Class 9 Structural Bioinformatics (Pt.1)

Brittney Hayes

2024-02-14

## Introduction to the RCSB PDB

Download a CSV file from the PDB site

Data\_Export\_Summary <- read.csv("Data Export Summary.csv")  
des <- Data\_Export\_Summary

# Q1: What percentage of structures in the PDB are solved by X-Ray and Electron Microscopy.

# I kept getting error messages for this part so I manually calculated the sum for total structures, xray structures, and em structures. At first the sum function worked but I had problems when knitting.  
  
# Calculate the total number of structures  
total\_structures <- 215908  
  
# Calculate the number of structures solved by X-Ray and Electron Microscopy  
xray\_structures <- 182491  
em\_structures <- 18889  
  
# Calculate the percentage of structures solved by X-Ray and Electron Microscopy  
percentage\_xray <- (xray\_structures / total\_structures) \* 100  
percentage\_em <- (em\_structures / total\_structures) \* 100  
  
# Display percentages  
percentage\_xray

## [1] 84.52257

percentage\_em

## [1] 8.748634

# Add the two percentages  
sum(percentage\_xray,percentage\_em)

## [1] 93.27121

Answer: For X-ray 84.52%, and for electron microscopy 8.75%. Together this is 93.27%.

# Q2: What proportion of structures in the PDB are protein?

# Again I had issues with the sum function when knitting.  
  
# Calculate the amount of protein  
protein <- 211299  
  
#Display amount of protein  
protein

## [1] 211299

# Find proportion or percent of structures that are proteins  
percentage\_protein <- (protein/total\_structures)\*100  
  
#Display percentage of proteins  
percentage\_protein

## [1] 97.86529

I included protein only, protein/oligosaccharide, and protein/NA and found that 97.9% of the structures are proteins.

# Q3: Type HIV in the PDB website search box on the home page and determine how many HIV-1 protease structures are in the current PDB?

The search for HIV displays 4,412 structures.

## Visualizing the HIV-1 Protease Structure

# Q4: Water molecules normally have 3 atoms. Why do we see just one atom per water molecule in this structure?

This simplification is done to reduce visual clutter and make it easier to interpret the structure since water molecules are so abundant.

# Q5: There is a critical “conserved” water molecule in the binding site. Can you identify this water molecule? What residue number does this water molecule have

Yes, the residue number is 313.

# Q6: Generate and save a figure clearly showing the two distinct chains of HIV-protease along with the ligand. You might also consider showing the catalytic residues ASP 25 in each chain and the critical water (we recommend “Ball & Stick” for these side-chains). Add this figure to your Quarto document.

Discussion Topic: Can you think of a way in which indinavir, or even larger ligands and substrates, could enter the binding site?

Ligands and substrates can enter the binding site by inducing conformational changes.

## Introduction to Bio3D in R

# Load the Bio3D package  
library(bio3d)  
  
# Read PDB file  
pdb <- read.pdb("1hsg")

## Note: Accessing on-line PDB file

# Quick summary of the contents of the pdb   
pdb

##   
## Call: read.pdb(file = "1hsg")  
##   
## Total Models#: 1  
## Total Atoms#: 1686, XYZs#: 5058 Chains#: 2 (values: A B)  
##   
## Protein Atoms#: 1514 (residues/Calpha atoms#: 198)  
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)  
##   
## Non-protein/nucleic Atoms#: 172 (residues: 128)  
## Non-protein/nucleic resid values: [ HOH (127), MK1 (1) ]  
##   
## Protein sequence:  
## PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD  
## QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE  
## ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP  
## VNIIGRNLLTQIGCTLNF  
##   
## + attr: atom, xyz, seqres, helix, sheet,  
## calpha, remark, call

# Q7 How many amino acid residues are there in this pdb object?

There are 198 amino acid residues.

# Q8 Name one of the two non-protein residues?

HOH is one of the non-protein residues.

# Q9 How many protein chains are in this structure?

There are 2 protein chains.

