wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 07:26 PM EDT

PDB ID	:	4B3L
Title	:	Family 1 6-phospho-beta-D glycosidase from Streptococcus pyogenes
Authors	:	Stepper, J.; Dabin, J.; Ekloef, J.M.; Thongpoo, P.; Kongsaeree, P.T.; Taylor, E.J.; Turkenburg, J.P.; Brumer, H.; Davies, G.J.
Deposited on	:	2012-07-24
Resolution	:	2.51 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i)) were used in the production of this report:

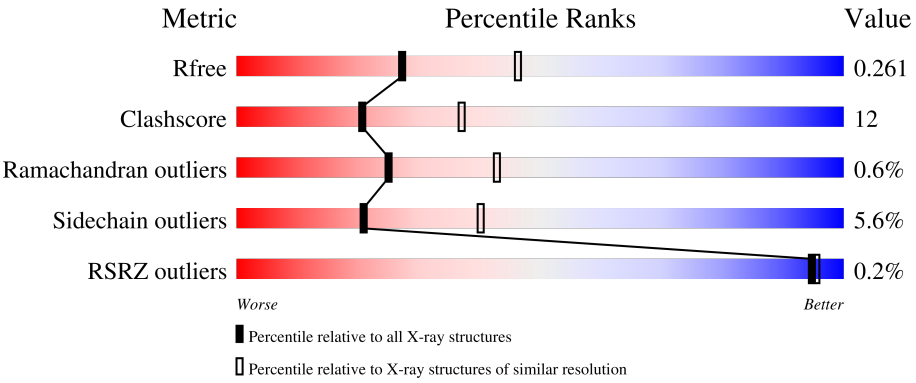
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.51 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	479	74% 21%
1	B	479	77% 17%
1	C	479	71% 23%
1	D	479	77% 17%
1	E	479	70% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	479	% 58% 33% 6%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 22619 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 BETA-GLUCOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			
1	B	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			
1	C	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			
1	D	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			
1	E	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			
1	F	461	Total	C	N	O	S	0	0	0
			3767	2435	636	686	10			

- Molecule 2 is water.

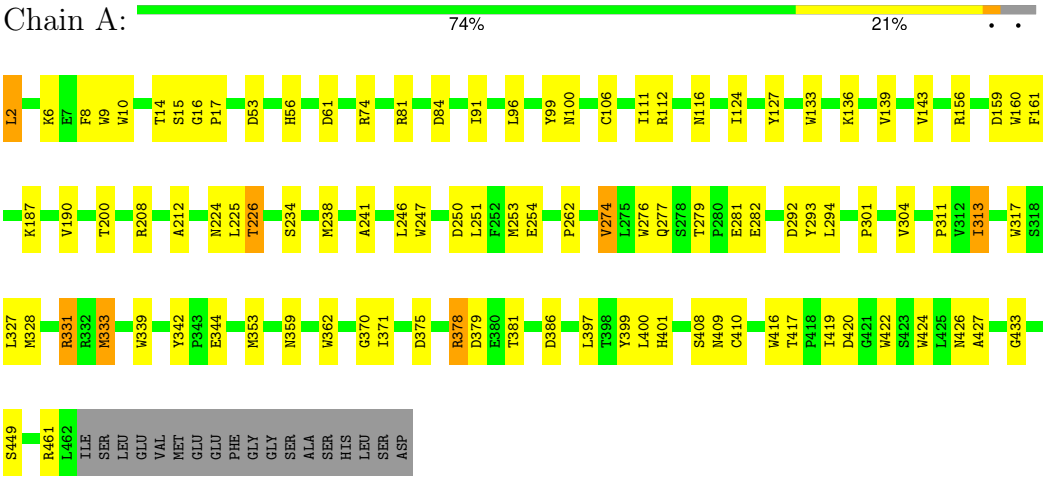
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	O	0	0
			1	1		
2	B	4	Total	O	0	0
			4	4		
2	C	3	Total	O	0	0
			3	3		
2	D	5	Total	O	0	0
			5	5		
2	E	1	Total	O	0	0
			1	1		
2	F	3	Total	O	0	0
			3	3		

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.

