

AlphaFold Analysis

Amanda Wilpitz | A17463962

Here we analyze our AlphaFold structure prediction models. The input directory/folder comes from the ColabFold server:

```
#Change this for YOUR results dir name
results_dir <- ""
```

```
#File names for all PDB models
pdb_files <- list.files(path = "/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer",
                        pattern = ".pdb",
                        full.names = TRUE)
```

```
#Print our PDB file names
basename(pdb_files)
```

```
[1] "hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000.pdb"
[2] "hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000.pdb"
[3] "hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000.pdb"
[4] "hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000.pdb"
[5] "hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000.pdb"
```

I will use the Bio3D package for analysis

```
library(bio3d)
```

Align and superpose

```
pdbbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

Reading PDB files:

```
/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_monomer_23119/hivpr_
/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_monomer_23119/hivpr_
/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_monomer_23119/hivpr_
/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_monomer_23119/hivpr_
/Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_monomer_23119/hivpr_
.....
```

Extracting sequences

```
pdb/seq: 1   name: /Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_r
pdb/seq: 2   name: /Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_r
pdb/seq: 3   name: /Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_r
pdb/seq: 4   name: /Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_r
pdb/seq: 5   name: /Users/awzme/OneDrive/Desktop/bimm 143/class11/hivpr_monomer_23119/hivpr_r
```

pdbs

```

1                                     .                               .                               50
[Truncated_Name:1]hivpr_mono  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:2]hivpr_mono  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:3]hivpr_mono  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:4]hivpr_mono  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
[Truncated_Name:5]hivpr_mono  PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGI
*****
1                                     .                               .                               50

51                                     .                               .                               100
[Truncated_Name:1]hivpr_mono  GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:2]hivpr_mono  GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:3]hivpr_mono  GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:4]hivpr_mono  GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
[Truncated_Name:5]hivpr_mono  GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFP
*****
51                                     .                               .                               100

101                                    .                               .                               150
[Truncated_Name:1]hivpr_mono  QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:2]hivpr_mono  QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:3]hivpr_mono  QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:4]hivpr_mono  QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
[Truncated_Name:5]hivpr_mono  QITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIG
```

```

*****
101      .      .      .      .      150

151      .      .      .      .      198
[Truncated_Name:1]hivpr_mono  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]hivpr_mono  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]hivpr_mono  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:4]hivpr_mono  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]hivpr_mono  GFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
*****
151      .      .      .      .      198

```

Call:

```
pdbaln(files = pdb_files, fit = TRUE, exefile = "msa")
```

Class:

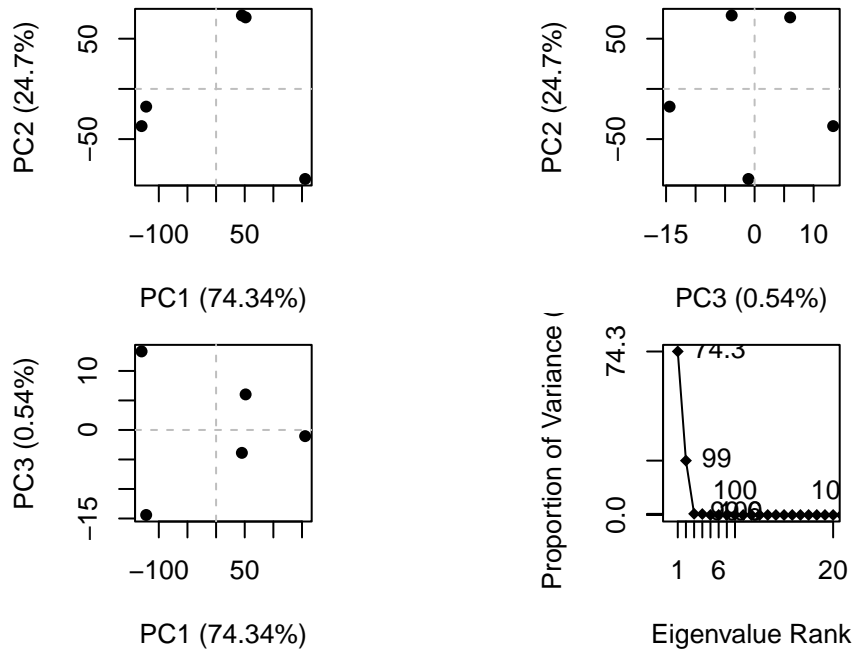
```
pdbs, fasta
```

Alignment dimensions:

```
5 sequence rows; 198 position columns (198 non-gap, 0 gap)
```

```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

```
pc <- pca(pdb)
plot(pc)
```



RMSD Analysis

RMSD is a common measure of structural distance used in structural biology

```
rd <- rmsd(pdbbs, fit = T)
```

Warning in rmsd(pdbbs, fit = T): No indices provided, using the 198 non NA positions

```
rd
```

```

hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000
hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000
hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000

```

```

hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000
hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000
hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000
hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000
hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000
hivpr_monomer_23119_unrelaxed_rank_001_alphafold2_multimer_v3_model_1_seed_000
hivpr_monomer_23119_unrelaxed_rank_002_alphafold2_multimer_v3_model_5_seed_000
hivpr_monomer_23119_unrelaxed_rank_003_alphafold2_multimer_v3_model_4_seed_000
hivpr_monomer_23119_unrelaxed_rank_004_alphafold2_multimer_v3_model_3_seed_000
hivpr_monomer_23119_unrelaxed_rank_005_alphafold2_multimer_v3_model_2_seed_000

```

```
range(rd)
```

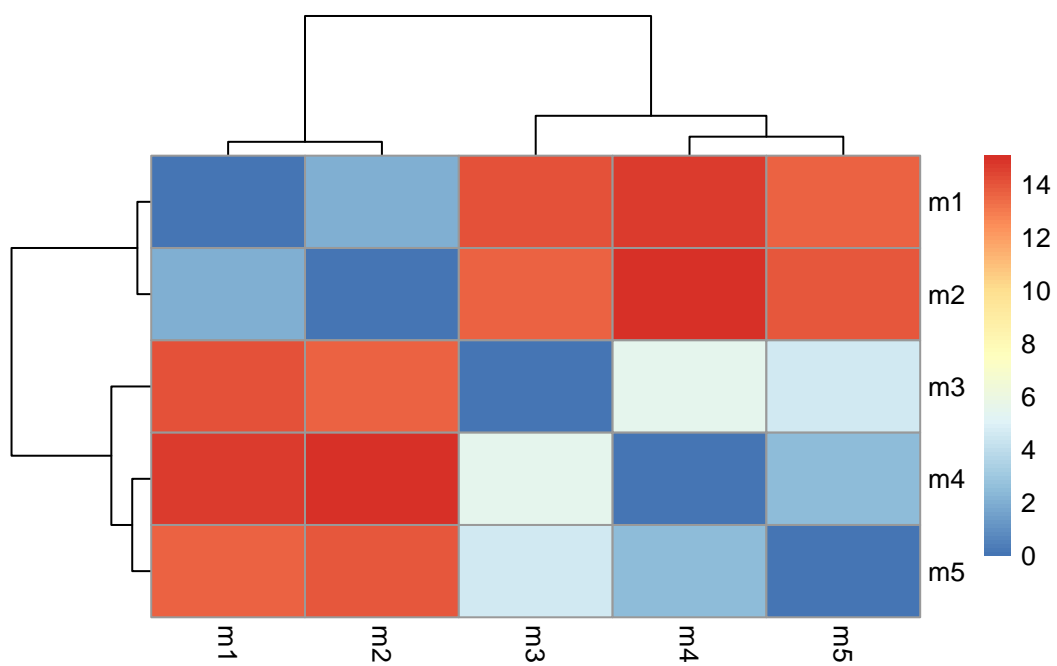
```
[1] 0.000 15.072
```

```

library(pheatmap)

colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)