# Find the attributes   
attributes(pdb)

## $names  
## [1] "atom" "xyz" "seqres" "helix" "sheet" "calpha" "remark" "call"   
##   
## $class  
## [1] "pdb" "sse"

# Acess the atom attribute   
head(pdb$atom)

## type eleno elety alt resid chain resno insert x y z o b  
## 1 ATOM 1 N <NA> PRO A 1 <NA> 29.361 39.686 5.862 1 38.10  
## 2 ATOM 2 CA <NA> PRO A 1 <NA> 30.307 38.663 5.319 1 40.62  
## 3 ATOM 3 C <NA> PRO A 1 <NA> 29.760 38.071 4.022 1 42.64  
## 4 ATOM 4 O <NA> PRO A 1 <NA> 28.600 38.302 3.676 1 43.40  
## 5 ATOM 5 CB <NA> PRO A 1 <NA> 30.508 37.541 6.342 1 37.87  
## 6 ATOM 6 CG <NA> PRO A 1 <NA> 29.296 37.591 7.162 1 38.40  
## segid elesy charge  
## 1 <NA> N <NA>  
## 2 <NA> C <NA>  
## 3 <NA> C <NA>  
## 4 <NA> O <NA>  
## 5 <NA> C <NA>  
## 6 <NA> C <NA>

# Read a new PDB structure of Adenylate Kinase and perform Normal mode analysis  
adk <- read.pdb("6s36")

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

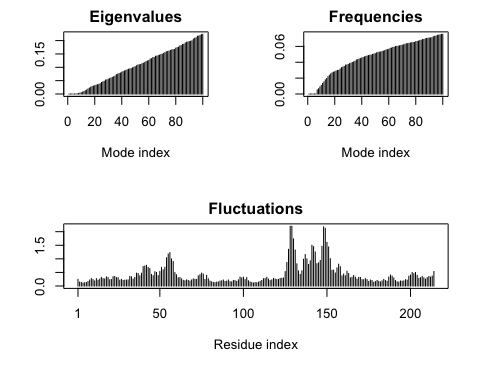
adk

##   
## Call: read.pdb(file = "6s36")  
##   
## Total Models#: 1  
## Total Atoms#: 1898, XYZs#: 5694 Chains#: 1 (values: A)  
##   
## Protein Atoms#: 1654 (residues/Calpha atoms#: 214)  
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)  
##   
## Non-protein/nucleic Atoms#: 244 (residues: 244)  
## Non-protein/nucleic resid values: [ CL (3), HOH (238), MG (2), NA (1) ]  
##   
## Protein sequence:  
## MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT  
## DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDKI  
## VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG  
## YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG  
##   
## + attr: atom, xyz, seqres, helix, sheet,  
## calpha, remark, call

# Perform flexiblity prediction  
m <- nma(adk)

## Building Hessian... Done in 0.017 seconds.  
## Diagonalizing Hessian... Done in 0.278 seconds.

plot(m)



# View a “movie” of these predicted motions/generate a molecular “trajectory” with the mktrj() function.  
mktrj(m, file="adk\_m7.pdb")

## Comparative structure analysis of Adenylate Kinase

# Install packages in the R console NOT your Rmd/Quarto file

# Q10. Which of the packages above is found only on BioConductor and not CRAN?

msa is found only on BioConductor.

# Q11. Which of the above packages is not found on BioConductor or CRAN?:

bio3d-view is not found on BioConductor or CRAN.

# Q12. True or False? Functions from the devtools package can be used to install packages from GitHub and BitBucket?

True.

## Search and retrieve ADK structures

# Search and retrieve ADK structures  
library(bio3d)  
aa <- get.seq("1ake\_A")

## Warning in get.seq("1ake\_A"): Removing existing file: seqs.fasta

## Fetching... Please wait. Done.

aa

## 1 . . . . . 60   
## pdb|1AKE|A MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT  
## 1 . . . . . 60   
##   
## 61 . . . . . 120   
## pdb|1AKE|A DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI  
## 61 . . . . . 120   
##   
## 121 . . . . . 180   
## pdb|1AKE|A VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG  
## 121 . . . . . 180   
##   
## 181 . . . 214   
## pdb|1AKE|A YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG  
## 181 . . . 214   
##   
## Call:  
## read.fasta(file = outfile)  
##   
## Class:  
## fasta  
##   
## Alignment dimensions:  
## 1 sequence rows; 214 position columns (214 non-gap, 0 gap)   
##   
## + attr: id, ali, call

# How many amino acids are in this sequence, i.e. how long is this sequence?

There are 214 amino acids in this sequence.

