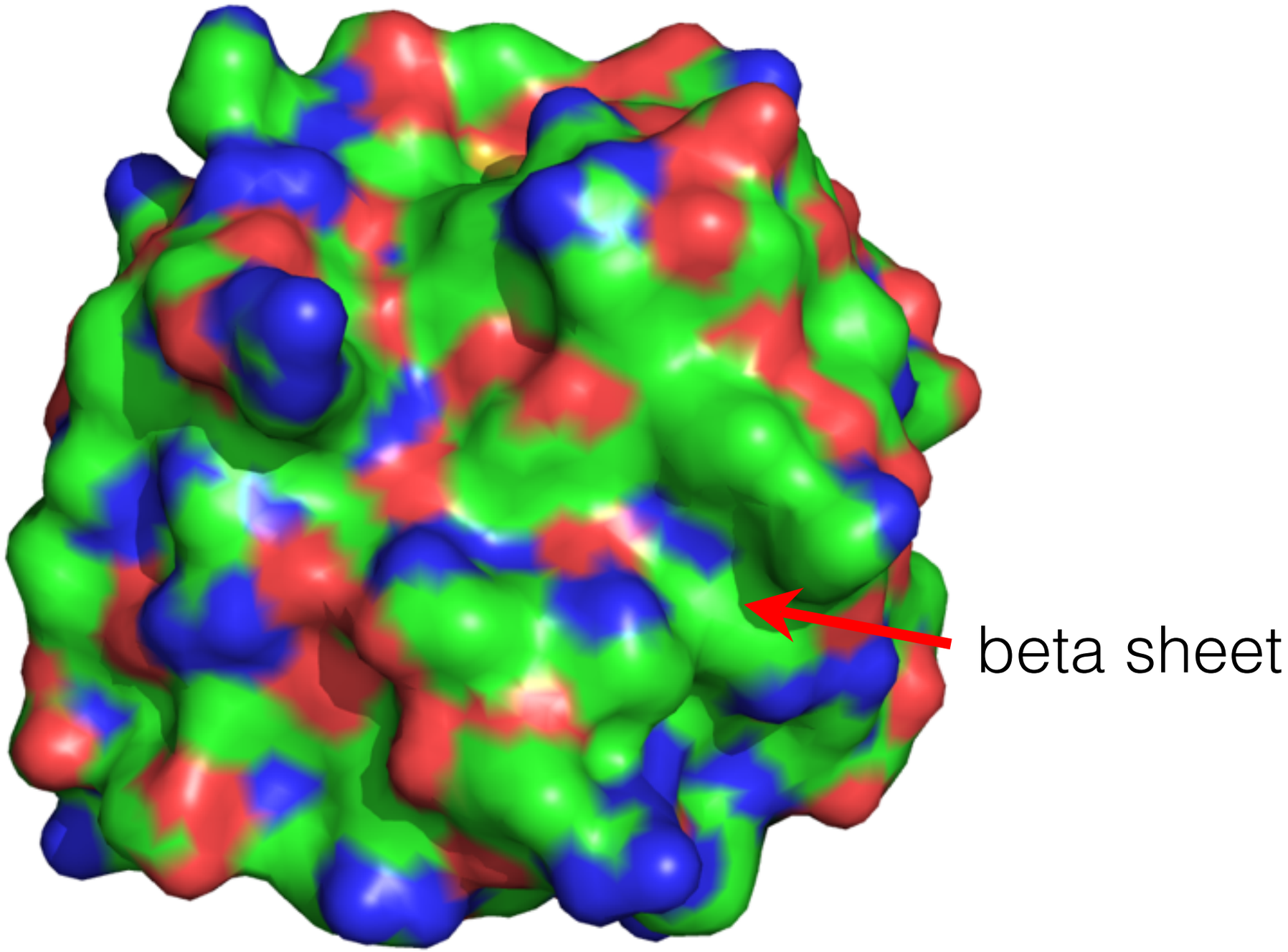
Different visualizations of a structure: Sticks



