W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 09:10 pm BST

PDB ID	:	2PJR
Title	:	HELICASE PRODUCT COMPLEX
Authors	:	Velankar, S.S.; Soultanas, P.; Dillingham, M.S.; Subramanya, H.S.; Wigley, D.B.
Deposited on	:	1999-03-12
Resolution	:	2.90 Å(reported)

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

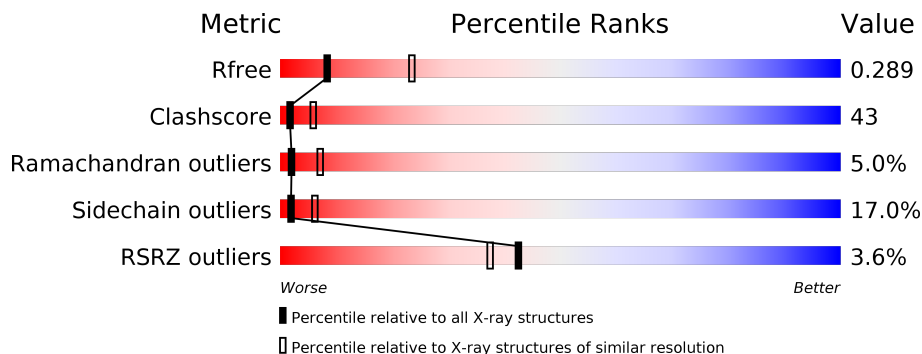|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.90 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.
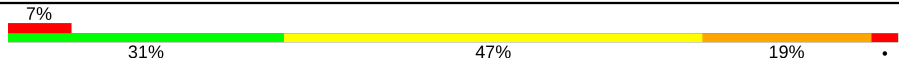
Metric	Percentile Ranks	Value
Rfree		0.289
Clashscore		43
Ramachandran outliers		5.0%
Sidechain outliers		17.0%
RSRZ outliers		3.6%

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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 on 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	C	5	100%
1	D	5	100%
2	H	2	100% / 50% / 50%
3	I	5	20% / 60% / 20%
4	A	548	2% / 40% / 50% / 7% ••
4	F	548	4% / 39% / 45% / 14% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	B	95	7% 31% 47% 19%
5	G	95	5% 31% 52% 16%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 10671 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 DNA chain called DNA (5'-D(*TP*TP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	5	Total	C	N	O	P	0	0	0
			97	50	10	33	4			
1	D	5	Total	C	N	O	P	0	0	0
			97	50	10	33	4			

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	2	Total	C	N	O	P	0	0	0
			38	19	8	10	1			

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*CP*TP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	5	Total	C	N	O	P	0	0	0
			98	48	18	28	4			

- Molecule 4 is a protein called PROTEIN (HELICASE PCRA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	542	Total	C	N	O	S	0	0	0
			4409	2792	775	829	13			
4	F	544	Total	C	N	O	S	0	0	0
			4424	2802	777	832	13			

- Molecule 5 is a protein called PROTEIN (HELICASE PCRA).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	B	95	Total	C	N	O	S	0	0	0
			749	471	125	147	6			
5	G	95	Total	C	N	O	S	0	0	0
			749	471	125	147	6			

• Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
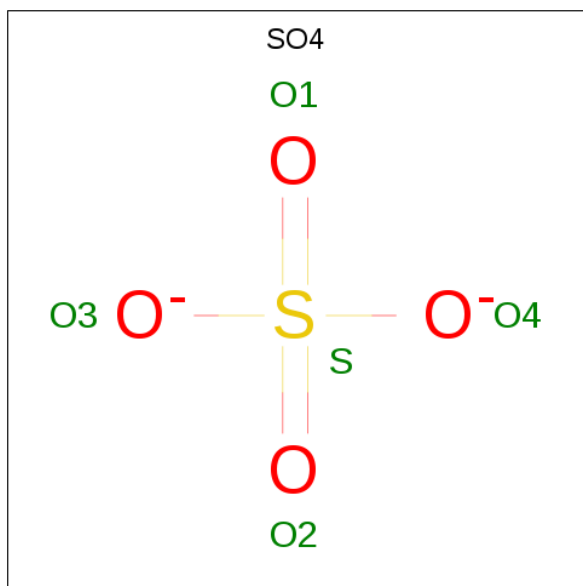
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total 5	O 4	S 1	0	0
6	F	1	Total 5	O 4	S 1	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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.

