wwPDB X-ray Structure Validation Summary Report (i)

Jan 8, 2024 – 01:26 pm GMT

PDB ID	:	5OWP
Title	:	Crystal structure of glycopeptide "GVTSAfPDT*RPAP" in complex with scFv-SM3
Authors	:	Bermejo, I.A.; Albuquerque, I.S.; Somovilla, V.J.; Martinez-Saez, N.; Castro-Lopez, J.; Garcia-Martin, F.; Hinou, H.; Nishimura, S.; Jimenez-Barbero, J.; Asensio, J.L.; Avenoza, A.; Busto, J.H.; Hurtado-Guerrero, R.; Peregrina, J.M.; Bernardes, G.J.; Corzana, F.
Deposited on	:	2017-09-02
Resolution	:	1.85 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:

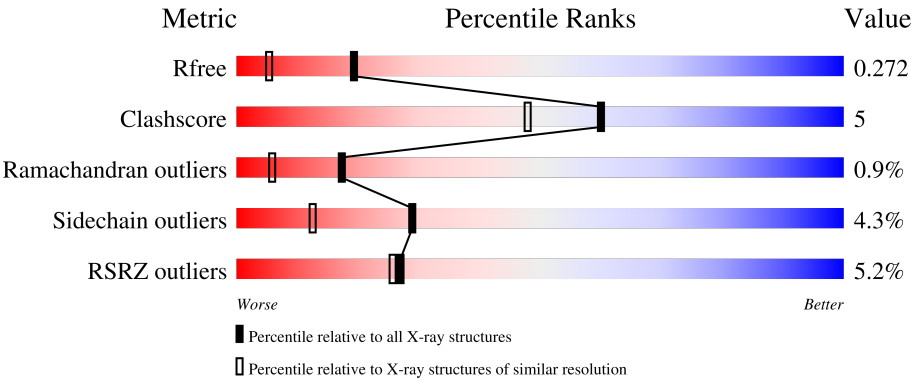
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	H	244	4% 82% 8% · 7%
2	D	8	38% 75% 25%

Validation Pipeline (wwPDB-VP) : 2.36

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 1946 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 Ig heavy chain V-III region J606,Ig lambda-1 chain V region H2020.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	226	Total	C	N	O	S	0	0	0
			1720	1083	292	338	7			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	1	GLN	-	expression tag	UNP P01801
H	2	VAL	-	expression tag	UNP P01801
H	3	GLN	-	expression tag	UNP P01801
H	4	LEU	-	expression tag	UNP P01801
H	5	GLN	-	expression tag	UNP P01801
H	100	GLY	-	insertion	UNP P01801
H	101	VAL	-	insertion	UNP P01801
H	102	GLY	THR	conflict	UNP P01801
H	103	GLN	GLY	conflict	UNP P01801
H	114	THR	LEU	conflict	UNP P01801
H	116	THR	-	linker	UNP P01801
H	117	VAL	-	linker	UNP P01801
H	118	SER	-	linker	UNP P01801
H	990	SER	-	linker	UNP P01801
H	991	SER	-	linker	UNP P01801
H	992	SER	-	linker	UNP P01801
H	993	GLY	-	linker	UNP P01801
H	994	GLY	-	linker	UNP P01801
H	995	GLY	-	linker	UNP P01801
H	996	GLY	-	linker	UNP P01801
H	997	SER	-	linker	UNP P01801
H	998	GLY	-	linker	UNP P01801
H	999	GLY	-	linker	UNP P01801
H	1000	GLY	-	linker	UNP P01801
H	1001	GLY	-	linker	UNP P01801
H	1002	GLY	-	linker	UNP P01801

Continued on next page...

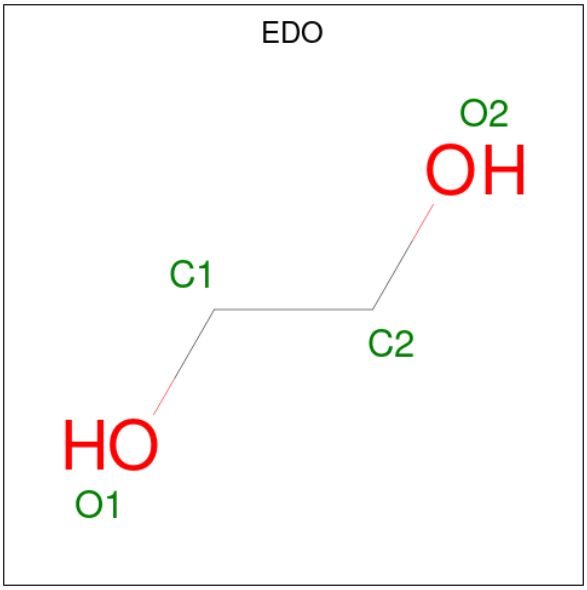
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	1003	SER	-	linker	UNP P01801
H	1004	SER	-	linker	UNP P01801
H	1005	GLY	-	linker	UNP P01801
H	1006	SER	-	linker	UNP P01801
H	1008	ASP	GLN	conflict	UNP P01726
H	1009	ILE	ALA	conflict	UNP P01726
H	1032	SER	THR	conflict	UNP P01726
H	1039	SER	GLY	conflict	UNP P01726

