

Class10,11_bggn13

AUTHOR

Lisanne Stouthart (PID A69036187)

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Barry told us uploading the class lecture is sufficient for class 11. In the class we learned the google document (ipynb) of alpha fold and we discussed the obtained figures. Please let me know if I need to provide more information to have a good submission.

PDB Intro

```
# Go to PDB
# Enter 1IEP
# Click on structure
# Display Files – PDB Format
# Search for atom, this data is used for the viewer of protein structure
# Meaning we dont have information for the first 224 residus.
```

Little exercise

```
# Download a CSV file from the PDB site (accessible from "Analyze" > "PDB Statist
# Save your input data file into your Project directory
pdbstat <- "~/Downloads/data_export_summary.csv"
pdbstats <- read.csv(pdbstat, row.names=1)
head(pdbstats)
```

	X.ray	EM	NMR	Multiple.methods	Neutron	Other
Protein (only)	167,317	15,698	12,534	208	77	32
Protein/Oligosaccharide	9,645	2,639	34	8	2	0
Protein/NA	8,735	4,718	286	7	0	0
Nucleic acid (only)	2,869	138	1,507	14	3	1
Other	170	10	33	0	0	0
Oligosaccharide (only)	11	0	6	1	0	4
Total						
Protein (only)	195,866					
Protein/Oligosaccharide	12,328					

Protein/NA	13,746
Nucleic acid (only)	4,532
Other	213
Oligosaccharide (only)	22

```
x <- pdbstats$Total
x
```

```
[1] "195,866" "12,328" "13,746" "4,532" "213" "22"
```

```
#as.numeric(x) #Warning: NAs introduced by coercion, the comma's are messing this
gsub(',', '', x)
```

```
[1] "195866" "12328" "13746" "4532" "213" "22"
```

```
convert_comma_numbers <- function(x) {
  #remove comma
  x <- gsub(',', '', x)
  #convert to numeric
  x <- as.numeric(x)

  return(x)
}
```

```
convert_comma_numbers(pdbstats$Total)
```

```
[1] 195866 12328 13746 4532 213 22
```

The 'apply()' function is very useful as it can take any function and "apply" it over either the ROWS or COLs of a data.frame

```
n.tot <- sum(convert_comma_numbers(pdbstats$Total))
n.tot
```

```
[1] 226707
```

```
195866 / 248838887 * 100
```

```
[1] 0.07871197
```

```
colSums(apply(pdbstats, 2, convert_comma_numbers))/n.tot
```

	X.ray	EM	NMR	Multiple.methods
0.8325592064	0.1023479646	0.0635181093	0.0010498132	
Neutron	Other	Total		
0.0003617003	0.0001632063	1.0000000000		

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

```
n.xray <- sum(convert_comma_numbers(pdbstats$X.ray))
n.em <- sum(convert_comma_numbers(pdbstats$EM))

n.xray/n.tot * 100
```

[1] 83.25592

```
n.em/n.tot * 100
```

[1] 10.2348

Q2 - What proportion of structures in the PDB are protein?

```
protein <- (n.xray+n.em)/n.tot * 100
protein
```

[1] 93.49072

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?

```
# Typing in "HIV-1 protease" gives 25,309 structures. This seems a bit much.
```

Visualizing the HIV-1 protease structure

```
# open Molstar (https://molstar.org/viewer/)
# 1HSG open -> apply
# Save your input data file into your Project directory
# Copy Paste it in the right folder

library(bio3d)
hsg_try <- read.pdb("~/Downloads/1hsg.pdb")
print(hsg_try)
```

Call: read.pdb(file = "~/Downloads/1hsg.pdb")

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:

```
PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGGFIKVRQYD  
QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE  
ALLDTGADDTVLEEMSLPGRWKPKMIGGIGGGFIKVRQYDQILIEICGHKAIGTVLVGPTP  
VNIIGRNLLTQIGCTLNF
```

+ attr: atom, xyz, seqres, helix, sheet,
calpha, remark, call

[summary\(hsg_try\)](#)

Call: read.pdb(file = "~/Downloads/1hsg.pdb")

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)]

+ attr: atom, xyz, seqres, helix, sheet,
calpha, remark, call

```
# Extract water molecules
water_indices <- which(hsg_try$atom$resid == "HOH" | hsg_try$atom$resid == "WAT")
water_molecules <- hsg_try$atom[water_indices, ]

# View water molecules
print(water_molecules)
```

	type	eleno	elety	alt	resid	chain	resno	insert	x	y	z	o
1560	HETATM	1562	0	<NA>	HOH	A	305	<NA>	20.857	43.192	21.450	1
1561	HETATM	1563	0	<NA>	HOH	A	307	<NA>	14.076	19.789	19.440	1
1562	HETATM	1564	0	<NA>	HOH	A	309	<NA>	28.075	21.177	7.222	1
1563	HETATM	1565	0	<NA>	HOH	A	314	<NA>	16.759	40.274	1.287	1
1564	HETATM	1566	0	<NA>	HOH	A	315	<NA>	13.997	22.233	21.468	1
1565	HETATM	1567	0	<NA>	HOH	A	324	<NA>	11.282	30.738	1.625	1
1566	HETATM	1568	0	<NA>	HOH	A	325	<NA>	16.774	42.740	2.296	1

1567	HETATM	1569	0 <NA>	HOH	A	327	<NA>	14.623	29.552	28.235	1
1568	HETATM	1570	0 <NA>	HOH	A	328	<NA>	1.651	36.463	19.459	1
1569	HETATM	1571	0 <NA>	HOH	A	329	<NA>	14.435	44.966	11.913	1
1570	HETATM	1572	0 <NA>	HOH	A	330	<NA>	19.877	40.160	21.917	1
1571	HETATM	1573	0 <NA>	HOH	A	331	<NA>	17.126	45.934	10.801	1
1572	HETATM	1574	0 <NA>	HOH	A	332	<NA>	8.840	28.026	4.860	1
1573	HETATM	1575	0 <NA>	HOH	A	335	<NA>	10.341	18.480	14.477	1
1574	HETATM	1576	0 <NA>	HOH	A	341	<NA>	19.233	16.711	9.027	1
1575	HETATM	1577	0 <NA>	HOH	A	342	<NA>	23.799	21.928	12.391	1
1576	HETATM	1578	0 <NA>	HOH	A	344	<NA>	9.953	37.934	4.548	1
1577	HETATM	1579	0 <NA>	HOH	A	345	<NA>	8.478	35.995	5.789	1
1578	HETATM	1580	0 <NA>	HOH	A	357	<NA>	3.960	19.389	17.384	1
1579	HETATM	1581	0 <NA>	HOH	A	373	<NA>	27.561	43.155	19.015	1
1580	HETATM	1582	0 <NA>	HOH	A	384	<NA>	1.245	19.292	18.124	1
1581	HETATM	1583	0 <NA>	HOH	A	386	<NA>	31.402	27.051	3.335	1
1582	HETATM	1584	0 <NA>	HOH	A	389	<NA>	32.446	31.200	4.417	1
1583	HETATM	1585	0 <NA>	HOH	A	391	<NA>	25.480	38.468	17.938	1
1584	HETATM	1586	0 <NA>	HOH	A	394	<NA>	23.940	41.721	0.346	1
1585	HETATM	1587	0 <NA>	HOH	A	401	<NA>	5.912	15.727	3.369	1
1586	HETATM	1588	0 <NA>	HOH	A	406	<NA>	9.272	33.891	12.681	1
1587	HETATM	1589	0 <NA>	HOH	A	408	<NA>	21.185	25.233	16.048	1
1588	HETATM	1590	0 <NA>	HOH	A	416	<NA>	18.474	26.012	21.664	1
1589	HETATM	1591	0 <NA>	HOH	A	420	<NA>	9.469	16.910	17.371	1
1590	HETATM	1592	0 <NA>	HOH	A	422	<NA>	13.074	17.786	16.615	1
1591	HETATM	1593	0 <NA>	HOH	A	439	<NA>	28.821	29.338	7.342	1
1592	HETATM	1594	0 <NA>	HOH	A	457	<NA>	23.284	23.107	15.132	1
1593	HETATM	1595	0 <NA>	HOH	A	468	<NA>	3.114	26.260	6.773	1
1594	HETATM	1596	0 <NA>	HOH	A	501	<NA>	6.382	26.424	5.893	1
1595	HETATM	1597	0 <NA>	HOH	A	503	<NA>	35.293	43.006	5.212	1
1596	HETATM	1598	0 <NA>	HOH	A	510	<NA>	21.891	49.715	7.192	1
1597	HETATM	1599	0 <NA>	HOH	A	524	<NA>	34.085	32.735	2.849	1
1598	HETATM	1600	0 <NA>	HOH	A	529	<NA>	31.491	41.524	6.678	1
1599	HETATM	1601	0 <NA>	HOH	A	553	<NA>	5.943	34.223	6.748	1
1600	HETATM	1602	0 <NA>	HOH	A	561	<NA>	0.934	40.259	19.405	1
1601	HETATM	1603	0 <NA>	HOH	A	567	<NA>	29.539	25.486	13.281	1
1602	HETATM	1604	0 <NA>	HOH	A	572	<NA>	24.552	17.352	10.295	1
1603	HETATM	1605	0 <NA>	HOH	A	575	<NA>	23.112	15.510	8.776	1
1604	HETATM	1606	0 <NA>	HOH	B	301	<NA>	20.445	8.036	-12.631	1
1605	HETATM	1607	0 <NA>	HOH	B	303	<NA>	20.044	14.822	4.638	1
1606	HETATM	1608	0 <NA>	HOH	B	304	<NA>	21.538	6.875	-10.099	1
1607	HETATM	1609	0 <NA>	HOH	B	306	<NA>	22.449	23.958	5.252	1
1608	HETATM	1610	0 <NA>	HOH	B	308	<NA>	11.720	21.289	7.190	1
1609	HETATM	1611	0 <NA>	HOH	B	312	<NA>	14.097	5.111	-11.638	1
1610	HETATM	1612	0 <NA>	HOH	B	313	<NA>	20.998	21.834	6.561	1
1611	HETATM	1613	0 <NA>	HOH	B	316	<NA>	22.659	14.583	-2.196	1
1612	HETATM	1614	0 <NA>	HOH	B	317	<NA>	28.724	15.629	-11.660	1
1613	HETATM	1615	0 <NA>	HOH	B	318	<NA>	16.539	45.207	0.079	1
1614	HETATM	1616	0 <NA>	HOH	B	319	<NA>	23.678	14.931	2.680	1
1615	HETATM	1617	0 <NA>	HOH	B	321	<NA>	20.718	15.976	-3.657	1
1616	HETATM	1618	0 <NA>	HOH	B	323	<NA>	31.249	26.796	-9.595	1
1617	HETATM	1619	0 <NA>	HOH	B	326	<NA>	28.813	28.445	-2.106	1

