Full wwPDB X-ray Structure Validation Report (i)

Jun 19, 2024 – 10:49 AM EDT

PDB ID	:	3W7U
Title	:	Escherichia coli K12 YgjK complexed with galactose
Authors	:	Miyazaki, T.; Kurakata, Y.; Uechi, A.; Yoshida, H.; Kamitori, S.; Sakano, Y.; Nishikawa, A.; Tonozuka, T.
Deposited on	:	2013-03-06
Resolution	:	1.99 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 (i)) were used in the production of this report:

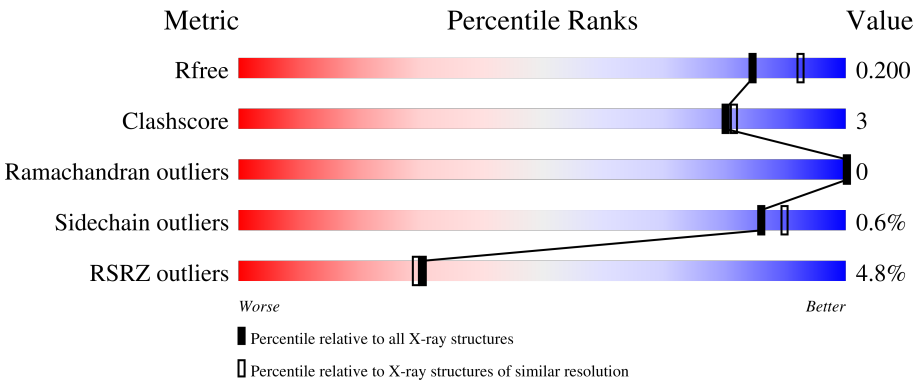
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.99 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	760	3% 94% 6%
1	B	760	7% 93% 7%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 13284 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.

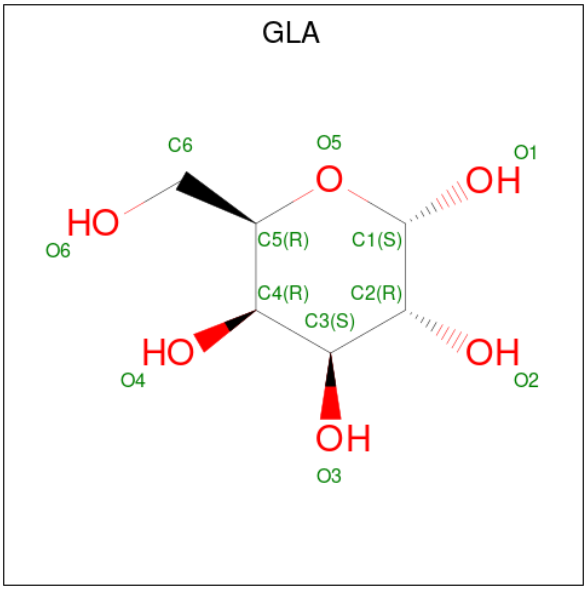
- Molecule 1 is a protein called Uncharacterized protein YgjK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	760	6101	3874	1047	1163	17	0	4	0
1	B	760	6091	3867	1047	1160	17	0	2	0

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
2	A	1	1	1	0	0
2	B	1	1	1	0	0

- Molecule 3 is alpha-D-galactopyranose (three-letter code: GLA) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	12	6	6	0	0
			Total	C	O		
3	B	1	12	6	6	0	0

- Molecule 4 is water.

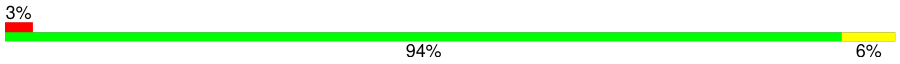
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	590	590	590	0	0
			Total	O		
4	B	476	476	476	0	0

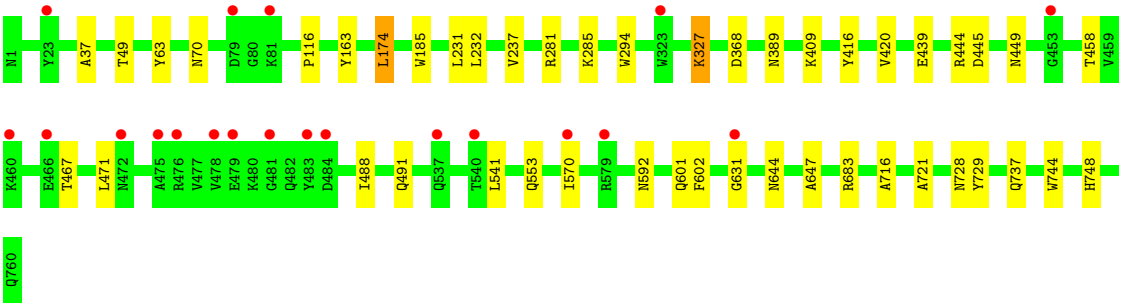
3 Residue-property plots (i)

