W O R L D W I D E

PDB
www

PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jan 8, 2018 – 09:39 PM EST

PDB ID	:	6BMV
Title	:	Non-receptor Protein Tyrosine Phosphatase SHP2 in Complex with Allosteric Inhibitor SHP504
Authors	:	Stams, T.; Fodor, M.
Deposited on	:	2017-11-15
Resolution	:	2.05 Å(reported)

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

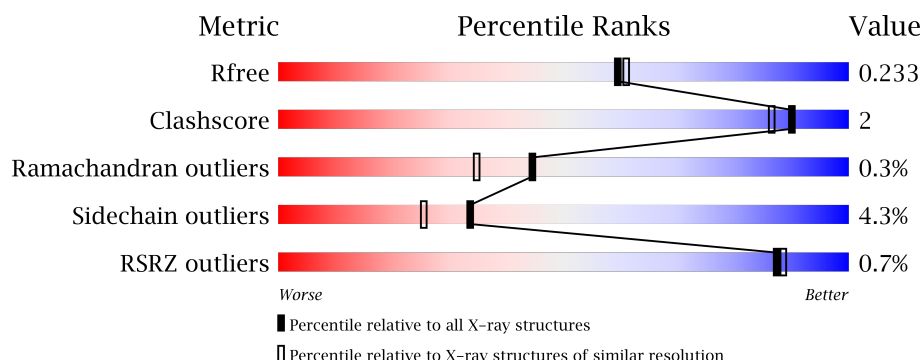|		
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

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 2.05 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		0.233
Clashscore		2
Ramachandran outliers		0.3%
Sidechain outliers		4.3%
RSRZ outliers		0.7%

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}	100719	2028 (2.08-2.04)
Clashscore	112137	2143 (2.08-2.04)
Ramachandran outliers	110173	2126 (2.08-2.04)
Sidechain outliers	110143	2126 (2.08-2.04)
RSRZ outliers	101464	2035 (2.08-2.04)

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. 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	526	.% 84% 7% 8%
1	B	526	.% 82% 9% 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8555 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.
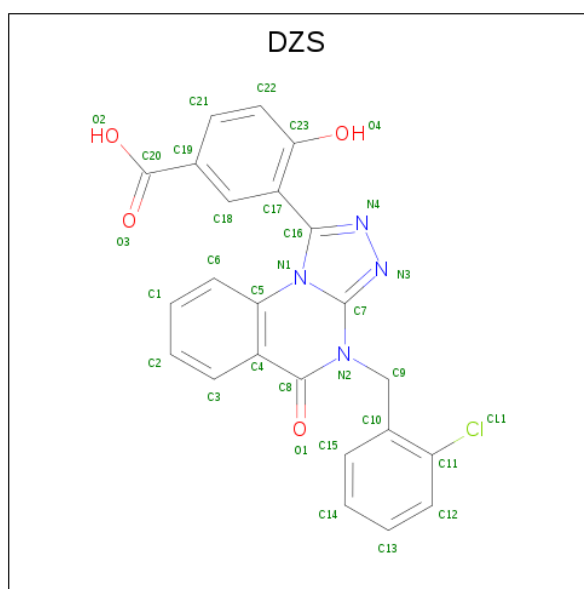
- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	483	Total 3917	C 2470	N 696	O 732	S 19	0	2	0
1	B	481	Total 3907	C 2466	N 694	O 727	S 20	0	2	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP Q06124
B	0	SER	-	expression tag	UNP Q06124

- Molecule 2 is 3-{4-[(2-chlorophenyl)methyl]-5-oxo-4,5-dihydro[1,2,4]triazolo[4,3-a]quinazolin -1-yl}-4-hydroxybenzoic acid (three-letter code: DZS) (formula: $C_{23}H_{15}ClN_4O_4$).