# Use this sequence as a query to BLAST search the PDB to find similar sequences and structures.  
  
# Blast or hmmer search   
#b <- blast.pdb(aa)

# Plot a summary of search results  
#hits <- plot(b)

# Visualize and filter blast results

# List out some 'top hits'  
#head(hits$pdb.id)

hits <- NULL  
hits$pdb.id <- c('1AKE\_A','6S36\_A','6RZE\_A','3HPR\_A','1E4V\_A','5EJE\_A','1E4Y\_A','3X2S\_A','6HAP\_A','6HAM\_A','4K46\_A','3GMT\_A','4PZL\_A')

# Download releated PDB files  
files <- get.pdb(hits$pdb.id, path="pdbs", split=TRUE, gzip=TRUE)

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/1AKE.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/6S36.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/6RZE.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/3HPR.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/1E4V.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/5EJE.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/1E4Y.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/3X2S.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/6HAP.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/6HAM.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/4K46.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/3GMT.pdb.gz exists. Skipping download

## Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):  
## pdbs/4PZL.pdb.gz exists. Skipping download

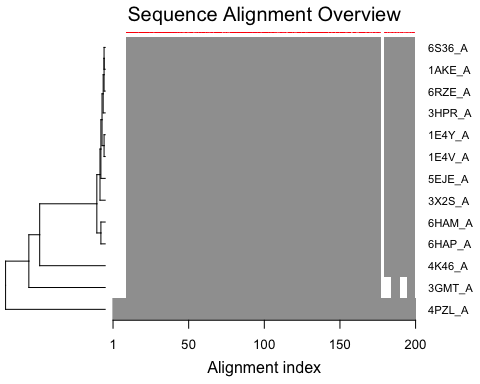
## | | | 0% | |===== | 8% | |=========== | 15% | |================ | 23% | |====================== | 31% | |=========================== | 38% | |================================ | 46% | |====================================== | 54% | |=========================================== | 62% | |================================================ | 69% | |====================================================== | 77% | |=========================================================== | 85% | |================================================================= | 92% | |======================================================================| 100%

## Align and superpose structures

# Align releated PDBs  
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")

## Reading PDB files:  
## pdbs/split\_chain/1AKE\_A.pdb  
## pdbs/split\_chain/6S36\_A.pdb  
## pdbs/split\_chain/6RZE\_A.pdb  
## pdbs/split\_chain/3HPR\_A.pdb  
## pdbs/split\_chain/1E4V\_A.pdb  
## pdbs/split\_chain/5EJE\_A.pdb  
## pdbs/split\_chain/1E4Y\_A.pdb  
## pdbs/split\_chain/3X2S\_A.pdb  
## pdbs/split\_chain/6HAP\_A.pdb  
## pdbs/split\_chain/6HAM\_A.pdb  
## pdbs/split\_chain/4K46\_A.pdb  
## pdbs/split\_chain/3GMT\_A.pdb  
## pdbs/split\_chain/4PZL\_A.pdb  
## PDB has ALT records, taking A only, rm.alt=TRUE  
## . PDB has ALT records, taking A only, rm.alt=TRUE  
## . PDB has ALT records, taking A only, rm.alt=TRUE  
## . PDB has ALT records, taking A only, rm.alt=TRUE  
## .. PDB has ALT records, taking A only, rm.alt=TRUE  
## .... PDB has ALT records, taking A only, rm.alt=TRUE  
## . PDB has ALT records, taking A only, rm.alt=TRUE  
## ...  
##   
## Extracting sequences  
##   
## pdb/seq: 1 name: pdbs/split\_chain/1AKE\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 2 name: pdbs/split\_chain/6S36\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 3 name: pdbs/split\_chain/6RZE\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 4 name: pdbs/split\_chain/3HPR\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 5 name: pdbs/split\_chain/1E4V\_A.pdb   
## pdb/seq: 6 name: pdbs/split\_chain/5EJE\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 7 name: pdbs/split\_chain/1E4Y\_A.pdb   
## pdb/seq: 8 name: pdbs/split\_chain/3X2S\_A.pdb   
## pdb/seq: 9 name: pdbs/split\_chain/6HAP\_A.pdb   
## pdb/seq: 10 name: pdbs/split\_chain/6HAM\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 11 name: pdbs/split\_chain/4K46\_A.pdb   
## PDB has ALT records, taking A only, rm.alt=TRUE  
## pdb/seq: 12 name: pdbs/split\_chain/3GMT\_A.pdb   
## pdb/seq: 13 name: pdbs/split\_chain/4PZL\_A.pdb

# Vector containing PDB codes for figure axis  
ids <- basename.pdb(pdbs$id)  
  
# Draw schematic alignment  
plot(pdbs, labels=ids)



## Annotate collected PDB structures

anno <- pdb.annotate(ids)  
unique(anno$source)

