AlphaFold Analysis

Emily Ignatoff (A16732102)

AlphaFold Analysis of find-a-gene sequence:

Here we analyze our AlphaFold structure prediction models for our find-a-gene sequence:

Note: To my understanding, AlphaFold only uses one reading frame for estimating protein structure. I therefore looked for which of my reading frames had the start codon (M) since reading frame 1 did not. (I used reading frame 2)

```
results_dir <- "hsp70_sardinops_b602d/"
```

First, I will read in each generated pdb file from the results directory:

```
pdb_files <- list.files(path=results_dir,
                        pattern="*.pdb",
                        full.names = TRUE)

basename(pdb_files)
```

```
[1] "hsp70_sardinops_b602d_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_000.pdb"
[2] "hsp70_sardinops_b602d_unrelaxed_rank_002_alphafold2_ptm_model_2_seed_000.pdb"
[3] "hsp70_sardinops_b602d_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "hsp70_sardinops_b602d_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "hsp70_sardinops_b602d_unrelaxed_rank_005_alphafold2_ptm_model_5_seed_000.pdb"
```

Now, I will superimpose each pdb on top of the others:

```
library(bio3d)
```

```
Warning: package 'bio3d' was built under R version 4.3.3
```

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

```
Reading PDB files:
hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_001_alphafold2_ptm_model_4_seed_00
hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_002_alphafold2_ptm_model_2_seed_00
hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_00
hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_00
hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_005_alphafold2_ptm_model_5_seed_00
.....

Extracting sequences

pdb/seq: 1   name: hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_001_alphafold2_
pdb/seq: 2   name: hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_002_alphafold2_
pdb/seq: 3   name: hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_003_alphafold2_
pdb/seq: 4   name: hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_004_alphafold2_
pdb/seq: 5   name: hsp70_sardinops_b602d/hsp70_sardinops_b602d_unrelaxed_rank_005_alphafold2_
```

```r
pdbs
```

```
                              1        .         .         .         .         50
[Truncated_Name:1]hsp70_sard  TPQTQSFDAKRLIGRKFDDGVVQSDMKHWPFTVIDDSTRPKVQVDYKGET
[Truncated_Name:2]hsp70_sard  TPQTQSFDAKRLIGRKFDDGVVQSDMKHWPFTVIDDSTRPKVQVDYKGET
[Truncated_Name:3]hsp70_sard  TPQTQSFDAKRLIGRKFDDGVVQSDMKHWPFTVIDDSTRPKVQVDYKGET
[Truncated_Name:4]hsp70_sard  TPQTQSFDAKRLIGRKFDDGVVQSDMKHWPFTVIDDSTRPKVQVDYKGET
[Truncated_Name:5]hsp70_sard  TPQTQSFDAKRLIGRKFDDGVVQSDMKHWPFTVIDDSTRPKVQVDYKGET
                              **************************************************
                              1        .         .         .         .         50

                             51        .         .         .         .         100
[Truncated_Name:1]hsp70_sard  KSFYPEEISSMVLVKMKEIAEAYLGKTVSNAVVTVPAYFNDSQRQATKDA
[Truncated_Name:2]hsp70_sard  KSFYPEEISSMVLVKMKEIAEAYLGKTVSNAVVTVPAYFNDSQRQATKDA
[Truncated_Name:3]hsp70_sard  KSFYPEEISSMVLVKMKEIAEAYLGKTVSNAVVTVPAYFNDSQRQATKDA
[Truncated_Name:4]hsp70_sard  KSFYPEEISSMVLVKMKEIAEAYLGKTVSNAVVTVPAYFNDSQRQATKDA