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: Uncharacterized protein YgjK

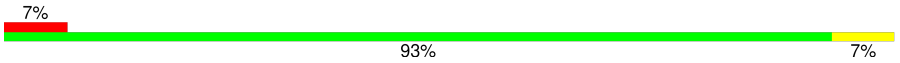
Chain A:

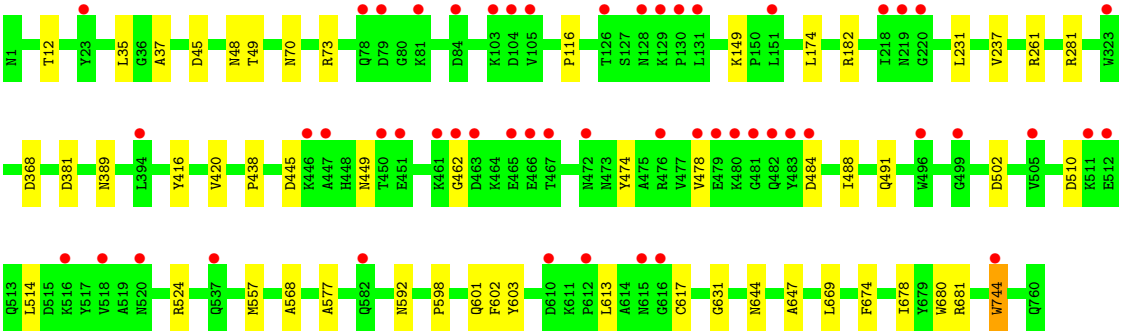




• Molecule 1: Uncharacterized protein YgjK

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.07Å 140.02Å 86.26Å 90.00° 97.66° 90.00°	Depositor
Resolution (Å)	36.25 – 1.99 40.88 – 1.99	Depositor EDS
% Data completeness (in resolution range)	92.4 (36.25-1.99) 92.4 (40.88-1.99)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.14 (at 1.98Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.191 , 0.236 0.192 , 0.200	Depositor DCC
R_{free} test set	10094 reflections (10.93%)	wwPDB-VP
Wilson B-factor (Å^2)	29.7	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 43.0	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	13284	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% 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: CA, GLA

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	0.44	0/6276	0.59	0/8529
1	B	0.43	0/6260	0.59	1/8507 (0.0%)
All	All	0.44	0/12536	0.59	1/17036 (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	744	TRP	CA-CB-CG	5.55	124.25	113.70