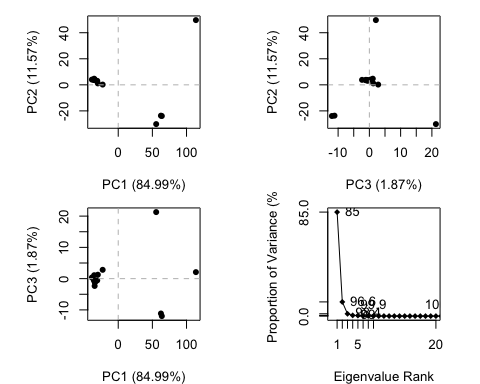
## [1] "Escherichia coli"   
## [2] "Escherichia coli K-12"   
## [3] "Escherichia coli O139:H28 str. E24377A"   
## [4] "Escherichia coli str. K-12 substr. MDS42"   
## [5] "Photobacterium profundum"   
## [6] "Burkholderia pseudomallei 1710b"   
## [7] "Francisella tularensis subsp. tularensis SCHU S4"

# View all available annotation data  
anno

## structureId chainId macromoleculeType chainLength experimentalTechnique  
## 1AKE\_A 1AKE A Protein 214 X-ray  
## 6S36\_A 6S36 A Protein 214 X-ray  
## 6RZE\_A 6RZE A Protein 214 X-ray  
## 3HPR\_A 3HPR A Protein 214 X-ray  
## 1E4V\_A 1E4V A Protein 214 X-ray  
## 5EJE\_A 5EJE A Protein 214 X-ray  
## 1E4Y\_A 1E4Y A Protein 214 X-ray  
## 3X2S\_A 3X2S A Protein 214 X-ray  
## 6HAP\_A 6HAP A Protein 214 X-ray  
## 6HAM\_A 6HAM A Protein 214 X-ray  
## 4K46\_A 4K46 A Protein 214 X-ray  
## 3GMT\_A 3GMT A Protein 230 X-ray  
## 4PZL\_A 4PZL A Protein 242 X-ray  
## resolution scopDomain pfam  
## 1AKE\_A 2.00 Adenylate kinase Adenylate kinase, active site lid (ADK\_lid)  
## 6S36\_A 1.60 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 6RZE\_A 1.69 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 3HPR\_A 2.00 <NA> Adenylate kinase (ADK)  
## 1E4V\_A 1.85 Adenylate kinase Adenylate kinase (ADK)  
## 5EJE\_A 1.90 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 1E4Y\_A 1.85 Adenylate kinase Adenylate kinase, active site lid (ADK\_lid)  
## 3X2S\_A 2.80 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 6HAP\_A 2.70 <NA> Adenylate kinase (ADK)  
## 6HAM\_A 2.55 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 4K46\_A 2.01 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 3GMT\_A 2.10 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## 4PZL\_A 2.10 <NA> Adenylate kinase, active site lid (ADK\_lid)  
## ligandId  
## 1AKE\_A AP5  
## 6S36\_A CL (3),NA,MG (2)  
## 6RZE\_A NA (3),CL (2)  
## 3HPR\_A AP5  
## 1E4V\_A AP5  
## 5EJE\_A AP5,CO  
## 1E4Y\_A AP5  
## 3X2S\_A JPY (2),AP5,MG  
## 6HAP\_A AP5  
## 6HAM\_A AP5  
## 4K46\_A ADP,AMP,PO4  
## 3GMT\_A SO4 (2)  
## 4PZL\_A CA,FMT,GOL  
## ligandName  
## 1AKE\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 6S36\_A CHLORIDE ION (3),SODIUM ION,MAGNESIUM ION (2)  
## 6RZE\_A SODIUM ION (3),CHLORIDE ION (2)  
## 3HPR\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 1E4V\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 5EJE\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE,COBALT (II) ION  
## 1E4Y\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 3X2S\_A N-(pyren-1-ylmethyl)acetamide (2),BIS(ADENOSINE)-5'-PENTAPHOSPHATE,MAGNESIUM ION  
## 6HAP\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 6HAM\_A BIS(ADENOSINE)-5'-PENTAPHOSPHATE  
## 4K46\_A ADENOSINE-5'-DIPHOSPHATE,ADENOSINE MONOPHOSPHATE,PHOSPHATE ION  
## 3GMT\_A SULFATE ION (2)  
## 4PZL\_A CALCIUM ION,FORMIC ACID,GLYCEROL  
## source  
## 1AKE\_A Escherichia coli  
## 6S36\_A Escherichia coli  
## 6RZE\_A Escherichia coli  
## 3HPR\_A Escherichia coli K-12  
## 1E4V\_A Escherichia coli  
## 5EJE\_A Escherichia coli O139:H28 str. E24377A  
## 