



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 11, 2022 – 02:08 PM EST

PDB ID : 7KPJ
Title : Crystal structure of Ruminococcus gnavus immunoglobulin binding protein in complex with 338E6 Fab
Authors : Borowska, M.T.; Adams, E.J.
Deposited on : 2020-11-11
Resolution : 2.10 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.25
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) : 2.25

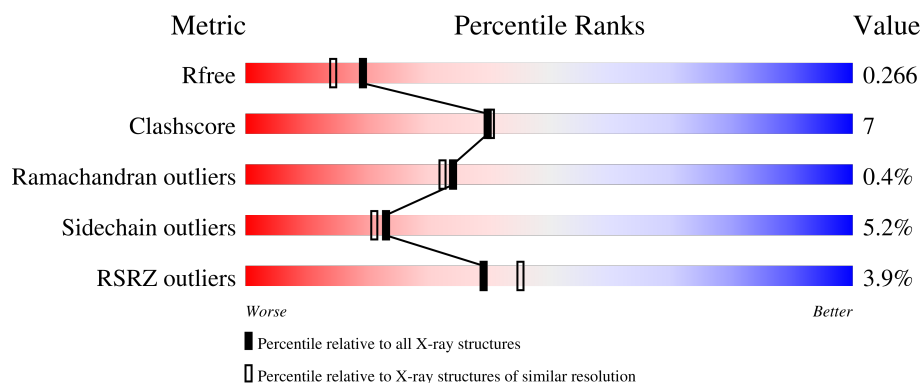
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 $\leq 5\%$. 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	B	214	<div> <div>2%</div> <div>82%</div> <div>15%</div> <div>.</div> </div>
1	D	214	<div> <div>9%</div> <div>71%</div> <div>20%</div> <div>6%</div> </div>
2	E	272	<div> <div>63%</div> <div>16%</div> <div>20%</div> </div>
2	F	272	<div> <div>6%</div> <div>61%</div> <div>16%</div> <div>21%</div> </div>
3	A	224	<div> <div>82%</div> <div>11%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	C	224	<div><div></div><div>3%</div><div>75%</div><div>17%</div><div>• 7%</div></div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 9999 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 338E6 Fab light chain kappa.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	213	Total	C	N	O	S	0	0	0
			1643	1024	276	338	5			
1	D	201	Total	C	N	O	S	0	0	0
			1535	959	253	318	5			

- Molecule 2 is a protein called GRAM_POS_ANCHORING domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	218	Total	C	N	O	S	0	0	0
			1699	1070	269	355	5			
2	F	215	Total	C	N	O	S	0	0	0
			1674	1054	265	350	5			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-6	HIS	-	expression tag	UNP A0A2N5NGA8
E	-5	HIS	-	expression tag	UNP A0A2N5NGA8
E	-4	HIS	-	expression tag	UNP A0A2N5NGA8
E	-3	HIS	-	expression tag	UNP A0A2N5NGA8
E	-2	HIS	-	expression tag	UNP A0A2N5NGA8
E	-1	HIS	-	expression tag	UNP A0A2N5NGA8
E	1	ALA	SER	conflict	UNP A0A2N5NGA8
E	2	GLU	VAL	conflict	UNP A0A2N5NGA8
F	-6	HIS	-	expression tag	UNP A0A2N5NGA8
F	-5	HIS	-	expression tag	UNP A0A2N5NGA8
F	-4	HIS	-	expression tag	UNP A0A2N5NGA8
F	-3	HIS	-	expression tag	UNP A0A2N5NGA8
F	-2	HIS	-	expression tag	UNP A0A2N5NGA8
F	-1	HIS	-	expression tag	UNP A0A2N5NGA8
F	1	ALA	SER	conflict	UNP A0A2N5NGA8
F	2	GLU	VAL	conflict	UNP A0A2N5NGA8

- Molecule 3 is a protein called 338E6 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	210	Total	C	N	O	S	0	0	0
			1593	1012	261	313	7			
3	C	208	Total	C	N	O	S	0	0	0
			1580	1004	258	311	7			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Ca	0	0
			1	1		
4	F	1	Total	Ca	0	0
			1	1		

