



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 26, 2019 – 02:11 AM EDT

PDB ID : 4APQ
Title : Crystal structure of autoreactive-Valpha14-Vbeta6 NKT TCR in complex with CD1d-sulfatide
Authors : Clarke, A.J.; Le Nours, J.; Rossjohn, J.
Deposited on : 2012-04-05
Resolution : 3.00 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
Mogul : 1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix) : 1.13
EDS : rb-20031633
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20031633

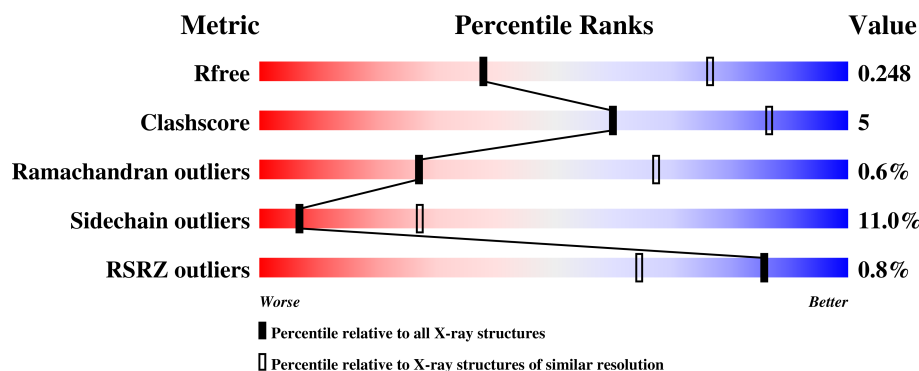
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	111664	1851 (3.00-3.00)
Clashscore	122126	2167 (3.00-3.00)
Ramachandran outliers	120053	2101 (3.00-3.00)
Sidechain outliers	120020	2104 (3.00-3.00)
RSRZ outliers	108989	1751 (3.00-3.00)

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. 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	A	302	
2	B	99	
3	C	207	
4	D	243	

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 6533 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 ANTIGEN-PRESENTING GLYCOPROTEIN CD1D1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	0	1	0
			2305	1471	402	418	14			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	HIS	ASP	variant	UNP P11609
A	280	GLY	-	expression tag	UNP P11609
A	281	SER	-	expression tag	UNP P11609
A	282	LEU	-	expression tag	UNP P11609
A	283	HIS	-	expression tag	UNP P11609
A	284	HIS	-	expression tag	UNP P11609
A	285	ILE	-	expression tag	UNP P11609
A	286	LEU	-	expression tag	UNP P11609
A	287	ASP	-	expression tag	UNP P11609
A	288	ALA	-	expression tag	UNP P11609
A	289	GLN	-	expression tag	UNP P11609
A	290	LYS	-	expression tag	UNP P11609
A	291	MET	-	expression tag	UNP P11609
A	292	VAL	-	expression tag	UNP P11609
A	293	TRP	-	expression tag	UNP P11609
A	294	ASN	-	expression tag	UNP P11609
A	295	HIS	-	expression tag	UNP P11609
A	296	ARG	-	expression tag	UNP P11609
A	297	HIS	-	expression tag	UNP P11609
A	298	HIS	-	expression tag	UNP P11609
A	299	HIS	-	expression tag	UNP P11609
A	300	HIS	-	expression tag	UNP P11609
A	301	HIS	-	expression tag	UNP P11609
A	302	HIS	-	expression tag	UNP P11609

- Molecule 2 is a protein called BETA-2-MICROGLOBULIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	99	Total	C	N	O	S	0	0	0
			806	514	136	149	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	85	ALA	ASP	variant	UNP P01887

