W O R L D W I D E

wwPDB

P R O T E I N D A T A B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Oct 22, 2023 – 08:11 PM EDT

PDB ID	:	3BWE
Title	:	Crystal structure of aggregated form of DJ1
Authors	:	Cha, S.S.
Deposited on	:	2008-01-09
Resolution	:	2.40 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

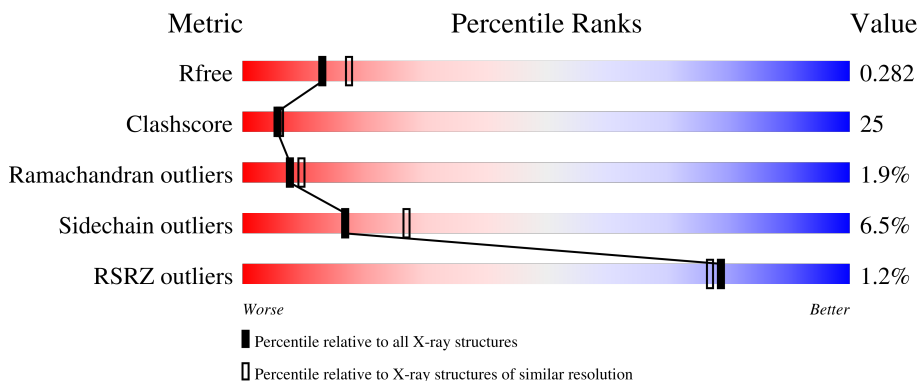|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.
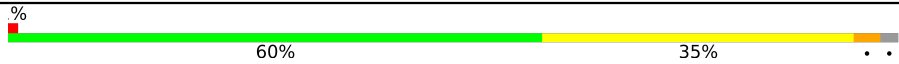
Metric	Percentile Ranks	Value
Rfree		0.282
Clashscore		25
Ramachandran outliers		1.9%
Sidechain outliers		6.5%
RSRZ outliers		1.2%

Worse — *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	189	% 62% 33% · ·
1	B	189	62% 33% · ·
1	C	189	% 56% 37% 7% ·
1	D	189	59% 34% 5% ·
1	E	189	3% 46% 48% 6% ·

Continued on next page...

Continued from previous page...

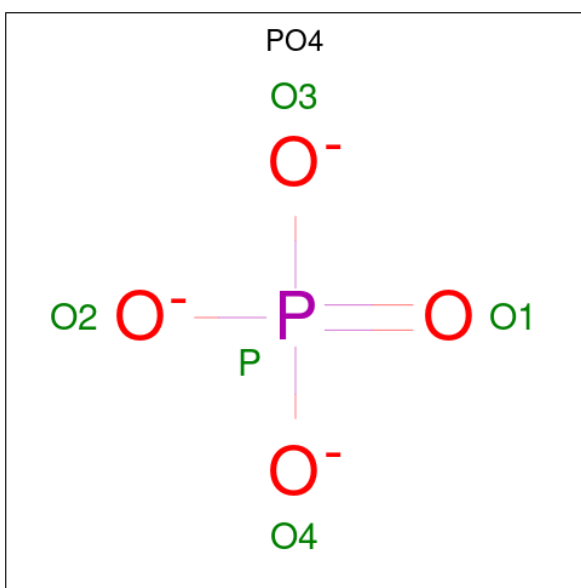| Mol	Chain	Length	Quality of chain
1	F	189	60% 35%
1	G	189	49% 45%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9808 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Protein DJ-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	188	Total	C	N	O	S	Se	0	0	0
			1368	859	238	264	3	4			
1	B	187	Total	C	N	O	S	Se	0	0	0
			1363	857	238	261	3	4			
1	C	188	Total	C	N	O	S	Se	0	0	0
			1368	859	238	264	3	4			
1	D	186	Total	C	N	O	S	Se	0	0	0
			1358	854	237	260	3	4			
1	E	188	Total	C	N	O	S	Se	0	0	0
			1351	849	238	257	3	4			
1	F	186	Total	C	N	O	S	Se	0	0	0
			1350	849	236	258	3	4			
1	G	186	Total	C	N	O	S	Se	0	0	0
			1332	838	231	256	3	4			

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).