Different visualizations of a structure: Spheres

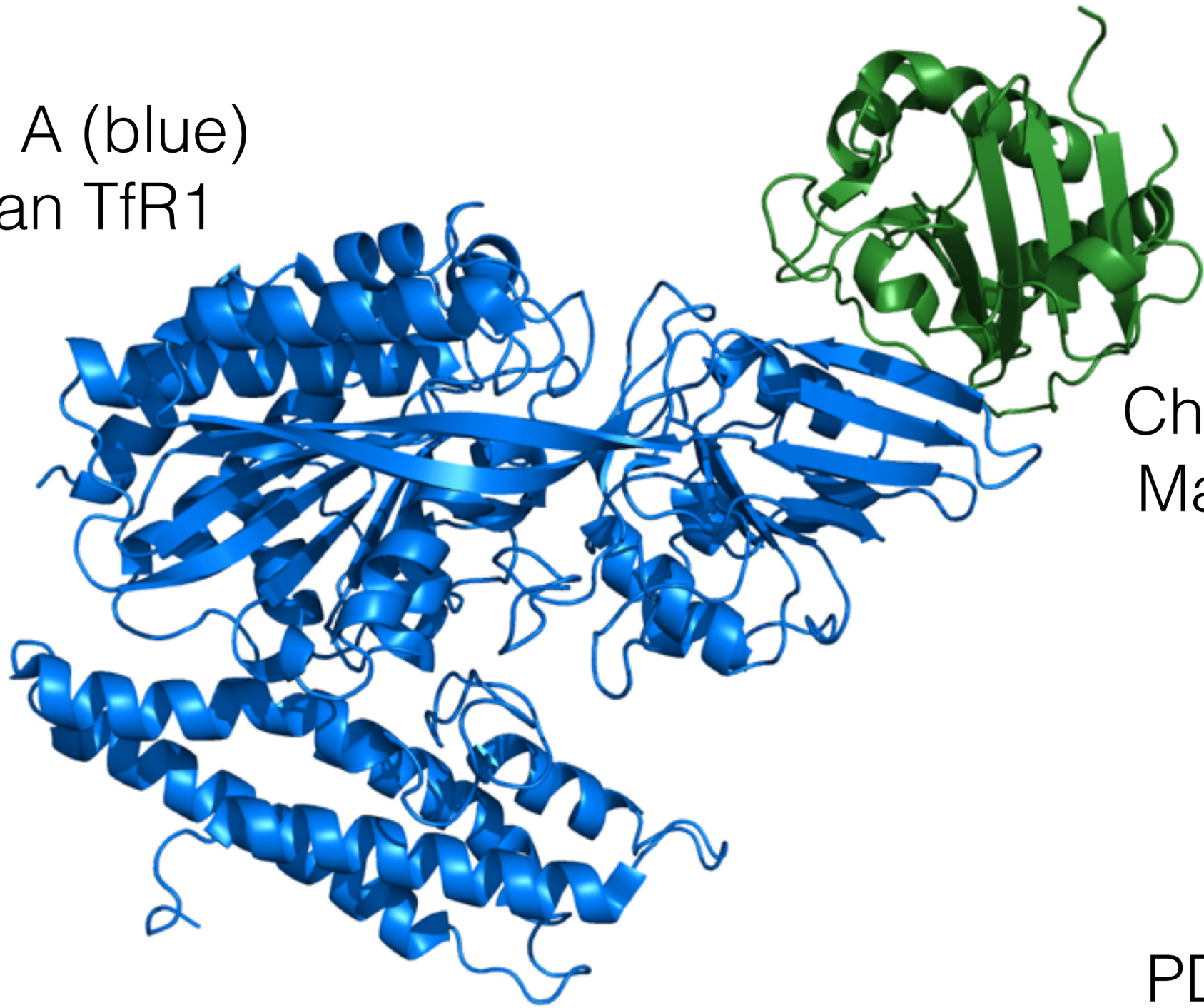


Different visualizations of a structure: Surface



PDB files can contain multiple chains

Chain A (blue)
human TfR1



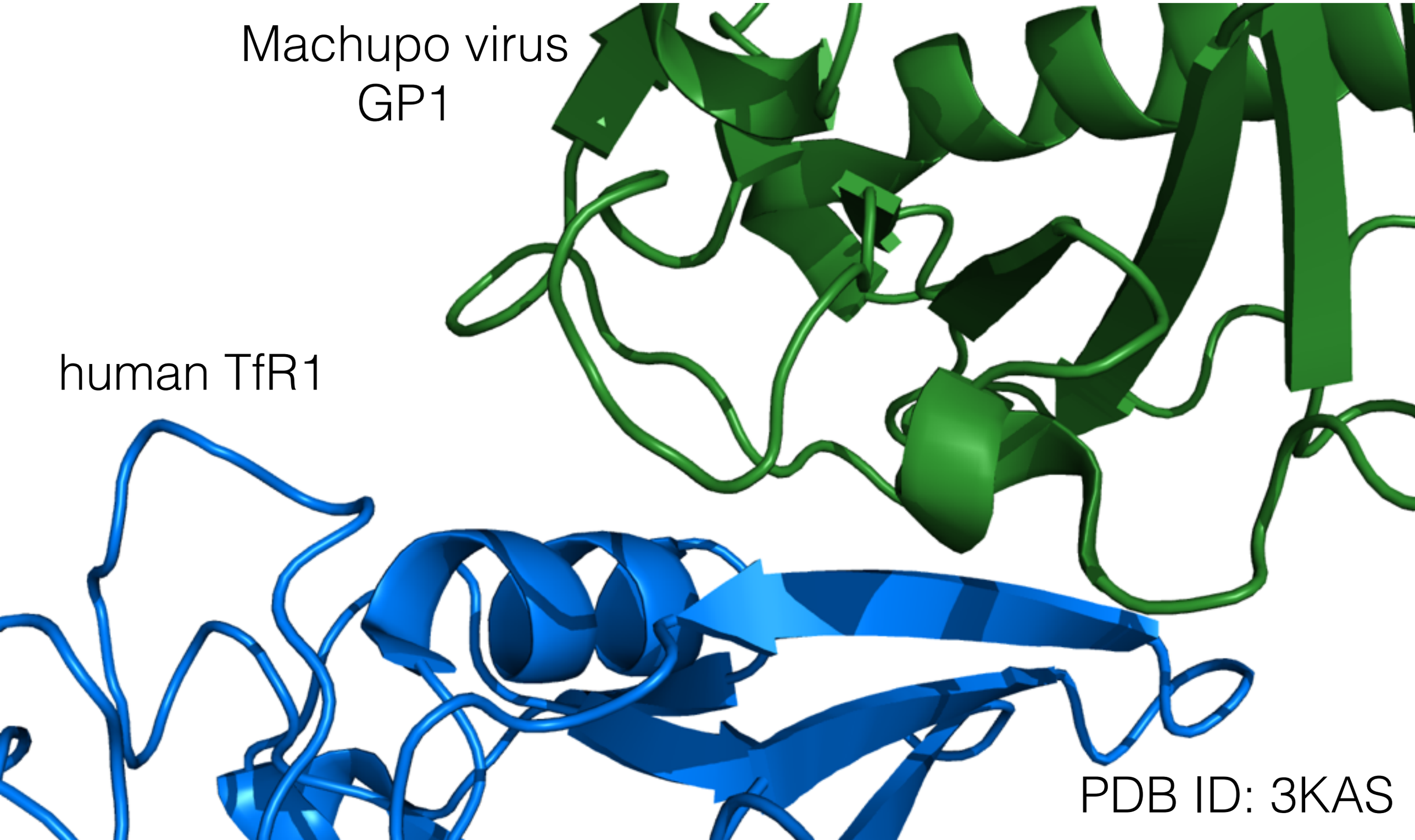
Chain B (green)
Machupo virus
GP1

PDB ID: 3KAS

PDB files can contain multiple chains

Machupo virus
GP1

human TfR1



PDB ID: 3KAS

Anatomy of a PDB file

```
HEADER    GROWTH FACTOR                                 15-APR-93   1BFG
TITLE     CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH FACTOR AT 1.6
TITLE    2 ANGSTROMS RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: BASIC FIBROBLAST GROWTH FACTOR;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606
KEYWDS    GROWTH FACTOR
EXPDTA    X-RAY DIFFRACTION
AUTHOR    Y.KITAGAWA,H.AGO,Y.KATSUBE,A.FUJISHIMA,Y.MATSUURA
REVDAT   3   24-FEB-09 1BFG    1       VERSN
REVDAT   2   01-APR-03 1BFG    1       JRNL
REVDAT   1   31-JAN-94 1BFG    0
JRNL        AUTH   H.AGO,Y.KITAGAWA,A.FUJISHIMA,Y.MATSUURA,Y.KATSUBE
JRNL        TITL   CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH
JRNL        TITL 2 FACTOR AT 1.6 A RESOLUTION.
JRNL        REF    J.BIOCHEM.(TOKYO)                 V. 110   360 1991
JRNL        REFN                   ISSN 0021-924X
JRNL        PMID   1769963
...
```

