Full wwPDB X-ray Structure Validation Report (i)

Sep 26, 2023 – 11:38 AM EDT

PDB ID	:	6CRF
Title	:	Crystal Structure of Shp2 E76K GOF Mutant in the Open Conformation
Authors	:	Stams, T.; Fodor, M.
Deposited on	:	2018-03-17
Resolution	:	2.62 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

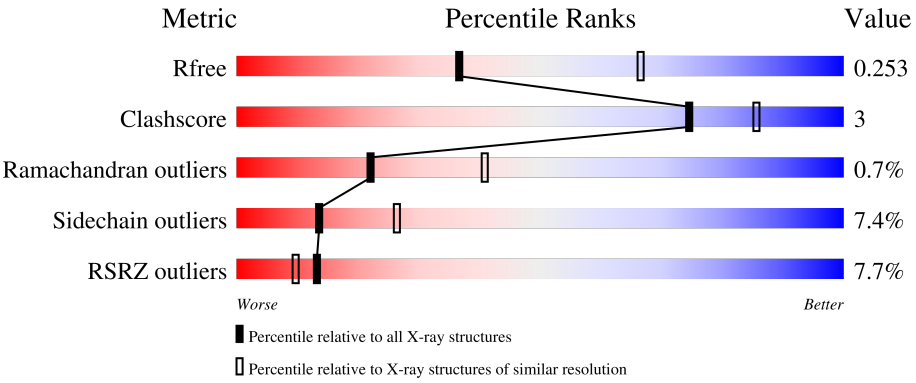
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.62 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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	526	6% 75% 15% • 10%
1	B	526	7% 77% 8% • 14%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7170 atoms, of which 1 is hydrogen 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.

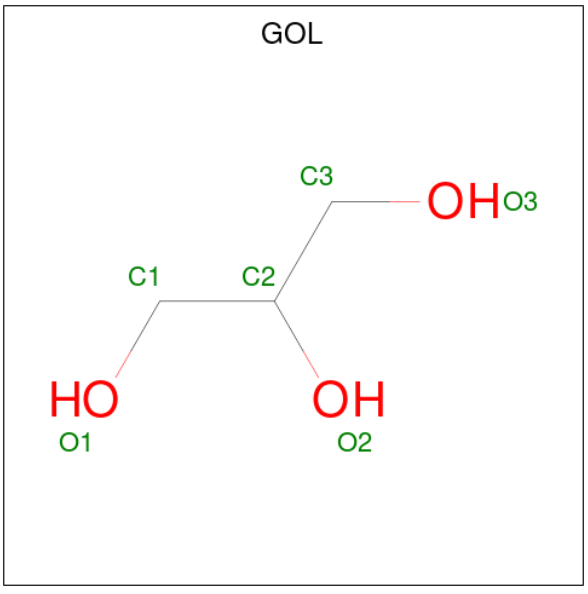
- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 11.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	475	Total C H N O S 3679 2316 1 659 687 16	0	0	0
1	B	451	Total C N O S 3277 2045 596 621 15	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q06124
A	76	LYS	GLU	engineered mutation	UNP Q06124
B	0	GLY	-	expression tag	UNP Q06124
B	76	LYS	GLU	engineered mutation	UNP Q06124

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total C O 6 3 3	0	0

- Molecule 3 is water.

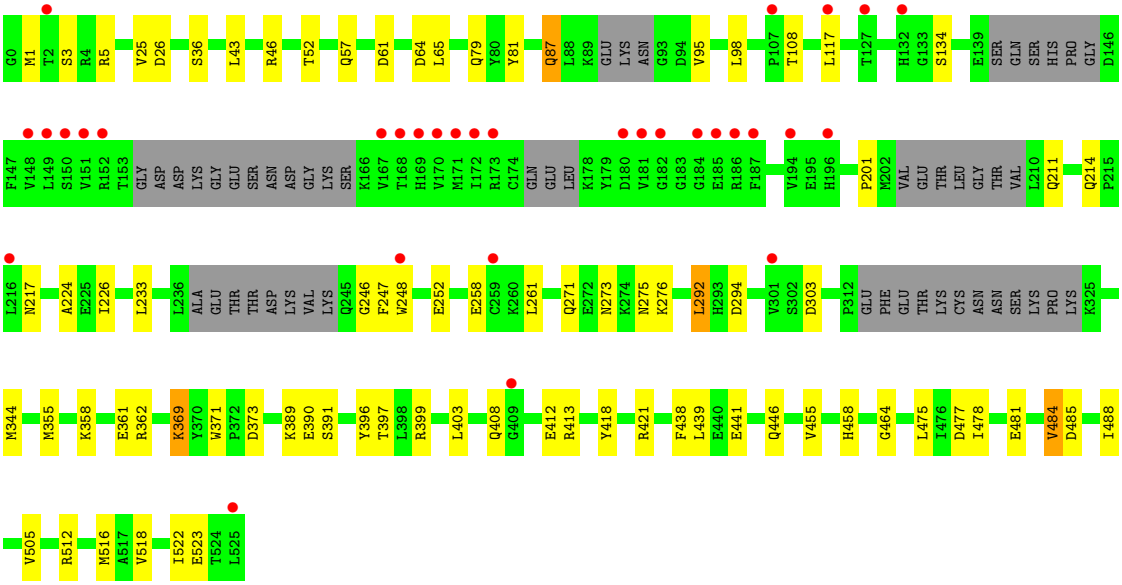
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	90	Total O 90 90	0	0
3	B	118	Total O 118 118	0	0