DZS

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 32	C 23	Cl 1	N 4	O 4	0	0

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total 32	C 23	Cl 1	N 4	O 4	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 5	O 4	P 1	0	0
3	A	1	Total 5	O 4	P 1	0	0
3	A	1	Total 5	O 4	P 1	0	0
3	B	1	Total 5	O 4	P 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	310	Total 310	O 310	0	0
4	B	337	Total 337	O 337	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: Tyrosine-protein phosphatase non-receptor type 11
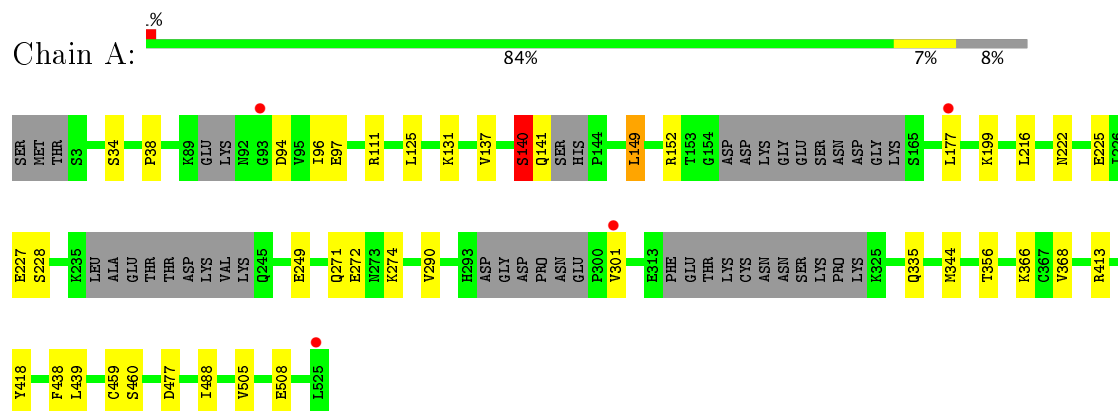
- Molecule 1: Tyrosine-protein phosphatase non-receptor type 11
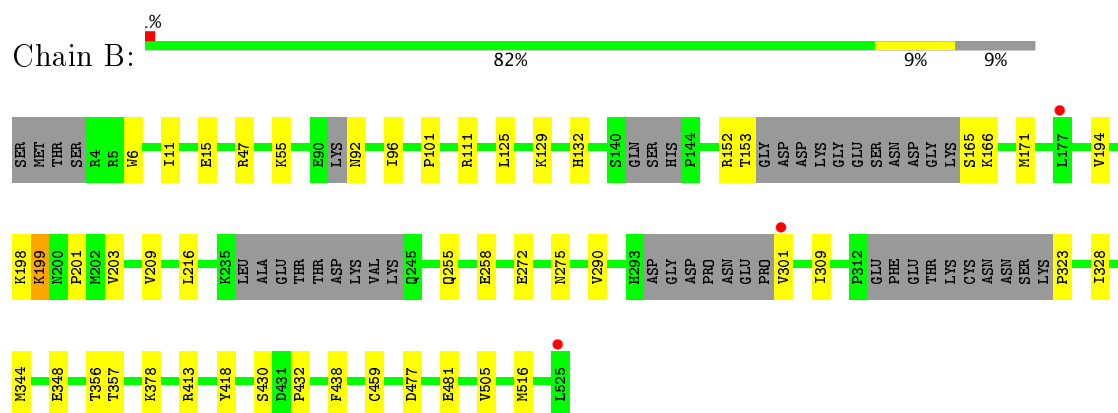
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	46.42Å 216.88Å 55.86Å 90.00° 97.30° 90.00°	Depositor
Resolution (Å)	108.44 – 2.05 108.44 – 2.05	Depositor EDS
% Data completeness (in resolution range)	97.4 (108.44-2.05) 97.8 (108.44-2.05)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.27 (at 2.05Å)	Xtriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.174 , 0.223 0.181 , 0.233	Depositor DCC
R_{free} test set	3272 reflections (5.17%)	DCC
Wilson B-factor (Å^2)	33.6	Xtriage
Anisotropy	0.485	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 56.5	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	0.96	EDS
Total number of atoms	8555	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.30% 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: PO4, DZS

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	0.52	0/4002	0.66	1/5392 (0.0%)
1	B	0.51	0/3992	0.64	0/5378
All	All	0.51	0/7994	0.65	1/10770 (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	A	140	SER	C-N-CA	6.25	137.32	121.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

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	3917	0	3867	15	0
1	B	3907	0	3864	18	0
2	A	32	0	0	0	0
2	B	32	0	0	0	0
3	A	15	0	0	0	0
3	B	5	0	0	0	0
4	A	310	0	0	0	0
4	B	337	0	0	1	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	8555	0	7731	33	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 2.