[Truncated_Name:5]hsp70_sard  KSFYPEEISSMVLVKMKEIAEAYLGKTVSNAVVTVPAYFNDSQRQATKDA
                              **************************************************
                             51        .         .         .         .         100

                            101        .         .         .         .         150
[Truncated_Name:1]hsp70_sard  GTISGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILT
[Truncated_Name:2]hsp70_sard  GTISGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILT
```

```
[Truncated_Name:3]hsp70_sard   GTISGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILT
[Truncated_Name:4]hsp70_sard   GTISGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILT
[Truncated_Name:5]hsp70_sard   GTISGLNVLRIINEPTAAAIAYGLDKKVGAERNVLIFDLGGGTFDVSILT
                               **************************************************
                           101        .         .         .         .         150

                           151        .         .         .         .         200
[Truncated_Name:1]hsp70_sard   IEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKFKKDISDNKRAVR
[Truncated_Name:2]hsp70_sard   IEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKFKKDISDNKRAVR
[Truncated_Name:3]hsp70_sard   IEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKFKKDISDNKRAVR
[Truncated_Name:4]hsp70_sard   IEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKFKKDISDNKRAVR
[Truncated_Name:5]hsp70_sard   IEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKFKKDISDNKRAVR
                               **************************************************
                           151        .         .         .         .         200

                           201        .         .         .         .         250
[Truncated_Name:1]hsp70_sard   RLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFR
[Truncated_Name:2]hsp70_sard   RLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFR
[Truncated_Name:3]hsp70_sard   RLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFR
[Truncated_Name:4]hsp70_sard   RLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFR
[Truncated_Name:5]hsp70_sard   RLRTACERAKRTLSSSTQASIEIDSLYEGVDFYTSITRARFEELNADLFR
                               **************************************************
                           201        .         .         .         .         250

                           251        .         .         .         .         300
[Truncated_Name:1]hsp70_sard   GTLDPVEKSLRDAKMDKSAVHDIVLVGGSTRIPKIQKLLQDFFNGKELNK
[Truncated_Name:2]hsp70_sard   GTLDPVEKSLRDAKMDKSAVHDIVLVGGSTRIPKIQKLLQDFFNGKELNK
[Truncated_Name:3]hsp70_sard   GTLDPVEKSLRDAKMDKSAVHDIVLVGGSTRIPKIQKLLQDFFNGKELNK
[Truncated_Name:4]hsp70_sard   GTLDPVEKSLRDAKMDKSAVHDIVLVGGSTRIPKIQKLLQDFFNGKELNK
[Truncated_Name:5]hsp70_sard   GTLDPVEKSLRDAKMDKSAVHDIVLVGGSTRIPKIQKLLQDFFNGKELNK
                               **************************************************
                           251        .         .         .         .         300

                           301        .         .         .         .         350
[Truncated_Name:1]hsp70_sard   SINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLSLGIETAGGVMT
[Truncated_Name:2]hsp70_sard   SINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLSLGIETAGGVMT
