Full wwPDB X-ray Structure Validation Report (i)

Aug 7, 2020 – 05:34 PM BST

PDB ID	:	5KSA
Title	:	Bel602-DQ8.5-glia-gamma1 complex
Authors	:	Petersen, J.; Rossjohn, J.; Reid, H.H.
Deposited on	:	2016-07-08
Resolution	:	2.00 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 following versions of software and data (see references (1)) were used in the production of this report:

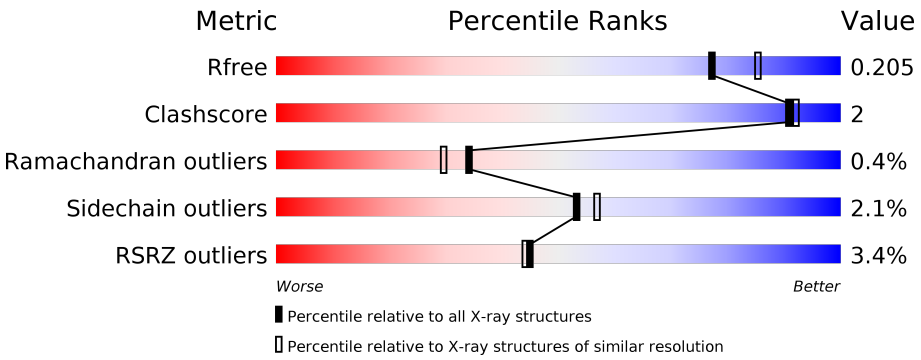
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.00 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 on 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	191	% 90% 6% 5%
2	B	225	4% 81% • 16%
3	C	206	3% 86% 7% • 5%
4	D	243	5% 94% 5%
5	J	11	91% 9%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 7226 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 HLA class II histocompatibility antigen, DQ alpha 1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	182	1453	935	237	279	2	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	SER	CYS	conflict	UNP P01909
A	182	THR	-	expression tag	UNP P01909
A	183	SER	-	expression tag	UNP P01909
A	184	GLY	-	expression tag	UNP P01909
A	185	ASP	-	expression tag	UNP P01909
A	186	ASP	-	expression tag	UNP P01909
A	187	ASP	-	expression tag	UNP P01909
A	188	ASP	-	expression tag	UNP P01909
A	189	LYS	-	expression tag	UNP P01909

- Molecule 2 is a protein called HLA class II histocompatibility antigen, DQ beta 1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	190	1569	992	277	293	7	0	3	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	SER	-	linker	PDB ?
B	-12	GLY	-	linker	PDB ?
B	-11	GLY	-	linker	PDB ?
B	-10	SER	-	linker	PDB ?
B	-9	ILE	-	linker	PDB ?
B	-8	GLU	-	linker	PDB ?
B	-7	GLY	-	linker	PDB ?

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-6	ARG	-	linker	PDB ?
B	-5	GLY	-	linker	PDB ?
B	-4	GLY	-	linker	PDB ?
B	-3	SER	-	linker	PDB ?
B	-2	GLY	-	linker	PDB ?
B	-1	ALA	-	linker	PDB ?
B	0	SER	-	linker	PDB ?
B	193	THR	-	expression tag	UNP O19707
B	194	GLY	-	expression tag	UNP O19707
B	195	GLY	-	expression tag	UNP O19707
B	196	ASP	-	expression tag	UNP O19707
B	197	ASP	-	expression tag	UNP O19707
B	198	ASP	-	expression tag	UNP O19707
B	199	ASP	-	expression tag	UNP O19707
B	200	LYS	-	expression tag	UNP O19707

- Molecule 3 is a protein called Bel602 alpha TRAV20*01.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	195	1526	960	251	307	8	0	0	0

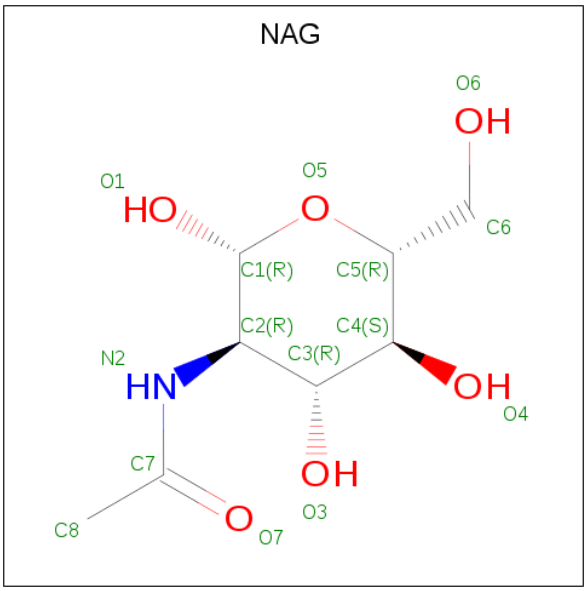
- Molecule 4 is a protein called Bel602 beta TRBV9*01.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	242	1940	1225	337	373	5	0	2	0

- Molecule 5 is a protein called DQ8.5-glia-gamma1 peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
5	J	11	90	55	15	20	0	0	0

- Molecule 6 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	A	1	14	8	1	5	0	0

- Molecule 7 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
7	A	1	1	1	0	0

- Molecule 8 is water.