PO4

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	P 1	0	0
2	A	1	Total 5	O 4	P 1	0	0
2	C	1	Total 5	O 4	P 1	0	0
2	D	1	Total 5	O 4	P 1	0	0
2	E	1	Total 5	O 4	P 1	0	0
2	F	1	Total 5	O 4	P 1	0	0
2	F	1	Total 5	O 4	P 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	37	Total 37	O 37	0	0
3	B	60	Total 60	O 60	0	0
3	C	56	Total 56	O 56	0	0
3	D	54	Total 54	O 54	0	0
3	E	42	Total 42	O 42	0	0
3	F	22	Total 22	O 22	0	0
3	G	12	Total 12	O 12	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Protein DJ-1

- Molecule 1: Protein DJ-1

- Molecule 1: Protein DJ-1

- Molecule 1: Protein DJ-1

MSE ALA S603 Q604 L607 V608 K612 G613 A614 E615 E616 M617 E618 T619 V620 I621 P622 M626 I631 K632 P643 V644 Q645 C646 V650 C653 P654 L658 K662 K663 E664 G665 P666 Y667 V671 L672 P673 G674 L677 G678 N681 L682 S683 E684 S685 A686 A687 V688
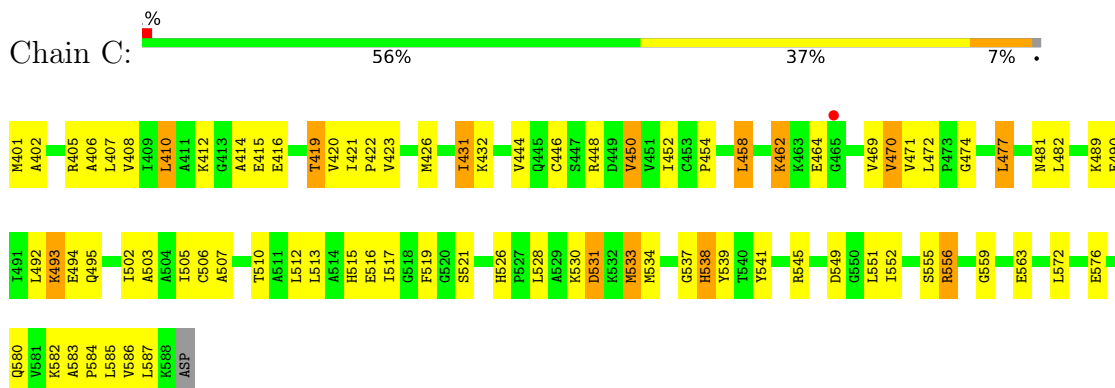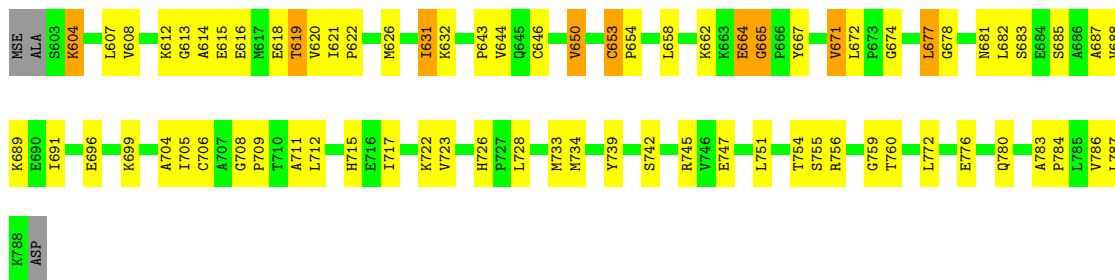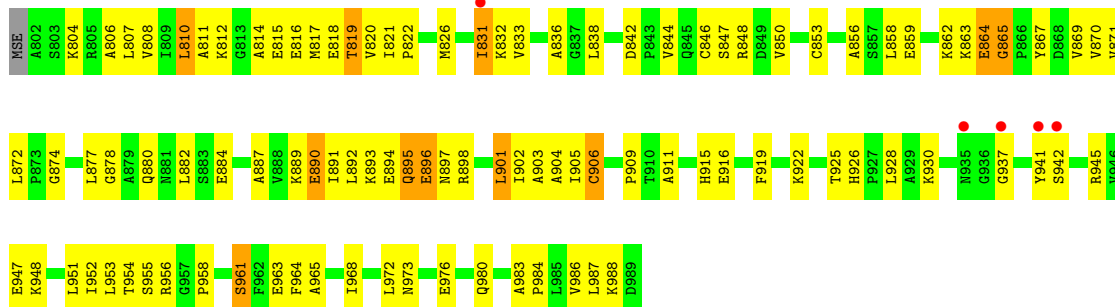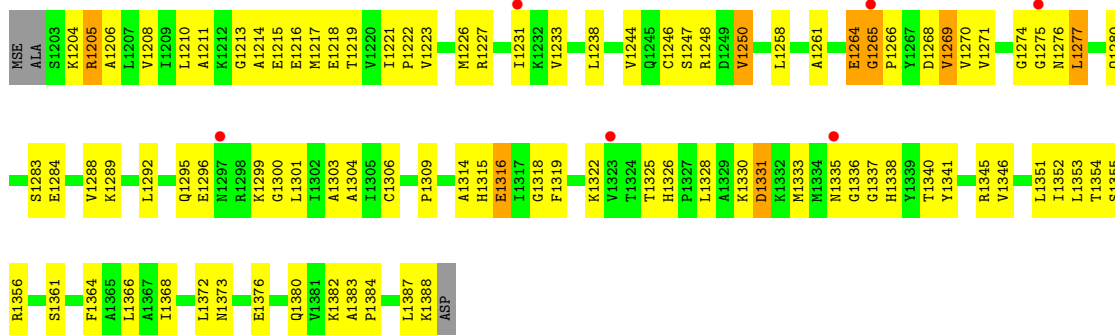
K689 E690 I691 E696 K699 A704 I705 C706 A707 G708 P709 T710 A711 L712 H715 E716 I717 K722 V723 H726 P727 L728 M733 M734 Y739 S742 R745 V746 E747 L751 T754 S755 R756 G759 T760 L772 E776 Q780 A783 P784 L785 V786 L787