Anatomy of a PDB file

```
HEADER     GROWTH FACTOR                                15-APR-93   1BFG
TITLE      CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH FACTOR AT 1.6
TITLE     2 ANGSTROMS RESOLUTION
COMPND     MOL_ID: 1;
COMPND    2 MOLECULE: BASIC FIBROBLAST GROWTH FACTOR;
COMPND    3 CHAIN: A;
COMPND    4 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE    3 ORGANISM_COMMON: HUMAN;
SOURCE    4 ORGANISM_TAXID: 9606
KEYWDS     GROWTH FACTOR
EXPDTA     X-RAY DIFFRACTION
AUTHOR     Y.KITAGAWA,H.AGO,Y.KATSUBE,A.FUJISHIMA,Y.MATSUURA
REVDAT    3   24-FEB-09 1BFG    1       VERSN
REVDAT    2   01-APR-03 1BFG    1       JRNL
REVDAT    1   31-JAN-94 1BFG    0
JRNL        AUTH   H.AGO,Y.KITAGAWA,A.FUJISHIMA,Y.MATSUURA,Y.KATSUBE
JRNL        TITL   CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH
JRNL        TITL 2 FACTOR AT 1.6 A RESOLUTION.
JRNL        REF    J.BIOCHEM.(TOKYO)                 V. 110   360 1991
JRNL        REFN                   ISSN 0021-924X
JRNL        PMID   1769963
...
```

Anatomy of a PDB file

```
HEADER    GROWTH FACTOR                                 15-APR-93   1BFG
TITLE     CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH FACTOR AT 1.6
TITLE    2 ANGSTROMS RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: BASIC FIBROBLAST GROWTH FACTOR;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606
KEYWDS    GROWTH FACTOR
EXPDTA    X-RAY DIFFRACTION
AUTHOR    Y.KITAGAWA,H.AGO,Y.KATSUBE,A.FUJISHIMA,Y.MATSUURA
REVDAT   3   24-FEB-09 1BFG    1       VERSN
REVDAT   2   01-APR-03 1BFG    1       JRNL
REVDAT   1   31-JAN-94 1BFG    0
JRNL        AUTH   H.AGO,Y.KITAGAWA,A.FUJISHIMA,Y.MATSUURA,Y.KATSUBE
JRNL        TITL   CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH
JRNL        TITL 2 FACTOR AT 1.6 A RESOLUTION.
JRNL        REF    J.BIOCHEM.(TOKYO)                 V. 110   360 1991
JRNL        REFN                   ISSN 0021-924X
JRNL        PMID   1769963
...
```

Anatomy of a PDB file

```
HEADER    GROWTH FACTOR                                 15-APR-93   1BFG
TITLE     CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH FACTOR AT 1.6
TITLE    2 ANGSTROMS RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: BASIC FIBROBLAST GROWTH FACTOR;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE   3 ORGANISM_COMMON: HUMAN;
SOURCE   4 ORGANISM_TAXID: 9606
KEYWDS    GROWTH FACTOR
EXPDTA    X-RAY DIFFRACTION
AUTHOR    Y.KITAGAWA,H.AGO,Y.KATSUBE,A.FUJISHIMA,Y.MATSUURA
REVDAT   3   24-FEB-09 1BFG    1       VERSN
REVDAT   2   01-APR-03 1BFG    1       JRNL
REVDAT   1   31-JAN-94 1BFG    0
JRNL        AUTH   H.AGO,Y.KITAGAWA,A.FUJISHIMA,Y.MATSUURA,Y.KATSUBE
JRNL        TITL   CRYSTAL STRUCTURE OF BASIC FIBROBLAST GROWTH
JRNL        TITL 2 FACTOR AT 1.6 A RESOLUTION.
JRNL        REF    J.BIOCHEM.(TOKYO)                 V. 110   360 1991
JRNL        REFN                   ISSN 0021-924X
JRNL        PMID   1769963
...
```

Anatomy of a PDB file

```
                                                ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                ...
```

Anatomy of a PDB file

atom number

```
                                                ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                ...
```

Anatomy of a PDB file

atom name

```
                                                ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                ...
```

Anatomy of a PDB file

residue name

```
                                                ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                ...
```

Anatomy of a PDB file

chain

```
                                                  ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                  ...
```

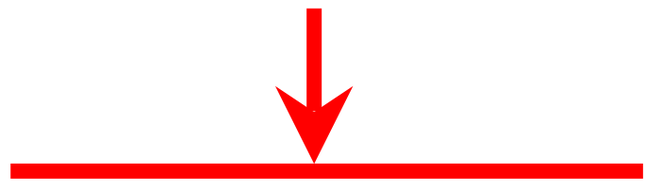
Anatomy of a PDB file

residue number

```
                                                  ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                  ...
```