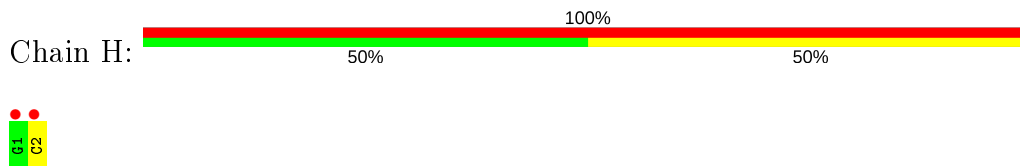- Molecule 1: DNA (5'-D(*TP*TP*TP*TP*T)-3')

Chain C:

- Molecule 1: DNA (5'-D(*TP*TP*TP*TP*T)-3')

Chain D:
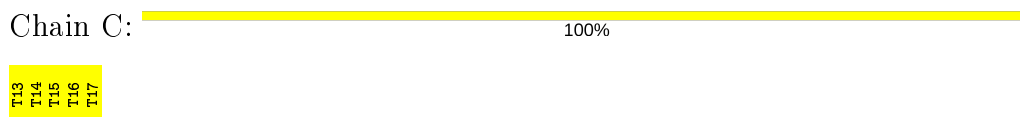
- Molecule 2: DNA (5'-D(*GP*C)-3')

Chain H:

- Molecule 3: DNA (5'-D(*AP*CP*TP*GP*C)-3')

Chain I:

- Molecule 4: PROTEIN (HELICASE PCRA)

Chain A:

[Residue color-coded validation chart for residues I148 through D548]

• Molecule 4: PROTEIN (HELICASE PCRA)

Chain F:

4%			
39%	45%	14%	

[Residue color-coded validation chart for Chain F, residues MET ASN PHE, L704 through F1215]

L1216 S1217 V1218 T1219 K1220 H1221 F1222 E1223 N1224 V1225 S1226 D1227 D1228 K1229 S1230 L1231 I1232 A1233 F1234 L1235 T1236 D1237 L1238 A1239 L1240 I1241 S1242 D1243 L1244 L1247 ASP
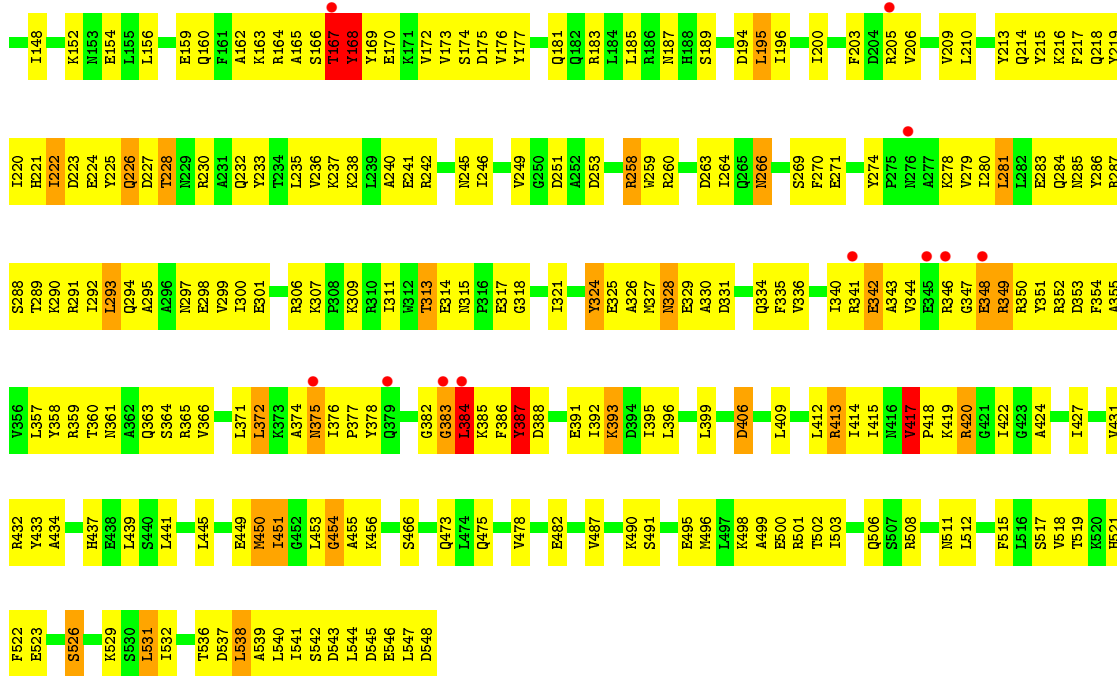
• Molecule 5: PROTEIN (HELICASE PCRA)