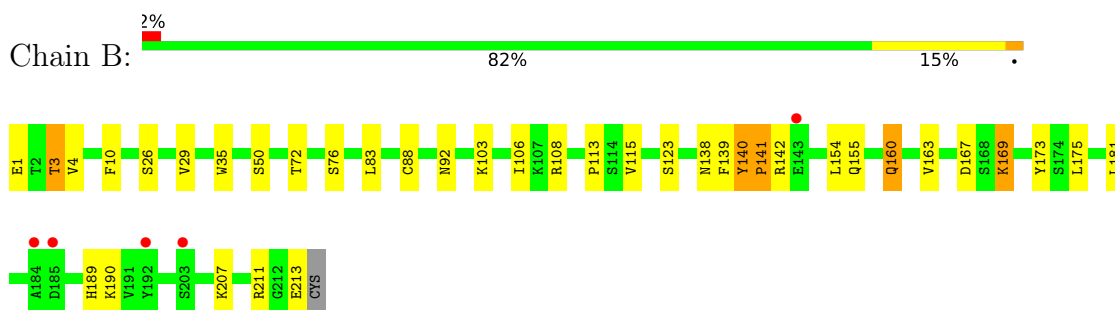
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	50	Total	O	0	0
			50	50		
5	E	85	Total	O	0	0
			85	85		
5	F	22	Total	O	0	0
			22	22		
5	A	86	Total	O	0	0
			86	86		
5	C	14	Total	O	0	0
			14	14		
5	D	16	Total	O	0	0
			16	16		

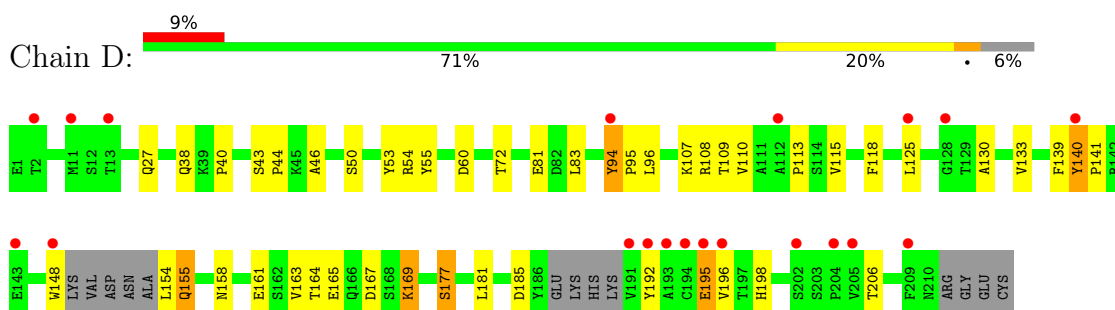
3 Residue-property plots

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.

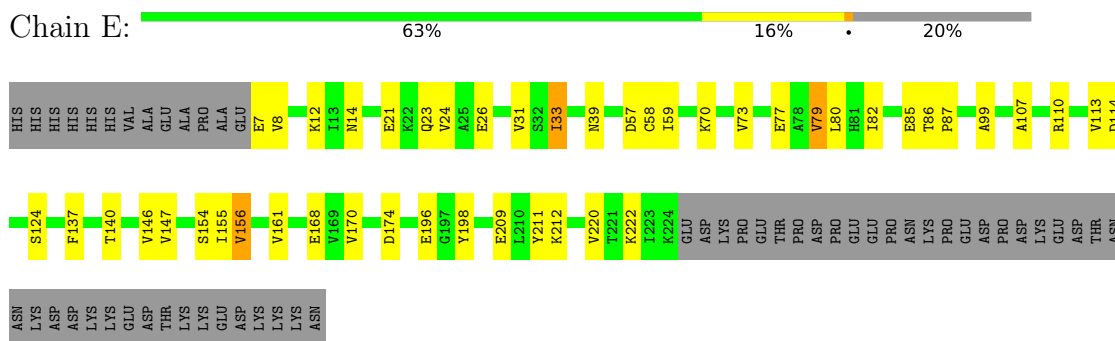
- Molecule 1: 338E6 Fab light chain kappa