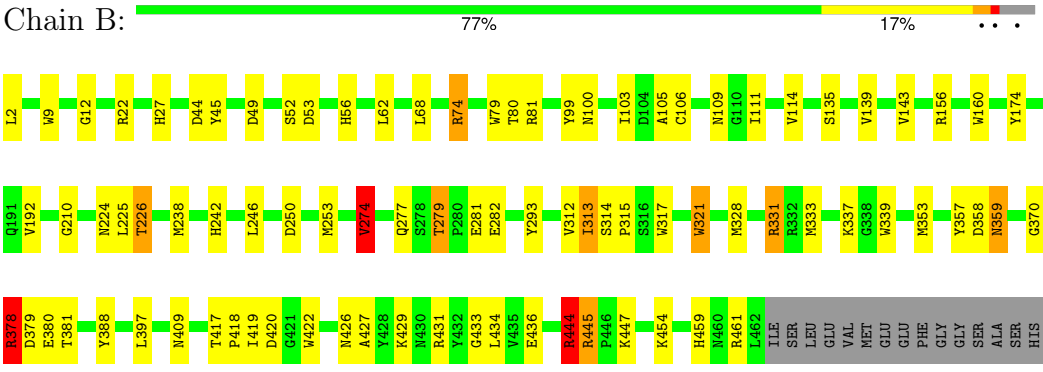
• Molecule 1: BETA-GLUCOSIDASE

Chain A:



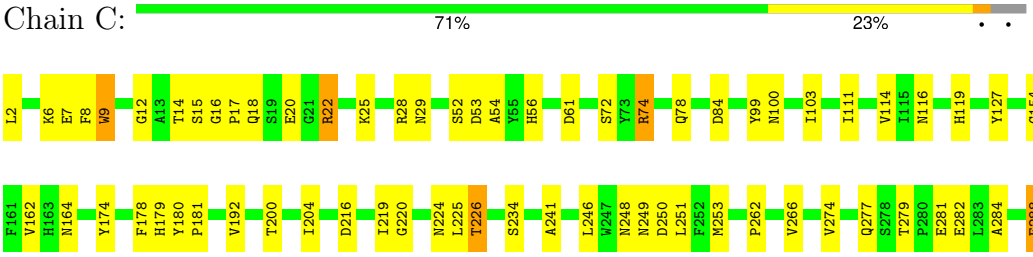
• Molecule 1: BETA-GLUCOSIDASE

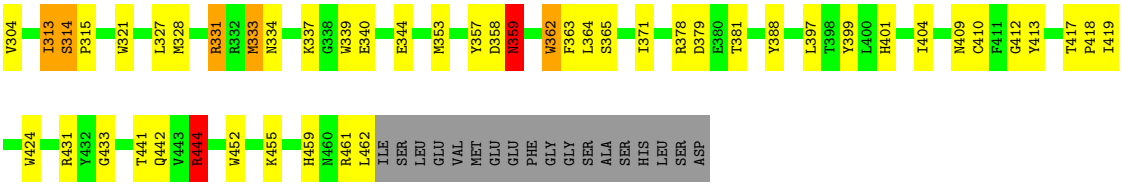
Chain B:



• Molecule 1: BETA-GLUCOSIDASE

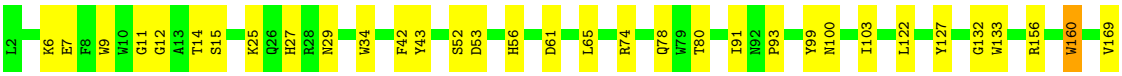
Chain C:

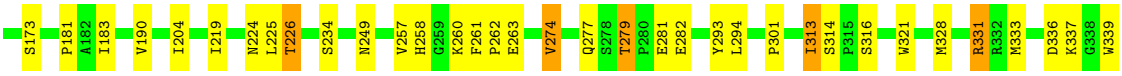


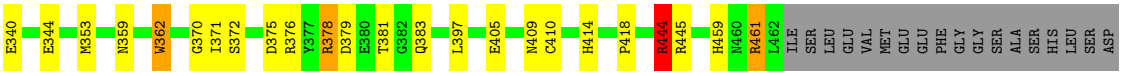


• Molecule 1: BETA-GLUCOSIDASE

Chain D:

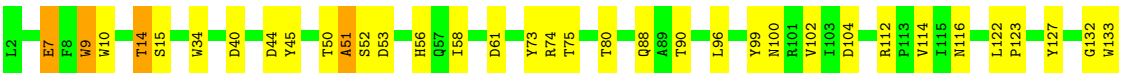


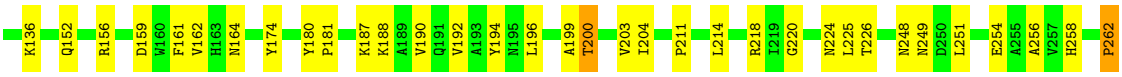


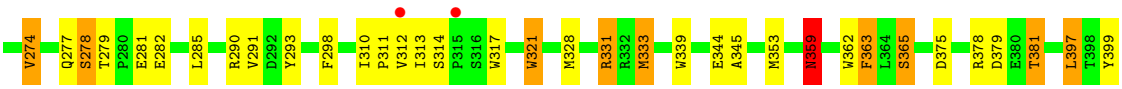


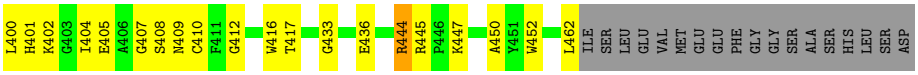
• Molecule 1: BETA-GLUCOSIDASE

Chain E:



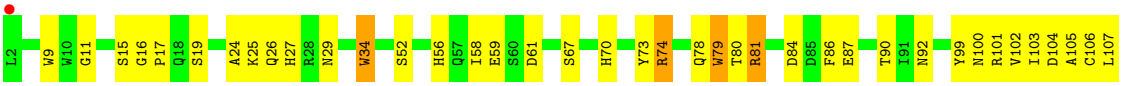


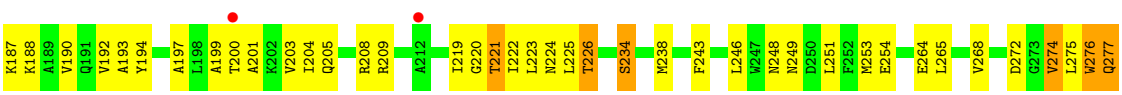


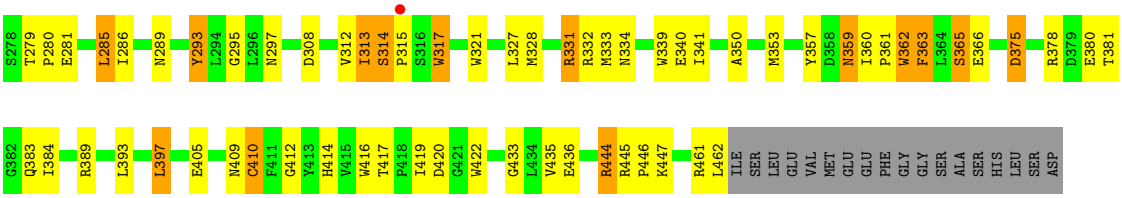


• Molecule 1: BETA-GLUCOSIDASE

Chain F:






S278 T279 P280 E281 L285 I286 N289 Y293 L294 G295 L296 N297 D308 V312 I313 S314 P315 S316 W317 W321 L327 M328 R331 R332 M333 N334 W339 E340 I341 A350 M353 Y357 D358 N359 I360 P361 W362 F363 L364 S365 E366 D375 R378 D379 E380 T381
G382 Q383 I384 R389 L393 L397 E405 N409 C410 F411 G412 Y413 H414 V415 W416 T417 P418 I419 D420 G421 W422 G433 L434 V435 E436 R444 R445 P446 K447 R461 L462 ILE SER LEU GLU VAL MET GLU GLU PHE GLY GLY SER ALA SER HIS LEU SER ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	107.83Å 198.04Å 107.88Å 90.00° 118.51° 90.00°	Depositor
Resolution (Å)	94.75 – 2.51 99.02 – 2.51	Depositor EDS
% Data completeness (in resolution range)	99.5 (94.75-2.51) 99.5 (99.02-2.51)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.32 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.7.0025	Depositor
R, R_{free}	0.205 , 0.261 0.205 , 0.261	Depositor DCC
R_{free} test set	6762 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	47.9	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 14.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.023 for -h-l,k,h 0.023 for l,k,-h-l 0.032 for h,-k,-h-l 0.034 for -h-l,-k,l 0.087 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	22619	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.11% 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

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	0.87	8/3891 (0.2%)	0.88	7/5304 (0.1%)
1	B	0.89	7/3891 (0.2%)	0.88	6/5304 (0.1%)
1	C	0.86	7/3891 (0.2%)	0.89	8/5304 (0.2%)
1	D	0.86	6/3891 (0.2%)	0.85	1/5304 (0.0%)
1	E	0.82	6/3891 (0.2%)	0.84	0/5304
1	F	0.91	8/3891 (0.2%)	0.91	2/5304 (0.0%)
All	All	0.87	42/23346 (0.2%)	0.87	24/31824 (0.1%)