Chain B:

7%

31% 47% 19%

G556 D557 A558 V559 M560 L561 M562 T563 L564 H565 A566 A567 K568 G569 L570 E571 F572 P573 F576 L577 I578 G579 M580 E581 E582 G583 I584 F585 P586 H587 N588 R589 S590 L591 E592 D593 D594 D595 E596 M597 E598 E599 R600 R601 R602 L603 A604 Y605 V606 T609 R610 A611 E612 E613 V616 L617 T618

Q621 N622 R623 T624 L625 F626 G627 N628 I629 Q630 S635 R636 F637 L638 N639 E640 I641 P642 L645 T648 A649 S650

• Molecule 5: PROTEIN (HELICASE PCRA)

Chain G:

5%

31% 52% 16%

G1248 D1249 A1250 V1251 M1252 L1253 M1254 T1255 L1256 H1257 K1260 G1261 F1264 P1265 V1266 V1267 F1268 L1269 E1273 E1274 G1275 I1276 F1277 P1278 H1279 N1280 R1281 S1282 L1283 E1284 D1285 D1286 D1287 E1288 M1289 E1290 E1291 E1292 R1293 R1294 L1295 A1296 Y1297 V1298 R1302 A1303 E1304 E1305 E1306 L1307 V1308 L1309 T1310 Q1313 M1314

R1315 T1316 L1317 F1318 G1319 N1320 I1321 Q1322 M1323 D1324 P1325 P1326 S1327 R1328 F1329 L1330 N1331 E1332 I1333 P1334 A1335 H1336 L1337 L1338 E1339 T1340 A1341 S1342

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	85.05Å 62.60Å 141.83Å 90.00° 95.84° 90.00°	Depositor
Resolution (Å)	15.00 – 2.90 14.98 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.9 (15.00-2.90) 97.9 (14.98-2.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.59 (at 2.91Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.240 , 0.296 0.240 , 0.289	Depositor DCC
R_{free} test set	1644 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	45.9	Xtriage
Anisotropy	0.295	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 67.4	EDS
L-test for twinning [2]	$<|L|>$ = 0.49, $<L^2>$ = 0.32	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	10671	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.57% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	C	0.62	0/106	0.72	0/162
1	D	0.67	0/106	1.13	1/162 (0.6%)
2	H	1.79	2/42 (4.8%)	1.56	1/63 (1.6%)
3	I	2.09	4/109 (3.7%)	1.79	2/166 (1.2%)
4	A	0.52	0/4485	0.77	3/6059 (0.0%)
4	F	0.49	0/4500	0.78	8/6079 (0.1%)
5	B	0.63	1/762 (0.1%)	0.83	1/1028 (0.1%)
5	G	0.56	0/762	0.82	1/1028 (0.1%)
All	All	0.57	7/10872 (0.1%)	0.81	17/14747 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
3	I	0	1
4	A	0	1
4	F	0	1
All	All	0	3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	I	32	DT	O3'-P	-10.64	1.48	1.61
5	B	650	SER	CA-CB	7.21	1.63	1.52
3	I	33	DG	O3'-P	5.79	1.68	1.61
3	I	34	DC	C4'-C3'	5.35	1.58	1.53
3	I	34	DC	C3'-O3'	5.32	1.50	1.44

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	17	DT	O4'-C4'-C3'	-7.29	101.58	104.50
3	I	30	DA	O5'-C5'-C4'	7.16	128.89	111.00
3	I	33	DG	OP2-P-O3'	6.65	119.84	105.20
4	F	1117	VAL	C-N-CD	6.57	142.21	128.40
4	A	417	VAL	C-N-CD	6.39	141.82	128.40