- Molecule 1: 338E6 Fab light chain kappa

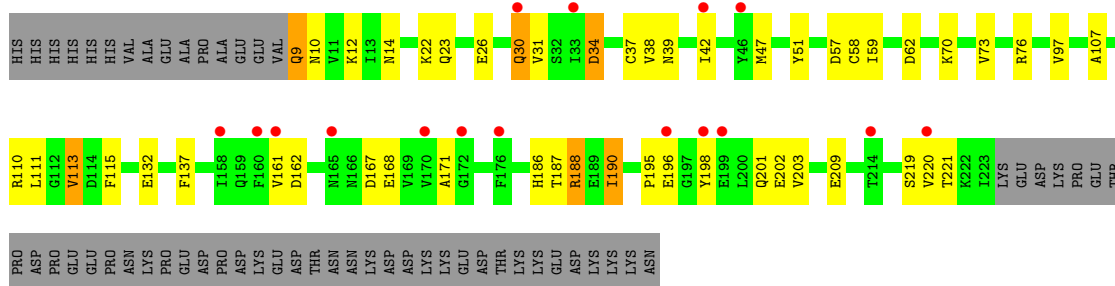


- Molecule 2: GRAM_POS_ANCHORING domain-containing protein



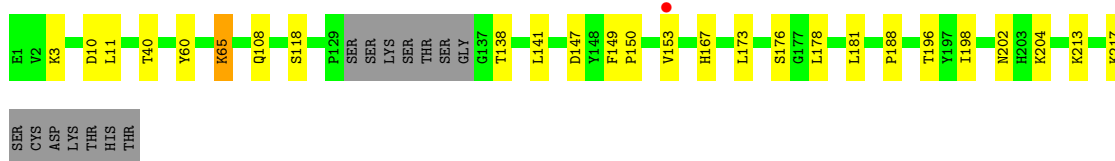
- Molecule 2: GRAM_POS_ANCHORING domain-containing protein





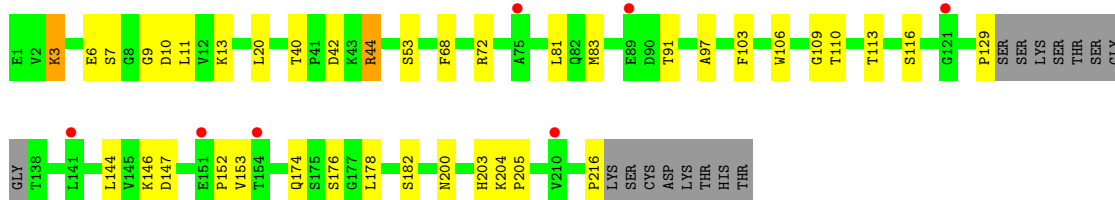
• Molecule 3: 338E6 Fab heavy chain

Chain A: 82% 11% 6%



• Molecule 3: 338E6 Fab heavy chain

Chain C: 3% 75% 17% 7%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	85.24Å 102.19Å 118.28Å 90.00° 102.05° 90.00°	Depositor
Resolution (Å)	83.36 – 2.10 83.36 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.1 (83.36-2.10) 99.1 (83.36-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.26 (at 2.10Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.229 , 0.266 0.230 , 0.266	Depositor DCC
R_{free} test set	1997 reflections (1.75%)	wwPDB-VP
Wilson B-factor (Å ²)	34.9	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 38.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	9999	wwPDB-VP
Average B, all atoms (Å ²)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.17% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: CA

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	B	0.46	0/1677	0.61	0/2276
1	D	0.45	0/1566	0.61	0/2128
2	E	0.60	1/1728 (0.1%)	0.67	0/2351
2	F	0.48	0/1703	0.60	0/2318
3	A	0.49	0/1633	0.67	0/2219
3	C	0.41	0/1620	0.58	0/2203
All	All	0.48	1/9927 (0.0%)	0.62	0/13495