The worst 5 of 42 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	339	TRP	CD2-CE2	6.69	1.49	1.41
1	D	339	TRP	CD2-CE2	6.69	1.49	1.41
1	E	339	TRP	CD2-CE2	6.51	1.49	1.41
1	F	34	TRP	CD2-CE2	6.50	1.49	1.41
1	B	339	TRP	CD2-CE2	6.48	1.49	1.41

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	444	ARG	NE-CZ-NH2	-9.72	115.44	120.30
1	C	444	ARG	NE-CZ-NH1	8.95	124.77	120.30
1	C	444	ARG	NE-CZ-NH2	-7.59	116.51	120.30
1	C	431	ARG	NE-CZ-NH1	7.30	123.95	120.30
1	B	429	LYS	CD-CE-NZ	-6.62	96.47	111.70

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	3767	0	3574	78	0
1	B	3767	0	3574	74	0
1	C	3767	0	3574	79	0
1	D	3767	0	3574	64	0
1	E	3767	0	3574	93	0
1	F	3767	0	3574	155	0
2	A	1	0	0	0	0
2	B	4	0	0	2	0
2	C	3	0	0	0	0
2	D	5	0	0	0	0
2	E	1	0	0	0	0
2	F	3	0	0	0	0
All	All	22619	0	21444	540	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 12.

The worst 5 of 540 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:224:ASN:OD1	1:B:226:THR:HG23	1.35	1.23
1:F:331:ARG:HH11	1:F:331:ARG:CG	1.47	1.22
1:B:331:ARG:CG	1:B:331:ARG:HH11	1.54	1.16
1:A:224:ASN:OD1	1:A:226:THR:HG23	1.46	1.16
1:F:331:ARG:HH11	1:F:331:ARG:HG2	1.10	1.15

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	459/479 (96%)	438 (95%)	20 (4%)	1 (0%)	47	68
1	B	459/479 (96%)	429 (94%)	26 (6%)	4 (1%)	17	31
1	C	459/479 (96%)	425 (93%)	32 (7%)	2 (0%)	34	54
1	D	459/479 (96%)	424 (92%)	34 (7%)	1 (0%)	47	68
1	E	459/479 (96%)	425 (93%)	30 (6%)	4 (1%)	17	31
1	F	459/479 (96%)	409 (89%)	46 (10%)	4 (1%)	17	31
All	All	2754/2874 (96%)	2550 (93%)	188 (7%)	16 (1%)	25	43

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	359	ASN
1	B	375	ASP
1	D	336	ASP
1	E	274	VAL
1	E	359	ASN

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	389/404 (96%)	373 (96%)	16 (4%)	30	55
1	B	389/404 (96%)	372 (96%)	17 (4%)	28	52
1	C	389/404 (96%)	369 (95%)	20 (5%)	24	45
1	D	389/404 (96%)	369 (95%)	20 (5%)	24	45
1	E	389/404 (96%)	370 (95%)	19 (5%)	25	47
1	F	389/404 (96%)	351 (90%)	38 (10%)	8	15
All	All	2334/2424 (96%)	2204 (94%)	130 (6%)	21	40

5 of 130 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	314	SER
1	F	359	ASN
1	C	378	ARG
1	C	365	SER
1	F	365	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 70 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	56	HIS
1	F	100	ASN
1	F	277	GLN
1	C	119	HIS
1	C	100	ASN

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)

There are no ligands in this entry.

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	461/479 (96%)	-0.32	0	100	100	24, 40, 63, 93	1 (0%)
1	B	461/479 (96%)	-0.43	0	100	100	25, 41, 63, 84	1 (0%)
1	C	461/479 (96%)	-0.29	0	100	100	26, 44, 67, 91	1 (0%)
1	D	461/479 (96%)	-0.41	0	100	100	25, 42, 64, 82	1 (0%)
1	E	461/479 (96%)	-0.32	2 (0%)	92	93	28, 48, 70, 102	2 (0%)
1	F	461/479 (96%)	-0.16	4 (0%)	84	86	27, 53, 75, 97	2 (0%)
All	All	2766/2874 (96%)	-0.32	6 (0%)	95	95	24, 45, 69, 102	8 (0%)

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	315	PRO	3.0
1	E	312	VAL	2.6
1	F	200	THR	2.5
1	F	2	LEU	2.5
1	F	315	PRO	2.1

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.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.