1618	HETATM	1620	0 <NA>	HOH	B	333	<NA>	12.251	39.551	-2.672	1
1619	HETATM	1621	0 <NA>	HOH	B	334	<NA>	25.465	12.592	-8.670	1
1620	HETATM	1622	0 <NA>	HOH	B	338	<NA>	12.998	36.205	-3.972	1
1621	HETATM	1623	0 <NA>	HOH	B	339	<NA>	17.541	17.060	-17.194	1
1622	HETATM	1624	0 <NA>	HOH	B	340	<NA>	5.321	14.325	-4.866	1
1623	HETATM	1625	0 <NA>	HOH	B	346	<NA>	9.314	17.330	-9.801	1
1624	HETATM	1626	0 <NA>	HOH	B	347	<NA>	7.435	26.652	-14.854	1
1625	HETATM	1627	0 <NA>	HOH	B	348	<NA>	4.405	16.704	-3.635	1
1626	HETATM	1628	0 <NA>	HOH	B	349	<NA>	19.414	7.026	4.428	1
1627	HETATM	1629	0 <NA>	HOH	B	350	<NA>	6.718	34.538	-1.322	1
1628	HETATM	1630	0 <NA>	HOH	B	354	<NA>	15.041	31.743	-13.235	1
1629	HETATM	1631	0 <NA>	HOH	B	355	<NA>	27.404	32.078	-10.860	1
1630	HETATM	1632	0 <NA>	HOH	B	356	<NA>	27.673	18.789	-6.155	1
1631	HETATM	1633	0 <NA>	HOH	B	358	<NA>	21.289	-1.161	-5.102	1
1632	HETATM	1634	0 <NA>	HOH	B	359	<NA>	6.973	36.523	1.489	1
1633	HETATM	1635	0 <NA>	HOH	B	360	<NA>	27.602	21.234	-0.635	1
1634	HETATM	1636	0 <NA>	HOH	B	362	<NA>	3.902	9.376	-0.027	1
1635	HETATM	1637	0 <NA>	HOH	B	364	<NA>	28.498	36.632	-7.529	1
1636	HETATM	1638	0 <NA>	HOH	B	366	<NA>	18.572	40.567	-10.042	1
1637	HETATM	1639	0 <NA>	HOH	B	367	<NA>	25.658	18.970	0.428	1
1638	HETATM	1640	0 <NA>	HOH	B	369	<NA>	20.843	1.263	-7.014	1
1639	HETATM	1641	0 <NA>	HOH	B	370	<NA>	13.975	15.741	12.070	1
1640	HETATM	1642	0 <NA>	HOH	B	374	<NA>	7.661	23.876	-6.324	1
1641	HETATM	1643	0 <NA>	HOH	B	375	<NA>	10.125	5.706	-1.458	1
1642	HETATM	1644	0 <NA>	HOH	B	376	<NA>	18.450	20.497	-18.728	1
1643	HETATM	1645	0 <NA>	HOH	B	377	<NA>	29.267	20.487	-3.497	1
1644	HETATM	1646	0 <NA>	HOH	B	379	<NA>	6.685	26.541	-5.608	1
1645	HETATM	1647	0 <NA>	HOH	B	381	<NA>	25.810	26.789	-19.106	1
1646	HETATM	1648	0 <NA>	HOH	B	383	<NA>	21.144	-4.428	-11.331	1
1647	HETATM	1649	0 <NA>	HOH	B	387	<NA>	16.904	27.594	-15.938	1
1648	HETATM	1650	0 <NA>	HOH	B	388	<NA>	23.926	45.612	-4.998	1
1649	HETATM	1651	0 <NA>	HOH	B	390	<NA>	25.300	17.493	3.076	1
1650	HETATM	1652	0 <NA>	HOH	B	392	<NA>	6.618	28.079	-3.427	1
1651	HETATM	1653	0 <NA>	HOH	B	393	<NA>	19.795	13.651	-16.606	1
1652	HETATM	1654	0 <NA>	HOH	B	395	<NA>	7.202	9.982	-4.103	1
1653	HETATM	1655	0 <NA>	HOH	B	400	<NA>	8.474	34.203	-4.893	1
1654	HETATM	1656	0 <NA>	HOH	B	405	<NA>	16.659	15.866	11.446	1
1655	HETATM	1657	0 <NA>	HOH	B	410	<NA>	26.400	10.057	-3.287	1
1656	HETATM	1658	0 <NA>	HOH	B	414	<NA>	9.503	3.489	-4.419	1
1657	HETATM	1659	0 <NA>	HOH	B	419	<NA>	15.438	12.973	-18.484	1
1658	HETATM	1660	0 <NA>	HOH	B	425	<NA>	11.428	19.956	-24.551	1
1659	HETATM	1661	0 <NA>	HOH	B	430	<NA>	18.725	43.171	-5.575	1
1660	HETATM	1662	0 <NA>	HOH	B	436	<NA>	32.141	29.620	-8.580	1
1661	HETATM	1663	0 <NA>	HOH	B	443	<NA>	8.811	13.667	-20.256	1
1662	HETATM	1664	0 <NA>	HOH	B	444	<NA>	4.071	26.169	-0.230	1
1663	HETATM	1665	0 <NA>	HOH	B	461	<NA>	11.425	44.636	-3.033	1
1664	HETATM	1666	0 <NA>	HOH	B	469	<NA>	6.902	23.686	-10.066	1
1665	HETATM	1667	0 <NA>	HOH	B	471	<NA>	5.749	25.785	-19.792	1
1666	HETATM	1668	0 <NA>	HOH	B	500	<NA>	25.592	16.404	-5.805	1
1667	HETATM	1669	0 <NA>	HOH	B	502	<NA>	4.040	15.516	-7.200	1
1668	HETATM	1670	0 <NA>	HOH	B	505	<NA>	28.640	34.232	-5.637	1

1669	HETATM	1671	0	<NA>	HOH	B	506	<NA>	8.979	11.173	11.112	1
1670	HETATM	1672	0	<NA>	HOH	B	509	<NA>	19.882	3.986	-18.136	1
1671	HETATM	1673	0	<NA>	HOH	B	514	<NA>	27.409	15.355	2.200	1
1672	HETATM	1674	0	<NA>	HOH	B	515	<NA>	17.222	39.766	-23.774	1
1673	HETATM	1675	0	<NA>	HOH	B	517	<NA>	28.742	24.158	-16.641	1
1674	HETATM	1676	0	<NA>	HOH	B	525	<NA>	22.694	-2.192	-12.589	1
1675	HETATM	1677	0	<NA>	HOH	B	526	<NA>	17.901	43.157	-14.082	1
1676	HETATM	1678	0	<NA>	HOH	B	531	<NA>	18.192	8.914	11.344	1
1677	HETATM	1679	0	<NA>	HOH	B	532	<NA>	19.507	45.215	1.709	1
1678	HETATM	1680	0	<NA>	HOH	B	548	<NA>	1.442	14.700	-6.128	1
1679	HETATM	1681	0	<NA>	HOH	B	549	<NA>	19.908	8.718	-19.215	1
1680	HETATM	1682	0	<NA>	HOH	B	556	<NA>	21.499	44.884	-1.280	1
1681	HETATM	1683	0	<NA>	HOH	B	564	<NA>	10.031	8.593	-22.052	1
1682	HETATM	1684	0	<NA>	HOH	B	568	<NA>	2.817	28.133	2.191	1
1683	HETATM	1685	0	<NA>	HOH	B	591	<NA>	15.835	40.105	-5.971	1
1684	HETATM	1686	0	<NA>	HOH	B	595	<NA>	4.515	36.451	-4.499	1
1685	HETATM	1687	0	<NA>	HOH	B	613	<NA>	24.127	-10.994	-0.982	1
1686	HETATM	1688	0	<NA>	HOH	B	617	<NA>	30.112	17.912	-4.791	1

b segid elesy charge

1560	63.07	<NA>	0	<NA>
1561	63.34	<NA>	0	<NA>
1562	66.96	<NA>	0	<NA>
1563	36.09	<NA>	0	<NA>
1564	64.67	<NA>	0	<NA>
1565	21.55	<NA>	0	<NA>
1566	26.65	<NA>	0	<NA>
1567	60.45	<NA>	0	<NA>
1568	25.82	<NA>	0	<NA>
1569	32.52	<NA>	0	<NA>
1570	41.02	<NA>	0	<NA>
1571	41.93	<NA>	0	<NA>
1572	27.94	<NA>	0	<NA>
1573	51.87	<NA>	0	<NA>
1574	66.74	<NA>	0	<NA>
1575	65.58	<NA>	0	<NA>
1576	67.74	<NA>	0	<NA>
1577	43.98	<NA>	0	<NA>
1578	37.23	<NA>	0	<NA>
1579	69.15	<NA>	0	<NA>
1580	70.78	<NA>	0	<NA>
1581	21.93	<NA>	0	<NA>
1582	46.57	<NA>	0	<NA>
1583	63.81	<NA>	0	<NA>
1584	47.08	<NA>	0	<NA>
1585	63.52	<NA>	0	<NA>
1586	31.73	<NA>	0	<NA>
1587	49.24	<NA>	0	<NA>
1588	65.44	<NA>	0	<NA>
1589	75.86	<NA>	0	<NA>
1590	67.42	<NA>	0	<NA>
1591	57.13	<NA>	0	<NA>

1592	60.42	<NA>	0	<NA>
1593	75.52	<NA>	0	<NA>
1594	38.21	<NA>	0	<NA>
1595	50.02	<NA>	0	<NA>
1596	53.78	<NA>	0	<NA>
1597	61.00	<NA>	0	<NA>
1598	73.78	<NA>	0	<NA>
1599	61.39	<NA>	0	<NA>
1600	49.60	<NA>	0	<NA>
1601	71.88	<NA>	0	<NA>
1602	66.74	<NA>	0	<NA>
1603	70.97	<NA>	0	<NA>
1604	63.94	<NA>	0	<NA>
1605	73.81	<NA>	0	<NA>
1606	42.37	<NA>	0	<NA>
1607	51.24	<NA>	0	<NA>
1608	18.18	<NA>	0	<NA>
1609	53.13	<NA>	0	<NA>
1610	47.68	<NA>	0	<NA>
1611	65.44	<NA>	0	<NA>
1612	38.53	<NA>	0	<NA>
1613	32.25	<NA>	0	<NA>
1614	61.86	<NA>	0	<NA>
1615	22.69	<NA>	0	<NA>
1616	59.93	<NA>	0	<NA>
1617	33.99	<NA>	0	<NA>
1618	79.22	<NA>	0	<NA>
1619	31.58	<NA>	0	<NA>
1620	47.41	<NA>	0	<NA>
1621	46.59	<NA>	0	<NA>
1622	48.25	<NA>	0	<NA>
1623	48.73	<NA>	0	<NA>
1624	54.68	<NA>	0	<NA>
1625	37.86	<NA>	0	<NA>
1626	68.44	<NA>	0	<NA>
1627	42.81	<NA>	0	<NA>
1628	60.62	<NA>	0	<NA>
1629	61.36	<NA>	0	<NA>
1630	35.03	<NA>	0	<NA>
1631	62.75	<NA>	0	<NA>
1632	71.64	<NA>	0	<NA>
1633	57.53	<NA>	0	<NA>
1634	50.97	<NA>	0	<NA>
1635	73.30	<NA>	0	<NA>
1636	62.30	<NA>	0	<NA>
1637	65.69	<NA>	0	<NA>
1638	61.76	<NA>	0	<NA>
1639	67.21	<NA>	0	<NA>
1640	61.89	<NA>	0	<NA>
1641	74.72	<NA>	0	<NA>
1642	48.75	<NA>	0	<NA>

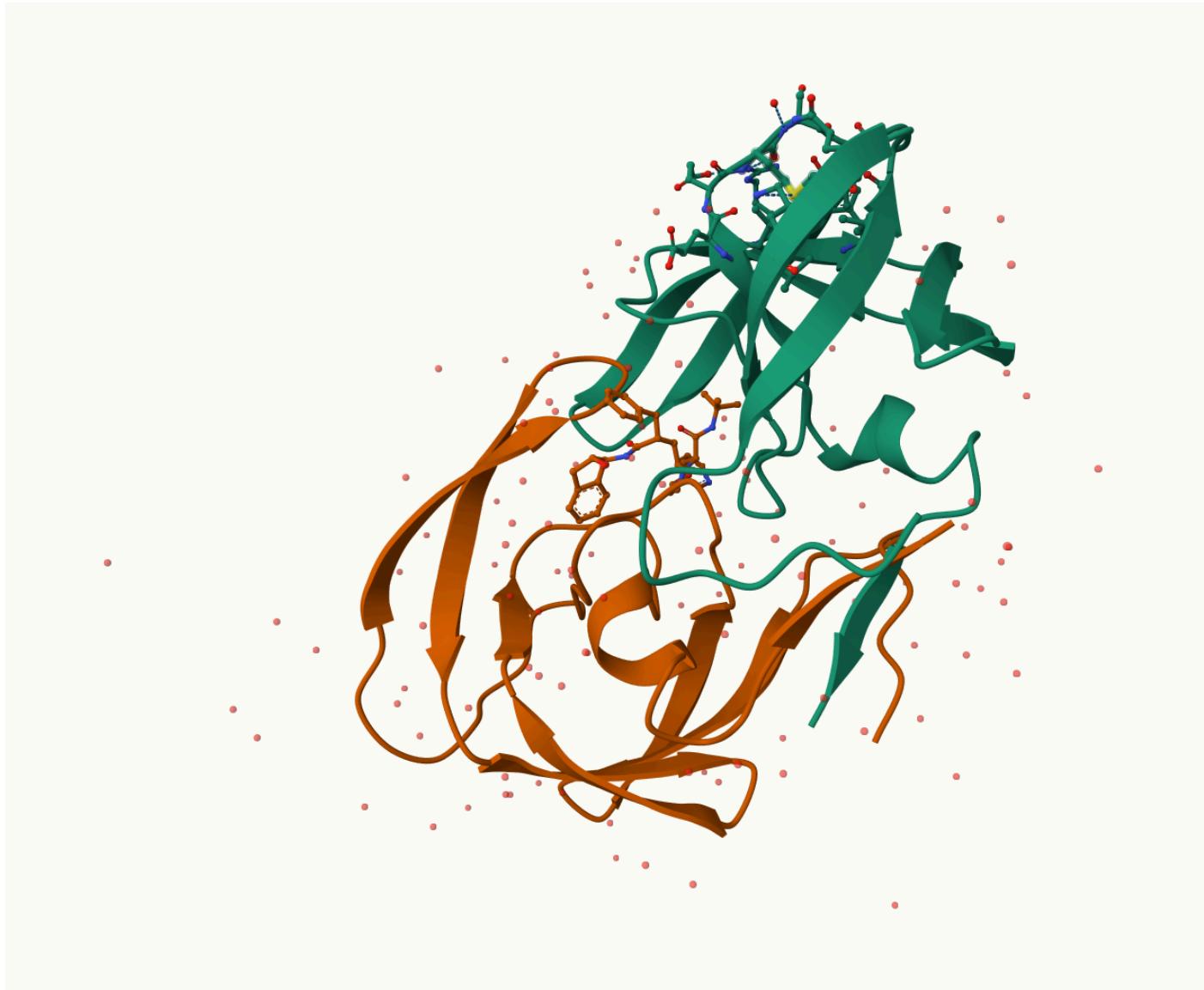
1643	60.17	<NA>	0	<NA>
1644	43.92	<NA>	0	<NA>
1645	70.16	<NA>	0	<NA>
1646	22.10	<NA>	0	<NA>
1647	27.84	<NA>	0	<NA>
1648	65.78	<NA>	0	<NA>
1649	67.04	<NA>	0	<NA>
1650	53.99	<NA>	0	<NA>
1651	54.21	<NA>	0	<NA>
1652	62.03	<NA>	0	<NA>
1653	63.64	<NA>	0	<NA>
1654	42.47	<NA>	0	<NA>
1655	65.50	<NA>	0	<NA>
1656	65.50	<NA>	0	<NA>
1657	73.55	<NA>	0	<NA>
1658	63.48	<NA>	0	<NA>
1659	52.97	<NA>	0	<NA>
1660	72.75	<NA>	0	<NA>
1661	75.75	<NA>	0	<NA>
1662	38.25	<NA>	0	<NA>
1663	68.43	<NA>	0	<NA>
1664	54.20	<NA>	0	<NA>
1665	63.96	<NA>	0	<NA>
1666	23.98	<NA>	0	<NA>
1667	52.93	<NA>	0	<NA>
1668	58.06	<NA>	0	<NA>
1669	64.79	<NA>	0	<NA>
1670	55.54	<NA>	0	<NA>
1671	61.69	<NA>	0	<NA>
1672	69.12	<NA>	0	<NA>
1673	78.93	<NA>	0	<NA>
1674	71.37	<NA>	0	<NA>
1675	78.14	<NA>	0	<NA>
1676	54.05	<NA>	0	<NA>
1677	72.78	<NA>	0	<NA>
1678	58.40	<NA>	0	<NA>
1679	58.78	<NA>	0	<NA>
1680	68.40	<NA>	0	<NA>
1681	64.90	<NA>	0	<NA>
1682	67.95	<NA>	0	<NA>
1683	53.68	<NA>	0	<NA>
1684	49.41	<NA>	0	<NA>
1685	64.49	<NA>	0	<NA>
1686	54.09	<NA>	0	<NA>

```
# Print the number of water molecules
cat("Total number of water molecules:", nrow(water_molecules), "\n")
```