```



Molstar superposed dimers (chain A superposed) structure:

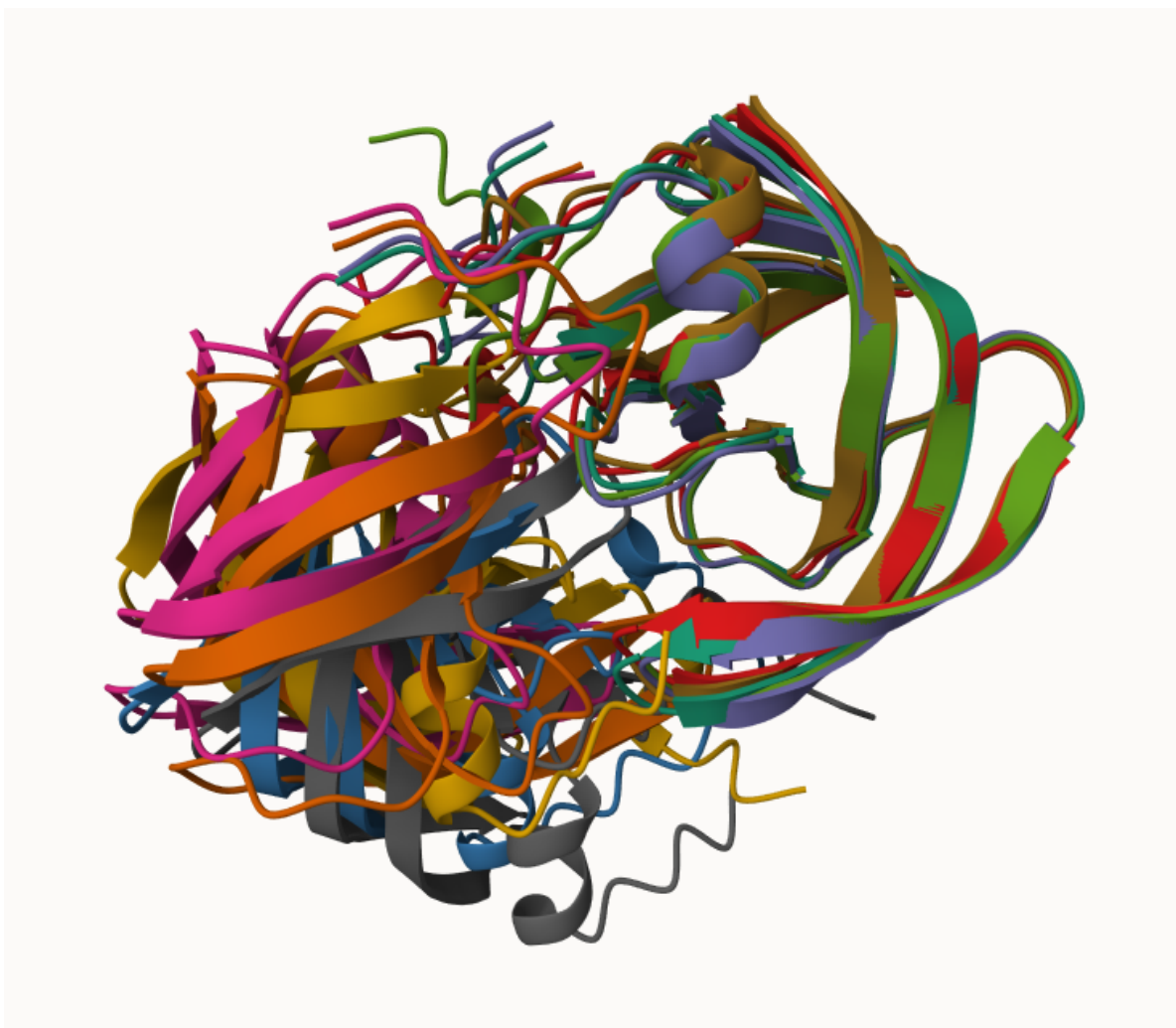
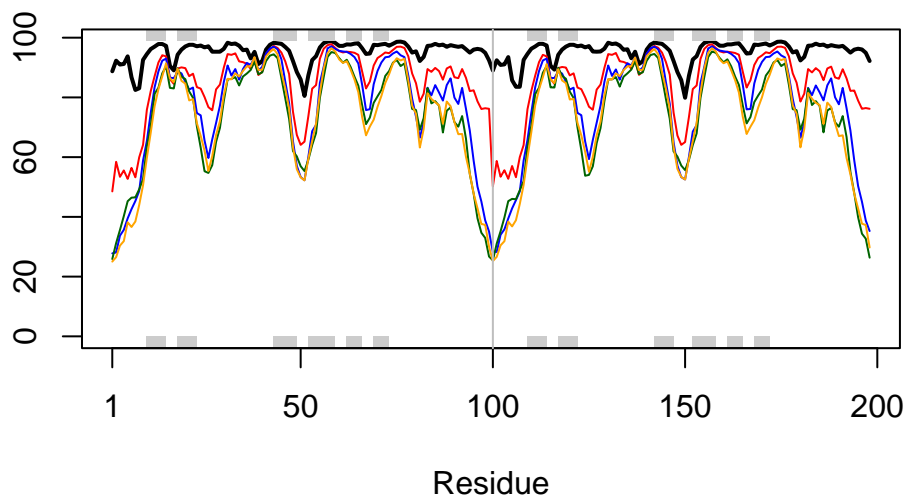