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1
3	A	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	124	SER	CA-CB	-6.72	1.42	1.52

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	150	PRO	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	140	TYR	Peptide

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	B	1643	0	1583	23	0
1	D	1535	0	1459	27	0
2	E	1699	0	1628	27	0
2	F	1674	0	1600	32	0
3	A	1593	0	1559	13	0
3	C	1580	0	1543	25	0
4	E	1	0	0	0	0
4	F	1	0	0	0	0
5	A	86	0	0	1	0
5	B	50	0	0	1	0
5	C	14	0	0	0	0
5	D	16	0	0	0	0
5	E	85	0	0	1	0
5	F	22	0	0	1	0
All	All	9999	0	9372	139	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 7.

The worst 5 of 139 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)
2:F:202:GLU:HG3	2:F:219:SER:H	1.51	0.76
2:F:39:ASN:HB3	2:F:113:VAL:HG12	1.69	0.74
3:C:9:GLY:H	3:C:110:THR:HG21	1.53	0.71
2:E:8:VAL:HG22	2:E:31:VAL:O	1.90	0.71
1:B:163:VAL:HG22	1:B:175:LEU:HD12	1.74	0.69

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	B	211/214 (99%)	200 (95%)	10 (5%)	1 (0%)	29	26
1	D	195/214 (91%)	184 (94%)	9 (5%)	2 (1%)	15	11
2	E	216/272 (79%)	208 (96%)	8 (4%)	0	100	100
2	F	213/272 (78%)	205 (96%)	8 (4%)	0	100	100
3	A	206/224 (92%)	200 (97%)	5 (2%)	1 (0%)	29	26
3	C	204/224 (91%)	193 (95%)	10 (5%)	1 (0%)	29	26
All	All	1245/1420 (88%)	1190 (96%)	50 (4%)	5 (0%)	34	32

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	141	PRO
3	A	149	PHE
1	D	95	PRO
3	C	152	PRO
1	D	140	TYR

5.3.2 Protein sidechains

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	B	188/189 (100%)	180 (96%)	8 (4%)	29	29
1	D	175/189 (93%)	164 (94%)	11 (6%)	18	15
2	E	190/241 (79%)	180 (95%)	10 (5%)	22	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	F	187/241 (78%)	174 (93%)	13 (7%)	15	12
3	A	179/192 (93%)	170 (95%)	9 (5%)	24	23
3	C	178/192 (93%)	172 (97%)	6 (3%)	37	39
All	All	1097/1244 (88%)	1040 (95%)	57 (5%)	23	21

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	190	ILE
1	D	177	SER
3	A	138	THR
1	D	169	LYS
1	D	81	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	174	GLN
3	C	174	GLN
1	D	155	GLN
1	B	155	GLN
2	E	14	ASN

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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	B	213/214 (99%)	0.52	5 (2%) 60 65	23, 43, 76, 97	0
1	D	201/214 (93%)	0.80	20 (9%) 7 9	36, 56, 93, 107	0
2	E	218/272 (80%)	0.57	0 100 100	26, 35, 52, 86	0
2	F	215/272 (79%)	0.69	16 (7%) 14 18	41, 56, 83, 99	0
3	A	210/224 (93%)	0.59	1 (0%) 91 92	23, 34, 53, 75	0
3	C	208/224 (92%)	0.49	7 (3%) 45 51	37, 52, 72, 96	0
All	All	1265/1420 (89%)	0.61	49 (3%) 39 45	23, 46, 82, 107	0

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	198	TYR	5.0
1	D	148	TRP	4.2
1	D	192	TYR	4.1
2	F	158	ILE	3.8
1	D	202	SER	3.3

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
4	CA	E	301	1/1	0.94	0.15	29,29,29,29	0
4	CA	F	301	1/1	0.95	0.14	53,53,53,53	0

6.5 Other polymers [i](#)

There are no such residues in this entry.