- Molecule 2 is a protein called 5,6-DIHYDRO-BENZO[H]CINNOLIN-3-YLAMINE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	8	Total	C	F	N	O	0	0	0
			57	33	1	11	12			

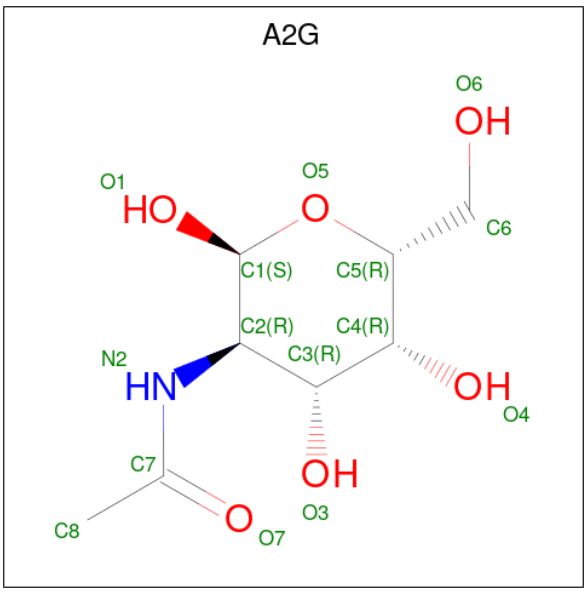
- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	H	1	Total	C	O	0	0
			4	2	2		
3	H	1	Total	C	O	0	0
			4	2	2		
3	H	1	Total	C	O	0	0
			4	2	2		

- Molecule 4 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G)

(formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	1	Total C N O 14 8 1 5	0	0

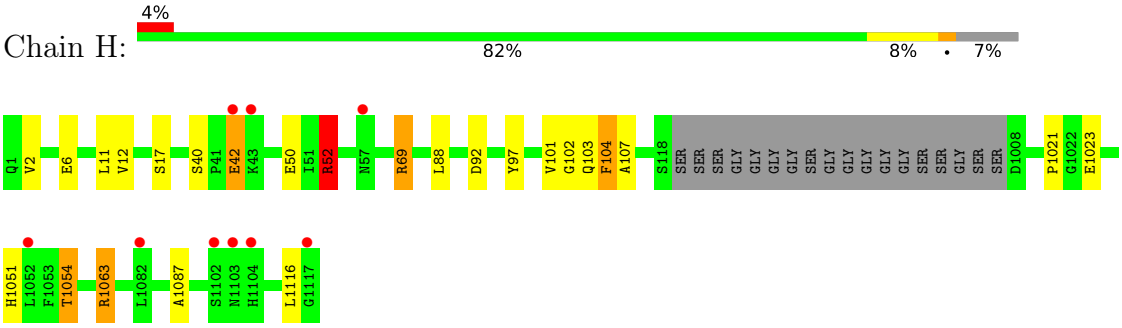
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	H	141	Total O 141 141	0	0
5	D	2	Total O 2 2	0	0

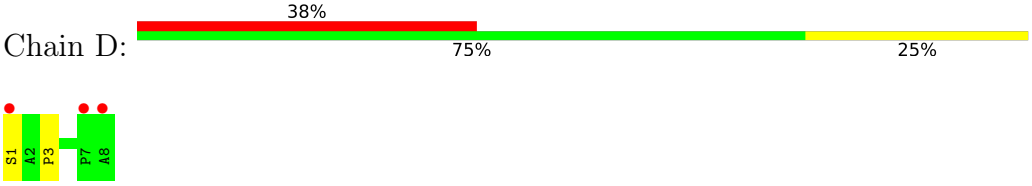
3 Residue-property plots (i)

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: Ig heavy chain V-III region J606,Ig lambda-1 chain V region H2020



• Molecule 2: 5,6-DIHYDRO-BENZO[H]CINNOLIN-3-YLAMINE



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	35.40Å 68.15Å 90.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	90.36 – 1.85 19.12 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.7 (90.36-1.85) 99.8 (19.12-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.99 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.194 , 0.256 0.202 , 0.272	Depositor DCC
R_{free} test set	562 reflections (2.91%)	wwPDB-VP
Wilson B-factor (Å^2)	24.1	Xtriage
Anisotropy	1.152	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 45.9	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	0.96	EDS
Total number of atoms	1946	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.70% 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: 4FB, EDO, A2G

