



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 16, 2017 – 11:55 AM EST

PDB ID : 5IDK
Title : Crystal structure of West Nile Virus NS2B-NS3 protease in complex with a capped dipeptide boronate inhibitor
Authors : Hilgenfeld, R.; Zhang, L.
Deposited on : 2016-02-24
Resolution : 1.50 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

<http://wwpdb.org/validation/2016/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 : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20028442
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20028442

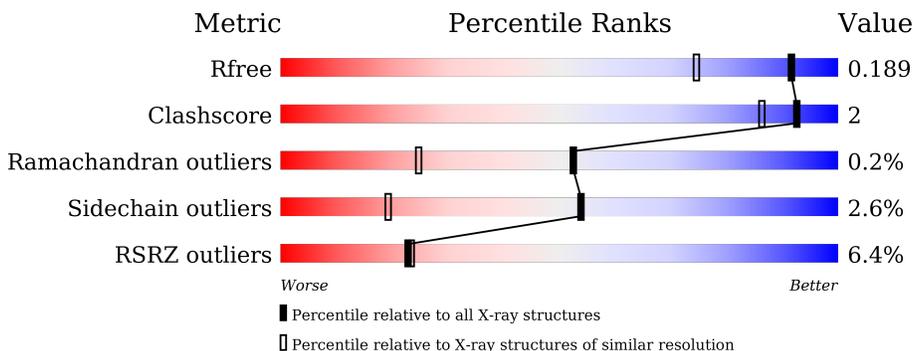
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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}	91344	2072 (1.50-1.50)
Clashscore	102246	2274 (1.50-1.50)
Ramachandran outliers	100387	2218 (1.50-1.50)
Sidechain outliers	100360	2216 (1.50-1.50)
RSRZ outliers	91569	2075 (1.50-1.50)

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	230	 3% 78% 9% • 11%
1	B	230	 9% 80% 8% • 11%
1	C	230	 5% 83% 5% • 11%

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 5466 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 Genome polyprotein,SERINE PROTEASE SUBUNIT NS2B, SERINE PROTEASE NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	204	1571	992	263	308	8	0	5	0
1	B	204	1563	984	263	308	8	8	2	0
1	C	205	1574	995	265	306	8	14	2	0

There are 42 discrepancies between the modelled and reference sequences:

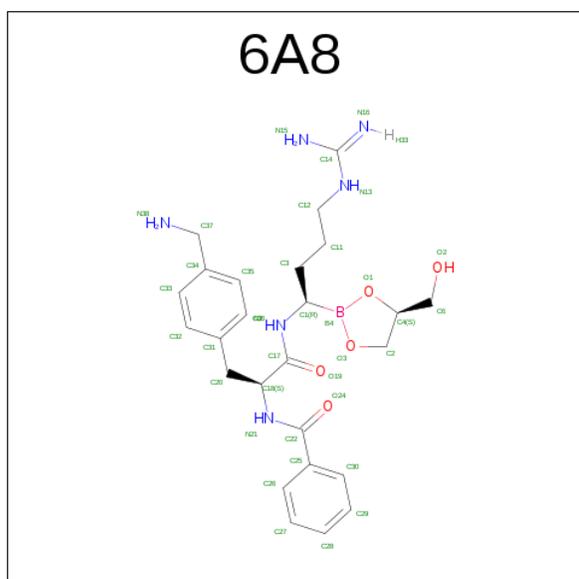
Chain	Residue	Modelled	Actual	Comment	Reference
A	46	GLY	-	expression tag	UNP P06935
A	47	SER	-	expression tag	UNP P06935
A	48	HIS	-	expression tag	UNP P06935
A	49	MET	-	expression tag	UNP P06935
A	991	ALA	-	linker	UNP P06935
A	992	GLY	-	linker	UNP P06935
A	993	GLY	-	linker	UNP P06935
A	994	GLY	-	linker	UNP P06935
A	995	GLY	-	linker	UNP P06935
A	996	SER	-	linker	UNP P06935
A	997	GLY	-	linker	UNP P06935
A	998	GLY	-	linker	UNP P06935
A	999	GLY	-	linker	UNP P06935
A	1000	GLY	-	linker	UNP P06935
B	46	GLY	-	expression tag	UNP P06935
B	47	SER	-	expression tag	UNP P06935
B	48	HIS	-	expression tag	UNP P06935
B	49	MET	-	expression tag	UNP P06935
B	991	ALA	-	linker	UNP P06935
B	992	GLY	-	linker	UNP P06935
B	993	GLY	-	linker	UNP P06935
B	994	GLY	-	linker	UNP P06935

Continued on next page...

Continued from previous page...