Anatomy of a PDB file

x, *y*, *z* coordinates



```
...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                  ...
```

Anatomy of a PDB file

occupancy

```
                                                                  ...
ATOM      1  N   ASP A  19       6.864  13.397  -2.220  1.00 40.02           N
ATOM      2  CA  ASP A  19       6.806  14.455  -1.186  1.00 38.62           C
ATOM      3  C   ASP A  19       7.688  14.002  -0.016  1.00 36.01           C
ATOM      4  O   ASP A  19       8.837  14.477   0.052  1.00 35.96           O
ATOM      5  CB  ASP A  19       5.372  14.834  -0.851  1.00 43.28           C
ATOM      6  CG  ASP A  19       5.264  16.239  -0.268  1.00 46.36           C
ATOM      7  OD1 ASP A  19       5.447  16.322   0.978  1.00 48.00           O
ATOM      8  OD2 ASP A  19       5.015  17.239  -0.970  1.00 48.45           O
ATOM      9  N   PRO A  20       7.165  13.125   0.822  1.00 33.63           N
ATOM     10  CA  PRO A  20       7.888  12.572   1.982  1.00 30.69           C
ATOM     11  C   PRO A  20       9.049  11.697   1.528  1.00 27.78           C
                                                                  ...
```

Anatomy of a PDB file

B factor

```
                                        ...
ATOM       1  N    ASP A  19        6.864  13.397  -2.220  1.00 40.02          N
ATOM       2  CA   ASP A  19        6.806  14.455  -1.186  1.00 38.62          C
ATOM       3  C    ASP A  19        7.688  14.002  -0.016  1.00 36.01          C
ATOM       4  O    ASP A  19        8.837  14.477   0.052  1.00 35.96          O
ATOM       5  CB   ASP A  19        5.372  14.834  -0.851  1.00 43.28          C
ATOM       6  CG   ASP A  19        5.264  16.239  -0.268  1.00 46.36          C
ATOM       7  OD1  ASP A  19        5.447  16.322   0.978  1.00 48.00          O
ATOM       8  OD2  ASP A  19        5.015  17.239  -0.970  1.00 48.45          O
ATOM       9  N    PRO A  20        7.165  13.125   0.822  1.00 33.63          N
ATOM      10  CA   PRO A  20        7.888  12.572   1.982  1.00 30.69          C
ATOM      11  C    PRO A  20        9.049  11.697   1.528  1.00 27.78          C
                                        ...
```