3 Residue-property plots (i)

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.

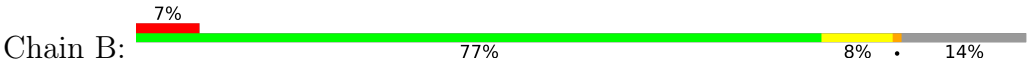
• Molecule 1: Tyrosine-protein phosphatase non-receptor type 11

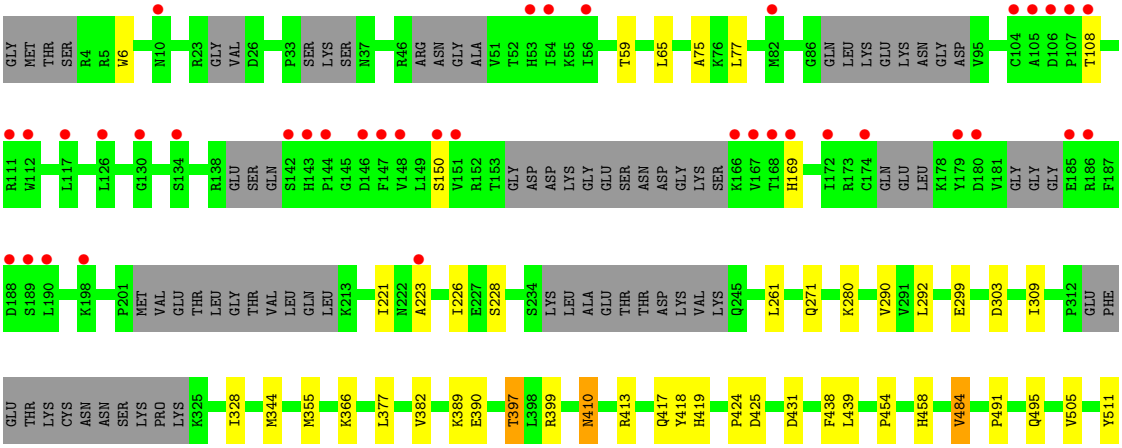
Chain A:



• Molecule 1: Tyrosine-protein phosphatase non-receptor type 11

Chain B:






L525

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	249.10Å 41.73Å 153.85Å 90.00° 125.76° 90.00°	Depositor
Resolution (Å)	124.84 – 2.62 120.02 – 2.62	Depositor EDS
% Data completeness (in resolution range)	99.5 (124.84-2.62) 99.8 (120.02-2.62)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.15 (at 2.62Å)	Xtriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.198 , 0.237 0.211 , 0.253	Depositor DCC
R_{free} test set	1936 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	56.1	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 72.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.012 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7170	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.02% 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: GOL

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	0.50	0/3751	0.70	0/5074
1	B	0.52	0/3332	0.69	1/4517 (0.0%)
All	All	0.51	0/7083	0.69	1/9591 (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	B	424	PRO	C-N-CA	5.58	135.65	121.70

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	3678	1	3461	27	0
1	B	3277	0	2840	17	0
2	B	6	0	8	0	0
3	A	90	0	0	0	0
3	B	118	0	0	0	0
All	All	7169	1	6309	43	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 3.

