Full wwPDB X-ray Structure Validation Report ⓘ

Oct 10, 2023 – 07:39 AM EDT

PDB ID	:	6WMZ
Title	:	Crystal structure of human SFPQ/NONO complex
Authors	:	Lee, M.; Bond, C.S.
Deposited on	:	2020-04-22
Resolution	:	2.85 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 ⓘ 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 ⓘ) were used in the production of this report:

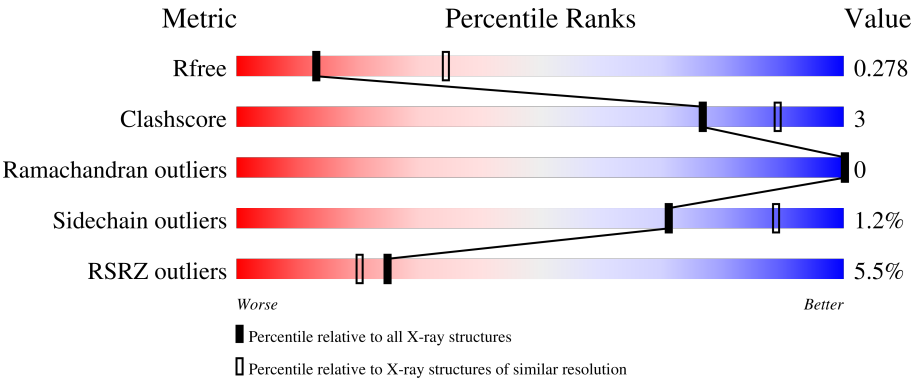
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.85 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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	388	7% 70% 7% 23%
1	C	388	6% 69% 6% 25%
2	B	261	3% 83% 9% 8%
2	D	261	% 87% 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8954 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 Splicing factor, proline- and glutamine-rich.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	299	Total	C	N	O	S	0	0	0
			2478	1534	455	474	15			
1	C	292	Total	C	N	O	S	0	0	0
			2415	1497	446	459	13			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	211	GLY	-	expression tag	UNP P23246
A	212	ALA	-	expression tag	UNP P23246
A	213	MET	-	expression tag	UNP P23246
C	211	GLY	-	expression tag	UNP P23246
C	212	ALA	-	expression tag	UNP P23246
C	213	MET	-	expression tag	UNP P23246

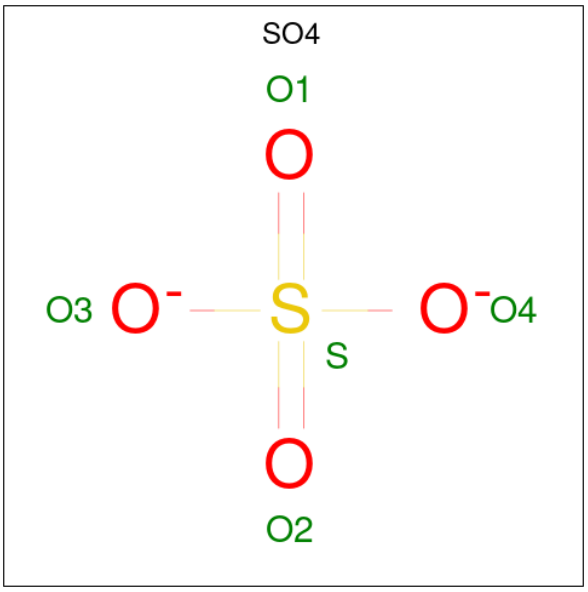
- Molecule 2 is a protein called Non-POU domain-containing octamer-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	240	Total	C	N	O	S	0	0	0
			1955	1231	351	364	9			
2	D	254	Total	C	N	O	S	0	0	0
			2062	1299	370	384	9			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	52	MET	-	initiating methionine	UNP Q15233
D	52	MET	-	initiating methionine	UNP Q15233