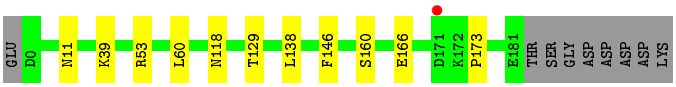
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	170	170	170	0	0
			Total	O		
8	B	142	142	142	0	0
			Total	O		
8	C	140	140	140	0	0
			Total	O		
8	D	167	167	167	0	0
			Total	O		
8	J	14	14	14	0	0

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.

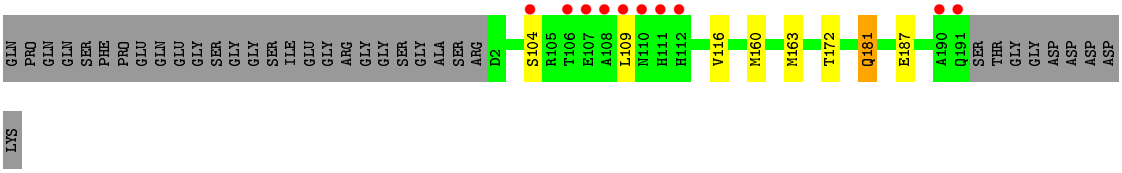
- Molecule 1: HLA class II histocompatibility antigen, DQ alpha 1 chain

Chain A:



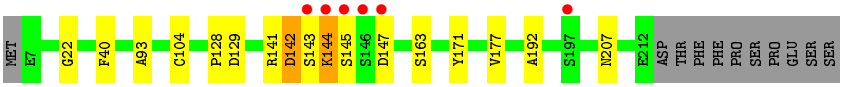
- Molecule 2: HLA class II histocompatibility antigen, DQ beta 1 chain

Chain B:



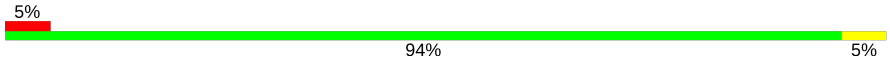
- Molecule 3: Bel602 alpha TRAV20*01

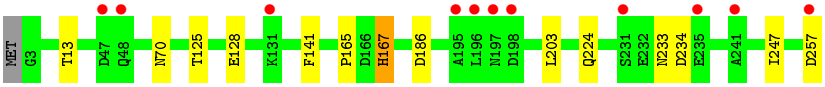
Chain C:



- Molecule 4: Bel602 beta TRBV9*01

Chain D:





- Molecule 5: DQ8.5-glia-gamma1 peptide

Chain J:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.55Å 98.83Å 80.17Å 90.00° 95.69° 90.00°	Depositor
Resolution (Å)	45.75 – 2.00 42.35 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.75-2.00) 100.0 (42.35-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.70 (at 2.00Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.169 , 0.206 0.170 , 0.205	Depositor DCC
R_{free} test set	3290 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	31.3	Xtriage
Anisotropy	0.407	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 53.0	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	0.97	EDS
Total number of atoms	7226	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 4.95% 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, NAG