K788 ASP

• Molecule 1: Protein DJ-1

Chain E:

MSE A802 S803 K804 R805 A806 L807 V808 I809 L810 A811 K812 G813 A814 E815 E816 M817 E818 T819 V820 I821 P822 M826 I831 K832 V833 A836 G837 L838 D842 P843 V844 Q845 C846 S847 R848 D849 V850 C853 A856 S857 L858 E859 K862 K863 E864 G865 P866 Y867 D868 V869 V870 V871

L872 P873 G874 L877 G878 A879 Q880 N881 L882 S883 E884 A887 V888 K889 E890 I891 L892 K893 E894 Q895 S896 N897 R898 L901 I902 A903 A904 I905 C906 P909 T910 A911 H915 E916 F919 K922 T925 H926 P927 L928 A929 K930 N935 G936 G937 Y941 S942 R945 V946

E947 K948 L951 I952 L953 T954 S955 R956 G957 P958 S961 F962 E963 F964 A965 I968 L972 N973 E976 Q980 A983 P984 L985 V986 L987 K988 D989

• Molecule 1: Protein DJ-1

Chain F:

MSE ALA S1003 K1004 L1007 V1008 I1009 L1010 G1013 A1014 E1015 E1016 T1019 V1020 I1021 P1022 V1023 D1024 V1025 R1028 A1036 V1044 R1048 D1049 V1050 I1051 I1052 L1058 E1059 K1062 G1065 P1066 Y1067 V1070 V1071 L1072 P1073 N1076 L1077 Q1080 E1084 K1089 K1093 E1094

Q1095 E1096 N1097 R1098 K1099 A1103 A1104 I1105 C1106 A1107 G1108 P1109 L1113 H1126 P1127 L1128 A1129 K1130 M1134 Y1141 R1145 K1148 D1149 G1150 L1151 L1152 L1153 T1154 S1155 T1160 S1161 F1162 E1163 F1164 A1165 L1172 N1173 E1176 V1177 Q1180 A1183 P1184 L1185 V1186 L1187 K1188 ASP