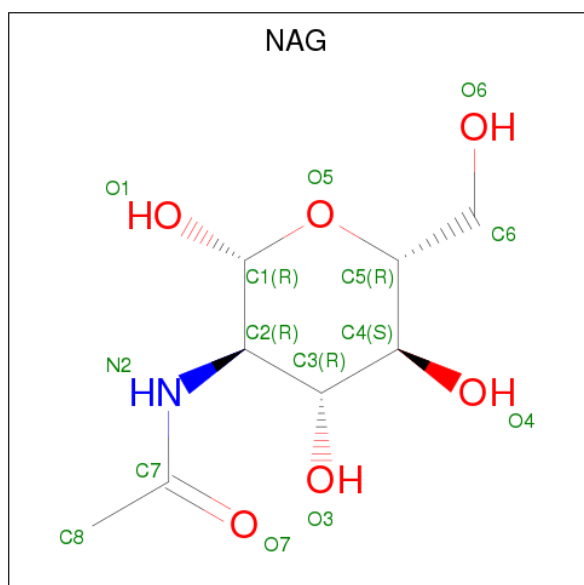
- Molecule 3 is a protein called MOUSE NKT TCR VALPHA14, HUMAN NKT TCR VALPHA14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	196	Total	C	N	O	S	0	0	0
			1451	899	246	300	6			

- Molecule 4 is a protein called MOUSE NKT TCR AUTOREACTIVE-VBETA6, HUMAN NKT TCR AUTOREACTIVE-VBETA6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	240	Total	C	N	O	S	0	0	0
			1831	1159	317	348	7			

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



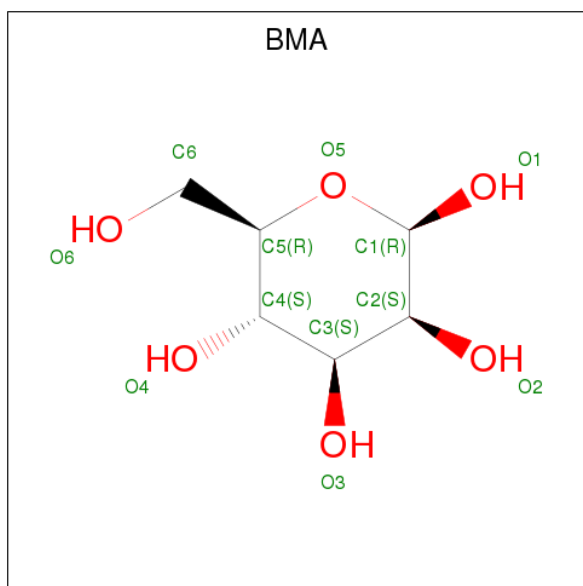
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			14	8	1	5		

Continued on next page...

Continued from previous page...

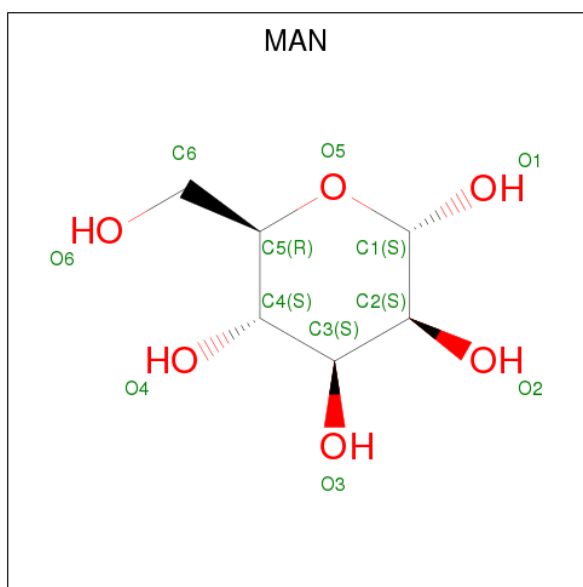
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 6 is BETA-D-MANNOSE (three-letter code: BMA) (formula: $C_6H_{12}O_6$).