Figure 1: Superposed Chain A

```
pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

```
plotb3(pdb$b[1,], typ="l", lwd=2, sse=pdb)
points(pdb$b[2,], typ="l", col="red")
points(pdb$b[3,], typ="l", col="blue")
points(pdb$b[4,], typ="l", col="darkgreen")
points(pdb$b[5,], typ="l", col="orange")
abline(v=100, col="gray")
```



```
core <- core.find(pdb)
```

```
core size 197 of 198 vol = 6150.058
core size 196 of 198 vol = 5289.578
core size 195 of 198 vol = 5032.237
core size 194 of 198 vol = 4816.278
core size 193 of 198 vol = 4673.449
core size 192 of 198 vol = 4551.972
core size 191 of 198 vol = 4459.583
core size 190 of 198 vol = 4392.07
core size 189 of 198 vol = 4311.863
core size 188 of 198 vol = 4274.237
core size 187 of 198 vol = 4251.962
core size 186 of 198 vol = 4269.423
core size 185 of 198 vol = 4359.797
core size 184 of 198 vol = 4494.55
core size 183 of 198 vol = 4651.457
core size 182 of 198 vol = 4791.189
core size 181 of 198 vol = 4891.373
core size 180 of 198 vol = 4894.872
core size 179 of 198 vol = 4838.574
core size 178 of 198 vol = 4725.681
```


core size 177 of 198	vol = 4621.603
core size 176 of 198	vol = 4519.772
core size 175 of 198	vol = 4382.958
core size 174 of 198	vol = 4266.471
core size 173 of 198	vol = 4152.827
core size 172 of 198	vol = 4004.711
core size 171 of 198	vol = 3887.774
core size 170 of 198	vol = 3776.659
core size 169 of 198	vol = 3671.538
core size 168 of 198	vol = 3562.535
core size 167 of 198	vol = 3445.149
core size 166 of 198	vol = 3349.929
core size 165 of 198	vol = 3244.437
core size 164 of 198	vol = 3132.773
core size 163 of 198	vol = 3020.467
core size 162 of 198	vol = 2919.849
core size 161 of 198	vol = 2827.843
core size 160 of 198	vol = 2736.156
core size 159 of 198	vol = 2639.435
core size 158 of 198	vol = 2546.882
core size 157 of 198	vol = 2458.998
core size 156 of 198	vol = 2379.319
core size 155 of 198	vol = 2286.432
core size 154 of 198	vol = 2196.565
core size 153 of 198	vol = 2102.329
core size 152 of 198	vol = 2014.076
core size 151 of 198	vol = 1944.541
core size 150 of 198	vol = 1869.772
core size 149 of 198	vol = 1798.513
core size 148 of 198	vol = 1730.23
core size 147 of 198	vol = 1679.175
core size 146 of 198	vol = 1616.075
core size 145 of 198	vol = 1557.923
core size 144 of 198	vol = 1505.449
core size 143 of 198	vol = 1450.194
core size 142 of 198	vol = 1396.764
core size 141 of 198	vol = 1343.397
core size 140 of 198	vol = 1283.985