• Molecule 1: Protein DJ-1

Chain G:

MSE ALA S1203 K1204 R1205 A1206 L1207 V1208 I1209 L1210 A1211 K1212 G1213 A1214 E1215 E1216 M1217 E1218 T1219 V1220 I1221 P1222 V1223 M1226 R1227 I1231 K1232 V1233 L1238 V1244 Q1245 C1246 S1247 R1248 D1249 V1250 L1258 A1261 E1264 G1265 P1266 Y1267 D1268 V1269 V1270 V1271 G1274 G1275 N1276 L1277 Q1280

S1283 E1284 V1288 K1289 L1292 Q1295 E1296 N1297 R1298 K1299 G1300 L1301 I1302 A1303 I1304 I1305 C1306 P1309 A1314 H1315 E1316 I1317 G1318 F1319 K1322 V1323 T1324 T1325 H1326 P1327 L1328 A1329 K1330 D1331 K1332 M1333 M1334 N1335 G1336 G1337 H1338 Y1339 T1340 Y1341 R1345 V1346 L1351 I1352 L1353 T1354 S1355

R1356 S1361 F1364 A1365 L1366 A1367 I1368 L1372 N1373 E1376 Q1380 V1381 K1382 A1383 P1384 L1387 K1388 ASP

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	193.62Å 102.58Å 85.53Å 90.00° 113.69° 90.00°	Depositor
Resolution (Å)	19.99 – 2.40 53.63 – 2.40	Depositor EDS
% Data completeness (in resolution range)	94.5 (19.99-2.40) 94.6 (53.63-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.05 (at 2.39Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.232 , 0.282 0.232 , 0.282	Depositor DCC
R_{free} test set	5761 reflections (10.16%)	wwPDB-VP
Wilson B-factor (Å^2)	27.8	Xtriage
Anisotropy	0.396	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.0	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.023 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	9808	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.36% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.38	0/1381	0.65	0/1863
1	B	0.39	0/1376	0.65	0/1855
1	C	0.40	0/1381	0.67	1/1863 (0.1%)
1	D	0.40	0/1371	0.69	0/1848
1	E	0.38	0/1364	0.66	0/1840
1	F	0.38	0/1363	0.65	0/1839
1	G	0.39	0/1345	0.65	0/1819
All	All	0.39	0/9581	0.66	1/12927 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	556	ARG	N-CA-C	5.06	124.66	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1368	0	1403	70	0
1	B	1363	0	1405	66	0
1	C	1368	0	1399	80	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	1358	0	1399	64	0
1	E	1351	0	1377	83	0
1	F	1350	0	1385	52	0
1	G	1332	0	1348	97	0
2	A	10	0	0	0	0
2	C	5	0	0	0	0
2	D	5	0	0	0	0
2	E	5	0	0	0	0
2	F	10	0	0	0	0
3	A	37	0	0	1	0
3	B	60	0	0	4	0
3	C	56	0	0	2	0
3	D	54	0	0	1	0
3	E	42	0	0	2	0
3	F	22	0	0	1	0
3	G	12	0	0	1	0
All	All	9808	0	9716	487	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 25.

The worst 5 of 487 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:113:LEU:HB2	1:A:133:MSE:HE1	1.23	1.15
1:B:203:SER:HB2	1:B:232:LYS:HE2	1.38	1.05
1:G:1205:ARG:HH11	1:G:1205:ARG:HB3	0.90	1.04
1:G:1205:ARG:HB3	1:G:1205:ARG:NH1	1.73	1.03
1:G:1205:ARG:HH11	1:G:1205:ARG:CB	1.75	0.98

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	186/189 (98%)	179 (96%)	3 (2%)	4 (2%)	6	7
1	B	185/189 (98%)	180 (97%)	3 (2%)	2 (1%)	14	20
1	C	186/189 (98%)	173 (93%)	10 (5%)	3 (2%)	9	13
1	D	184/189 (97%)	177 (96%)	6 (3%)	1 (0%)	29	41
1	E	186/189 (98%)	172 (92%)	9 (5%)	5 (3%)	5	5
1	F	184/189 (97%)	176 (96%)	6 (3%)	2 (1%)	14	20
1	G	184/189 (97%)	158 (86%)	19 (10%)	7 (4%)	3	2
All	All	1295/1323 (98%)	1215 (94%)	56 (4%)	24 (2%)	8	10