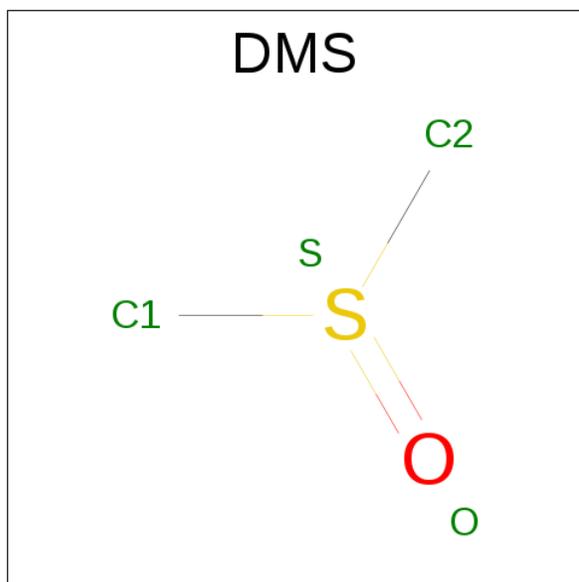
Chain	Residue	Modelled	Actual	Comment	Reference
B	995	GLY	-	linker	UNP P06935
B	996	SER	-	linker	UNP P06935
B	997	GLY	-	linker	UNP P06935
B	998	GLY	-	linker	UNP P06935
B	999	GLY	-	linker	UNP P06935
B	1000	GLY	-	linker	UNP P06935
C	46	GLY	-	expression tag	UNP P06935
C	47	SER	-	expression tag	UNP P06935
C	48	HIS	-	expression tag	UNP P06935
C	49	MET	-	expression tag	UNP P06935
C	991	ALA	-	linker	UNP P06935
C	992	GLY	-	linker	UNP P06935
C	993	GLY	-	linker	UNP P06935
C	994	GLY	-	linker	UNP P06935
C	995	GLY	-	linker	UNP P06935
C	996	SER	-	linker	UNP P06935
C	997	GLY	-	linker	UNP P06935
C	998	GLY	-	linker	UNP P06935
C	999	GLY	-	linker	UNP P06935
C	1000	GLY	-	linker	UNP P06935

- Molecule 2 is ((R)-1-((S)-3-(4-(aminomethyl)phenyl)-2-benzamidopropaneamido)-4-guanidinobutyl)boronic acid, cyclic double ester with glycerol (three-letter code: 6A8) (formula: C₂₅H₃₅BN₆O₅).



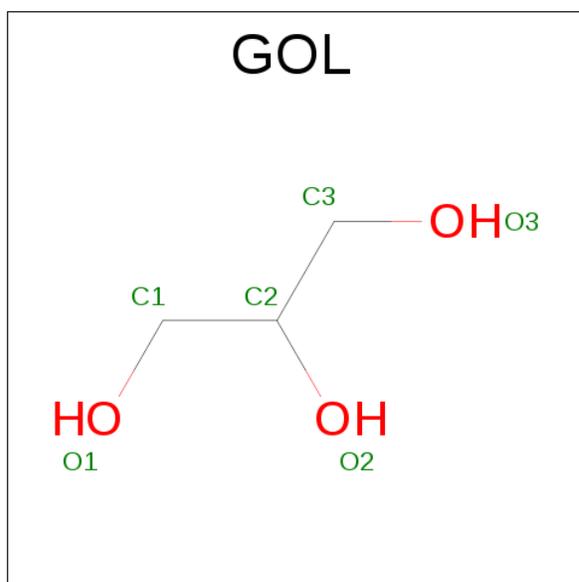
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	B	C	N	O		
2	A	1	37	1	25	6	5	0	0
2	B	1	37	1	25	6	5	0	0
2	C	1	37	1	25	6	5	0	0

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C₂H₆OS).



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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			6	3	3		

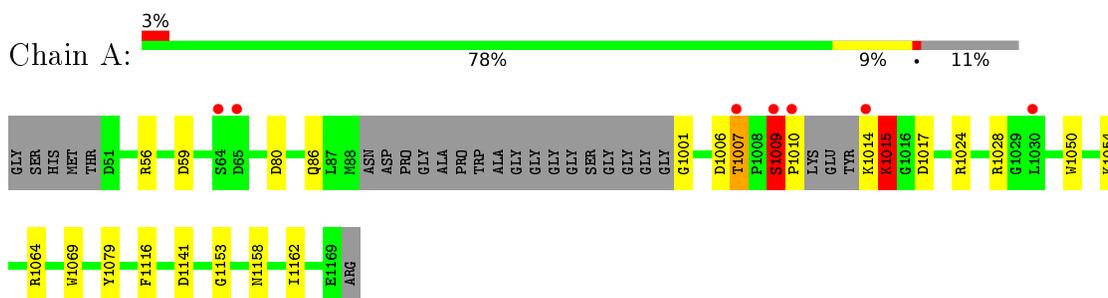
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	236	Total	O	0	0
			236	236		
5	B	200	Total	O	0	0
			200	200		
5	C	201	Total	O	0	0
			201	201		