core size 139 of 198	vol = 1225.993
core size 138 of 198	vol = 1183.503
core size 137 of 198	vol = 1144.717
core size 136 of 198	vol = 1092.995
core size 135 of 198	vol = 1039.465

core size 134 of 198 vol = 996.98
core size 133 of 198 vol = 949.606
core size 132 of 198 vol = 897.06
core size 131 of 198 vol = 844.035
core size 130 of 198 vol = 797.708
core size 129 of 198 vol = 755.472
core size 128 of 198 vol = 709.173
core size 127 of 198 vol = 680.362
core size 126 of 198 vol = 646.137
core size 125 of 198 vol = 622.37
core size 124 of 198 vol = 596.398
core size 123 of 198 vol = 564.067
core size 122 of 198 vol = 532.656
core size 121 of 198 vol = 509.606
core size 120 of 198 vol = 488.504
core size 119 of 198 vol = 467.28
core size 118 of 198 vol = 446.176
core size 117 of 198 vol = 419.563
core size 116 of 198 vol = 387.708
core size 115 of 198 vol = 356.775
core size 114 of 198 vol = 333.148
core size 113 of 198 vol = 310.579
core size 112 of 198 vol = 282.042
core size 111 of 198 vol = 259.287
core size 110 of 198 vol = 242.539
core size 109 of 198 vol = 223.145
core size 108 of 198 vol = 206.021
core size 107 of 198 vol = 186.686
core size 106 of 198 vol = 171.634
core size 105 of 198 vol = 154.665
core size 104 of 198 vol = 139.73
core size 103 of 198 vol = 122.851
core size 102 of 198 vol = 111.9
core size 101 of 198 vol = 97.221
core size 100 of 198 vol = 89.554
core size 99 of 198 vol = 83.468
core size 98 of 198 vol = 74.154
core size 97 of 198 vol = 63.707
core size 96 of 198 vol = 54.758
core size 95 of 198 vol = 50.385
core size 94 of 198 vol = 40.115
core size 93 of 198 vol = 26.004
core size 92 of 198 vol = 17.202

```

core size 91 of 198 vol = 8.936
core size 90 of 198 vol = 5.492
core size 89 of 198 vol = 3.714
core size 88 of 198 vol = 2.894
core size 87 of 198 vol = 2.285
core size 86 of 198 vol = 1.886
core size 85 of 198 vol = 1.443
core size 84 of 198 vol = 1.179
core size 83 of 198 vol = 0.91
core size 82 of 198 vol = 0.773
core size 81 of 198 vol = 0.627
core size 80 of 198 vol = 0.526
core size 79 of 198 vol = 0.43
FINISHED: Min vol ( 0.5 ) reached

```

```
core.inds <- print(core, vol=0.51)
```

```

# 80 positions (cumulative volume <= 0.51 Angstrom^3)
  start end length
1    10  25     16
2    27  48     22
3    53  94     42

```

```
xyz <- pdbfit(pdb, core.inds, outpath = "corefit_structures")
```

```
rf <- rmsf(xyz)
```

```

plotb3(rf, sse=pdb)
abline(v=100, col = "grey", ylab = "RMSF")

```