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	6101	0	5846	32	0
1	B	6091	0	5833	32	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	12	0	12	0	0
3	B	12	0	12	0	0
4	A	590	0	0	4	0
4	B	476	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	13284	0	11703	64	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 (64) 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:474:TYR:CE1	1:B:488:ILE:HD11	2.04	0.93
1:B:601:GLN:HB3	1:B:647:ALA:HB1	1.73	0.70
1:A:458:THR:OG1	1:A:467[A]:THR:HG22	1.98	0.63
1:A:592:ASN:HD22	1:A:644:ASN:HD22	1.49	0.60
1:B:445:ASP:H	1:B:449:ASN:HD21	1.49	0.58
1:A:285:LYS:CE	4:A:1682:HOH:O	2.53	0.56
1:A:439:GLU:C	1:A:553:GLN:HE22	2.09	0.56
1:A:174:LEU:HD23	1:A:174:LEU:N	2.21	0.56
1:A:416:TYR:O	1:A:420:VAL:HG23	2.06	0.55
1:A:327:LYS:HZ2	1:A:327:LYS:HB3	1.72	0.54
1:B:49:THR:HA	1:B:70:ASN:HD21	1.73	0.52
1:A:683:ARG:HD2	1:A:728:ASN:OD1	2.10	0.52
1:B:488:ILE:HD12	1:B:491:GLN:NE2	2.25	0.52
1:B:182:ARG:HD3	1:B:381:ASP:OD2	2.11	0.51
1:A:368:ASP:HB2	1:A:389:ASN:O	2.11	0.51
1:B:438:PRO:HG2	1:B:557:MET:SD	2.51	0.51
1:B:35:LEU:HD13	1:B:261:ARG:CZ	2.41	0.51
1:A:281:ARG:NH2	4:A:1234:HOH:O	2.44	0.51
1:A:231:LEU:HD13	1:A:237:VAL:HA	1.94	0.50
1:A:602:PHE:CE2	1:A:631:GLY:HA3	2.47	0.50
1:B:231:LEU:HD13	1:B:237:VAL:HA	1.95	0.49
1:B:514:LEU:HD21	1:B:524:ARG:HG3	1.95	0.49
1:B:613:LEU:HD12	1:B:617:CYS:HB2	1.95	0.49
1:B:281:ARG:NH2	4:B:1351:HOH:O	2.44	0.49
1:A:445:ASP:H	1:A:449:ASN:HD21	1.61	0.49
1:B:502:ASP:OD1	1:B:680:TRP:HB2	2.13	0.49
1:B:592:ASN:ND2	1:B:603:TYR:OH	2.45	0.49
1:B:37:ALA:HA	1:B:116:PRO:O	2.13	0.48
1:B:568:ALA:HB3	1:B:577:ALA:HB2	1.96	0.48
1:B:12:THR:HB	1:B:45:ASP:HB3	1.95	0.48
1:A:683:ARG:NH2	1:A:737:GLN:HE21	2.12	0.48
1:A:471:LEU:HD21	4:A:1685:HOH:O	2.13	0.48
1:A:232:LEU:HD12	1:A:232:LEU:N	2.28	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:445:ASP:H	1:B:449:ASN:ND2	2.12	0.47
1:A:716:ALA:HB1	1:A:729:TYR:CE1	2.51	0.46
1:A:49:THR:HA	1:A:70:ASN:HD21	1.81	0.46
1:A:439:GLU:CA	1:A:553:GLN:HE22	2.29	0.46
1:B:474:TYR:CZ	1:B:478:VAL:HG21	2.51	0.45
1:A:683:ARG:CZ	1:A:737:GLN:HG3	2.46	0.45
1:B:602:PHE:CE2	1:B:631:GLY:HA3	2.51	0.45
1:A:285:LYS:HE3	4:A:1682:HOH:O	2.16	0.45
1:B:598:PRO:O	1:B:601:GLN:NE2	2.50	0.45
1:B:368:ASP:HB2	1:B:389:ASN:O	2.17	0.44
1:B:678:ILE:HG21	1:B:681:ARG:HB2	1.99	0.43
1:A:444:ARG:HD2	1:A:541:LEU:HD23	2.01	0.43
1:A:409:LYS:HE2	1:A:570:ILE:HG23	2.01	0.43
1:A:285:LYS:NZ	1:A:721:ALA:O	2.52	0.42
1:A:445:ASP:H	1:A:449:ASN:ND2	2.17	0.42
1:A:488:ILE:HD12	1:A:491:GLN:HE21	1.85	0.42
1:B:474:TYR:HE1	1:B:488:ILE:HD11	1.75	0.42
1:B:416:TYR:O	1:B:420:VAL:HG23	2.20	0.42
1:B:462:GLY:N	1:B:484:ASP:OD2	2.52	0.42
1:B:510:ASP:OD1	1:B:510:ASP:C	2.58	0.42
1:B:478:VAL:O	1:B:478:VAL:HG12	2.20	0.42
1:A:327:LYS:HZ2	1:A:748:HIS:HE1	1.68	0.41
1:A:63:TYR:OH	1:A:185:TRP:HA	2.20	0.41
1:B:488:ILE:HD12	1:B:491:GLN:HE21	1.85	0.41
1:B:48:ASN:O	1:B:73:ARG:NH1	2.53	0.41
1:B:592:ASN:HD22	1:B:644:ASN:HD22	1.68	0.41
1:B:669:LEU:HD23	1:B:674:PHE:CZ	2.56	0.41
1:A:601:GLN:HB3	1:A:647:ALA:HB1	2.02	0.41
1:A:488:ILE:HD12	1:A:491:GLN:NE2	2.35	0.41
1:A:592:ASN:ND2	1:A:644:ASN:HD22	2.16	0.41
1:A:37:ALA:HA	1:A:116:PRO:O	2.21	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	761/760 (100%)	734 (96%)	27 (4%)	0	100	100
1	B	759/760 (100%)	729 (96%)	30 (4%)	0	100	100
All	All	1520/1520 (100%)	1463 (96%)	57 (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	638/634 (101%)	633 (99%)	5 (1%)	81	86
1	B	636/634 (100%)	633 (100%)	3 (0%)	88	92
All	All	1274/1268 (100%)	1266 (99%)	8 (1%)	86	90