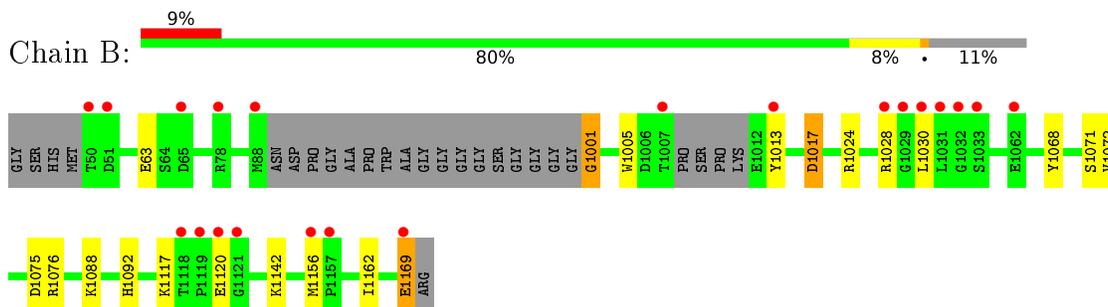
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

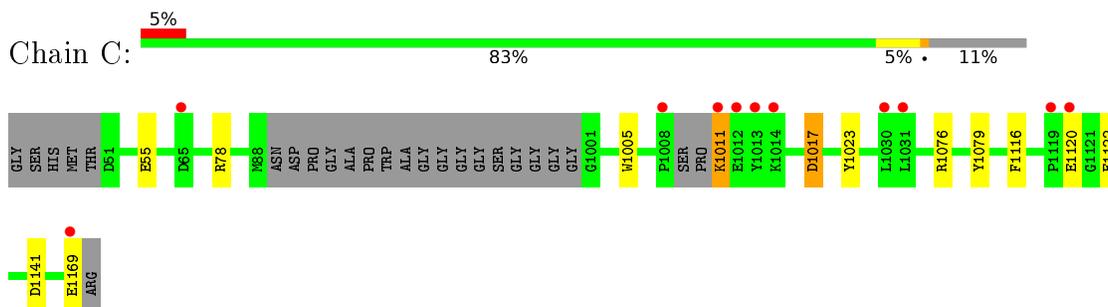
- Molecule 1: Genome polyprotein,SERINE PROTEASE SUBUNIT NS2B, SERINE PROTEASE NS3



- Molecule 1: Genome polyprotein,SERINE PROTEASE SUBUNIT NS2B, SERINE PROTEASE NS3



- Molecule 1: Genome polyprotein,SERINE PROTEASE SUBUNIT NS2B, SERINE PROTEASE NS3



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	36.42Å 96.97Å 187.86Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.94 – 1.50 35.75 – 1.50	Depositor EDS
% Data completeness (in resolution range)	98.2 (46.94-1.50) 98.2 (35.75-1.50)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.08 (at 1.50Å)	Xtrriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.159 , 0.186 0.168 , 0.189	Depositor DCC
R_{free} test set	5227 reflections (5.19%)	DCC
Wilson B-factor (Å ²)	16.1	Xtrriage
Anisotropy	0.408	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 47.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	5466	wwPDB-VP
Average B, all atoms (Å ²)	22.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.75% 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: GOL, 6A8, DMS

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	1.28	5/1621 (0.3%)	1.37	18/2203 (0.8%)
1	B	1.38	7/1603 (0.4%)	1.26	9/2174 (0.4%)
1	C	1.50	5/1615 (0.3%)	1.18	9/2190 (0.4%)
All	All	1.39	17/4839 (0.4%)	1.27	36/6567 (0.5%)

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	C	1	0

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1011	LYS	CA-CB	-32.57	0.82	1.53
1	B	1001	GLY	N-CA	14.00	1.67	1.46
1	B	1030	LEU	CA-CB	-11.80	1.26	1.53
1	A	1050	TRP	CG-CD2	-8.37	1.29	1.43
1	B	1001	GLY	C-O	7.92	1.36	1.23

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1028	ARG	NE-CZ-NH2	-19.91	110.34	120.30
1	A	1028	ARG	NE-CZ-NH1	19.39	129.99	120.30
1	B	1001	GLY	N-CA-C	12.08	143.30	113.10
1	C	1011	LYS	CB-CA-C	11.80	133.99	110.40
1	A	1024	ARG	NE-CZ-NH1	9.95	125.28	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	C	1011	LYS	CA

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	1571	0	1526	12	0
1	B	1563	0	1500	6	0
1	C	1574	0	1521	1	0
2	A	37	0	0	0	0
2	B	37	0	0	0	0
2	C	37	0	0	0	0
3	B	4	0	6	0	0
4	C	6	0	8	1	0
5	A	236	0	0	7	3
5	B	200	0	0	2	0
5	C	201	0	0	0	2
All	All	5466	0	4561	19	3

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.