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	A	387	TYR	Sidechain
4	F	1193	TYR	Sidechain
3	I	34	DC	Sidechain

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	C	97	0	62	10	0
1	D	97	0	62	9	0
2	H	38	0	24	0	0
3	I	98	0	58	11	0
4	A	4409	0	4432	355	0
4	F	4424	0	4450	447	0
5	B	749	0	731	104	0
5	G	749	0	731	89	0
6	A	5	0	0	1	0
6	F	5	0	0	0	0
All	All	10671	0	10550	919	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 43.

The worst 5 of 919 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)
4:F:1096:LEU:HD21	4:F:1238:LEU:HD23	1.23	1.15

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:G:1316:THR:HG23	5:G:1320:ASN:HA	1.24	1.14
4:A:326:ALA:O	5:B:621:GLN:HG2	1.55	1.06
4:A:327:MET:HA	5:B:621:GLN:HG3	1.33	1.05
4:A:396:LEU:HD11	4:A:538:LEU:HD23	1.38	1.04

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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	
4	A	540/548 (98%)	456 (84%)	60 (11%)	24 (4%)	2	10
4	F	542/548 (99%)	448 (83%)	63 (12%)	31 (6%)	1	5
5	B	93/95 (98%)	80 (86%)	8 (9%)	5 (5%)	2	6
5	G	93/95 (98%)	81 (87%)	8 (9%)	4 (4%)	2	10
All	All	1268/1286 (99%)	1065 (84%)	139 (11%)	64 (5%)	2	7

5 of 64 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	134	ILE
4	A	139	PHE
4	A	167	THR
4	A	168	TYR
4	A	348	GLU

5.3.2 Protein sidechains (i)

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	
4	A	472/478 (99%)	406 (86%)	66 (14%)	3	10
4	F	474/478 (99%)	390 (82%)	84 (18%)	2	5
5	B	81/81 (100%)	62 (76%)	19 (24%)	1	2
5	G	81/81 (100%)	62 (76%)	19 (24%)	1	2
All	All	1108/1118 (99%)	920 (83%)	188 (17%)	2	6

5 of 188 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	F	705	SER
4	F	839	PHE
5	G	1277	PHE
4	F	713	ASN
4	F	778	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
4	F	878	GLN
4	F	926	GLN
5	G	1257	HIS
4	F	882	GLN
4	F	888	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6   Ligand geometry (i)

2 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	$\#|Z| > 2$
6	SO4	F	1249	-	4,4,4	0.52	0	6,6,6	0.78	0
6	SO4	A	901	-	4,4,4	0.60	0	6,6,6	0.50	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	901	SO4	1	0

5.7   Other polymers (i)

There are no such residues in this entry.

5.8   Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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(Å^2)	Q<0.9
1	C	5/5 (100%)	-0.52	0	100	100	22, 25, 37, 37	0
1	D	5/5 (100%)	-0.32	0	100	100	29, 32, 41, 55	0
2	H	2/2 (100%)	2.97	2 (100%)	0	0	53, 53, 53, 64	0
3	I	5/5 (100%)	0.63	0	100	100	39, 39, 43, 47	0
4	A	542/548 (98%)	-0.22	12 (2%)	62	59	7, 31, 68, 100	0
4	F	544/548 (99%)	-0.00	20 (3%)	41	37	16, 42, 74, 101	0
5	B	95/95 (100%)	0.08	7 (7%)	14	11	11, 39, 79, 91	0
5	G	95/95 (100%)	0.14	5 (5%)	26	22	19, 41, 79, 91	0
All	All	1293/1303 (99%)	-0.07	46 (3%)	42	37	7, 37, 74, 101	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	345	GLU	9.7
5	G	1318	PHE	5.7
5	B	557	ASP	5.4
5	B	650	SER	5.1
4	F	836	PRO	4.4

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

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

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
6	SO4	A	901	5/5	0.96	0.13	44,49,52,52	0
6	SO4	F	1249	5/5	0.97	0.10	48,49,53,54	0

6.5 Other polymers (i)

There are no such residues in this entry.