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	163	TYR
1	A	174	LEU
1	A	294	TRP
1	A	327	LYS
1	A	744	TRP
1	B	149	LYS
1	B	174	LEU
1	B	744	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (20) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	48	ASN
1	A	70	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	208	ASN
1	A	384	ASN
1	A	449	ASN
1	A	491	GLN
1	A	553	GLN
1	A	592	ASN
1	A	737	GLN
1	A	741	ASN
1	A	748	HIS
1	A	754	ASN
1	B	48	ASN
1	B	70	ASN
1	B	384	ASN
1	B	449	ASN
1	B	491	GLN
1	B	592	ASN
1	B	741	ASN
1	B	748	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	GLA	B	1002	-	12,12,12	0.40	0	17,17,17	0.62	0
3	GLA	A	1002	-	12,12,12	0.47	0	17,17,17	0.57	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
3	GLA	B	1002	-	-	0/2/22/22	0/1/1/1
3	GLA	A	1002	-	-	0/2/22/22	0/1/1/1

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 (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	760/760 (100%)	0.11	20 (2%)	56	54	21, 28, 43, 67	0
1	B	760/760 (100%)	0.35	53 (6%)	16	15	19, 32, 53, 70	0
All	All	1520/1520 (100%)	0.23	73 (4%)	30	29	19, 30, 50, 70	0

All (73) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	481	GLY	6.3
1	A	475	ALA	6.1
1	A	479	GLU	5.2
1	B	23	TYR	4.6
1	A	481	GLY	4.6
1	B	479	GLU	4.6
1	B	478	VAL	4.4
1	B	465	GLU	4.2
1	A	23	TYR	4.0
1	A	472	ASN	4.0
1	B	482	GLN	3.9
1	B	472	ASN	3.8
1	B	616	GLY	3.7
1	B	103	LYS	3.5
1	B	582	GLN	3.5
1	B	466	GLU	3.4
1	B	480	LYS	3.4
1	B	483	TYR	3.3
1	B	105	VAL	3.3
1	B	151	LEU	3.2
1	B	511	LYS	3.2
1	A	323	TRP	3.2
1	B	219	ASN	3.1
1	B	520	ASN	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	128	ASN	3.1
1	B	78	GLN	3.1
1	B	467	THR	3.1
1	B	512	GLU	3.0
1	B	484	ASP	3.0
1	A	537	GLN	3.0
1	B	79	ASP	3.0
1	A	476	ARG	2.9
1	B	537	GLN	2.9
1	A	570	ILE	2.9
1	B	461	LYS	2.8
1	A	81	LYS	2.8
1	B	612	PRO	2.7
1	B	220	GLY	2.7
1	B	129	LYS	2.7
1	B	104	ASP	2.7
1	B	463	ASP	2.7
1	B	451	GLU	2.7
1	A	79	ASP	2.7
1	B	131	LEU	2.6
1	B	462	GLY	2.6
1	B	476	ARG	2.6
1	B	130	PRO	2.6
1	A	478	VAL	2.6
1	A	483	TYR	2.6
1	B	516	LYS	2.5
1	B	610	ASP	2.4
1	B	505	VAL	2.4
1	B	81	LYS	2.3
1	B	323	TRP	2.3
1	B	615	ASN	2.3
1	B	499	GLY	2.3
1	B	126	THR	2.3
1	A	540	THR	2.3
1	B	84	ASP	2.3
1	A	579	ARG	2.2
1	A	484	ASP	2.2
1	B	450	THR	2.1
1	A	631	GLY	2.1
1	B	744	TRP	2.1
1	A	460	LYS	2.1
1	B	218	ILE	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	466	GLU	2.1
1	B	394	LEU	2.1
1	B	446	LYS	2.1
1	B	518	VAL	2.0
1	B	496	TRP	2.0
1	A	453	GLY	2.0
1	B	447	ALA	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, 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
3	GLA	B	1002	12/12	0.91	0.16	38,39,40,41	0
3	GLA	A	1002	12/12	0.92	0.14	32,33,34,37	0
2	CA	A	1001	1/1	0.98	0.07	28,28,28,28	0
2	CA	B	1001	1/1	0.98	0.06	37,37,37,37	0

6.5 Other polymers (i)

There are no such residues in this entry.