1E4Y\_A Escherichia coli  
## 3X2S\_A Escherichia coli str. K-12 substr. MDS42  
## 6HAP\_A Escherichia coli O139:H28 str. E24377A  
## 6HAM\_A Escherichia coli K-12  
## 4K46\_A Photobacterium profundum  
## 3GMT\_A Burkholderia pseudomallei 1710b  
## 4PZL\_A Francisella tularensis subsp. tularensis SCHU S4  
## structureTitle  
## 1AKE\_A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIBITOR AP5A REFINED AT 1.9 ANGSTROMS RESOLUTION: A MODEL FOR A CATALYTIC TRANSITION STATE  
## 6S36\_A Crystal structure of E. coli Adenylate kinase R119K mutant  
## 6RZE\_A Crystal structure of E. coli Adenylate kinase R119A mutant  
## 3HPR\_A Crystal structure of V148G adenylate kinase from E. coli, in complex with Ap5A  
## 1E4V\_A Mutant G10V of adenylate kinase from E. coli, modified in the Gly-loop  
## 5EJE\_A Crystal structure of E. coli Adenylate kinase G56C/T163C double mutant in complex with Ap5a  
## 1E4Y\_A Mutant P9L of adenylate kinase from E. coli, modified in the Gly-loop  
## 3X2S\_A Crystal structure of pyrene-conjugated adenylate kinase  
## 6HAP\_A Adenylate kinase  
## 6HAM\_A Adenylate kinase  
## 4K46\_A Crystal Structure of Adenylate Kinase from Photobacterium profundum  
## 3GMT\_A Crystal structure of adenylate kinase from burkholderia pseudomallei  
## 4PZL\_A The crystal structure of adenylate kinase from Francisella tularensis subsp. tularensis SCHU S4  
## citation rObserved rFree  
## 1AKE\_A Muller, C.W., et al. J Mol Biol (1992) 0.19600 NA  
## 6S36\_A Rogne, P., et al. Biochemistry (2019) 0.16320 0.23560  
## 6RZE\_A Rogne, P., et al. Biochemistry (2019) 0.18650 0.23500  
## 3HPR\_A Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009) 0.21000 0.24320  
## 1E4V\_A Muller, C.W., et al. Proteins (1993) 0.19600 NA  
## 5EJE\_A Kovermann, M., et al. Proc Natl Acad Sci U S A (2017) 0.18890 0.23580  
## 1E4Y\_A Muller, C.W., et al. Proteins (1993) 0.17800 NA  
## 3X2S\_A Fujii, A., et al. Bioconjug Chem (2015) 0.20700 0.25600  
## 6HAP\_A Kantaev, R., et al. J Phys Chem B (2018) 0.22630 0.27760  
## 6HAM\_A Kantaev, R., et al. J Phys Chem B (2018) 0.20511 0.24325  
## 4K46\_A Cho, Y.-J., et al. To be published 0.17000 0.22290  
## 3GMT\_A Buchko, G.W., et al. Biochem Biophys Res Commun (2010) 0.23800 0.29500  
## 4PZL\_A Tan, K., et al. To be published 0.19360 0.23680  
## rWork spaceGroup  
## 1AKE\_A 0.19600 P 21 2 21  
## 6S36\_A 0.15940 C 1 2 1  
## 6RZE\_A 0.18190 C 1 2 1  
## 3HPR\_A 0.20620 P 21 21 2  
## 1E4V\_A 0.19600 P 21 2 21  
## 5EJE\_A 0.18630 P 21 2 21  
## 1E4Y\_A 0.17800 P 1 21 1  
## 3X2S\_A 0.20700 P 21 21 21  
## 6HAP\_A 0.22370 I 2 2 2  
## 6HAM\_A 0.20311 P 43  
## 4K46\_A 0.16730 P 21 21 21  
## 3GMT\_A 0.23500 P 1 21 1  
## 4PZL\_A 0.19130 P 32

## Principal component analysis

# Perform PCA  
pc.xray <- pca(pdbs)  
plot(pc.xray)



# Calculate all pairwise RMSD values of the structural ensemble

# Calculate RMSD  
rd <- rmsd(pdbs)

## Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions

# Structure-based clustering  
hc.rd <- hclust(dist(rd))  
grps.rd <- cutree(hc.rd, k=3)  
  
plot(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)