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

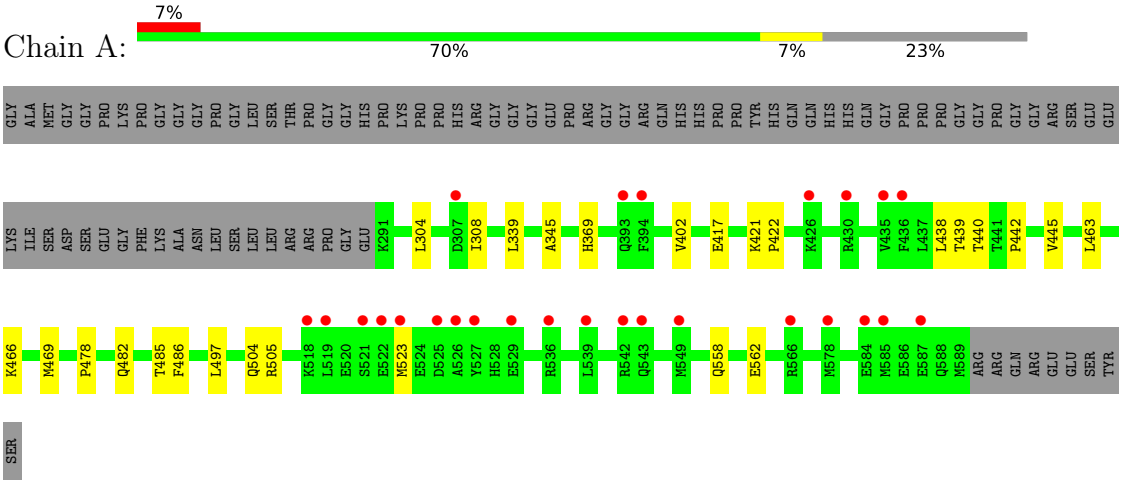
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	5	Total O 5 5	0	0
4	B	17	Total O 17 17	0	0
4	C	4	Total O 4 4	0	0
4	D	13	Total O 13 13	0	0

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.

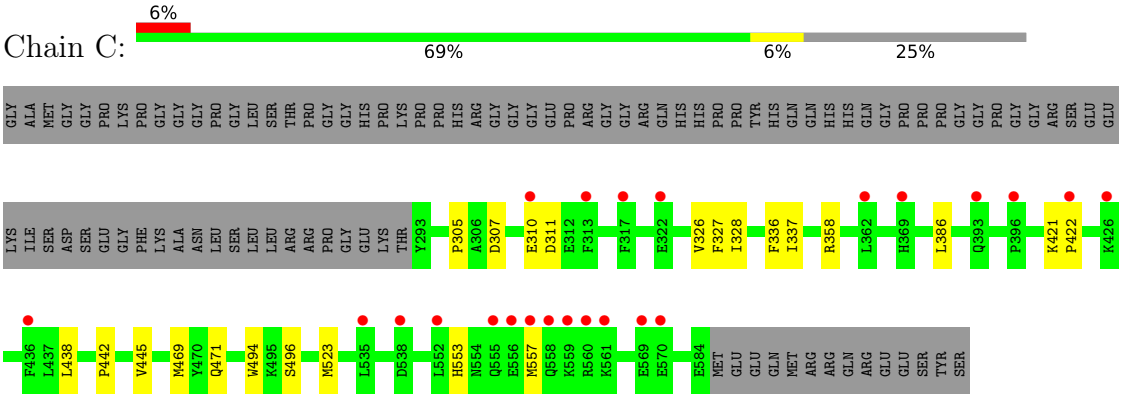
• Molecule 1: Splicing factor, proline- and glutamine-rich

Chain A:



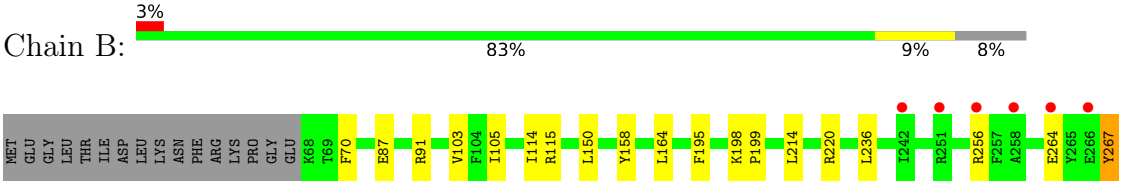
• Molecule 1: Splicing factor, proline- and glutamine-rich

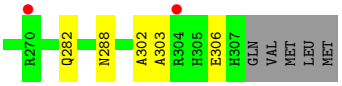
Chain C:



• Molecule 2: Non-POU domain-containing octamer-binding protein

Chain B:





• Molecule 2: Non-POU domain-containing octamer-binding protein

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	66.63Å 112.94Å 204.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	63.44 – 2.85 63.36 – 2.85	Depositor EDS
% Data completeness (in resolution range)	99.2 (63.44-2.85) 99.2 (63.36-2.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.27 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.236 , 0.283 0.238 , 0.278	Depositor DCC
R_{free} test set	1827 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	59.6	Xtriage
Anisotropy	0.338	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 55.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	8954	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.84% 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: 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	A	0.63	0/2518	0.74	0/3367
1	C	0.64	0/2455	0.73	0/3285
2	B	0.62	0/1993	0.75	0/2675
2	D	0.63	0/2102	0.75	0/2822
All	All	0.63	0/9068	0.74	0/12149

There are no bond length outliers.