Total number of water molecules: 127

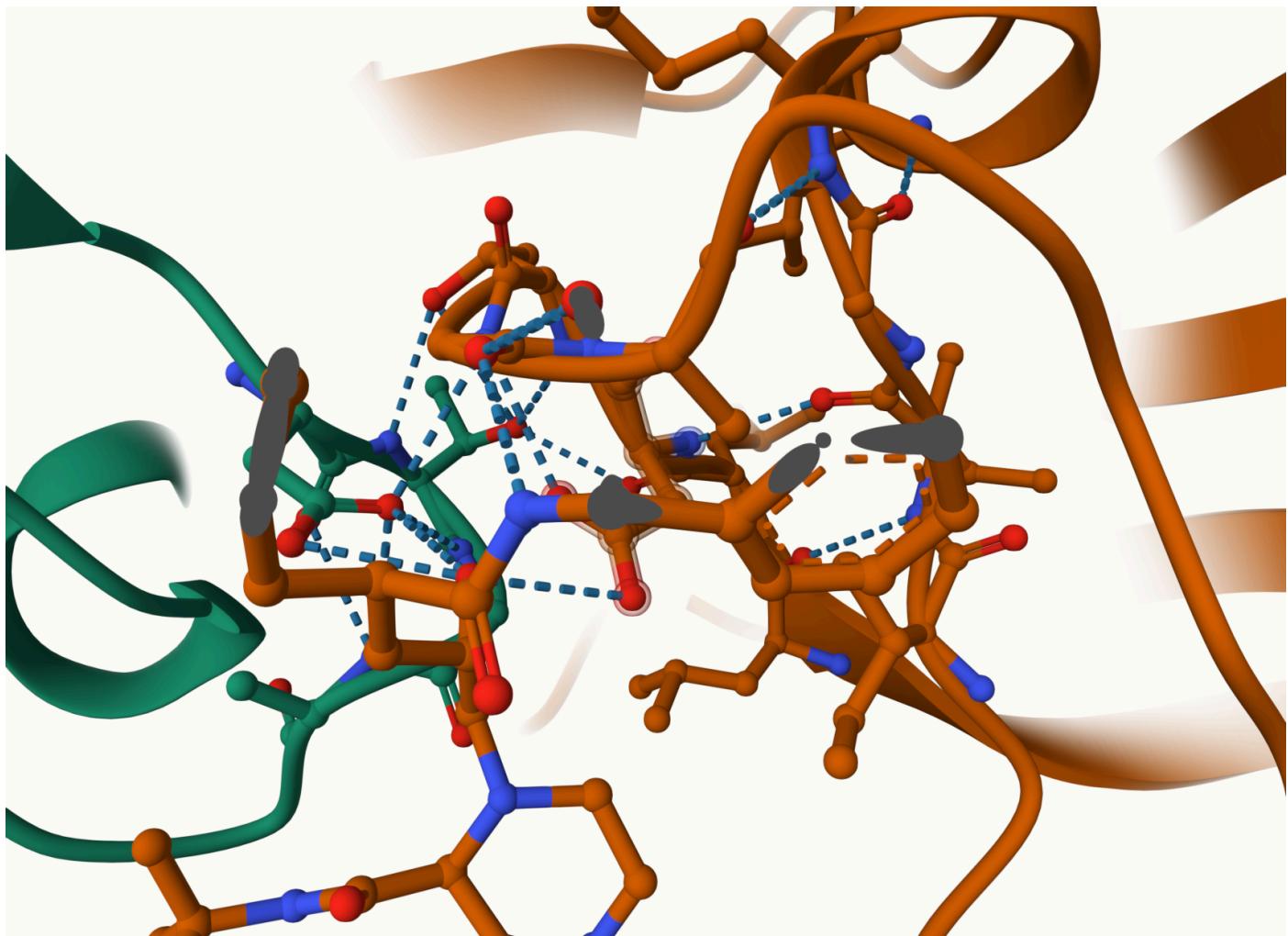
#127 water molecules

Using Mol*



My first image from Mol-star

Another Cool image



Another Cool image



The dot above is the water molecule.