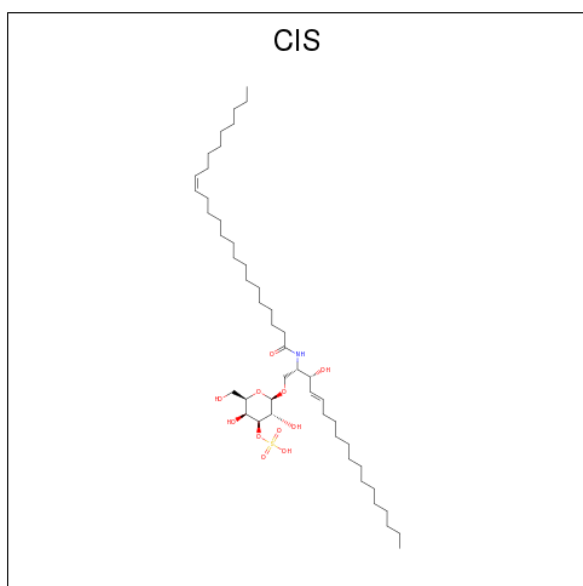
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			11	6	5		
6	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 7 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			11	6	5		

- Molecule 8 is (15Z)-N-((1S,2R,3E)-2-HYDROXY-1-[(3-O-SULFO-BETA-D-GALACTOPYRANOSYL)OXY]METHYL}HEPTADEC-3-ENYL)TETRACOS-15-ENAMIDE (three-letter code: CIS) (formula: C₄₈H₉₁NO₁₁S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
8	A	1	Total	C	N	O	S	0	0
			21	9	1	10	1		

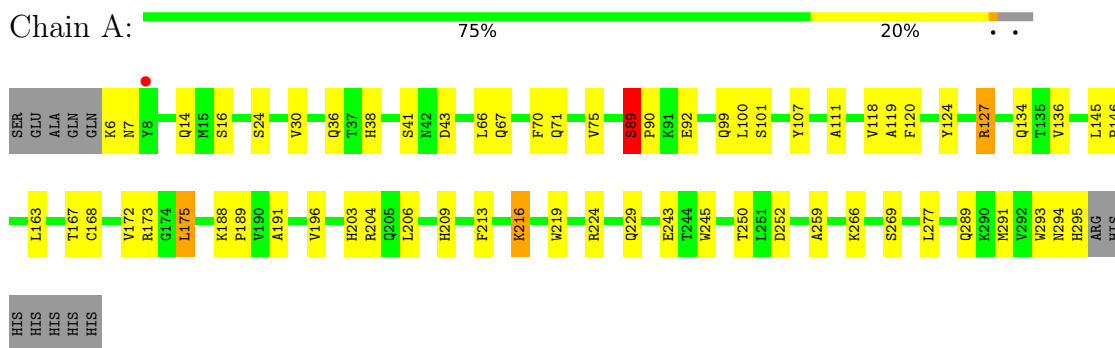
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	1	Total 1	O 1	0	0
9	B	1	Total 1	O 1	0	0

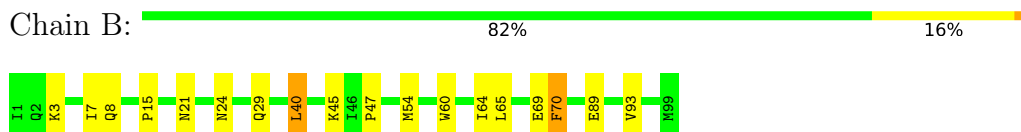
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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.

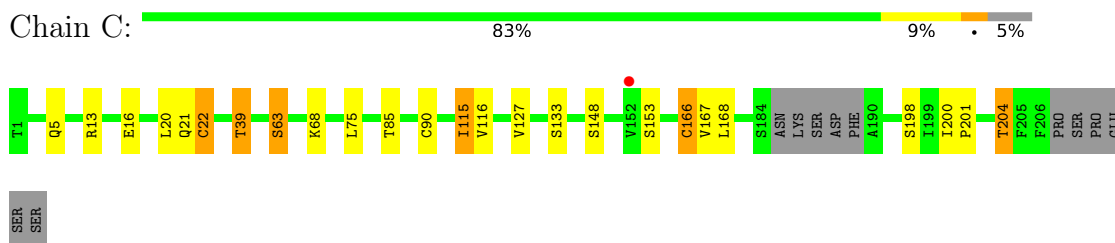
- Molecule 1: ANTIGEN-PRESENTING GLYCOPROTEIN CD1D1