There are no bond angle outliers.

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	2478	0	2444	18	0
1	C	2415	0	2382	12	0
2	B	1955	0	1948	17	0
2	D	2062	0	2054	19	0
3	D	5	0	0	0	0
4	A	5	0	0	0	0
4	B	17	0	0	0	0
4	C	4	0	0	0	0
4	D	13	0	0	0	0
All	All	8954	0	8828	54	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 (54) 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:C:307:ASP:OD1	1:C:358:ARG:NH1	2.27	0.68
1:A:442:PRO:HG2	2:B:267:TYR:CE2	2.33	0.64
2:D:103:VAL:HG22	2:D:114:ILE:HD12	1.81	0.62
2:B:87:GLU:O	2:B:91:ARG:HG3	2.00	0.61
1:C:310:GLU:HA	1:C:328:ILE:HD13	1.84	0.59
2:D:152:VAL:CG1	2:D:222:VAL:CG1	2.83	0.57
2:D:150:LEU:HD12	2:D:195:PHE:HE2	1.71	0.56
2:D:55:LEU:HD22	2:D:124:ILE:HG23	1.87	0.55
2:D:90:MET:HE3	2:D:114:ILE:HB	1.90	0.54
1:C:421:LYS:N	1:C:422:PRO:HD2	2.23	0.53
2:D:152:VAL:CG1	2:D:222:VAL:HG11	2.38	0.53
1:C:386:LEU:HD12	2:D:251:ARG:CZ	2.39	0.53
1:C:438:LEU:HD13	1:C:445:VAL:HG21	1.89	0.53
1:A:402:VAL:CG2	2:B:236:LEU:HB3	2.40	0.51
2:D:198:LYS:N	2:D:199:PRO:CD	2.74	0.51
1:A:339:LEU:HD12	1:A:345:ALA:HA	1.93	0.51
2:B:150:LEU:HD12	2:B:195:PHE:HE2	1.77	0.50
1:C:469:MET:HG2	2:D:158:TYR:CE2	2.46	0.50
2:D:90:MET:HG3	2:D:105:ILE:HD11	1.94	0.50
2:B:198:LYS:N	2:B:199:PRO:CD	2.75	0.49
1:A:504:GLN:HE22	2:B:288:ASN:HD22	1.59	0.49
2:D:214:LEU:HD23	2:D:221:PRO:HA	1.95	0.49
1:A:463:LEU:HD12	1:A:466:LYS:HD2	1.94	0.48
1:C:326:VAL:HG22	1:C:337:ILE:HG13	1.96	0.47
2:B:70:PHE:O	2:B:115:ARG:HD3	2.14	0.47
2:D:152:VAL:HG13	2:D:222:VAL:CG1	2.45	0.47
2:D:90:MET:HG3	2:D:105:ILE:CD1	2.45	0.46
1:C:553:HIS:O	1:C:557:MET:HG2	2.16	0.46
1:C:305:PRO:CG	1:C:358:ARG:HD2	2.46	0.46
1:C:327:PHE:HB3	1:C:336:PHE:CZ	2.51	0.46
1:A:523:MET:HB3	2:B:267:TYR:CD1	2.52	0.45
1:A:369:HIS:NE2	1:A:417:GLU:HB3	2.32	0.45
1:A:438:LEU:HD11	1:A:445:VAL:HG21	1.99	0.45
2:B:103:VAL:HG22	2:B:114:ILE:HD12	1.99	0.45
1:A:438:LEU:CD1	1:A:445:VAL:HG21	2.47	0.44
2:B:256:ARG:NH2	2:B:264:GLU:OE2	2.51	0.44
2:D:116:LEU:HD12	2:D:122:ALA:HA	1.99	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:478:PRO:HA	2:B:214:LEU:O	2.17	0.43
1:A:497:LEU:HD23	1:A:497:LEU:HA	1.91	0.43
2:B:220:ARG:O	2:B:220:ARG:HG3	2.18	0.43
2:B:302:ALA:HB1	2:D:113:PHE:CE2	2.53	0.43
2:B:306:GLU:HG2	2:B:306:GLU:O	2.18	0.42
1:A:469:MET:HG2	2:B:158:TYR:CE2	2.55	0.42
1:A:505:ARG:HE	2:B:282:GLN:HE21	1.67	0.42
1:A:421:LYS:HB2	1:A:422:PRO:HD3	2.01	0.42
1:A:482:GLN:O	1:A:485:THR:OG1	2.35	0.41
1:C:442:PRO:HG2	1:C:523:MET:SD	2.61	0.41
1:C:494:TRP:CD2	2:D:220:ARG:HD2	2.56	0.41
2:D:55:LEU:HD22	2:D:124:ILE:CG2	2.51	0.41
2:D:290:LYS:HE3	2:D:294:GLU:OE2	2.21	0.41
2:B:303:ALA:HA	2:D:111:PHE:CZ	2.56	0.40
1:A:558:GLN:O	1:A:562:GLU:HB2	2.21	0.40
1:A:439:THR:HG22	1:A:440:THR:N	2.36	0.40
1:A:304:LEU:HB3	1:A:308:ILE:HD13	2.03	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	297/388 (76%)	284 (96%)	13 (4%)	0	100	100
1	C	290/388 (75%)	282 (97%)	8 (3%)	0	100	100
2	B	238/261 (91%)	226 (95%)	12 (5%)	0	100	100
2	D	252/261 (97%)	241 (96%)	11 (4%)	0	100	100
All	All	1077/1298 (83%)	1033 (96%)	44 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	264/330 (80%)	263 (100%)	1 (0%)	91	96
1	C	256/330 (78%)	253 (99%)	3 (1%)	71	89
2	B	210/229 (92%)	207 (99%)	3 (1%)	67	86
2	D	221/229 (96%)	217 (98%)	4 (2%)	59	82
All	All	951/1118 (85%)	940 (99%)	11 (1%)	71	89