All (43) 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:B:390:GLU:HG2	1:B:399:ARG:HG2	1.73	0.69
1:A:201:PRO:HB3	1:A:211:GLN:HA	1.75	0.68
1:A:518:VAL:O	1:A:522:ILE:HG12	2.00	0.61
1:A:1:MET:HB3	1:A:79:GLN:OE1	2.03	0.58
1:A:418:TYR:HB3	1:A:438:PHE:CE1	2.39	0.58
1:A:224:ALA:HB2	1:A:484:VAL:HG13	1.87	0.57
1:A:57:GLN:HE22	1:B:271:GLN:HE21	1.53	0.55
1:A:358:LYS:HB2	1:A:361:GLU:HG3	1.90	0.54
1:A:397:THR:HG22	1:A:421:ARG:HD3	1.90	0.53
1:A:273:ASN:HA	1:A:276:LYS:HD2	1.91	0.51
1:B:491:PRO:HG3	1:B:511:TYR:OH	2.10	0.51
1:B:410:ASN:ND2	1:B:410:ASN:H	2.08	0.51
1:B:399:ARG:HD2	1:B:417:GLN:OE1	2.11	0.50
1:B:418:TYR:HB3	1:B:438:PHE:CE1	2.46	0.50
1:A:26:ASP:HA	1:A:46:ARG:HG2	1.94	0.50
1:A:369:LYS:HD3	1:A:371:TRP:O	2.11	0.49
1:B:6:TRP:HZ2	1:B:75:ALA:HA	1.78	0.49
1:A:134:SER:HA	1:A:214:GLN:O	2.13	0.49
1:A:390:GLU:HG2	1:A:399:ARG:HG2	1.93	0.49
1:B:150:SER:HA	1:B:169:HIS:HA	1.94	0.48
1:B:397:THR:HG23	1:B:419:HIS:HB3	1.95	0.48
1:B:290:VAL:HG11	1:B:344:MET:HG3	1.96	0.48
1:B:309:ILE:HD13	1:B:328:ILE:HG12	1.96	0.48
1:B:355:MET:HG3	1:B:458:HIS:CE1	2.49	0.47
1:A:475:LEU:HA	1:A:478:ILE:HD12	1.96	0.47
1:A:247:PHE:CG	1:A:512:ARG:HG2	2.50	0.46
1:A:418:TYR:HB3	1:A:438:PHE:HE1	1.79	0.46
1:B:418:TYR:HB3	1:B:438:PHE:HE1	1.81	0.46
1:B:328:ILE:HG13	1:B:454:PRO:HB2	1.98	0.45
1:A:43:LEU:HD21	1:A:98:LEU:HD21	1.99	0.44
1:A:248:TRP:CE2	1:A:252:GLU:HG3	2.53	0.44
1:A:403:LEU:O	1:A:412:GLU:HA	2.17	0.44
1:A:3:SER:HB2	1:A:5:ARG:HG3	2.00	0.43
1:A:81:TYR:CD1	1:A:87:GLN:HG2	2.52	0.43
1:B:261:LEU:HD22	1:B:495:GLN:NE2	2.33	0.43
1:B:223:ALA:HB3	1:B:484:VAL:O	2.19	0.42
1:A:226:ILE:HD13	1:A:488:ILE:HD11	2.01	0.42
1:A:233:LEU:HA	1:A:246:GLY:HA3	2.01	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:292:LEU:HD11	1:A:344:MET:HB2	2.02	0.41
1:B:65:LEU:HD13	1:B:77:LEU:HD21	2.02	0.41
1:A:355:MET:HG3	1:A:458:HIS:CE1	2.55	0.41
1:A:396:TYR:OH	1:A:441:GLU:HG2	2.22	0.40
1:A:446:GLN:HG3	1:A:455:VAL:CG2	2.52	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	459/526 (87%)	440 (96%)	16 (4%)	3 (1%)	22	41
1	B	425/526 (81%)	395 (93%)	27 (6%)	3 (1%)	22	41
All	All	884/1052 (84%)	835 (94%)	43 (5%)	6 (1%)	22	41

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	505	VAL
1	B	108	THR
1	A	217	ASN
1	B	59	THR
1	B	505	VAL
1	A	464	GLY

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	373/467 (80%)	342 (92%)	31 (8%)	11 21
1	B	292/467 (62%)	274 (94%)	18 (6%)	18 35
All	All	665/934 (71%)	616 (93%)	49 (7%)	13 26