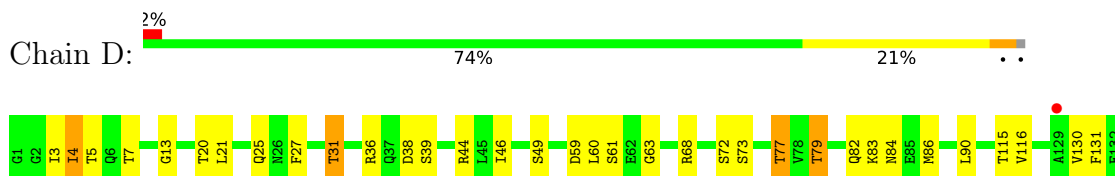
- Molecule 2: BETA-2-MICROGLOBULIN

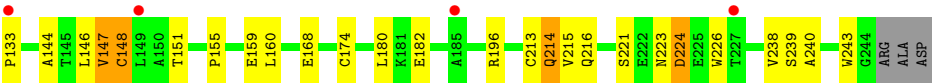


- Molecule 3: MOUSE NKT TCR VALPHA14, HUMAN NKT TCR VALPHA14



- Molecule 4: MOUSE NKT TCR AUTOREACTIVE-VBETA6, HUMAN NKT TCR AUTOREACTIVE-VBETA6





4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	95.11Å 95.11Å 291.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.49 – 3.00 72.85 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (36.49-3.00) 99.5 (72.85-3.00)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.03 (at 3.01Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.177 , 0.232 0.196 , 0.248	Depositor DCC
R_{free} test set	1392 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	81.8	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 70.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	6533	wwPDB-VP
Average B, all atoms (Å ²)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CIS, BMA, NAG, MAN

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	0.53	0/2379	0.69	0/3240
2	B	0.53	0/832	0.68	0/1132
3	C	0.55	0/1475	0.74	0/2015
4	D	0.52	0/1878	0.74	0/2559
All	All	0.53	0/6564	0.71	0/8946

There are no bond length outliers.

There are no bond angle outliers.

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	2305	0	2179	29	0
2	B	806	0	766	7	0
3	C	1451	0	1340	8	0
4	D	1831	0	1697	18	0
5	A	84	0	73	1	0
6	A	22	0	19	1	0
7	A	11	0	10	0	0
8	A	21	0	13	1	0
9	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	B	1	0	0	0	0
All	All	6533	0	6097	61	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 61 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)
4:D:148:CYS:HG	4:D:213:CYS:HG	1.36	0.72
4:D:20:THR:HG22	4:D:77:THR:HB	1.73	0.71
1:A:219:TRP:HB3	1:A:266:LYS:HB2	1.74	0.70
1:A:168:CYS:O	1:A:172:VAL:HG12	1.91	0.69
3:C:166:CYS:HB3	4:D:174:CYS:SG	2.40	0.61

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	289/302 (96%)	278 (96%)	10 (4%)	1 (0%)	43	80
2	B	97/99 (98%)	93 (96%)	3 (3%)	1 (1%)	17	56
3	C	192/207 (93%)	177 (92%)	14 (7%)	1 (0%)	31	71
4	D	238/243 (98%)	220 (92%)	16 (7%)	2 (1%)	21	61
All	All	816/851 (96%)	768 (94%)	43 (5%)	5 (1%)	27	67

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	89	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	D	83	LYS
4	D	224	ASP
2	B	47	PRO
3	C	39	THR

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	A	245/264 (93%)	224 (91%)	21 (9%)	11	40
2	B	90/93 (97%)	83 (92%)	7 (8%)	14	45
3	C	158/186 (85%)	142 (90%)	16 (10%)	8	31
4	D	187/209 (90%)	156 (83%)	31 (17%)	2	12
All	All	680/752 (90%)	605 (89%)	75 (11%)	7	27

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	68	LYS
3	C	168	LEU
4	D	196	ARG
3	C	115	ILE
3	C	148	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	229	GLN
3	C	193	ASN