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	0/1495	0.62	0/2042
2	B	0.52	0/1618	0.66	0/2208
3	C	0.50	0/1558	0.73	1/2109 (0.0%)
4	D	0.46	0/2000	0.62	0/2724
5	J	0.68	0/92	0.74	0/124
All	All	0.50	0/6763	0.66	1/9207 (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})
3	C	144	LYS	C-N-CA	5.04	134.31	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	1453	0	1401	5	0
2	B	1569	0	1501	3	0
3	C	1526	0	1459	10	0
4	D	1940	0	1839	5	0
5	J	90	0	79	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	14	0	13	0	0
7	A	1	0	0	0	0
8	A	170	0	0	0	0
8	B	142	0	0	0	0
8	C	140	0	0	0	0
8	D	167	0	0	0	0
8	J	14	0	0	0	0
All	All	7226	0	6292	22	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 (22) 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)
3:C:141:ARG:HG3	3:C:142:ASP:H	1.40	0.86
2:B:181:GLN:CD	2:B:181:GLN:H	1.94	0.70
3:C:128:PRO:HG3	3:C:177:VAL:HG11	1.78	0.64
3:C:163:SER:H	3:C:207:ASN:CB	2.15	0.59
2:B:172:THR:HG22	2:B:187:GLU:HG2	1.83	0.59
4:D:125:THR:HG21	4:D:165:PRO:HB3	1.83	0.59
3:C:141:ARG:HG3	3:C:142:ASP:N	2.16	0.57
1:A:39:LYS:HG2	1:A:60:LEU:HD11	1.93	0.51
2:B:116:VAL:HG22	2:B:160:MET:HG3	1.94	0.49
3:C:141:ARG:HH12	4:D:141:PHE:HE2	1.61	0.49
3:C:163:SER:OG	3:C:207:ASN:HB3	2.13	0.49
4:D:224:GLN:HG3	4:D:247:ILE:HG23	1.94	0.48
1:A:118:ASN:HB2	1:A:166:GLU:HB2	1.96	0.47
1:A:138:LEU:HD12	1:A:146:PHE:CE2	2.49	0.47
3:C:145:SER:O	3:C:147:ASP:HA	2.17	0.44
4:D:125:THR:HG1	4:D:167[A]:HIS:CE1	2.36	0.43
1:A:53:ARG:O	5:J:-1:GLN:HB3	2.18	0.43
3:C:40:PHE:O	3:C:104:CYS:HA	2.21	0.41
3:C:171:TYR:O	3:C:192:ALA:HA	2.20	0.41
1:A:166:GLU:HG2	1:A:173:PRO:HB3	2.02	0.41
3:C:22:GLY:HA2	3:C:93:ALA:HA	2.03	0.41
4:D:186:ASP:HB2	4:D:203:LEU:HD12	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	180/191 (94%)	177 (98%)	3 (2%)	0	100	100
2	B	191/225 (85%)	182 (95%)	8 (4%)	1 (0%)	29	23
3	C	193/206 (94%)	182 (94%)	9 (5%)	2 (1%)	15	9
4	D	242/243 (100%)	238 (98%)	4 (2%)	0	100	100
5	J	9/11 (82%)	9 (100%)	0	0	100	100
All	All	815/876 (93%)	788 (97%)	24 (3%)	3 (0%)	34	30

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	142	ASP
3	C	143	SER
2	B	109	LEU

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	166/174 (95%)	163 (98%)	3 (2%)	59	63
2	B	172/200 (86%)	169 (98%)	3 (2%)	60	65
3	C	170/185 (92%)	168 (99%)	2 (1%)	71	76
4	D	212/211 (100%)	204 (96%)	8 (4%)	33	31
5	J	10/10 (100%)	10 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	730/780 (94%)	714 (98%)	16 (2%)	53 55

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	11	ASN
1	A	129	THR
1	A	160	SER
2	B	104	SER
2	B	163	MET
2	B	181	GLN
3	C	129	ASP
3	C	144	LYS
4	D	13	THR
4	D	70	ASN
4	D	128	GLU
4	D	167[A]	HIS
4	D	167[B]	HIS
4	D	233	ASN
4	D	234	ASP
4	D	257	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	$\#\|Z\| > 2$
6	NAG	A	201	1	14,14,15	0.30	0	17,19,21	0.96	1 (5%)

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
6	NAG	A	201	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	A	201	NAG	C1-O5-C5	3.79	117.32	112.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	201	NAG	C4-C5-C6-O6
6	A	201	NAG	O5-C5-C6-O6

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	182/191 (95%)	-0.30	1 (0%) 91 90	22, 38, 67, 88	0
2	B	190/225 (84%)	0.23	10 (5%) 26 25	23, 36, 77, 119	0
3	C	195/206 (94%)	-0.11	6 (3%) 49 48	26, 45, 78, 99	0
4	D	242/243 (99%)	-0.02	11 (4%) 33 32	28, 47, 86, 110	0
5	J	11/11 (100%)	-0.10	0 100 100	24, 27, 56, 73	0
All	All	820/876 (93%)	-0.04	28 (3%) 45 44	22, 41, 81, 119	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	106	THR	10.0
2	B	109	LEU	8.5
2	B	108	ALA	8.2
2	B	111	HIS	7.6
2	B	110	ASN	6.2
2	B	112	HIS	5.5
3	C	145	SER	5.1
2	B	191	GLN	4.4
2	B	107	GLU	4.2
3	C	143	SER	3.3
4	D	257	ASP	3.3
4	D	198	ASP	3.1
4	D	197	ASN	3.1
2	B	104	SER	3.1
3	C	197	SER	3.0
3	C	146	SER	2.9
2	B	190	ALA	2.9
4	D	47	ASP	2.9
3	C	147	ASP	2.8
3	C	144	LYS	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
4	D	196	LEU	2.7
4	D	195	ALA	2.7
4	D	48	GLN	2.6
4	D	231	SER	2.5
4	D	131	LYS	2.4
4	D	241	ALA	2.2
4	D	235	GLU	2.1
1	A	171	ASP	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
6	NAG	A	201	14/15	0.86	0.21	56,69,82,82	0
7	CA	A	202	1/1	0.95	0.04	59,59,59,59	0

6.5 Other polymers (i)

There are no such residues in this entry.