Another Cool image

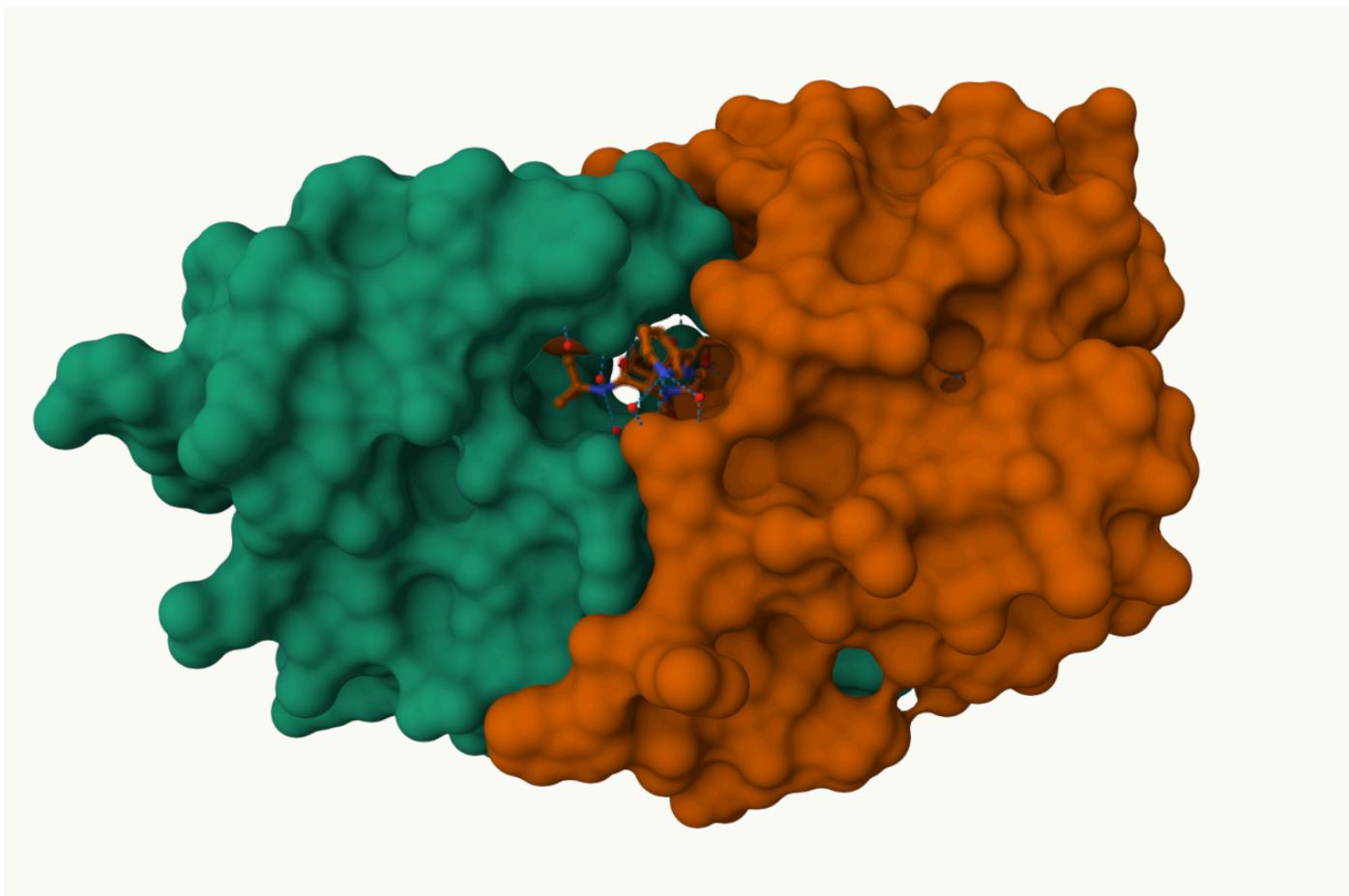


image 3

Another Cool image

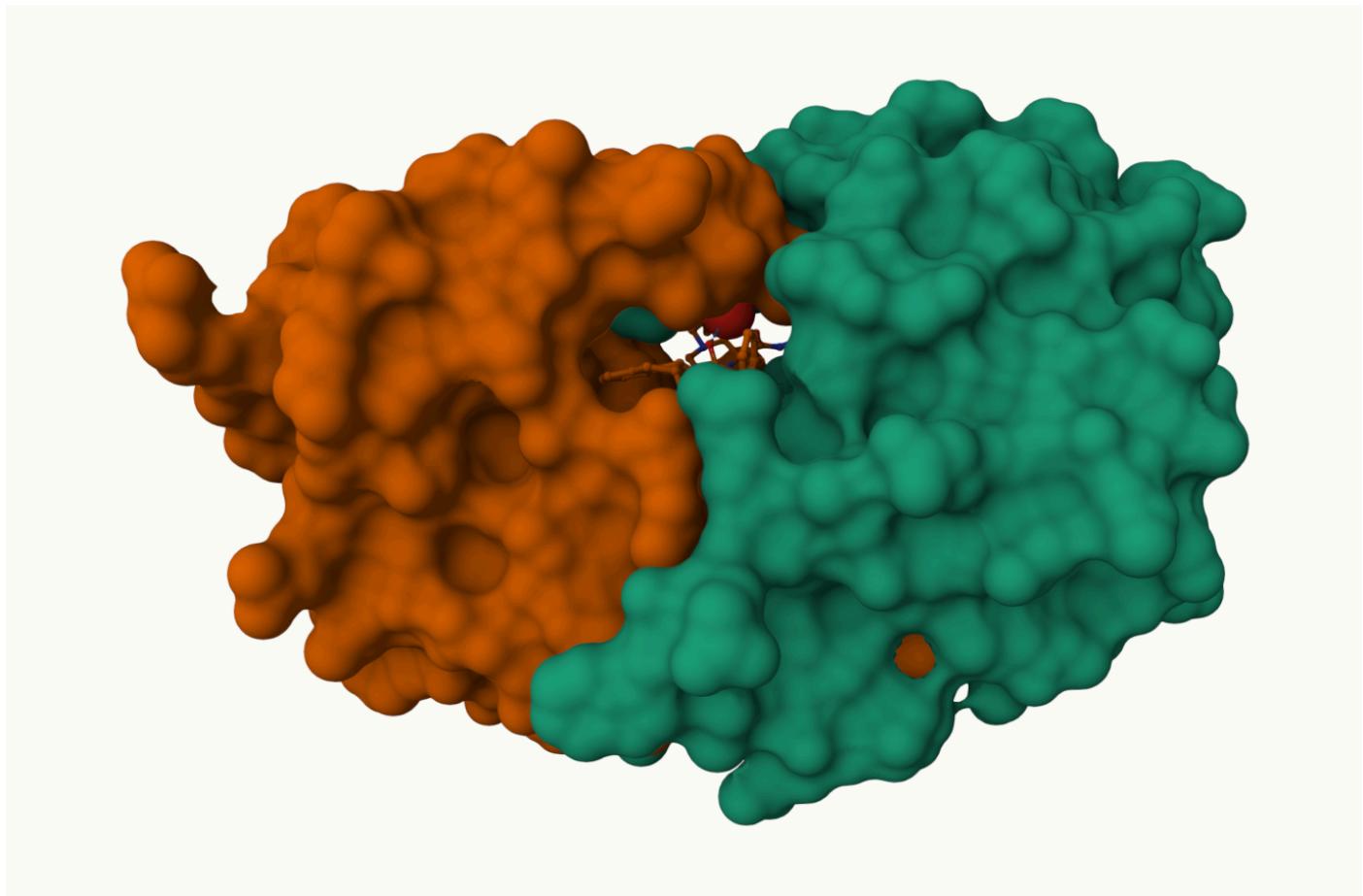


image 3

Another Cool image

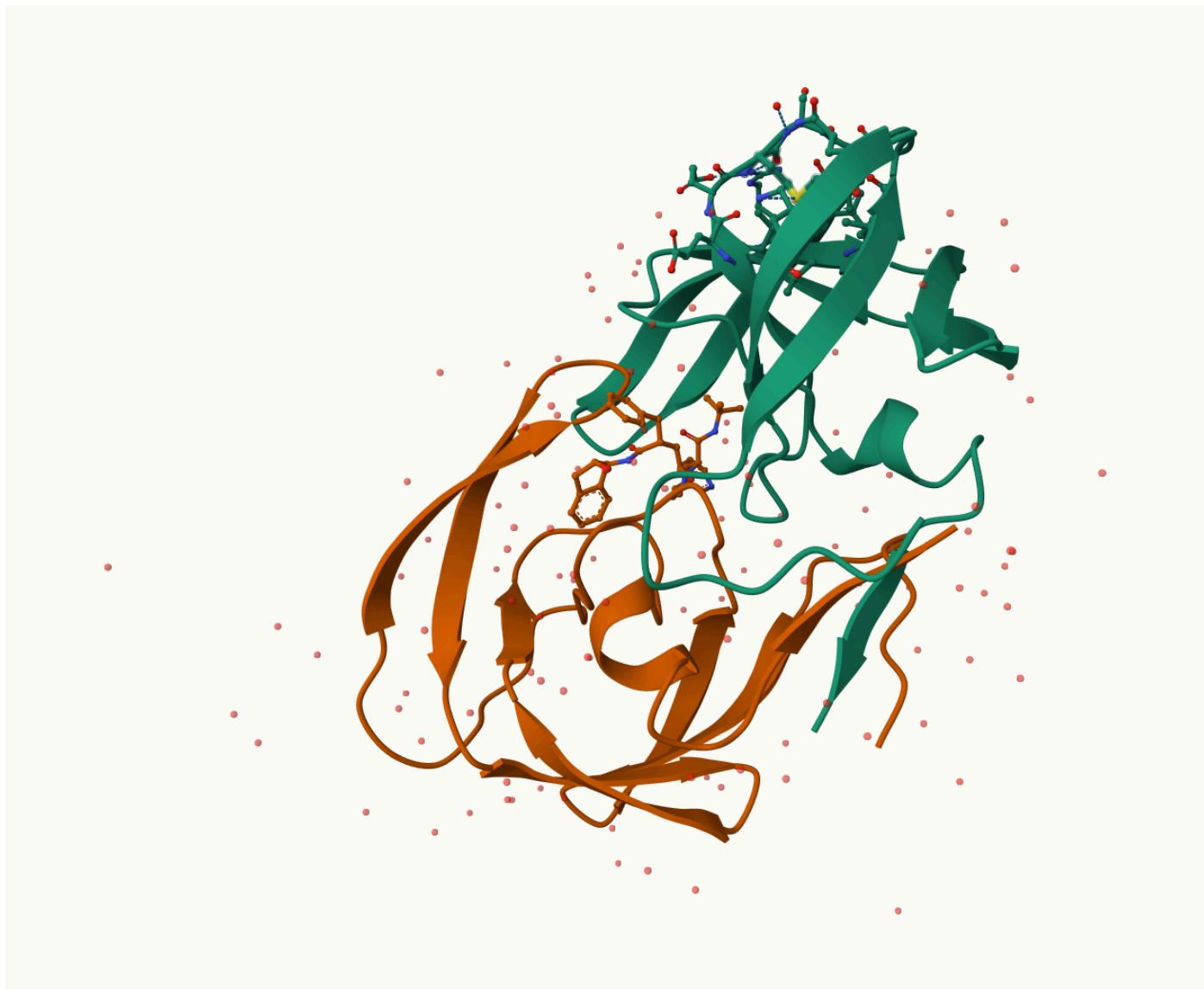


image 3

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

#H atom is on the beginning and it is very tiny. It is too small to see at this r

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

The watermolecule sits in the binding site.
Turn the water off, so that it is not visible.

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?

See previous uploaded

Q7 - [Optional] As you have hopefully observed HIV protease is a homodimer (i.e. it is composed of two identical chains). With the aid of the graphic display can you identify secondary structure elements that are likely to only form in the dimer rather than the monomer?

Optional

Interim

```
#Bio3D package for structural bioinformatics  
library(bio3d)  
pdb <- read.pdb("1hsg")
```

Note: Accessing on-line PDB file

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:

```
PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPMIGGIGGFVKVRQYD
QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKE
ALLDTGADDTVLEEMSLPGRWPKPMIGGIGGFVKVRQYDQILIEICGHKAIGTVLGPTP
VNIIGRNLLTQIGCTLNF
```

+ attr: atom, xyz, seqres, helix, sheet,
calpha, remark, call

Interim 2

`attributes(pdb)`

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

`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>										

`pdbseq(pdb)`

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
"P"	"Q"	"I"	"T"	"L"	"W"	"Q"	"R"	"P"	"L"	"V"	"T"	"I"	"K"	"I"	"G"	"G"	"Q"	"L"	"K"
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
"E"	"A"	"L"	"L"	"D"	"T"	"G"	"A"	"D"	"D"	"T"	"V"	"L"	"E"	"E"	"M"	"S"	"L"	"P"	"G"
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
"R"	"W"	"K"	"P"	"K"	"M"	"I"	"G"	"G"	"I"	"G"	"G"	"F"	"I"	"K"	"V"	"R"	"Q"	"Y"	"D"
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
"Q"	"I"	"L"	"I"	"E"	"I"	"C"	"G"	"H"	"K"	"A"	"I"	"G"	"T"	"V"	"L"	"V"	"G"	"P"	"T"

```

  81   82   83   84   85   86   87   88   89   90   91   92   93   94   95   96   97   98   99   1
"P" "V" "N" "I" "I" "G" "R" "N" "L" "L" "T" "Q" "I" "G" "C" "T" "L" "N" "F" "P"
  2   3   4   5   6   7   8   9   10  11  12  13  14  15  16  17  18  19  20  21
"Q" "I" "T" "L" "W" "Q" "R" "P" "L" "V" "T" "I" "K" "I" "G" "G" "Q" "L" "K" "E"
 22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41
"A" "L" "L" "D" "T" "G" "A" "D" "D" "T" "V" "L" "E" "E" "M" "S" "L" "P" "G" "R"
 42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61
"W" "K" "P" "K" "M" "I" "G" "G" "I" "G" "G" "F" "I" "K" "V" "R" "Q" "Y" "D" "Q"
 62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81
"I" "L" "I" "E" "I" "C" "G" "H" "K" "A" "I" "G" "T" "V" "L" "V" "G" "P" "T" "P"
 82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99
"V" "N" "I" "I" "G" "R" "N" "L" "L" "T" "Q" "I" "G" "C" "T" "L" "N" "F"

```

`pdbseq(pdb) [25]`

```

25
"D"

```

Interim 3

Functional dynamics prediction

Predicting functional motions of a single

```

source("https://tinyurl.com/viewpdb")
library(r3dmol)
library(shiny)

```

Warning: package 'shiny' was built under R version 4.3.3

`view.pdb(pdb, backgroundColor = "pink")`





Interim 4

```
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:

```
MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGMLRAAVKSGSELGKQAKDIDAGKLVT  
DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELVDKI  
VGRRVHAPSGRVYHVKNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG  
YYSKAEAGNTKYAKVDTGPVAEVRADLEKILG
```

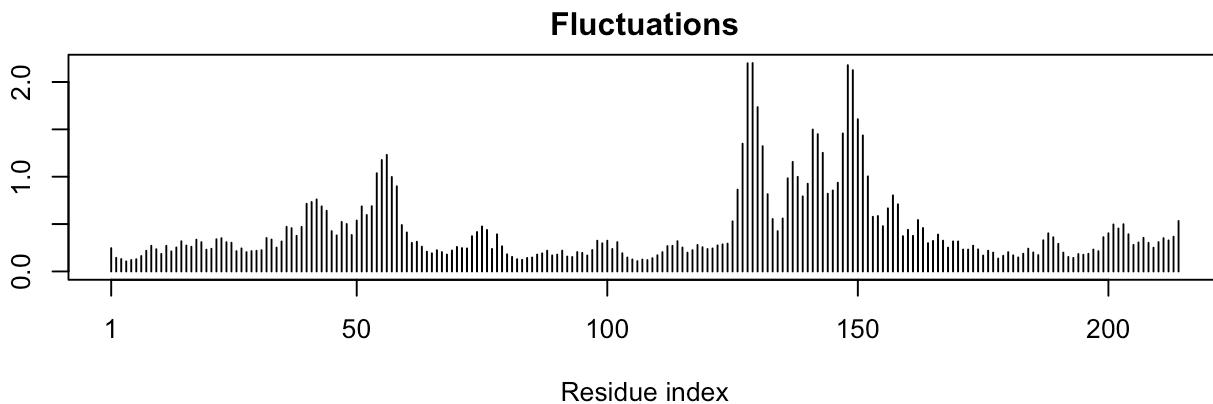
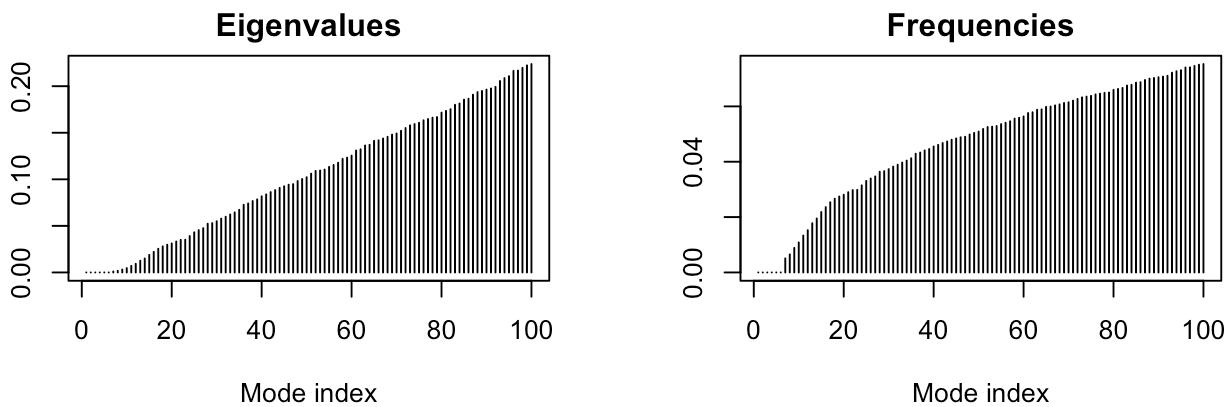
+ attr: atom, xyz, seqres, helix, sheet,
calpha, remark, call

```
modes <- nma(adk)
```

Building Hessian... Done in 0.018 seconds.

Diagonalizing Hessian... Done in 0.378 seconds.

```
plot(modes)
```



```
mktrj(modes, file='adk.pdb') #this is only A amino acids  
mktrj(modes, pdb=adk, file='adk.pdb')
```

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

```
length(pdbseq(pdb)) #answer = 198
```

[1] 198

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

```
#In the summary output of the pdb object, it shows that there are non-protein res
```

Q9 - How many protein chains are in this structure?

```
# Number of unique protein chains
num_chains <- length(unique(pdb$atom$chain))
print(num_chains) #answer is 2
```

[1] 2

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

```
#msa
# not needed, no time in class
```

Q11 - Which of the above packages is not found on BioConductor or CRAN?