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	665	GLY
1	E	864	GLU
1	G	1265	GLY
1	A	2	ALA
1	B	265	GLY

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	142/142 (100%)	134 (94%)	8 (6%)	21	34
1	B	142/142 (100%)	133 (94%)	9 (6%)	18	28
1	C	142/142 (100%)	128 (90%)	14 (10%)	8	11
1	D	142/142 (100%)	133 (94%)	9 (6%)	18	28
1	E	137/142 (96%)	125 (91%)	12 (9%)	10	15
1	F	140/142 (99%)	133 (95%)	7 (5%)	24	40
1	G	136/142 (96%)	131 (96%)	5 (4%)	34	53

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	981/994 (99%)	917 (94%)	64 (6%)	17 27

5 of 64 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	1077	LEU
1	G	1205	ARG
1	C	458	LEU
1	C	450	VAL
1	G	1250	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	726	HIS
1	G	1326	HIS
1	E	926	HIS
1	G	1373	ASN
1	G	1245	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

7 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The

Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
2	PO4	A	190	-	4,4,4	2.22	3 (75%)	6,6,6	0.39	0
2	PO4	F	4	-	4,4,4	1.90	1 (25%)	6,6,6	0.44	0
2	PO4	C	3	-	4,4,4	2.09	3 (75%)	6,6,6	0.42	0
2	PO4	A	191	-	4,4,4	2.25	3 (75%)	6,6,6	0.47	0
2	PO4	F	5	-	4,4,4	1.87	3 (75%)	6,6,6	0.42	0
2	PO4	E	6	-	4,4,4	1.99	3 (75%)	6,6,6	0.39	0
2	PO4	D	1	-	4,4,4	1.95	2 (50%)	6,6,6	0.38	0

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	191	PO4	P-O3	-2.58	1.46	1.54
2	A	191	PO4	P-O4	-2.50	1.47	1.54
2	C	3	PO4	P-O2	-2.49	1.47	1.54
2	A	190	PO4	P-O3	-2.48	1.47	1.54
2	F	4	PO4	P-O4	-2.37	1.47	1.54

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7　Other polymers ⓘ

There are no such residues in this entry.

5.8　Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6   Fit of model and data   ⓘ

6.1   Protein, DNA and RNA chains   ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2		OWAB(Å²)	Q<0.9
1	A	183/189 (96%)	-0.03	1 (0%)	91 89	13, 24, 37, 46	0
1	B	183/189 (96%)	-0.05	0	100 100	13, 23, 34, 42	0
1	C	183/189 (96%)	0.02	1 (0%)	91 89	13, 25, 34, 44	0
1	D	182/189 (96%)	-0.01	0	100 100	14, 26, 36, 40	0
1	E	184/189 (97%)	0.24	5 (2%)	54 52	20, 34, 43, 47	0
1	F	182/189 (96%)	0.08	2 (1%)	80 79	19, 30, 38, 45	0
1	G	182/189 (96%)	0.21	6 (3%)	46 45	26, 36, 45, 50	0
All	All	1279/1323 (96%)	0.06	15 (1%)	79 77	13, 28, 41, 50	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	942	SER	4.3
1	E	935	ASN	3.2
1	F	1065	GLY	3.0
1	E	831	ILE	2.9
1	G	1275	GLY	2.9

6.2   Non-standard residues in protein, DNA, RNA chains   ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates   ⓘ

There are no monosaccharides in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	PO4	A	191	5/5	0.96	0.22	33,33,34,35	0
2	PO4	A	190	5/5	0.97	0.19	29,30,32,33	0
2	PO4	D	1	5/5	0.98	0.19	39,40,41,43	0
2	PO4	F	4	5/5	0.98	0.14	37,37,38,38	0
2	PO4	F	5	5/5	0.98	0.17	43,44,44,45	0
2	PO4	C	3	5/5	0.99	0.21	32,33,35,35	0
2	PO4	E	6	5/5	0.99	0.14	35,35,35,36	0

6.5 Other polymers ⓘ

There are no such residues in this entry.