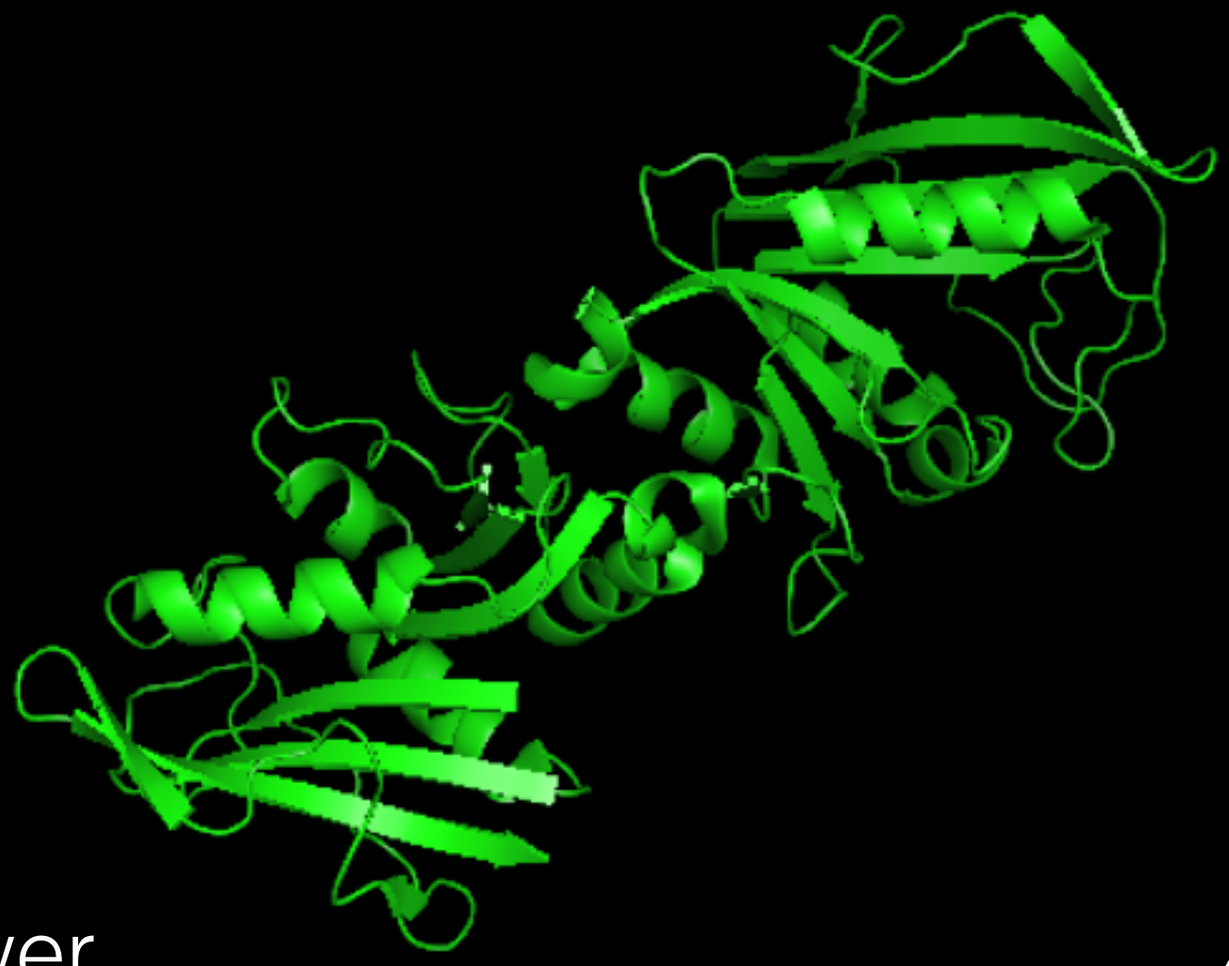
MacPyMOL

COMPND MOL_ID: 1;
COMPND 2 MOLECULE: DIHYDROFOLATE REDUCTASE;
COMPND 3 CHAIN: A, B;
COMPND 4 EC: 1.5.1.3;
COMPND 5 ENGINEERED: YES
ObjectMolecule: Read secondary structure assignments.
ObjectMolecule: Read crystal symmetry information.
Symmetry: Found 1 symmetry operators.
CmdLoad: "./1dhf.pdb" loaded as "1DHF".

command line/ console

Reset | Zoom | Orient | Draw | Ray
Unpick | Deselect | Rock | Get View
|< | < | Stop | Play | > | >| | MClear

PyMOL>



viewer

all A S H L C
1DHF 1/1 A S H L C

object control panel

mouse controls

Mouse Mode 3-Button Viewing
Buttons L M R Wheel
& Keys Rota Move MovZ Slab
Shft +Box -Box Clip MovS
Ctrl +/- PkAt Pk1 MvSZ
CtSh Sele Orig Clip MovZ
SnglClk +/- Cent Menu
DblClk Menu - PkAt
Selecting Residues
State 1/ 1

PyMOL>_

Object Control Panel

A (Actions): Rename, duplicate, remove, apply presets (like "ball-and-stick" or "publication"), perform computations

S (Show): Change the way things appear, e.g. change to stick or cartoon view.

H (Hide): Things that are shown using **S** accumulate, and don't automatically replace the last view. **H** is the opposite of **S** and hides unwanted representations.

L (Label): Label atoms, residues, etc.

C (Color): Change the color of atoms and groups.

PyMOL exercises

1. Download & open structure 3KAS:
   `fetch 3KAS`

   - Display in various forms (cartoon, stick, spheres, …)
   - Color different chains
   - Zoom in to display protein-protein interface

2. Download & open structure 1DLW

   - Display as cartoon
   - Show heme as sticks
   - Make a ray-traced image using the “Ray” button

Anything you can do with a menu, you can do with a command

`fetch 1DHF` - download a PDB file

`select chain A` - selects chain A

`select 1DHF` - selects whole structure

(http://pymol.sourceforge.net/newman/user/S0220commands.html)

`hide everything` - hides everything

`show ribbon` - shows current selection as ribbon

`show_as ribbon, chain A` - shows chain A as ribbon

Every command has a python counterpart

PyMOL: `select chain A`

Python: `cmd.select("chain A")`

Need help? (PyMOL command line) `help select`

Putting it all together:
Running python scripts in PyMOL

my_script.py:

```
cmd.fetch(“12AS”)
```

PyMOL command line:

```
run my_script.py
```