M2BP (1by2) and Hepsin (1p57 – Chain A)

From practical session 2
Structural Bioinformatics
Msc Bioinformatics
Birkbeck College

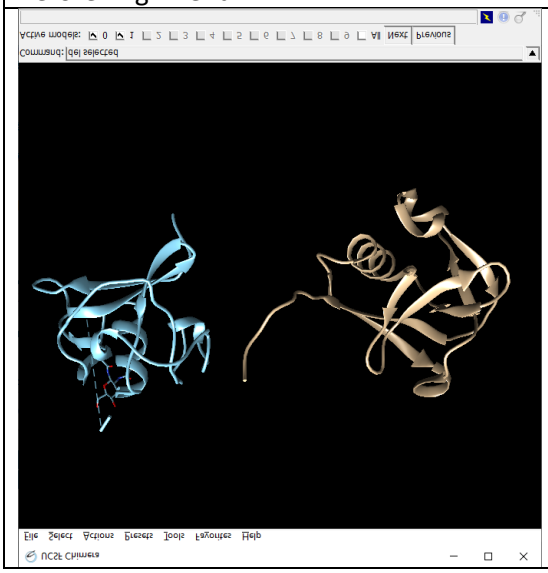
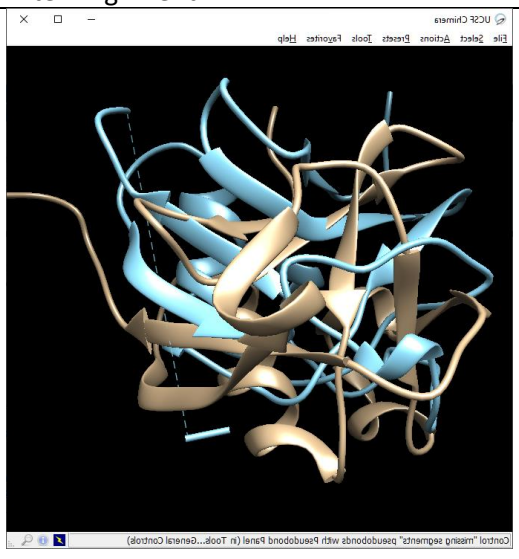
Version still not fully tested and working.

1) Alignment Report – checking only 1-10 conformations in all dimensions:

```
Value calculated = Optimised report
Initial Value=228127
Optimising for 1:1:1:1
Tripod 1=A=(21.492,12.511,2.01) B=(18.043,38.829,25.066) C=(31.099,20.559,15.54)
Tripod 2=A=(-29.537,-1.125,17.562) B=(10.417,12.67,-1.893) C=(-23.269,12.988,-8.071)
Transformation 1=
Translate:(-21.492,-12.511,-2.01)
Rotate XY:97.4661
Rotate YZ:90
Rotate XZ:-6.27744e+66
Transformation 2=
Translate:(29.537,1.125,-17.562)
Rotate XY:19.0484
Rotate YZ:234.661
Rotate XZ:-6.27744e+66
RMSD Value = 20751.8

Best RMSD Chosen as :
Optimising for 5:1:7:1
Tripod 1=A=(18.267,13.326,3.867) B=(18.043,38.829,25.066) C=(31.099,20.559,15.54)
Tripod 2=A=(-28.242,-2.366,12.448) B=(10.417,12.67,-1.893) C=(-23.269,12.988,-8.071)
Transformation 1=
Translate:(-18.267,-13.326,-3.867)
Rotate XY:90.5032
Rotate YZ:90
Rotate XZ:-6.27744e+66
Transformation 2=
Translate:(28.242,2.366,-12.448)
Rotate XY:21.253
Rotate YZ:223.645
Rotate XZ:-6.27744e+66
RMSD Value = 17248.5
Optimised report: RMSD Value=17248.5
```

Chimera Pictures

Before Alignment	After Alignment
	

2) An improvement in speed means more conformation checks – 160,000 conformations

This took 23 minutes ☹

And still didn't get the secondary structures completely on top of each other – but very close and just a translation away.

Log file checking 1-20 conformations in all directions

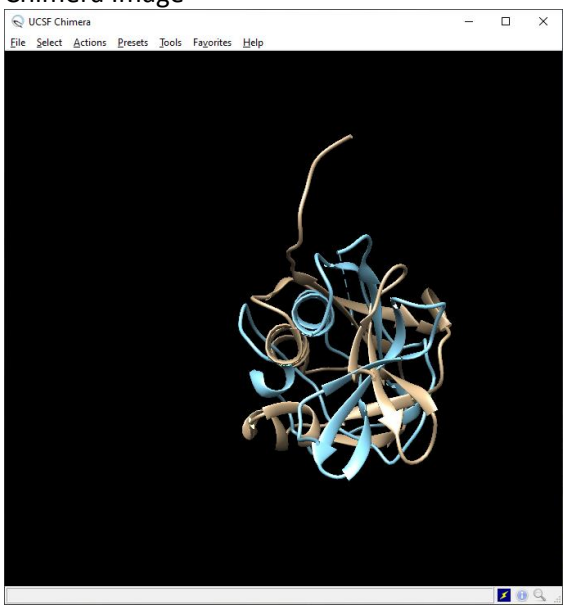
```
Best RMSD Chosen as :
Optimising for 18:1:13:1
Tripod 1=A=(20.355,10.411,12.419) B=(18.043,38.829,25.066) C=(31.099,20.559,15.54)
Tripod 2=A=(-26.799,-6.473,6.082) B=(10.417,12.67,-1.893) C=(-25.98,10.365,-8.33)
Transformation 1=
Translate:(-20.355,-10.411,-12.419)
Rotate XY:94.6512
Rotate YZ:90
Rotate XZ:33.342
Transformation 2=
Translate:(26.799,6.473,-6.082)
Rotate XY:27.2202
Rotate YZ:202.617
Rotate XZ:3.29139e-302
RMSD Value = 15906.5
Optimised report: RMSD Value=15906.5
```

3) **Given the above were both 1 in the orthogonal dimension, trying now 500 on the main axis and only 2 on the perpendicular axis. 1,000,000 conformation checks at 500 per second.**

Results file

```
Best RMSD Chosen as :
Optimising for 212:1:57:1
Tripod 1=A=(16.37,14.405,11.252) B=(15.358,36.438,26.233) C=(29.564,27.229,24.969)
Tripod 2=A=(-24.217,-5.115,8.614) B=(9.835,12.293,1.864) C=(-25.98,10.365,-8.33)
Transformation 1=
Translate:(-16.37,-14.405,-11.252)
Rotate XY:92.6298
Rotate YZ:90
Rotate XZ:159.165
Transformation 2=
Translate:(24.217,5.115,-8.614)
Rotate XY:27.0769
Rotate YZ:201.194
Rotate XZ:1.11264e-302
RMSD Value = 15510.9
Optimised report: RMSD Value=15510.9
```

Chimera Image



Conclusion

This is without sequence alignment, purely on CAlpha match to minimum. No amount of searching, probably, will line up the secondary structures. I need to know that the secondary structures are matched. I can do this with a good sequence alignment (which I have not implemented yet though I have the code for it in some other project). But, I would not be happy if for example, there was a reflection in structures so they should map functionally but never will geometrically, or if say the sequence was reversed but there were functional areas that matched up.

Next steps:

1) Pass in a sequence alignment in FASTA format (TCoffee) for these proteins – see first how good it is at getting the secondary structures matched.
2) Imply an alignment from the structure (including reflection) enabling potentially non sequential alignment (???)