The worst 5 of 19 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:1142:LYS:CG	5:B:1476:HOH:O	2.05	1.01
1:A:1007[B]:THR:HG22	5:A:1303:HOH:O	1.91	0.70
1:A:1014:LYS:CA	1:A:1015:LYS:HG2	2.28	0.64
1:B:1156:MET:SD	1:B:1162:ILE:HD11	2.41	0.61
1:A:1010:PRO:HG2	5:A:1493:HOH:O	2.01	0.59

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:1301:HOH:O	5:A:1415:HOH:O[1_455]	1.85	0.35
5:A:1515:HOH:O	5:C:1449:HOH:O[3_655]	1.89	0.31
5:A:1477:HOH:O	5:C:1368:HOH:O[3_655]	2.18	0.02

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	203/230 (88%)	196 (97%)	6 (3%)	1 (0%)	34 10
1	B	200/230 (87%)	197 (98%)	3 (2%)	0	100 100
1	C	201/230 (87%)	197 (98%)	4 (2%)	0	100 100
All	All	604/690 (88%)	590 (98%)	13 (2%)	1 (0%)	52 25

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1009	SER

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	170/180 (94%)	166 (98%)	4 (2%)	57 22
1	B	166/180 (92%)	159 (96%)	7 (4%)	36 7
1	C	167/180 (93%)	164 (98%)	3 (2%)	66 35

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	503/540 (93%)	489 (97%)	14 (3%)	54 17

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1028	ARG
1	B	1088	LYS
1	C	1011	LYS
1	B	1017	ASP
1	B	1169	GLU

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	1087	HIS
1	C	1087	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

5 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
2	6A8	A	1201	1	37,39,39	2.31	6 (16%)	41,51,51	2.38	9 (21%)
2	6A8	B	1201	1	37,39,39	1.85	6 (16%)	41,51,51	1.75	6 (14%)
3	DMS	B	1202	-	3,3,3	0.64	0	3,3,3	1.54	1 (33%)
2	6A8	C	1201	1	37,39,39	2.09	5 (13%)	41,51,51	2.05	7 (17%)
4	GOL	C	1202	-	5,5,5	0.56	0	5,5,5	0.91	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
2	6A8	A	1201	1	-	0/31/44/44	0/3/3/3
2	6A8	B	1201	1	-	0/31/44/44	0/3/3/3
3	DMS	B	1202	-	-	0/0/0/0	0/0/0/0
2	6A8	C	1201	1	-	0/31/44/44	0/3/3/3
4	GOL	C	1202	-	-	0/4/4/4	0/0/0/0

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1201	6A8	C30-C25	-2.82	1.34	1.39
2	B	1201	6A8	C1-N2	-2.36	1.41	1.47
2	C	1201	6A8	C18-C17	-2.16	1.46	1.52
2	A	1201	6A8	O1-C4	-2.12	1.40	1.45
2	B	1201	6A8	C12-N13	2.14	1.51	1.46

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1201	6A8	O1-B4-O3	-10.25	102.28	113.53
2	C	1201	6A8	O1-B4-O3	-9.79	102.78	113.53
2	B	1201	6A8	O1-B4-O3	-6.53	106.35	113.53
2	B	1201	6A8	O2-C6-C4	-4.84	97.77	111.65
2	B	1201	6A8	O24-C22-C25	-3.35	115.20	120.95

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	1202	GOL	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	204/230 (88%)	0.14	7 (3%) 49 51	10, 16, 38, 75	0
1	B	204/230 (88%)	0.54	21 (10%) 9 8	9, 17, 52, 77	2 (0%)
1	C	205/230 (89%)	0.10	11 (5%) 29 30	10, 17, 40, 70	3 (1%)
All	All	613/690 (88%)	0.26	39 (6%) 23 23	9, 16, 46, 77	5 (0%)

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1030	LEU	9.5
1	B	1013	TYR	7.7
1	B	1032	GLY	7.5
1	B	50	THR	6.9
1	B	1007	THR	6.3

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 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
4	GOL	C	1202	6/6	0.90	0.16	1.84	26,33,36,52	0
3	DMS	B	1202	4/4	0.92	0.13	1.56	30,41,41,62	0
2	6A8	C	1201	37/37	0.92	0.09	0.78	14,19,30,36	0
2	6A8	A	1201	37/37	0.94	0.09	0.35	11,15,24,30	0
2	6A8	B	1201	37/37	0.92	0.09	0.35	15,19,27,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.