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	H	0.92	1/1759 (0.1%)	0.98	5/2393 (0.2%)
2	D	1.17	0/48	1.61	0/63
All	All	0.93	1/1807 (0.1%)	1.01	5/2456 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	104	PHE	C-N	16.54	1.72	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	69	ARG	NE-CZ-NH2	-10.32	115.14	120.30
1	H	1063	ARG	NE-CZ-NH1	9.94	125.27	120.30
1	H	69	ARG	NE-CZ-NH1	9.65	125.13	120.30
1	H	1063	ARG	NE-CZ-NH2	-8.58	116.01	120.30
1	H	52	ARG	NE-CZ-NH1	7.66	124.13	120.30

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	H	1720	0	1645	15	0
2	D	57	0	53	0	0
3	H	12	0	18	3	0
4	D	14	0	12	0	0
5	D	2	0	0	0	0
5	H	141	0	0	3	0
All	All	1946	0	1728	16	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 5.

The worst 5 of 16 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:H:104:PHE:C	1:H:107:ALA:N	1.72	1.43
1:H:69:ARG:NH2	1:H:92:ASP:OD2	2.23	0.66
1:H:1054:THR:HG23	5:H:1391:HOH:O	2.00	0.60
1:H:104:PHE:C	1:H:107:ALA:CA	2.66	0.58
1:H:104:PHE:CA	1:H:107:ALA:N	2.67	0.51

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	H	222/244 (91%)	211 (95%)	9 (4%)	2 (1%)	17 6
2	D	5/8 (62%)	4 (80%)	1 (20%)	0	100 100
All	All	227/252 (90%)	215 (95%)	10 (4%)	2 (1%)	17 6

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	101	VAL
1	H	2	VAL

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	H	183/192 (95%)	176 (96%)	7 (4%)	33 16
2	D	5/5 (100%)	4 (80%)	1 (20%)	1 0
All	All	188/197 (95%)	180 (96%)	8 (4%)	29 12

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	1	SER
1	H	1116	LEU
1	H	1054	THR
1	H	52	ARG
1	H	1063	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	H	39	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue 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	$\#\|Z\| > 2$
2	4FB	D	3	2	6,8,9	1.25	1 (16%)	3,10,12	1.40	1 (33%)

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
2	4FB	D	3	2	-	0/0/11/13	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	3	4FB	FGX-CGX	-2.42	1.31	1.40

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	3	4FB	O-C-CA	-2.28	118.81	124.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

4 ligands are 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	$\#\|Z\| > 2$
4	A2G	D	101	2	14,14,15	0.44	0	17,19,21	0.82	0
3	EDO	H	1201	-	3,3,3	0.58	0	2,2,2	0.20	0
3	EDO	H	1202	-	3,3,3	0.51	0	2,2,2	0.42	0
3	EDO	H	1203	-	3,3,3	0.30	0	2,2,2	0.18	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
4	A2G	D	101	2	-	1/6/23/26	0/1/1/1
3	EDO	H	1201	-	-	1/1/1/1	-
3	EDO	H	1202	-	-	1/1/1/1	-
3	EDO	H	1203	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	H	1202	EDO	O1-C1-C2-O2
3	H	1201	EDO	O1-C1-C2-O2
4	D	101	A2G	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	H	1203	EDO	3	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	H	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	H	104:PHE	C	107:ALA	N	1.72

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	H	226/244 (92%)	0.14	9 (3%) 38 36	21, 32, 46, 69	5 (2%)
2	D	7/8 (87%)	2.43	3 (42%) 0 0	42, 58, 67, 72	0
All	All	233/252 (92%)	0.21	12 (5%) 27 26	21, 32, 51, 72	5 (2%)

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	1	SER	5.1
2	D	7	PRO	5.0
2	D	8	ALA	3.7
1	H	1102	SER	3.6
1	H	43	LYS	2.9

6.2 Non-standard residues in protein, DNA, RNA chains

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
2	4FB	D	3	8/9	0.90	0.14	40,42,44,45	0

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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
4	A2G	D	101	14/15	0.59	0.36	70,75,79,81	0
3	EDO	H	1201	4/4	0.82	0.37	52,57,58,59	0
3	EDO	H	1202	4/4	0.89	0.18	38,43,44,46	0
3	EDO	H	1203	4/4	0.95	0.24	40,42,42,45	0

6.5 Other polymers

There are no such residues in this entry.