```
#bitbucket
# not needed, no time in class
```

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

```
#True
# not needed, no time in class
```

Previous Class, finishing it up

```
library(bio3d)

id <- "1ake_A"
aa <- get.seq(id)
```

Warning in get.seq(id): Removing existing file: seqs.fasta

Fetching... Please wait. Done.

aa

1 pdb 1AKE A	.	60
		MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGMLRAAVKSGSELGKQAKDIMALGKLVT

1 60
61 120
pdb|1AKE|A DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVVDYLEFDVPDELIVDRI
61 120

121 180
pdb|1AKE|A VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTRKDDQEETVRKRLVEYHQMTAPLIG
121 180

181 214
pdb|1AKE|A YYSKAEAGNTKYAKVDGTPVAEVRADLEKILG
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

```
b <- blast.ncbi(aa)
```

Searching ... please wait (updates every 5 seconds) RID = JX1895U7016

Digitized by srujanika@gmail.com

Reporting 85 hits

b

\$hit.tbl

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens
1	Query_3264839	1AKE_A	100.000	214	0	0
2	Query_3264839	8BQF_A	99.533	214	1	0
3	Query_3264839	4X8M_A	99.533	214	1	0
4	Query_3264839	6S36_A	99.533	214	1	0
5	Query_3264839	8Q2B_A	99.533	214	1	0
6	Query_3264839	8RJ9_A	99.533	214	1	0
7	Query_3264839	6RZE_A	99.533	214	1	0
8	Query_3264839	4X8H_A	99.533	214	1	0
9	Query_3264839	3HPR_A	99.533	214	1	0
10	Query_3264839	1E4V_A	99.533	214	1	0
11	Query_3264839	5EJE_A	99.065	214	2	0
12	Query_3264839	1E4Y_A	99.065	214	2	0
13	Query_3264839	3X2S_A	98.598	214	3	0
14	Query_3264839	6HAP_A	98.131	214	4	0

15	Query_3264839	6HAM_A	97.196	214	6	0
16	Query_3264839	4K46_A	73.239	213	57	0
17	Query_3264839	4NP6_A	72.642	212	58	0
18	Query_3264839	3GMT_A	62.500	216	75	1
19	Query_3264839	4PZL_A	57.346	211	86	2
20	Query_3264839	5G3Y_A	55.505	218	88	2
21	Query_3264839	5G3Z_A	50.459	218	99	2
22	Query_3264839	5G40_A	49.541	218	101	2
23	Query_3264839	5X6J_A	50.000	218	98	3
24	Query_3264839	2C9Y_A	53.723	188	83	1
25	Query_3264839	1S3G_A	49.541	218	99	3
26	Query_3264839	1AK2_A	52.660	188	85	1
27	Query_3264839	3BE4_A	48.611	216	102	3
28	Query_3264839	1AKY_A	46.119	219	108	3
29	Query_3264839	3AKY_A	46.119	219	108	3
30	Query_3264839	3FB4_A	48.165	218	104	2
31	Query_3264839	4QBI_A	47.248	218	106	2
32	Query_3264839	1DVR_A	45.205	219	110	3
33	Query_3264839	3DKV_A	49.772	219	99	3
34	Query_3264839	3DL0_A	48.165	218	104	2
35	Query_3264839	1ZIN_A	45.413	218	110	2
36	Query_3264839	2P3S_A	47.248	218	106	2
37	Query_3264839	2EU8_A	47.248	218	106	2
38	Query_3264839	1P3J_A	47.248	218	106	2
39	Query_3264839	4QBF_A	49.772	219	99	3
40	Query_3264839	20RI_A	47.248	218	106	2
41	Query_3264839	5X6I_A	46.789	218	107	2
42	Query_3264839	2QAJ_A	47.005	217	106	2
43	Query_3264839	2007_A	46.789	218	107	2
44	Query_3264839	20SB_A	46.789	218	107	2
45	Query_3264839	4MKF_A	46.789	218	107	2
46	Query_3264839	3TLX_A	44.393	214	106	3
47	Query_3264839	4MKH_A	48.624	218	101	3
48	Query_3264839	4QBH_A	45.872	218	109	2
49	Query_3264839	4TYQ_A	48.165	218	102	3
50	Query_3264839	4QBG_B	47.248	218	104	3
51	Query_3264839	4TYP_A	47.248	218	104	3
52	Query_3264839	4JKY_A	44.037	218	103	5
53	Query_3264839	2RGX_A	43.578	218	104	4
54	Query_3264839	4JL0_A	43.578	218	104	5
55	Query_3264839	1ZAK_A	42.326	215	112	3
56	Query_3264839	1ZD8_A	43.915	189	96	3
57	Query_3264839	2AK3_A	44.324	185	101	2
58	Query_3264839	4NTZ_A	38.532	218	119	4
59	Query_3264839	2AR7_A	41.304	184	102	3
60	Query_3264839	3NDP_A	40.761	184	103	3
61	Query_3264839	1P4S_A	39.785	186	77	2
62	Query_3264839	2CDN_A	39.785	186	77	2
63	Query_3264839	3L0P_A	32.735	223	131	7
64	Query_3264839	5X6L_A	35.784	204	98	3
65	Query_3264839	2XB4_A	32.735	223	131	7

66	Query_3264839	5XRU_A	35.294		204	99	3
67	Query_3264839	5YCC_A	35.294		204	99	3
68	Query_3264839	5X6K_A	35.294		204	99	3
69	Query_3264839	5XZ2_A	33.962		212	107	3
70	Query_3264839	5YCF_A	34.906		212	105	3
71	Query_3264839	5YCB_A	34.804		204	100	3
72	Query_3264839	5YCD_A	36.464		181	87	2
73	Query_3264839	3ADK_A	36.066		183	89	3
74	Query_3264839	3UMF_A	33.333		186	92	3
75	Query_3264839	1Z83_A	34.973		183	91	3
76	Query_3264839	7X7S_A	34.973		183	91	3
77	Query_3264839	3CM0_A	34.434		212	106	5
78	Query_3264839	7DE3_A	36.066		183	89	5
79	Query_3264839	7N6G_6M	33.333		180	108	4
80	Query_3264839	1UKY_A	27.962		211	117	5
81	Query_3264839	1TEV_A	31.963		219	109	7
82	Query_3264839	7E9V_A	31.963		219	109	7
83	Query_3264839	2BWJ_A	30.851		188	98	4
84	Query_3264839	1QF9_A	27.907		215	117	5
85	Query_3264839	7N6G_6A	24.852		169	84	6

		q.start	q.end	s.start	s.end	evalue	bitscore	positives	mlog.evalue	pdb.id
1		1	214	1	214	1.58e-156	432.0	100.00	358.74585	1AKE_A
2		1	214	21	234	2.58e-156	433.0	100.00	358.25549	8BQF_A
3		1	214	1	214	2.82e-156	432.0	100.00	358.16654	4X8M_A
4		1	214	1	214	4.14e-156	432.0	100.00	357.78258	6S36_A
5		1	214	1	214	1.10e-155	431.0	99.53	356.80538	8Q2B_A
6		1	214	1	214	1.10e-155	431.0	99.53	356.80538	8RJ9_A
7		1	214	1	214	1.19e-155	431.0	99.53	356.72674	6RZE_A
8		1	214	1	214	1.56e-155	430.0	99.53	356.45600	4X8H_A
9		1	214	1	214	2.22e-155	430.0	99.53	356.10318	3HPR_A
10		1	214	1	214	2.34e-155	430.0	99.53	356.05054	1E4V_A
11		1	214	1	214	7.03e-155	429.0	99.07	354.95050	5EJE_A
12		1	214	1	214	4.07e-154	427.0	99.07	353.19446	1E4Y_A
13		1	214	1	214	6.81e-154	426.0	98.60	352.67971	3X2S_A
14		1	214	1	214	2.02e-153	425.0	98.60	351.59242	6HAP_A
15		1	214	1	214	4.12e-153	424.0	98.60	350.87967	6HAM_A
16		1	213	1	213	1.83e-115	329.0	84.98	264.19297	4K46_A
17		2	213	5	216	1.01e-113	325.0	84.43	260.18217	4NP6_A
18		2	211	10	225	7.93e-90	265.0	71.30	205.16201	3GMT_A
19		2	209	26	235	1.89e-86	256.0	74.41	197.38574	4PZL_A
20		1	214	1	213	2.80e-76	230.0	68.81	173.96685	5G3Y_A
21		1	214	1	213	5.36e-73	221.0	69.27	166.40975	5G3Z_A
22		1	214	1	213	1.10e-70	216.0	68.35	161.08565	5G40_A
23		1	213	1	212	2.01e-68	210.0	65.60	155.87765	5X6J_A
24		1	184	17	204	1.01e-67	209.0	69.68	154.26325	2C9Y_A
25		1	213	1	212	1.04e-67	208.0	65.14	154.23398	1S3G_A
26		1	184	17	204	2.61e-67	207.0	70.21	153.31385	1AK2_A
27		2	213	7	217	7.02e-67	206.0	68.06	152.32444	3BE4_A
28		1	214	5	218	9.92e-67	206.0	65.75	151.97865	1AKY_A
29		1	214	5	218	1.21e-66	205.0	65.30	151.78000	3AKY_A
30		1	214	1	213	4.43e-66	204.0	65.14	150.48222	3FB4_A

31	1	214	1	213	2.15e-64	199.0	65.60	146.59998	4QBI_A
32	1	214	5	218	4.63e-64	199.0	64.84	145.83289	1DVR_A
33	1	214	1	213	2.49e-63	197.0	66.67	144.15058	3DKV_A
34	1	214	1	213	1.10e-62	195.0	66.97	142.66497	3DL0_A
35	1	214	1	213	1.22e-62	195.0	65.60	142.56142	1ZIN_A
36	1	214	1	213	1.85e-62	194.0	66.97	142.14509	2P3S_A
37	1	214	1	213	1.93e-62	194.0	66.97	142.10276	2EU8_A
38	1	214	1	213	2.62e-62	194.0	66.97	141.79710	1P3J_A
39	1	214	1	213	4.18e-62	194.0	66.21	141.32996	4QBF_A
40	1	214	1	213	6.55e-62	193.0	66.97	140.88081	20RI_A
41	1	214	1	213	1.24e-61	192.0	66.51	140.24258	5X6I_A
42	1	213	1	212	1.34e-61	192.0	66.82	140.16502	2QAJ_A
43	1	214	1	213	1.56e-61	192.0	66.51	140.01300	2007_A
44	1	214	1	213	2.67e-61	192.0	66.51	139.47561	20SB_A
45	1	214	1	213	2.97e-61	191.0	66.51	139.36913	4MKF_A
46	2	211	31	235	1.88e-60	190.0	64.95	137.52383	3TLX_A
47	1	213	3	214	1.92e-60	189.0	65.60	137.50278	4MKH_A
48	1	214	1	213	3.84e-60	189.0	64.68	136.80963	4QBH_A
49	1	213	1	212	4.72e-60	188.0	65.60	136.60330	4TYQ_A
50	1	213	1	212	8.26e-59	185.0	65.60	133.74110	4QBG_B
51	1	213	1	212	1.55e-58	184.0	65.14	133.11168	4TYP_A
52	1	214	1	203	7.13e-56	177.0	66.51	126.98045	4JKY_A
53	1	214	1	203	8.17e-56	177.0	65.60	126.84430	2RGX_A
54	1	214	1	203	2.44e-55	176.0	66.51	125.75018	4JL0_A
55	1	214	6	209	4.91e-54	173.0	63.72	122.74832	1ZAK_A
56	1	185	8	190	3.82e-50	164.0	64.55	113.78900	1ZD8_A
57	1	185	7	189	5.25e-50	163.0	65.41	113.47103	2AK3_A
58	1	213	6	213	1.87e-46	154.0	62.39	105.29298	4NTZ_A
59	1	182	28	207	6.61e-45	150.0	64.13	101.72775	2AR7_A
60	1	182	6	185	5.52e-44	148.0	63.59	99.60537	3NDP_A
61	1	182	1	155	9.17e-39	133.0	56.99	87.58488	1P4S_A
62	1	182	21	175	1.65e-38	133.0	56.99	86.99746	2CDN_A
63	1	209	1	218	3.53e-31	115.0	54.26	70.11884	3L0P_A
64	3	205	13	184	4.30e-31	114.0	52.45	69.92152	5X6L_A
65	1	209	1	218	4.66e-31	114.0	54.26	69.84112	2XB4_A
66	3	205	11	182	5.17e-31	113.0	52.45	69.73727	5XRU_A
67	3	205	11	182	6.07e-31	113.0	52.45	69.57678	5YCC_A
68	3	205	13	184	7.20e-31	113.0	52.45	69.40606	5X6K_A
69	3	213	13	192	8.50e-31	113.0	52.83	69.24007	5XZ2_A
70	3	213	11	190	8.65e-31	113.0	51.89	69.22258	5YCF_A
71	3	205	11	182	9.33e-31	113.0	52.45	69.14690	5YCB_A
72	3	182	11	164	2.47e-30	112.0	54.70	68.17333	5YCD_A
73	3	184	12	167	2.96e-30	111.0	54.10	67.99236	3ADK_A
74	3	185	32	188	1.57e-29	110.0	56.45	66.32389	3UMF_A
75	3	184	12	167	1.84e-29	109.0	54.64	66.16520	1Z83_A
76	3	184	16	171	2.18e-29	109.0	54.64	65.99564	7X7S_A
77	3	214	7	185	8.86e-29	107.0	50.94	64.59342	3CM0_A
78	3	184	11	166	1.17e-28	107.0	56.28	64.31538	7DE3_A
79	1	168	1239	1418	1.52e-25	105.0	53.33	57.14592	7N6G_6
80	3	210	18	196	9.58e-25	97.8	53.55	55.30495	1UKY_A
81	3	213	6	192	6.79e-24	95.5	49.77	53.34659	1TEV_A

82	3	213	24	210	1.05e-23	95.5	49.77	52.91067	7E9V_A
83	3	187	15	173	7.01e-22	90.1	49.47	48.70953	2BWJ_A
84	3	213	9	189	2.73e-20	85.9	47.44	45.04740	1QF9_A
85	76	207	760	922	1.27e-05	47.0	42.01	11.27391	7N6G_6
		acc							
1		1AKE_A							
2		8BQF_A							
3		4X8M_A							
4		6S36_A							
5		8Q2B_A							
6		8RJ9_A							
7		6RZE_A							
8		4X8H_A							
9		3HPR_A							
10		1E4V_A							
11		5EJE_A							
12		1E4Y_A							
13		3X2S_A							
14		6HAP_A							
15		6HAM_A							
16		4K46_A							
17		4NP6_A							
18		3GMT_A							
19		4PZL_A							
20		5G3Y_A							
21		5G3Z_A							
22		5G40_A							
23		5X6J_A							
24		2C9Y_A							
25		1S3G_A							
26		1AK2_A							
27		3BE4_A							
28		1AKY_A							
29		3AKY_A							
30		3FB4_A							
31		4QBI_A							
32		1DVR_A							
33		3DKV_A							
34		3DL0_A							
35		1ZIN_A							
36		2P3S_A							
37		2EU8_A							
38		1P3J_A							
39		4QBF_A							
40		20RI_A							
41		5X6I_A							
42		20AJ_A							
43		2007_A							
44		20SB_A							
45		4MKF_A							
46		3TLX_A							