[Truncated_Name:3]hsp70_sard   SINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLSLGIETAGGVMT
[Truncated_Name:4]hsp70_sard   SINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLSLGIETAGGVMT
[Truncated_Name:5]hsp70_sard   SINPDEAVAYGAAVQAAILSGDKSENVQDLLLLDVTPLSLGIETAGGVMT
                               **************************************************
                           301        .         .         .         .         350

                           351        .         .         .         .         400
```

```
[Truncated_Name:1]hsp70_sard   VLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFEL
[Truncated_Name:2]hsp70_sard   VLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFEL
[Truncated_Name:3]hsp70_sard   VLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFEL
[Truncated_Name:4]hsp70_sard   VLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFEL
[Truncated_Name:5]hsp70_sard   VLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFEL
                               **************************************************
                             351       .         .         .         .         400

                             401       .         .         .         .         450
[Truncated_Name:1]hsp70_sard   TGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLS
[Truncated_Name:2]hsp70_sard   TGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLS
[Truncated_Name:3]hsp70_sard   TGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLS
[Truncated_Name:4]hsp70_sard   TGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLS
[Truncated_Name:5]hsp70_sard   TGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLS
                               **************************************************
                             401       .         .         .         .         450

                             451       .         .         .         .         500
[Truncated_Name:1]hsp70_sard   KEDIERMVQEAEQYKAEDDVQRDKVSSKNALESYSFSMKATVEDEKLQGK
[Truncated_Name:2]hsp70_sard   KEDIERMVQEAEQYKAEDDVQRDKVSSKNALESYSFSMKATVEDEKLQGK
[Truncated_Name:3]hsp70_sard   KEDIERMVQEAEQYKAEDDVQRDKVSSKNALESYSFSMKATVEDEKLQGK
[Truncated_Name:4]hsp70_sard   KEDIERMVQEAEQYKAEDDVQRDKVSSKNALESYSFSMKATVEDEKLQGK
[Truncated_Name:5]hsp70_sard   KEDIERMVQEAEQYKAEDDVQRDKVSSKNALESYSFSMKATVEDEKLQGK
                               **************************************************
                             451       .         .         .         .         500

                             501       .         .         .         .        549
[Truncated_Name:1]hsp70_sard   ISDEDKAKILEKCNEVISWLDKNQTAEKEEYEHQQKELEKLCNPIITKL
[Truncated_Name:2]hsp70_sard   ISDEDKAKILEKCNEVISWLDKNQTAEKEEYEHQQKELEKLCNPIITKL
[Truncated_Name:3]hsp70_sard   ISDEDKAKILEKCNEVISWLDKNQTAEKEEYEHQQKELEKLCNPIITKL
[Truncated_Name:4]hsp70_sard   ISDEDKAKILEKCNEVISWLDKNQTAEKEEYEHQQKELEKLCNPIITKL
[Truncated_Name:5]hsp70_sard   ISDEDKAKILEKCNEVISWLDKNQTAEKEEYEHQQKELEKLCNPIITKL
                               **************************************************
                             501       .         .         .         .        549

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

Class:
  pdbs, fasta

Alignment dimensions:
  5 sequence rows; 549 position columns (549 non-gap, 0 gap)
```