All (33) 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:272:GLU:CB	1:A:301:VAL:HG11	2.10	0.81
1:A:356:THR:OG1	1:A:459:CYS:HB3	1.85	0.77
1:A:272:GLU:HB3	1:A:301:VAL:HG11	1.70	0.73
1:A:272:GLU:HB2	1:A:301:VAL:HG11	1.82	0.61
1:B:356:THR:OG1	1:B:459:CYS:HB3	2.00	0.60
1:A:140:SER:HA	1:A:141:GLN:HB3	1.82	0.60
1:B:357:THR:HG22	4:B:850:HOH:O	2.03	0.59
1:A:290:VAL:HG11	1:A:344:MET:HG3	1.85	0.58
1:A:125:LEU:HB3	1:A:216:LEU:HD21	1.86	0.57
1:A:140:SER:HA	1:A:141:GLN:CB	2.34	0.56
1:A:271:GLN:HA	1:A:274:LYS:HD2	1.91	0.53
1:B:290:VAL:HG11	1:B:344:MET:HG3	1.93	0.51
1:B:6:TRP:HB3	1:B:101:PRO:HB3	1.93	0.51
1:A:222:ASN:O	1:A:225:GLU:HG2	2.11	0.51
1:B:309:ILE:HD13	1:B:328:ILE:HG12	1.93	0.50
1:B:132:HIS:HD1	1:B:153:THR:HA	1.77	0.49
1:B:255:GLN:O	1:B:258:GLU:HG2	2.12	0.48
1:B:418:TYR:HB3	1:B:438:PHE:CE1	2.49	0.47
1:B:11:ILE:HG12	1:B:15:GLU:HB2	1.96	0.47
1:B:203:VAL:HG22	1:B:209:VAL:HG22	1.97	0.47
1:B:432:PRO:HG3	1:B:516[A]:MET:HG2	1.96	0.47
1:A:366:LYS:HD3	1:A:460:SER:HB3	1.97	0.46
1:A:140:SER:CA	1:A:141:GLN:HB3	2.44	0.46
1:B:199:LYS:O	1:B:201:PRO:HD3	2.15	0.46
1:B:125:LEU:HB3	1:B:216:LEU:HD21	1.99	0.45
1:B:129:LYS:HB2	1:B:216:LEU:HD11	1.98	0.44
1:A:137:VAL:HG22	1:A:149:LEU:HD12	2.00	0.44
1:A:272:GLU:HB3	1:A:301:VAL:CG1	2.44	0.44
1:B:194:VAL:O	1:B:198:LYS:HB2	2.19	0.42
1:B:328:ILE:HD11	1:B:348:GLU:HG3	2.01	0.42
1:A:418:TYR:HB3	1:A:438:PHE:CE1	2.54	0.42
1:B:272:GLU:HG3	1:B:301:VAL:HB	2.03	0.40
1:B:111:ARG:HG3	1:B:111:ARG:HH11	1.87	0.40

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	
1	A	471/526 (90%)	451 (96%)	18 (4%)	2 (0%)	38	27
1	B	469/526 (89%)	456 (97%)	12 (3%)	1 (0%)	51	42
All	All	940/1052 (89%)	907 (96%)	30 (3%)	3 (0%)	44	35

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	505	VAL
1	B	505	VAL
1	A	38	PRO

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	
1	A	431/468 (92%)	410 (95%)	21 (5%)	29	20
1	B	430/468 (92%)	414 (96%)	16 (4%)	39	31
All	All	861/936 (92%)	824 (96%)	37 (4%)	33	25

All (37) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	SER
1	A	94	ASP
1	A	96	ILE
1	A	97	GLU
1	A	111	ARG
1	A	131	LYS
1	A	140	SER
1	A	149	LEU
1	A	152	ARG
1	A	177	LEU
1	A	199	LYS
1	A	227	GLU
1	A	228	SER
1	A	249	GLU
1	A	335	GLN
1	A	368	VAL
1	A	413	ARG
1	A	439	LEU
1	A	477	ASP
1	A	488	ILE
1	A	508	GLU
1	B	47	ARG
1	B	55	LYS
1	B	92	ASN
1	B	96	ILE
1	B	152	ARG
1	B	165	SER
1	B	166	LYS
1	B	171	MET
1	B	199	LYS
1	B	275	ASN
1	B	323	PRO
1	B	378	LYS
1	B	413	ARG
1	B	430	SER
1	B	477	ASP
1	B	481	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	141	GLN
1	A	211	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	269	GLN
1	B	114	HIS
1	B	256	GLN
1	B	275	ASN
1	B	281	ASN
1	B	394	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)