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	486	PHE
2	B	105	ILE
2	B	164	LEU
2	B	267	TYR
1	C	311	ASP
1	C	471	GLN
1	C	496	SER
2	D	145	CYS
2	D	237	PRO
2	D	243	LYS
2	D	275	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	492	GLN
1	A	504	GLN
1	A	564	GLN
2	B	282	GLN
2	B	307	HIS
1	C	567	GLN
2	D	246	GLN
2	D	282	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	305	HIS

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

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	$\#\|Z\| > 2$
3	SO4	D	401	-	4,4,4	0.38	0	6,6,6	0.07	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.

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

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	299/388 (77%)	0.68	26 (8%) 10 7	39, 80, 118, 138	0
1	C	292/388 (75%)	0.69	23 (7%) 12 9	40, 86, 125, 147	0
2	B	240/261 (91%)	0.44	8 (3%) 46 41	31, 58, 115, 135	0
2	D	254/261 (97%)	0.35	3 (1%) 79 78	33, 63, 107, 123	0
All	All	1085/1298 (83%)	0.55	60 (5%) 25 20	31, 74, 119, 147	0

All (60) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	525	ASP	5.0
2	B	266	GLU	4.2
1	A	518	LYS	4.2
2	B	258	ALA	3.9
1	A	543	GLN	3.8
1	A	522	GLU	3.7
1	A	542	ARG	3.6
1	C	396	PRO	3.5
1	A	529	GLU	3.5
2	B	270	ARG	3.4
2	B	264	GLU	3.4
1	C	556	GLU	3.3
1	A	523	MET	3.3
1	C	322	GLU	3.3
1	A	566	ARG	3.2
1	A	539	LEU	3.1
1	C	422	PRO	3.0
1	A	436	PHE	3.0
1	C	317	PHE	2.9
1	A	430	ARG	2.9
1	C	559	LYS	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	304	ARG	2.8
1	A	526	ALA	2.8
1	A	536	ARG	2.7
1	A	578	MET	2.7
1	C	569	GLU	2.7
1	C	393	GLN	2.7
1	A	527	TYR	2.7
1	C	538	ASP	2.7
1	A	394	PHE	2.6
1	A	393	GLN	2.5
1	C	362	LEU	2.5
1	C	426	LYS	2.5
1	A	587	GLU	2.4
1	C	369	HIS	2.4
1	C	310	GLU	2.4
1	C	535	LEU	2.4
1	C	570	GLU	2.4
2	D	252	GLU	2.4
1	A	521	SER	2.4
1	C	560	ARG	2.3
1	A	519	LEU	2.3
1	C	555	GLN	2.3
1	A	585	MET	2.3
1	A	307	ASP	2.3
2	B	256	ARG	2.2
1	A	426	LYS	2.2
1	C	313	PHE	2.2
2	B	242	ILE	2.1
2	D	105	ILE	2.1
1	C	557	MET	2.1
1	C	561	LYS	2.1
1	A	549	MET	2.1
2	D	107	LYS	2.1
1	C	552	LEU	2.0
1	A	584	GLU	2.0
1	A	435	VAL	2.0
1	C	436	PHE	2.0
2	B	251	ARG	2.0
1	C	558	GLN	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
3	SO4	D	401	5/5	0.81	0.21	114,120,124,125	0

6.5 Other polymers (i)

There are no such residues in this entry.