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

Now, I will generate the structural difference between these coordinate sets using `rmsd()`:

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

```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 549 non NA positions
```

```
range(rd)
```

```
[1]  0.000 33.897
```

Heatmap of RMSD values:

```
library(pheatmap)
```

```
Warning: package 'pheatmap' was built under R version 4.3.3
```

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

I will be using a human HSP70 protein from the PDB database as my reference protein for the following steps:

```
pdb <- read.pdb("4po2")
```

```
  Note: Accessing on-line PDB file
   PDB has ALT records, taking A only, rm.alt=TRUE
```

Now, I will plot the pLDDT values of each model against each other:

```
plotb3(pdbs$b[1,], typ="l", lwd=2, sse=pdb)
```

```
Warning in plotb3(pdbs$b[1, ], typ = "l", lwd = 2, sse = pdb): Length of input
'sse' does not equal the length of input 'x'; Ignoring 'sse'
```

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

I will now find the the most "rigid" or consistent portion across models using the `core.find()` function:

```r
core <- core.find(pdbs)
```

```
core size 548 of 549  vol = 7109.547
core size 547 of 549  vol = 7088.297
core size 546 of 549  vol = 7085.065
core size 545 of 549  vol = 7120.348
core size 544 of 549  vol = 7180.386
core size 543 of 549  vol = 7236.389
core size 542 of 549  vol = 7290.173
core size 541 of 549  vol = 7347.306
core size 540 of 549  vol = 7402.5
core size 539 of 549  vol = 7476.46
core size 538 of 549  vol = 7544.681
core size 537 of 549  vol = 7566.823
core size 536 of 549  vol = 7601.172
core size 535 of 549  vol = 7617.761
core size 534 of 549  vol = 7588.411
core size 533 of 549  vol = 7557.263
core size 532 of 549  vol = 7506.871
core size 531 of 549  vol = 7443.589
core size 530 of 549  vol = 7351.131
core size 529 of 549  vol = 7274.492
core size 528 of 549  vol = 7155.502
core size 527 of 549  vol = 7024.994
core size 526 of 549  vol = 6885.366
core size 525 of 549  vol = 6752.533
core size 524 of 549  vol = 6594.844
core size 523 of 549  vol = 6436.804
core size 522 of 549  vol = 6259.679
core size 521 of 549  vol = 6093.047
core size 520 of 549  vol = 5932.441
core size 519 of 549  vol = 5768.099
core size 518 of 549  vol = 5580.357
core size 517 of 549  vol = 5399.174
core size 516 of 549  vol = 5199.004
core size 515 of 549  vol = 5030.25
core size 514 of 549  vol = 4852.857
core size 513 of 549  vol = 4697.384
core size 512 of 549  vol = 4536.706
core size 511 of 549  vol = 4363.227
core size 510 of 549  vol = 4198.556
core size 509 of 549  vol = 4031.465
```

```
core size 508 of 549  vol = 3872.506
core size 507 of 549  vol = 3716.383
core size 506 of 549  vol = 3567.801
core size 505 of 549  vol = 3403.628
core size 504 of 549  vol = 3255.569
core size 503 of 549  vol = 3119.522
core size 502 of 549  vol = 2975.903
core size 501 of 549  vol = 2844.798
core size 500 of 549  vol = 2718.69
core size 499 of 549  vol = 2594.277
core size 498 of 549  vol = 2472.983
core size 497 of 549  vol = 2355.744
core size 496 of 549  vol = 2231.516
core size 495 of 549  vol = 2124.295
core size 494 of 549  vol = 2025.638
core size 493 of 549  vol = 1918.124
core size 492 of 549  vol = 1816.685
core size 491 of 549  vol = 1715.384
core size 490 of 549  vol = 1610.873
core size 489 of 549  vol = 1514.188
core size 488 of 549  vol = 1422.637
core size 487 of 549  vol = 1327.722
core size 486 of 549  vol = 1232.572
core size 485 of 549  vol = 1149.631
core size 484 of 549  vol = 1078.993
core size 483 of 549  vol = 1002.957
core size 482 of 549  vol = 933.057
core size 481 of 549  vol = 865.281
core size 480 of 549  vol = 796.623
core size 479 of 549  vol = 727.854