```

47 4MKH_A
48 4QBH_A
49 4TYQ_A
50 4QBG_B
51 4TYP_A
52 4JKY_A
53 2RGX_A
54 4JL0_A
55 1ZAK_A
56 1ZD8_A
57 2AK3_A
58 4NTZ_A
59 2AR7_A
60 3NDP_A
61 1P4S_A
62 2CDN_A
63 3L0P_A
64 5X6L_A
65 2XB4_A
66 5XRU_A
67 5YCC_A
68 5X6K_A
69 5XZ2_A
70 5YCF_A
71 5YCB_A
72 5YCD_A
73 3ADK_A
74 3UMF_A
75 1Z83_A
76 7X7S_A
77 3CM0_A
78 7DE3_A
79 7N6G_6M
80 1UKY_A
81 1TEV_A
82 7E9V_A
83 2BWJ_A
84 1QF9_A
85 7N6G_6A

```

\$raw

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens
1	Query_3264839	1AKE_A	100.000	214	0	0
2	Query_3264839	8BQF_A	99.533	214	1	0
3	Query_3264839	4X8M_A	99.533	214	1	0
4	Query_3264839	6S36_A	99.533	214	1	0
5	Query_3264839	8Q2B_A	99.533	214	1	0
6	Query_3264839	8RJ9_A	99.533	214	1	0
7	Query_3264839	6RZE_A	99.533	214	1	0
8	Query_3264839	4X8H_A	99.533	214	1	0
9	Query_3264839	3HPR_A	99.533	214	1	0

10	Query_3264839	1E4V_A	99.533	214	1	0
11	Query_3264839	5EJE_A	99.065	214	2	0
12	Query_3264839	1E4Y_A	99.065	214	2	0
13	Query_3264839	3X2S_A	98.598	214	3	0
14	Query_3264839	6HAP_A	98.131	214	4	0
15	Query_3264839	6HAM_A	97.196	214	6	0
16	Query_3264839	4K46_A	73.239	213	57	0
17	Query_3264839	4NP6_A	72.642	212	58	0
18	Query_3264839	3GMT_A	62.500	216	75	1
19	Query_3264839	4PZL_A	57.346	211	86	2
20	Query_3264839	5G3Y_A	55.505	218	88	2
21	Query_3264839	5G3Z_A	50.459	218	99	2
22	Query_3264839	5G40_A	49.541	218	101	2
23	Query_3264839	5X6J_A	50.000	218	98	3
24	Query_3264839	2C9Y_A	53.723	188	83	1
25	Query_3264839	1S3G_A	49.541	218	99	3
26	Query_3264839	1AK2_A	52.660	188	85	1
27	Query_3264839	3BE4_A	48.611	216	102	3
28	Query_3264839	1AKY_A	46.119	219	108	3
29	Query_3264839	3AKY_A	46.119	219	108	3
30	Query_3264839	3FB4_A	48.165	218	104	2
31	Query_3264839	4QBI_A	47.248	218	106	2
32	Query_3264839	1DVR_A	45.205	219	110	3
33	Query_3264839	3DKV_A	49.772	219	99	3
34	Query_3264839	3DL0_A	48.165	218	104	2
35	Query_3264839	1ZIN_A	45.413	218	110	2
36	Query_3264839	2P3S_A	47.248	218	106	2
37	Query_3264839	2EU8_A	47.248	218	106	2
38	Query_3264839	1P3J_A	47.248	218	106	2
39	Query_3264839	4QBF_A	49.772	219	99	3
40	Query_3264839	20RI_A	47.248	218	106	2
41	Query_3264839	5X6I_A	46.789	218	107	2
42	Query_3264839	2QAJ_A	47.005	217	106	2
43	Query_3264839	2007_A	46.789	218	107	2
44	Query_3264839	20SB_A	46.789	218	107	2
45	Query_3264839	4MKF_A	46.789	218	107	2
46	Query_3264839	3TLX_A	44.393	214	106	3
47	Query_3264839	4MKH_A	48.624	218	101	3
48	Query_3264839	4QBH_A	45.872	218	109	2
49	Query_3264839	4TYQ_A	48.165	218	102	3
50	Query_3264839	4QBG_B	47.248	218	104	3
51	Query_3264839	4TYP_A	47.248	218	104	3
52	Query_3264839	4JKY_A	44.037	218	103	5
53	Query_3264839	2RGX_A	43.578	218	104	4
54	Query_3264839	4JL0_A	43.578	218	104	5
55	Query_3264839	1ZAK_A	42.326	215	112	3
56	Query_3264839	1ZD8_A	43.915	189	96	3
57	Query_3264839	2AK3_A	44.324	185	101	2
58	Query_3264839	4NTZ_A	38.532	218	119	4
59	Query_3264839	2AR7_A	41.304	184	102	3
60	Query_3264839	3NDP_A	40.761	184	103	3

61	Query_3264839	1P4S_A	39.785	186	77	2
62	Query_3264839	2CDN_A	39.785	186	77	2
63	Query_3264839	3L0P_A	32.735	223	131	7
64	Query_3264839	5X6L_A	35.784	204	98	3
65	Query_3264839	2XB4_A	32.735	223	131	7
66	Query_3264839	5XRU_A	35.294	204	99	3
67	Query_3264839	5YCC_A	35.294	204	99	3
68	Query_3264839	5X6K_A	35.294	204	99	3
69	Query_3264839	5XZ2_A	33.962	212	107	3
70	Query_3264839	5YCF_A	34.906	212	105	3
71	Query_3264839	5YCB_A	34.804	204	100	3
72	Query_3264839	5YCD_A	36.464	181	87	2
73	Query_3264839	3ADK_A	36.066	183	89	3
74	Query_3264839	3UMF_A	33.333	186	92	3
75	Query_3264839	1Z83_A	34.973	183	91	3
76	Query_3264839	7X7S_A	34.973	183	91	3
77	Query_3264839	3CM0_A	34.434	212	106	5
78	Query_3264839	7DE3_A	36.066	183	89	5
79	Query_3264839	7N6G_6M	33.333	180	108	4
80	Query_3264839	1UKY_A	27.962	211	117	5
81	Query_3264839	1TEV_A	31.963	219	109	7
82	Query_3264839	7E9V_A	31.963	219	109	7
83	Query_3264839	2BWJ_A	30.851	188	98	4
84	Query_3264839	1QF9_A	27.907	215	117	5
85	Query_3264839	7N6G_6A	24.852	169	84	6

		q.start	q.end	s.start	s.end	evalue	bitscore	positives
1		1	214	1	214	1.58e-156	432.0	100.00
2		1	214	21	234	2.58e-156	433.0	100.00
3		1	214	1	214	2.82e-156	432.0	100.00
4		1	214	1	214	4.14e-156	432.0	100.00
5		1	214	1	214	1.10e-155	431.0	99.53
6		1	214	1	214	1.10e-155	431.0	99.53
7		1	214	1	214	1.19e-155	431.0	99.53
8		1	214	1	214	1.56e-155	430.0	99.53
9		1	214	1	214	2.22e-155	430.0	99.53
10		1	214	1	214	2.34e-155	430.0	99.53
11		1	214	1	214	7.03e-155	429.0	99.07
12		1	214	1	214	4.07e-154	427.0	99.07
13		1	214	1	214	6.81e-154	426.0	98.60
14		1	214	1	214	2.02e-153	425.0	98.60
15		1	214	1	214	4.12e-153	424.0	98.60
16		1	213	1	213	1.83e-115	329.0	84.98
17		2	213	5	216	1.01e-113	325.0	84.43
18		2	211	10	225	7.93e-90	265.0	71.30
19		2	209	26	235	1.89e-86	256.0	74.41
20		1	214	1	213	2.80e-76	230.0	68.81
21		1	214	1	213	5.36e-73	221.0	69.27
22		1	214	1	213	1.10e-70	216.0	68.35
23		1	213	1	212	2.01e-68	210.0	65.60
24		1	184	17	204	1.01e-67	209.0	69.68
25		1	213	1	212	1.04e-67	208.0	65.14

26	1	184	17	204	2.61e-67	207.0	70.21
27	2	213	7	217	7.02e-67	206.0	68.06
28	1	214	5	218	9.92e-67	206.0	65.75
29	1	214	5	218	1.21e-66	205.0	65.30
30	1	214	1	213	4.43e-66	204.0	65.14
31	1	214	1	213	2.15e-64	199.0	65.60
32	1	214	5	218	4.63e-64	199.0	64.84
33	1	214	1	213	2.49e-63	197.0	66.67
34	1	214	1	213	1.10e-62	195.0	66.97
35	1	214	1	213	1.22e-62	195.0	65.60
36	1	214	1	213	1.85e-62	194.0	66.97
37	1	214	1	213	1.93e-62	194.0	66.97
38	1	214	1	213	2.62e-62	194.0	66.97
39	1	214	1	213	4.18e-62	194.0	66.21
40	1	214	1	213	6.55e-62	193.0	66.97
41	1	214	1	213	1.24e-61	192.0	66.51
42	1	213	1	212	1.34e-61	192.0	66.82
43	1	214	1	213	1.56e-61	192.0	66.51
44	1	214	1	213	2.67e-61	192.0	66.51
45	1	214	1	213	2.97e-61	191.0	66.51
46	2	211	31	235	1.88e-60	190.0	64.95
47	1	213	3	214	1.92e-60	189.0	65.60
48	1	214	1	213	3.84e-60	189.0	64.68
49	1	213	1	212	4.72e-60	188.0	65.60
50	1	213	1	212	8.26e-59	185.0	65.60
51	1	213	1	212	1.55e-58	184.0	65.14
52	1	214	1	203	7.13e-56	177.0	66.51
53	1	214	1	203	8.17e-56	177.0	65.60
54	1	214	1	203	2.44e-55	176.0	66.51
55	1	214	6	209	4.91e-54	173.0	63.72
56	1	185	8	190	3.82e-50	164.0	64.55
57	1	185	7	189	5.25e-50	163.0	65.41
58	1	213	6	213	1.87e-46	154.0	62.39
59	1	182	28	207	6.61e-45	150.0	64.13
60	1	182	6	185	5.52e-44	148.0	63.59
61	1	182	1	155	9.17e-39	133.0	56.99
62	1	182	21	175	1.65e-38	133.0	56.99
63	1	209	1	218	3.53e-31	115.0	54.26
64	3	205	13	184	4.30e-31	114.0	52.45
65	1	209	1	218	4.66e-31	114.0	54.26
66	3	205	11	182	5.17e-31	113.0	52.45
67	3	205	11	182	6.07e-31	113.0	52.45
68	3	205	13	184	7.20e-31	113.0	52.45
69	3	213	13	192	8.50e-31	113.0	52.83
70	3	213	11	190	8.65e-31	113.0	51.89
71	3	205	11	182	9.33e-31	113.0	52.45
72	3	182	11	164	2.47e-30	112.0	54.70
73	3	184	12	167	2.96e-30	111.0	54.10
74	3	185	32	188	1.57e-29	110.0	56.45
75	3	184	12	167	1.84e-29	109.0	54.64
76	3	184	16	171	2.18e-29	109.0	54.64

77	3	214	7	185	8.86e-29	107.0	50.94
78	3	184	11	166	1.17e-28	107.0	56.28
79	1	168	1239	1418	1.52e-25	105.0	53.33
80	3	210	18	196	9.58e-25	97.8	53.55
81	3	213	6	192	6.79e-24	95.5	49.77
82	3	213	24	210	1.05e-23	95.5	49.77
83	3	187	15	173	7.01e-22	90.1	49.47
84	3	213	9	189	2.73e-20	85.9	47.44
85	76	207	760	922	1.27e-05	47.0	42.01

\$url

JX1895U7016