All (49) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	25	VAL
1	A	36	SER
1	A	52	THR
1	A	61	ASP
1	A	64	ASP
1	A	65	LEU
1	A	87	GLN
1	A	95	VAL
1	A	108	THR
1	A	117	LEU
1	A	258	GLU
1	A	261	LEU
1	A	271	GLN
1	A	275	ASN
1	A	292	LEU
1	A	294	ASP
1	A	303	ASP
1	A	362	ARG
1	A	369	LYS
1	A	373	ASP
1	A	389	LYS
1	A	391	SER
1	A	408	GLN
1	A	413	ARG
1	A	439	LEU
1	A	477	ASP
1	A	481	GLU
1	A	484	VAL
1	A	485	ASP
1	A	516	MET
1	A	523	GLU
1	B	221	ILE
1	B	226	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	228	SER
1	B	280	LYS
1	B	292	LEU
1	B	299	GLU
1	B	303	ASP
1	B	366	LYS
1	B	377	LEU
1	B	382	VAL
1	B	389	LYS
1	B	397	THR
1	B	410	ASN
1	B	413	ARG
1	B	425	ASP
1	B	431	ASP
1	B	439	LEU
1	B	484	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	57	GLN
1	A	114	HIS
1	A	410	ASN
1	A	519	GLN
1	B	114	HIS
1	B	495	GLN
1	B	506	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	GOL	B	601	-	5,5,5	0.10	0	5,5,5	0.21	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	GOL	B	601	-	-	0/4/4/4	-

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.

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	117:LEU	C	118:SER	N	3.28

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, 95^{th} 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	475/526 (90%)	0.18	32 (6%)	17 13	38, 66, 135, 164	0
1	B	451/526 (85%)	0.48	39 (8%)	10 7	29, 62, 164, 203	0
All	All	926/1052 (88%)	0.32	71 (7%)	13 10	29, 66, 150, 203	0

All (71) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	167	VAL	14.5
1	B	168	THR	13.1
1	B	107	PRO	12.0
1	A	168	THR	10.3
1	A	184	GLY	7.3
1	B	169	HIS	7.1
1	B	142	SER	6.9
1	B	143	HIS	6.3
1	B	174	CYS	6.2
1	B	108	THR	5.9
1	B	188	ASP	5.8
1	A	186	ARG	5.7
1	B	10	ASN	5.3
1	A	196	HIS	5.2
1	B	105	ALA	5.1
1	A	167	VAL	5.0
1	B	148	VAL	4.8
1	B	151	VAL	4.8
1	B	56	ILE	4.6
1	B	185	GLU	4.4
1	B	186	ARG	4.4
1	B	117	LEU	4.1
1	A	170	VAL	4.1
1	B	144	PRO	4.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	185	GLU	4.0
1	A	181	VAL	4.0
1	A	409	GLY	3.9
1	B	54	ILE	3.7
1	A	169	HIS	3.7
1	B	172	ILE	3.7
1	B	53	HIS	3.6
1	B	166	LYS	3.2
1	B	198	LYS	3.2
1	A	152	ARG	3.2
1	A	182	GLY	3.1
1	B	179	TYR	3.1
1	A	127	THR	3.1
1	B	190	LEU	3.0
1	B	147	PHE	3.0
1	A	172	ILE	2.9
1	A	151	VAL	2.9
1	A	2	THR	2.9
1	A	171	MET	2.9
1	B	223	ALA	2.9
1	A	148	VAL	2.8
1	A	301	VAL	2.8
1	B	112	TRP	2.8
1	B	189	SER	2.8
1	B	150	SER	2.7
1	B	180	ASP	2.7
1	A	150	SER	2.6
1	A	173	ARG	2.6
1	B	130	GLY	2.6
1	A	149	LEU	2.6
1	A	180	ASP	2.5
1	B	146	ASP	2.5
1	A	194	VAL	2.4
1	A	132	HIS	2.4
1	A	259	CYS	2.4
1	A	117	LEU	2.3
1	A	248	TRP	2.3
1	B	106	ASP	2.2
1	B	134	SER	2.2
1	A	107	PRO	2.1
1	B	82	MET	2.1
1	A	525	LEU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	126	LEU	2.1
1	A	187	PHE	2.0
1	A	216	LEU	2.0
1	B	111	ARG	2.0
1	B	104	CYS	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 monosaccharides 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, 95^{th} 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	GOL	B	601	6/6	0.85	0.18	77,80,83,85	0

6.5 Other polymers (i)

There are no such residues in this entry.