core size 478 of 549  vol = 662.113
core size 477 of 549  vol = 599.622
core size 476 of 549  vol = 543.322
core size 475 of 549  vol = 501.317
core size 474 of 549  vol = 459.711
core size 473 of 549  vol = 421.175
core size 472 of 549  vol = 388.641
core size 471 of 549  vol = 357.07
core size 470 of 549  vol = 326.254
core size 469 of 549  vol = 304.955
core size 468 of 549  vol = 282.522
core size 467 of 549  vol = 259.896
core size 466 of 549  vol = 239.949
```

```
core size 465 of 549  vol = 218.181
core size 464 of 549  vol = 199.36
core size 463 of 549  vol = 183.502
core size 462 of 549  vol = 169.928
core size 461 of 549  vol = 155.682
core size 460 of 549  vol = 143.38
core size 459 of 549  vol = 128.872
core size 458 of 549  vol = 114.483
core size 457 of 549  vol = 100.528
core size 456 of 549  vol = 87.377
core size 455 of 549  vol = 75.125
core size 454 of 549  vol = 64.363
core size 453 of 549  vol = 53.88
core size 452 of 549  vol = 48.94
core size 451 of 549  vol = 44.077
core size 450 of 549  vol = 37.183
core size 449 of 549  vol = 32.138
core size 448 of 549  vol = 28.935
core size 447 of 549  vol = 25.528
core size 446 of 549  vol = 21.466
core size 445 of 549  vol = 17.982
core size 444 of 549  vol = 16.124
core size 443 of 549  vol = 14.606
core size 442 of 549  vol = 13.269
core size 441 of 549  vol = 12.847
core size 440 of 549  vol = 12.494
core size 439 of 549  vol = 12.221
core size 438 of 549  vol = 11.941
core size 437 of 549  vol = 11.69
core size 436 of 549  vol = 11.461
core size 435 of 549  vol = 11.244
core size 434 of 549  vol = 11.04
core size 433 of 549  vol = 10.836
core size 432 of 549  vol = 10.612
core size 431 of 549  vol = 10.437
core size 430 of 549  vol = 10.277
core size 429 of 549  vol = 10.098
core size 428 of 549  vol = 9.937
core size 427 of 549  vol = 9.779
core size 426 of 549  vol = 9.634
core size 425 of 549  vol = 9.469
core size 424 of 549  vol = 9.304
core size 423 of 549  vol = 9.194
```

```
core size 422 of 549  vol = 9.031
core size 421 of 549  vol = 8.873
core size 420 of 549  vol = 8.744
core size 419 of 549  vol = 8.621
core size 418 of 549  vol = 8.445
core size 417 of 549  vol = 8.302
core size 416 of 549  vol = 8.133
core size 415 of 549  vol = 7.996
core size 414 of 549  vol = 7.871
core size 413 of 549  vol = 7.697
core size 412 of 549  vol = 7.557
core size 411 of 549  vol = 7.412
core size 410 of 549  vol = 7.282
core size 409 of 549  vol = 7.16
core size 408 of 549  vol = 7.069
core size 407 of 549  vol = 6.948
core size 406 of 549  vol = 6.849
core size 405 of 549  vol = 6.761
core size 404 of 549  vol = 6.678
core size 403 of 549  vol = 6.578
core size 402 of 549  vol = 6.465
core size 401 of 549  vol = 6.363
core size 400 of 549  vol = 6.29
core size 399 of 549  vol = 6.201
core size 398 of 549  vol = 6.06
core size 397 of 549  vol = 5.948
core size 396 of 549  vol = 5.832
core size 395 of 549  vol = 5.711
core size 394 of 549  vol = 5.592
core size 393 of 549  vol = 5.479
core size 392 of 549  vol = 5.379
core size 391 of 549  vol = 5.278
core size 390 of 549  vol = 5.16
core size 389 of 549  vol = 5.025
core size 388 of 549  vol = 4.874
core size 387 of 549  vol = 4.78
core size 386 of 549  vol = 4.683
core size 385 of 549  vol = 4.592
core size 384 of 549  vol = 4.473
core size 383 of 549  vol = 4.38
core size 382 of 549  vol = 4.29
core size 381 of 549  vol = 4.156
core size 380 of 549  vol = 4.04
```

```
core size 379 of 549  vol = 3.916
core size 378 of 549  vol = 3.819
core size 377 of 549  vol = 3.725
core size 376 of 549  vol = 3.623
core size 375 of 549  vol = 3.531
core size 374 of 549  vol = 3.447
core size 373 of 549  vol = 3.359
core size 372 of 549  vol = 3.295
core size 371 of 549  vol = 3.209
core size 370 of 549  vol = 3.138
core size 369 of 549  vol = 3.067