6 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	DZS	A	601	-	28,36,36	1.85	4 (14%)	31,53,53	1.55	4 (12%)
3	PO4	A	602	-	4,4,4	2.62	1 (25%)	6,6,6	0.51	0
3	PO4	A	603	-	4,4,4	2.66	1 (25%)	6,6,6	0.73	0
3	PO4	A	604	-	4,4,4	2.60	1 (25%)	6,6,6	0.45	0
2	DZS	B	601	-	28,36,36	1.96	5 (17%)	31,53,53	1.40	4 (12%)
3	PO4	B	602	-	4,4,4	2.59	1 (25%)	6,6,6	0.28	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DZS	A	601	-	-	0/8/12/12	0/5/5/5
3	PO4	A	602	-	-	0/0/0/0	0/0/0/0
3	PO4	A	603	-	-	0/0/0/0	0/0/0/0
3	PO4	A	604	-	-	0/0/0/0	0/0/0/0
2	DZS	B	601	-	-	0/8/12/12	0/5/5/5
3	PO4	B	602	-	-	0/0/0/0	0/0/0/0

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	DZS	C15-C10	2.00	1.43	1.39
2	B	601	DZS	C14-C15	2.01	1.42	1.38
2	B	601	DZS	C15-C10	2.05	1.43	1.39
2	A	601	DZS	C1-C6	2.06	1.41	1.36
2	B	601	DZS	C1-C6	2.45	1.42	1.36
2	A	601	DZS	C4-C5	3.18	1.43	1.41
2	B	601	DZS	C4-C5	3.31	1.44	1.41
3	B	602	PO4	P-O1	4.53	1.60	1.50
3	A	604	PO4	P-O1	4.59	1.60	1.50
3	A	602	PO4	P-O1	4.63	1.60	1.50
3	A	603	PO4	P-O1	4.67	1.60	1.50
2	A	601	DZS	C8-C4	6.19	1.51	1.41
2	B	601	DZS	C8-C4	6.81	1.52	1.41

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	DZS	C4-C5-N1	-4.06	117.14	119.50
2	A	601	DZS	C4-C5-N1	-3.70	117.35	119.50
2	A	601	DZS	C9-C10-C11	-2.16	116.28	120.53
2	B	601	DZS	C9-N2-C7	-2.11	116.55	122.56
2	B	601	DZS	C9-N2-C8	2.62	120.96	117.79
2	A	601	DZS	C10-C9-N2	3.07	118.40	113.31
2	B	601	DZS	C3-C4-C5	3.24	119.53	118.04
2	A	601	DZS	C3-C4-C5	4.72	120.22	118.04

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 (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	A	483/526 (91%)	-0.19	4 (0%)	86	87	23, 43, 70, 93	0
1	B	481/526 (91%)	-0.23	3 (0%)	89	90	22, 42, 68, 96	0
All	All	964/1052 (91%)	-0.21	7 (0%)	87	88	22, 43, 69, 96	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	177	LEU	3.1
1	A	301	VAL	2.4
1	B	525	LEU	2.2
1	A	525	LEU	2.1
1	A	93	GLY	2.1
1	B	177	LEU	2.1
1	B	301	VAL	2.0

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
3	PO4	A	603	5/5	0.93	0.15	0.99	65,66,67,69	0
3	PO4	A	602	5/5	0.95	0.13	-0.32	74,75,77,79	0
2	DZS	A	601	32/32	0.96	0.10	-0.43	32,37,47,49	0
3	PO4	B	602	5/5	0.98	0.11	-0.69	48,49,54,54	0
2	DZS	B	601	32/32	0.96	0.09	-0.93	26,34,42,45	0
3	PO4	A	604	5/5	0.89	0.13	-	121,122,122,123	0

6.5 Other polymers (i)

There are no such residues in this entry.