5.3.3 RNA ⓘ

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

10 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	$\# Z > 2$
8	CIS	A	1296	-	21,21,61	1.14	2 (9%)	23,29,72	1.61	3 (13%)
5	NAG	A	304	1,5	14,14,15	0.25	0	17,19,21	0.80	1 (5%)
5	NAG	A	305	5	14,14,15	0.29	0	17,19,21	0.63	0
5	NAG	A	306	1,5	14,14,15	0.29	0	17,19,21	1.36	1 (5%)
5	NAG	A	307	5,6	14,14,15	0.35	0	17,19,21	1.12	2 (11%)
6	BMA	A	308	5,7	11,11,12	0.29	0	15,15,17	0.67	0
5	NAG	A	309	1,5	14,14,15	0.30	0	17,19,21	0.56	0
5	NAG	A	310	5,6	14,14,15	0.33	0	17,19,21	1.32	2 (11%)
6	BMA	A	311	5	11,11,12	0.41	0	15,15,17	0.52	0
7	MAN	A	312	6	11,11,12	0.46	0	15,15,17	1.08	2 (13%)

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
8	CIS	A	1296	-	-	0/13/33/78	0/1/1/1
5	NAG	A	304	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	305	5	-	0/6/23/26	0/1/1/1
5	NAG	A	306	1,5	-	0/6/23/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	NAG	A	307	5,6	-	0/6/23/26	0/1/1/1
6	BMA	A	308	5,7	-	0/2/19/22	0/1/1/1
5	NAG	A	309	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	310	5,6	-	0/6/23/26	0/1/1/1
6	BMA	A	311	5	-	0/2/19/22	0/1/1/1
7	MAN	A	312	6	-	0/2/19/22	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	A	1296	CIS	O8-S	2.08	1.53	1.44
8	A	1296	CIS	O-C43	3.26	1.45	1.40

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	A	1296	CIS	C45-O4-S	-3.65	111.82	118.88
8	A	1296	CIS	O2-C18-N	-3.58	121.09	124.85
5	A	307	NAG	C1-C2-N2	-2.62	106.01	110.49
5	A	307	NAG	O4-C4-C3	-2.54	104.45	110.34
5	A	310	NAG	C1-O5-C5	2.20	115.19	112.20

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	A	1296	CIS	1	0
5	A	307	NAG	1	0
6	A	308	BMA	1	0

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	290/302 (96%)	-0.01	1 (0%) 93 83	57, 79, 113, 136	1 (0%)
2	B	99/99 (100%)	0.05	0 100 100	54, 76, 103, 112	0
3	C	196/207 (94%)	-0.09	1 (0%) 90 74	60, 98, 176, 185	0
4	D	240/243 (98%)	0.05	5 (2%) 63 34	62, 105, 153, 168	0
All	All	825/851 (96%)	-0.01	7 (0%) 86 64	54, 87, 156, 185	1 (0%)

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	149	LEU	2.6
4	D	133	PRO	2.6
4	D	227	THR	2.5
1	A	8	TYR	2.2
3	C	152	VAL	2.2

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 carbohydrates 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(\AA^2)	Q<0.9
6	BMA	A	308	11/12	0.68	0.21	162,171,175,179	0
6	BMA	A	311	11/12	0.75	0.24	167,172,178,181	0
7	MAN	A	312	11/12	0.76	0.22	184,188,194,194	0
8	CIS	A	1296	21/61	0.83	0.39	130,165,175,177	0
5	NAG	A	305	14/15	0.85	0.32	117,129,136,138	0
5	NAG	A	307	14/15	0.91	0.27	123,132,143,153	0
5	NAG	A	309	14/15	0.92	0.19	103,108,113,121	0
5	NAG	A	306	14/15	0.92	0.18	93,100,110,117	0
5	NAG	A	304	14/15	0.92	0.17	94,109,116,116	0
5	NAG	A	310	14/15	0.93	0.21	133,141,155,159	0

6.5 Other polymers [i](#)

There are no such residues in this entry.