core size 368 of 549  vol = 3.002
core size 367 of 549  vol = 2.917
core size 366 of 549  vol = 2.848
core size 365 of 549  vol = 2.777
core size 364 of 549  vol = 2.717
core size 363 of 549  vol = 2.648
core size 362 of 549  vol = 2.597
core size 361 of 549  vol = 2.548
core size 360 of 549  vol = 2.498
core size 359 of 549  vol = 2.447
core size 358 of 549  vol = 2.403
core size 357 of 549  vol = 2.358
core size 356 of 549  vol = 2.314
core size 355 of 549  vol = 2.262
core size 354 of 549  vol = 2.217
core size 353 of 549  vol = 2.174
core size 352 of 549  vol = 2.13
core size 351 of 549  vol = 2.081
core size 350 of 549  vol = 2.042
core size 349 of 549  vol = 2
core size 348 of 549  vol = 1.96
core size 347 of 549  vol = 1.924
core size 346 of 549  vol = 1.887
core size 345 of 549  vol = 1.848
core size 344 of 549  vol = 1.813
core size 343 of 549  vol = 1.778
core size 342 of 549  vol = 1.74
core size 341 of 549  vol = 1.705
core size 340 of 549  vol = 1.674
core size 339 of 549  vol = 1.638
core size 338 of 549  vol = 1.604
core size 337 of 549  vol = 1.572
```

```
core size 336 of 549  vol = 1.543
core size 335 of 549  vol = 1.517
core size 334 of 549  vol = 1.487
core size 333 of 549  vol = 1.46
core size 332 of 549  vol = 1.434
core size 331 of 549  vol = 1.411
core size 330 of 549  vol = 1.385
core size 329 of 549  vol = 1.362
core size 328 of 549  vol = 1.342
core size 327 of 549  vol = 1.322
core size 326 of 549  vol = 1.301
core size 325 of 549  vol = 1.277
core size 324 of 549  vol = 1.259
core size 323 of 549  vol = 1.24
core size 322 of 549  vol = 1.219
core size 321 of 549  vol = 1.2
core size 320 of 549  vol = 1.181
core size 319 of 549  vol = 1.159
core size 318 of 549  vol = 1.138
core size 317 of 549  vol = 1.119
core size 316 of 549  vol = 1.102
core size 315 of 549  vol = 1.084
core size 314 of 549  vol = 1.065
core size 313 of 549  vol = 1.045
core size 312 of 549  vol = 1.026
core size 311 of 549  vol = 1.007
core size 310 of 549  vol = 0.989
core size 309 of 549  vol = 0.973
core size 308 of 549  vol = 0.953
core size 307 of 549  vol = 0.935
core size 306 of 549  vol = 0.921
core size 305 of 549  vol = 0.906
core size 304 of 549  vol = 0.892
core size 303 of 549  vol = 0.88
core size 302 of 549  vol = 0.866
core size 301 of 549  vol = 0.85
core size 300 of 549  vol = 0.835
core size 299 of 549  vol = 0.82
core size 298 of 549  vol = 0.805
core size 297 of 549  vol = 0.792
core size 296 of 549  vol = 0.776
core size 295 of 549  vol = 0.764
core size 294 of 549  vol = 0.751
```

```
 core size 293 of 549  vol = 0.739
 core size 292 of 549  vol = 0.726
 core size 291 of 549  vol = 0.715
 core size 290 of 549  vol = 0.702
 core size 289 of 549  vol = 0.691
 core size 288 of 549  vol = 0.676
 core size 287 of 549  vol = 0.663
 core size 286 of 549  vol = 0.653
 core size 285 of 549  vol = 0.64
 core size 284 of 549  vol = 0.63
 core size 283 of 549  vol = 0.619
 core size 282 of 549  vol = 0.607
 core size 281 of 549  vol = 0.597
 core size 280 of 549  vol = 0.585
 core size 279 of 549  vol = 0.573
 core size 278 of 549  vol = 0.562
 core size 277 of 549  vol = 0.55
 core size 276 of 549  vol = 0.541
 core size 275 of 549  vol = 0.531
 core size 274 of 549  vol = 0.519
 core size 273 of 549  vol = 0.508
 core size 272 of 549  vol = 0.497
 FINISHED: Min vol ( 0.5 ) reached
```

```r
core.inds <- print(core, vol=0.5)
```

```
# 273 positions (cumulative volume <= 0.5 Angstrom^3)
   start end length
1     35  35      1
2     39  39      1
3     43  43      1
4     51  51      1
5     56  70     15
6     75  75      1
7     77  87     11
8     94  94      1
9     97  98      2
10   100 101      2
11   104 126     23
12   130 163     34
13   167 336    170
14   354 359      6
```

```
15   421 424      4
```