```
"https://blast.ncbi.nlm.nih.gov/Blast.cgi?
CMD=Get&FORMAT_OBJECT=Alignment&ALIGNMENT_VIEW=Tabular&RESULTS_FILE=on&FORMAT_TYPE=CSV&ALIGNMENTS=20000&DESCRIPTIONS=20000&RID=JX1895U7016"
```

```
attr(),"class")
[1] "blast"
```

attributes(b)

```
$names
[1] "hit.tbl" "raw"      "url"
```

```
$class
[1] "blast"
```

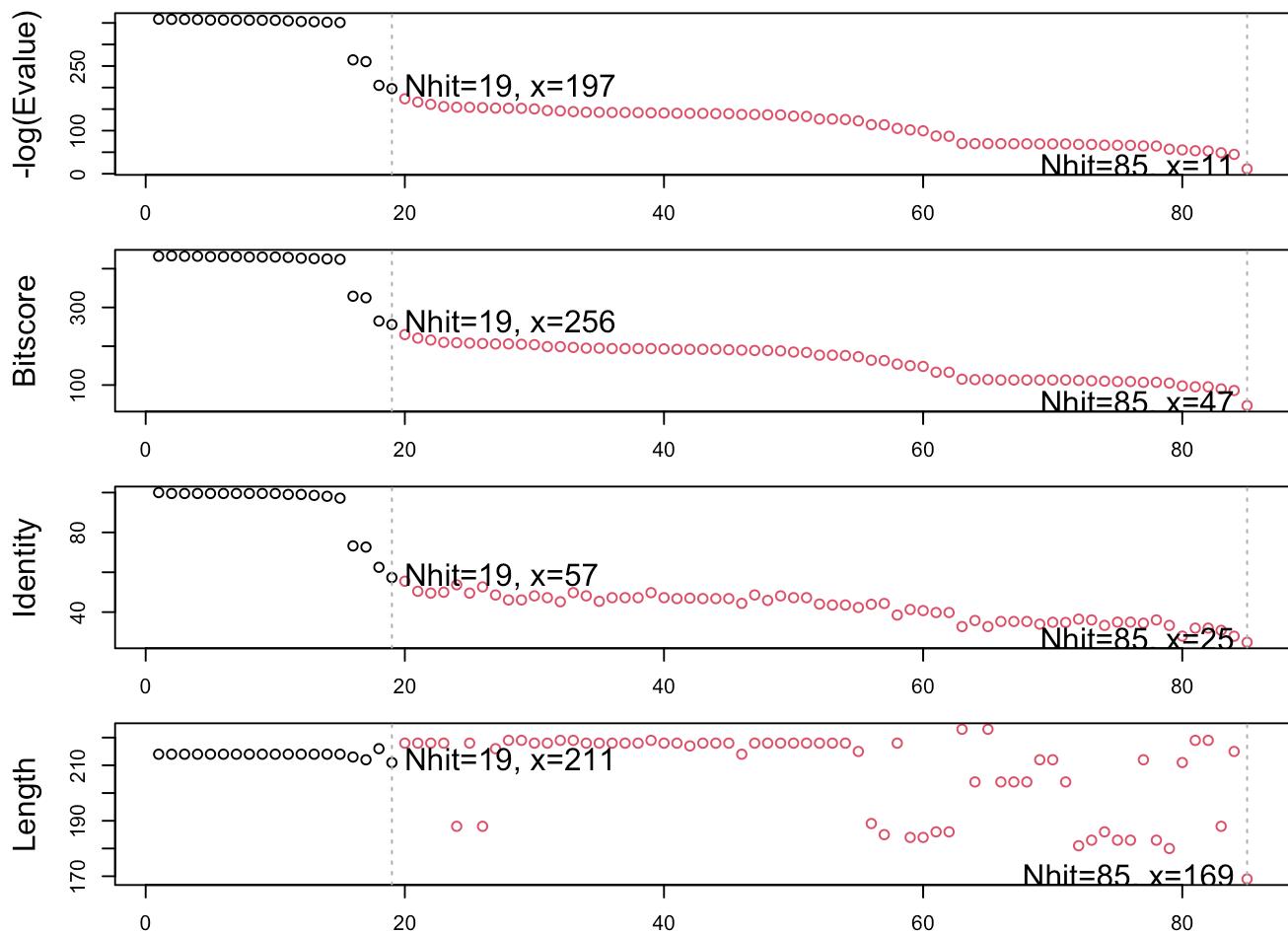
head(b\$hit.tbl)

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start		
1	Query_3264839	1AKE_A	100.000	214	0	0	1		
2	Query_3264839	8BQF_A	99.533	214	1	0	1		
3	Query_3264839	4X8M_A	99.533	214	1	0	1		
4	Query_3264839	6S36_A	99.533	214	1	0	1		
5	Query_3264839	8Q2B_A	99.533	214	1	0	1		
6	Query_3264839	8RJ9_A	99.533	214	1	0	1		
	q.end	s.start	s.end	evalue	bitscore	positives	mlog.evalue	pdb.id	acc
1	214	1	214	1.58e-156	432	100.00	358.7458	1AKE_A	1AKE_A
2	214	21	234	2.58e-156	433	100.00	358.2555	8BQF_A	8BQF_A
3	214	1	214	2.82e-156	432	100.00	358.1665	4X8M_A	4X8M_A
4	214	1	214	4.14e-156	432	100.00	357.7826	6S36_A	6S36_A
5	214	1	214	1.10e-155	431	99.53	356.8054	8Q2B_A	8Q2B_A
6	214	1	214	1.10e-155	431	99.53	356.8054	8RJ9_A	8RJ9_A

hits <- plot(b) #it takes the blast results and shows a quick and dirty overview.

```
* Possible cutoff values:    197 11
Yielding Nhits:    19 85
```

```
* Chosen cutoff value of: 197
Yielding Nhits: 19
```



```
attributes(hits) #you use attributes if you want an idea what is in the dataset.
```

```
$names
[1] "hits"    "pdb.id"   "acc"      "inds"
```

```
$class
[1] "blast"
```

Top hits that we like from our blast results:

```
hits$pdb.id
```

```
[1] "1AKE_A" "8BQF_A" "4X8M_A" "6S36_A" "8Q2B_A" "8RJ9_A" "6RZE_A" "4X8H_A"
[9] "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A" "6HAP_A" "6HAM_A" "4K46_A"
[17] "4NP6_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/8BQF.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/4X8M.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/8Q2B.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/8RJ9.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/4X8H.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

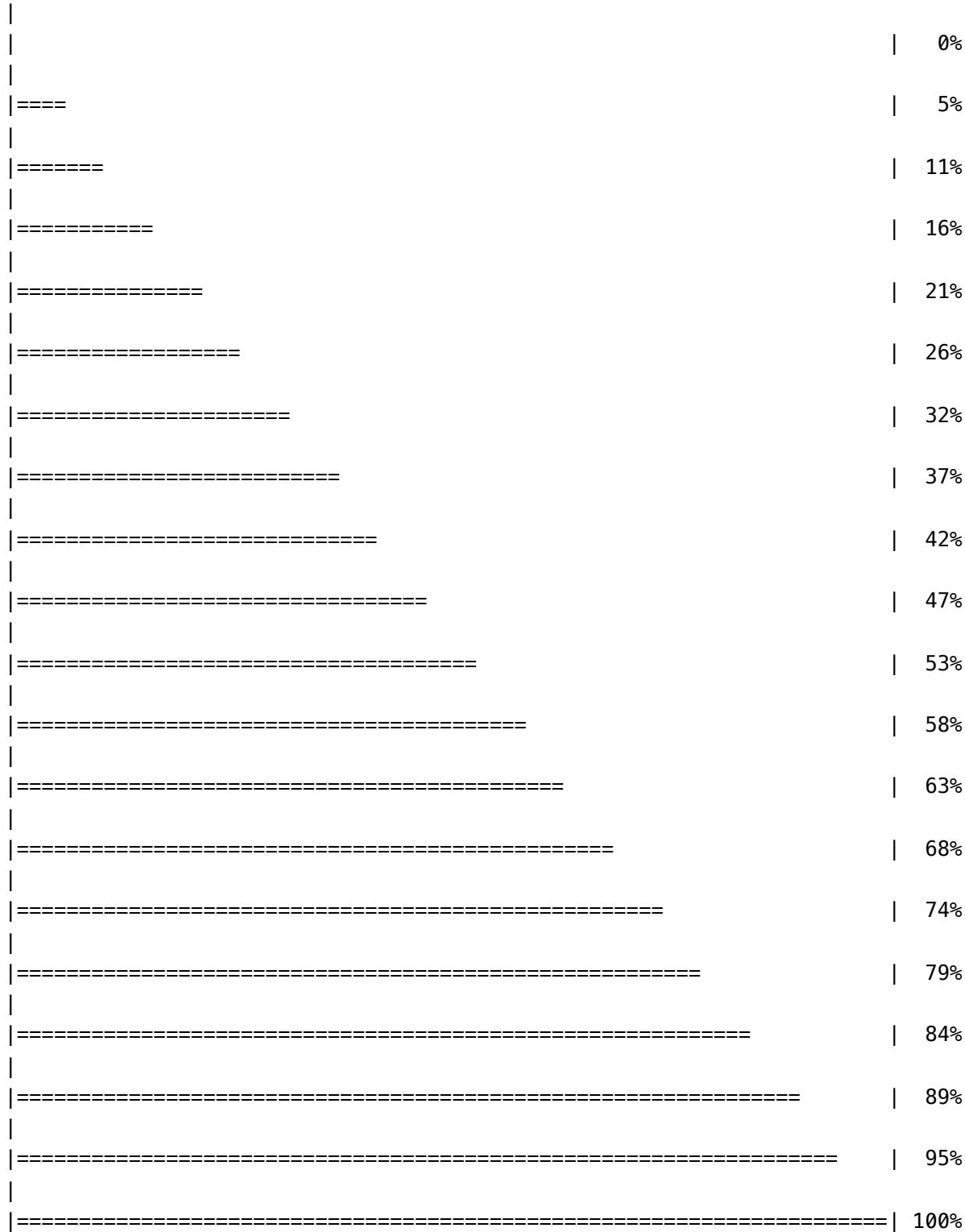
Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6HAM.pdb.gz exists. Skipping download

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/4NP6.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
```



I have now found and downloaded all ADK structures in the PDB database but viewing them is difficult as they need to be aligned and superposed.

I am going to install BiocManager package from CRAN. Then I can use 'BiocManager::install()' to install any bioconductor package.

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

Reading PDB files:

```
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/8BQF_A.pdb
pdbs/split_chain/4X8M_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/8Q2B_A.pdb
pdbs/split_chain/8RJ9_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/4X8H_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/4NP6_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
.. 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/8BQF_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 3 name: pdbs/split_chain/4X8M_A.pdb
pdb/seq: 4 name: pdbs/split_chain/6S36_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 5 name: pdbs/split_chain/8Q2B_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 6 name: pdbs/split_chain/8RJ9_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
```

pdb/seq: 7 name: pdbs/split_chain/6RZE_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 8 name: pdbs/split_chain/4X8H_A.pdb
pdb/seq: 9 name: pdbs/split_chain/3HPR_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 10 name: pdbs/split_chain/1E4V_A.pdb
pdb/seq: 11 name: pdbs/split_chain/5EJE_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 12 name: pdbs/split_chain/1E4Y_A.pdb
pdb/seq: 13 name: pdbs/split_chain/3X2S_A.pdb
pdb/seq: 14 name: pdbs/split_chain/6HAP_A.pdb
pdb/seq: 15 name: pdbs/split_chain/6HAM_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 16 name: pdbs/split_chain/4K46_A.pdb
PDB has ALT records, taking A only, rm.alt=TRUE
pdb/seq: 17 name: pdbs/split_chain/4NP6_A.pdb
pdb/seq: 18 name: pdbs/split_chain/3GMT_A.pdb
pdb/seq: 19 name: pdbs/split_chain/4PZL_A.pdb

pdbs

[Truncated_Name:1]1AKE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:2]8BQF_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:3]4X8M_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:4]6S36_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:5]8Q2B_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:6]8RJ9_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:7]6RZE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:8]4X8H_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:9]3HPR_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:10]1E4V_A.pdb	-----	MRIILLGAPVAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:11]5EJE_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:12]1E4Y_A.pdb	-----	MRIILLGALVAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:13]3X2S_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:14]6HAP_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:15]6HAM_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMEKYGIPQIS					
[Truncated_Name:16]4K46_A.pdb	-----	MRIILLGAPGAGKGTQAQFIMAKFGIPQIS					
[Truncated_Name:17]4NP6_A.pdb	-----	NAMRIILLGAPGAGKGTQAQFIMEKFGIPQIS					
[Truncated_Name:18]3GMT_A.pdb	-----	MRLILLLGAPGAGKGTQANFIKEKFHIPQIS					
[Truncated_Name:19]4PZL_A.pdb	TENLYFQSNAMEIILLGAPGAGKGTQAKIIEQKYNIAHIS						
	^**	*****	*	*^*	**		
	1	40

	41	80
[Truncated_Name:1]1AKE_A.pdb	TGDM	LRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:2]8BQF_A.pdb	TGDM	LRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:3]4X8M_A.pdb	TGDM	LRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated_Name:4]6S36_A.pdb	TGDM	LRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
[Truncated Name:5]802B_A.pdb	TGDM	LRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE

[Truncated_Name:6]8RJ9_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:7]6RZE_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:8]4X8H_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:9]3HPR_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:10]1E4V_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:11]5EJE_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDACKLVTDELVIALVKE
 [Truncated_Name:12]1E4Y_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVKE
 [Truncated_Name:13]3X2S_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDCGKLVTDELVIALVKE
 [Truncated_Name:14]6HAP_A.pdb TGDMRLRAAVKSGSELGKQAKDIMDAGKLVTDELVIALVRE
 [Truncated_Name:15]6HAM_A.pdb TGDMRLRAAIKGSSSELGKQAKDIMDAGKLVTDEIIIALVKE
 [Truncated_Name:16]4K46_A.pdb TGDMRLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGLVKE
 [Truncated_Name:17]4NP6_A.pdb TGDMRLRAAIKAGTELGKQAKAVIDAGQLVSDDIILGLIKE
 [Truncated_Name:18]3GMT_A.pdb TGDMRLRAAVKAGTPLGVVEAKTYMDEGKLVPDSLIIIGLVKE
 [Truncated_Name:19]4PZL_A.pdb TGDMIRETICKSGSALGQELKKVLDAGELVSDEFIIKIVKD
 ****^* ^* *^ ** * ^* ** * ^ ^ ^ ^
 41 80

81 120
 [Truncated_Name:1]1AKE_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:2]8BQF_A.pdb RIAQE---GFLLDGFPRTIPQADAMKEAGINVDYVIEFD
 [Truncated_Name:3]4X8M_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:4]6S36_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:5]8Q2B_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:6]8RJ9_A.pdb RIAQEDCRNGFLLAGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:7]6RZE_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:8]4X8H_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:9]3HPR_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:10]1E4V_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:11]5EJE_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:12]1E4Y_A.pdb RIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:13]3X2S_A.pdb RIAQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:14]6HAP_A.pdb RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:15]6HAM_A.pdb RICQEDSRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFD
 [Truncated_Name:16]4K46_A.pdb RIAQDDCAKGFLDGFPRTIPQADGLKEVGVVVDYVIEFD
 [Truncated_Name:17]4NP6_A.pdb RIAQADCEKGFLDGFPRTIPQADGLKEMGINVDYVIEFD
 [Truncated_Name:18]3GMT_A.pdb RLKEADCANGYLFDGFPRTIAQADAMKEAGVAIDYVLEID
 [Truncated_Name:19]4PZL_A.pdb RISKNDCNNGFLLDGVPRTIPQAQELDKLGVNIDYIVEVD
 *^ *^* * **** ** ^ *^ ^*^*^* *
 81 120

121 160
 [Truncated_Name:1]1AKE_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:2]8BQF_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:3]4X8M_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:4]6S36_A.pdb VPDELIVDKIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:5]8Q2B_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:6]8RJ9_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:7]6RZE_A.pdb VPDELIVDAIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:8]4X8H_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
 [Truncated_Name:9]3HPR_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDGTG
 [Truncated_Name:10]1E4V_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG

[Truncated_Name:11]5EJE_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:12]1E4Y_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:13]3X2S_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:14]6HAP_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:15]6HAM_A.pdb	VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG
[Truncated_Name:16]4K46_A.pdb	VADSVIVERMAGRRAHLASGRTYHNVNPPKVEGKDDVTG
[Truncated_Name:17]4NP6_A.pdb	VADDVIVERMAGRRAHLSGRTYHVVNPPKVEGKDDVTG
[Truncated_Name:18]3GMT_A.pdb	VPFSEIIERMSRRTHPASGRTYHVKFNPPKVEGKDDVTG
[Truncated_Name:19]4PZL_A.pdb	VADNLILLERITGRRIHPASGRTYHTKFNPPKVADKDDVTG
	* ^^^ ^ *** * *** ** ^***** *** **
121	160
 161	200
[Truncated_Name:1]1AKE_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:2]8BQF_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:3]4X8M_A.pdb	EELTRRKDDQEETVRKRLVEWHQMTAPLIGYYSKAEAGN
[Truncated_Name:4]6S36_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:5]8Q2B_A.pdb	EELTRRKADQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:6]8RJ9_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:7]6RZE_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:8]4X8H_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAALIGYYSKAEAGN
[Truncated_Name:9]3HPR_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:10]1E4V_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:11]5EJE_A.pdb	EELTRRKDDQEECVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:12]1E4Y_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:13]3X2S_A.pdb	EELTRRKDDQEETVRKRLCEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:14]6HAP_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:15]6HAM_A.pdb	EELTRRKDDQEETVRKRLVEYHQMTAPLIGYYSKAEAGN
[Truncated_Name:16]4K46_A.pdb	EDLVIREDDKEETVLARLGVYHNQTAPIIAYYGKEAEAGN
[Truncated_Name:17]4NP6_A.pdb	EDLVIREDDKEETVRARLNVYHTQTAPIIEYYGKEAAAGK
[Truncated_Name:18]3GMT_A.pdb	EPLVQRDDDKEETVKKRLDVYEAQTKPLITYGDWARRGA
[Truncated_Name:19]4PZL_A.pdb	EPLITRTDDNEDTVQQLRSVYHAQTAKLIDFYRNFSSTNT
	* * * * *^ * ** ^ * * ** ^*
161	200
 201	227
[Truncated_Name:1]1AKE_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:2]8BQF_A.pdb	T--KYAKVDGTPVAEVRADLEKIL--
[Truncated_Name:3]4X8M_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:4]6S36_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:5]8Q2B_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:6]8RJ9_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:7]6RZE_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:8]4X8H_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:9]3HPR_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:10]1E4V_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:11]5EJE_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:12]1E4Y_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:13]3X2S_A.pdb	T--KYAKVDGTPVAEVRADLEKILG-
[Truncated_Name:14]6HAP_A.pdb	T--KYAKVDGTPVCEVRADLEKILG-
[Truncated_Name:15]6HAM_A.pdb	T--KYAKVDGTPVCEVRADLEKILG-

```
[Truncated_Name:16]4K46_A.pdb T--QYLKFDTKAVAEVSAELEKALA-
[Truncated_Name:17]4NP6_A.pdb T--QYLKFDTKQVSEVSADIAKALA-
[Truncated_Name:18]3GMT_A.pdb E-----NGLKAPA-----YRKISG-
[Truncated_Name:19]4PZL_A.pdb KIPKYIKINGDQAVEKVSQDIFDQLNK
```

*
201 . . 227

Call:

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

Class:

```
pdbs, fasta
```

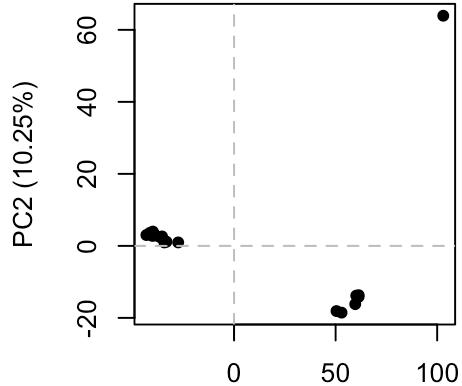
Alignment dimensions:

```
19 sequence rows; 227 position columns (199 non-gap, 28 gap)
```

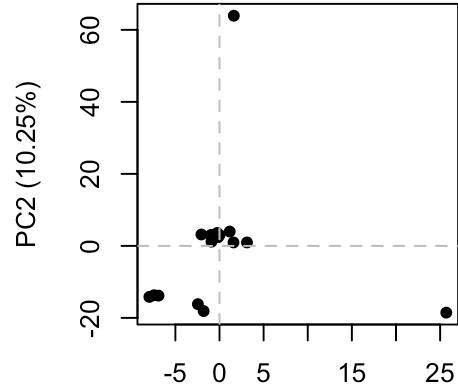
```
+ attr: xyz, resno, b, chain, id, ali, resid, sse, call
```

Principal Component Analysis

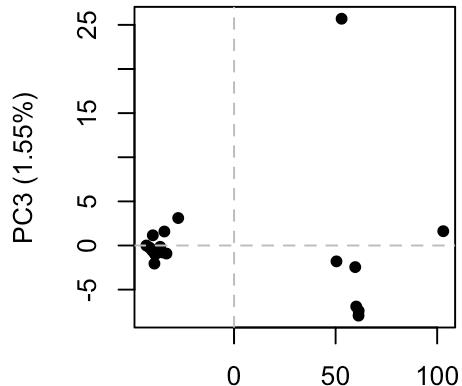
```
pc <- pca(pdbs)
plot(pc)
```



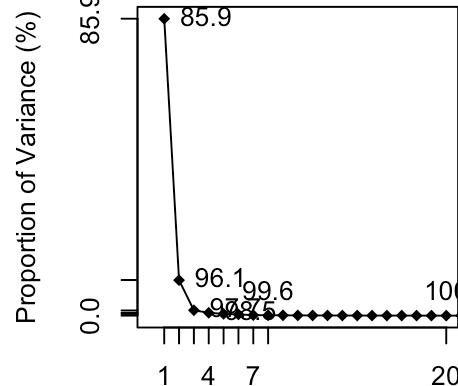
PC1 (85.89%)



PC3 (1.55%)

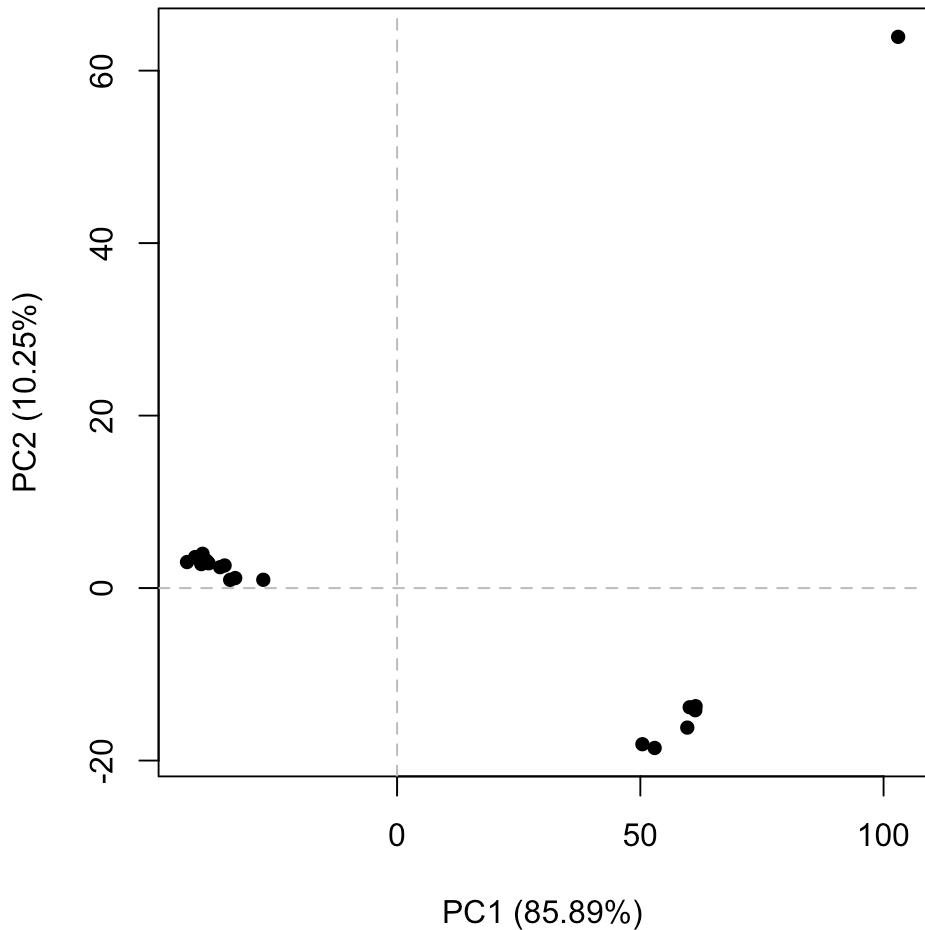


PC1 (85.89%)



Eigenvalue Rank

```
plot(pc, pc.axes = c(1:2))
```



To examine in more detail what PC1 (or any PC) is capturing here we can plot the loadings or make a wee movie (trajectory) of moving along PC1.

```
mktrj(pc, pc=1, file="pc1.pdb")
```

Q13 - How many amino acids are in this sequence, i.e. how long is this sequence?

```
# 214
```

Q14 - What do you note about this plot? Are the black and colored lines similar or different? Where do you think they differ most and why?

```
# not needed, no time in class
```