Writing the superimposed coordinates to a new file for use in Mol*

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

Figure 1: Uncertainty/disorder of strucutres in Mol*

I will now use RMSF to observe the conformational variance across the structure:

```
rf <- rmsf(xyz)

plotb3(rf, sse=pdb)
```

```
Warning in plotb3(rf, sse = pdb): Length of input 'sse' does not equal the
length of input 'x'; Ignoring 'sse'
```

```
abline(v=100, col="gray", ylab="RMSF")
```

Note: I am unsure why the sse function is not working using an existing HSP70 PDB file as my reference protein :(

Predicted Alignment Error for Domains

```
library(jsonlite)
```

```
Warning: package 'jsonlite' was built under R version 4.3.3
```

```
pae_files <- list.files(path=results_dir,
                        pattern=".*model.*\\.json",
                        full.names = TRUE)
```

```
pae1 <- read_json(pae_files[1],simplifyVector = TRUE)
pae2 <- read_json(pae_files[2],simplifyVector = TRUE)
pae3 <- read_json(pae_files[3],simplifyVector = TRUE)
pae4 <- read_json(pae_files[4],simplifyVector = TRUE)
pae5 <- read_json(pae_files[5],simplifyVector = TRUE)

attributes(pae1)
```

```
$names
[1] "plddt"   "max_pae" "pae"     "ptm"
```

```r
head(pae1$plddt)
```

```
[1] 54.72 58.44 64.50 71.56 77.12 84.31
```

Which model has the best (lowest) max PAE?

```r
pae1$max_pae
```

```
[1] 30.375
```

```r
pae2$max_pae
```

```
[1] 29.6875
```

```r
pae3$max_pae
```

```
[1] 30.46875
```

```r
pae4$max_pae
```

```
[1] 30.75
```

```r
pae5$max_pae
```

```
[1] 30.67188
```

PAE 2 is lowest in this case!

I will plot each of these PAE in the same range:

```r
plot.dmat(pae1$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
           grid.col = "black",
          zlim=c(0,30))
```

```
plot.dmat(pae2$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```

```
plot.dmat(pae3$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```

```
plot.dmat(pae4$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```

```
plot.dmat(pae5$pae,
          xlab="Residue Position (i)",
          ylab="Residue Position (j)",
          grid.col = "black",
          zlim=c(0,30))
```

Residue Conservation from Alignment File

Generate alignment file:

```
aln_file <- list.files(path=results_dir,
                       pattern=".a3m$",
                        full.names = TRUE)
aln_file
```

```
[1] "hsp70_sardinops_b602d/hsp70_sardinops_b602d.a3m"
```

```
aln <- read.fasta(aln_file[1], to.upper = TRUE)
```

```
[1] " ** Duplicated sequence id's: 101 **"
```

How many sequences in this alignment?

```
dim(aln$ali)
```

```
[1] 33473   994
```

I will score residue conservation using the `conserv()` function:

```
sim <- conserv(aln)
```

```
plotb3(sim[1:99], sse=trim.pdb(pdb, chain="A"),
       ylab="Conservation Score")
```

```
Warning in plotb3(sim[1:99], sse = trim.pdb(pdb, chain = "A"), ylab =
"Conservation Score"): Length of input 'sse' does not equal the length of input
'x'; Ignoring 'sse'
```

Let's look at the sequences which will stand out in a consensus sequence

```
con <- consensus(aln, cutoff = 0.7)
con$seq
```

```
  [1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
 [19] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
 [37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
 [55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
 [73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
```

```
 [91] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[127] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[145] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[163] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[181] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[199] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[217] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[235] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[253] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[271] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[289] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[307] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[325] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[343] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[361] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[379] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[397] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[415] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[433] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[451] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[469] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[487] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[505] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[523] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[541] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[559] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[577] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[595] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[613] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[631] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[649] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[667] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[685] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[703] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[721] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[739] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[757] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[775] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[793] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[811] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[829] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[847] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
```

```
[865] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[883] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[901] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[919] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[937] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[955] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[973] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
[991] "-" "-" "-" "-"
```

```
m1.pdb <- read.pdb(pdb_files[1])
occ <- vec2resno(rep(sim[1:99], length.out=549), m1.pdb$atom$resno)
write.pdb(m1.pdb, o=occ, file="m1_conserv.pdb")
```

There are not many colors present in the following representation by occupancy:

Figure 2: Color by Occupancy in Mol*