



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 1, 2016 – 01:48 PM EDT

PDB ID : 5LC0
Title : Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor
Authors : Lei, J.; Hansen, G.; Zhang, L.L.; Hilgenfeld, R.
Deposited on : 2016-06-18
Resolution : 2.70 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

<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-20027939
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-20027939

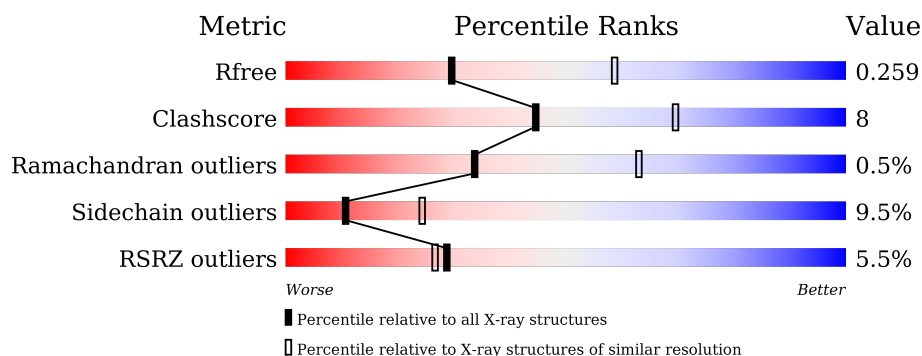
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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	<div> <div>5%</div> <div> <div></div> <div>66%</div> <div>16%</div> <div>•</div> <div>16%</div> </div> </div>
1	B	230	<div> <div>4%</div> <div> <div></div> <div>66%</div> <div>16%</div> <div>•</div> <div>17%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2981 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 NS2B-NS3 protease, NS2B-NS3 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	193	Total	C	N	O	S	0	0	0
			1456	915	250	284	7			
1	B	190	Total	C	N	O	S	0	0	0
			1436	902	246	281	7			

There are 30 discrepancies between the modelled and reference sequences:

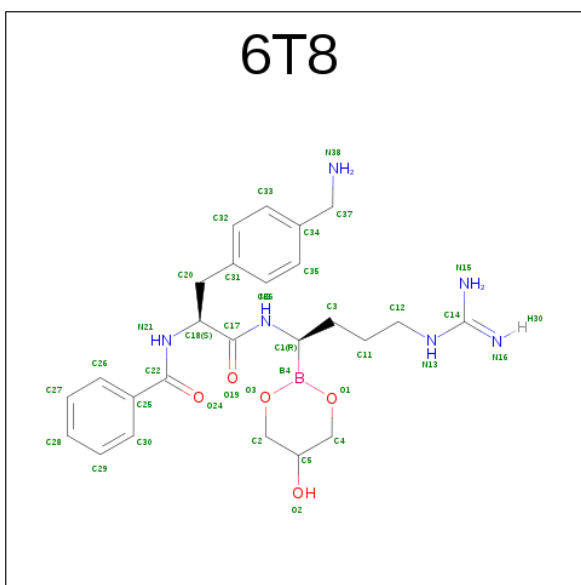
Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP A0A140FB25
A	-3	SER	-	expression tag	UNP A0A140FB25
A	-2	HIS	-	expression tag	UNP A0A140FB25
A	-1	MET	-	expression tag	UNP A0A140FB25
A	991	ALA	ARG	engineered mutation	UNP A0A140FB25
A	992	GLY	-	linker	UNP A0A140FB25
A	993	GLY	-	linker	UNP A0A140FB25
A	994	GLY	-	linker	UNP A0A140FB25
A	995	GLY	-	linker	UNP A0A140FB25
A	996	SER	-	linker	UNP A0A140FB25
A	997	GLY	-	linker	UNP A0A140FB25
A	998	GLY	-	linker	UNP A0A140FB25
A	999	GLY	-	linker	UNP A0A140FB25
A	1000	GLY	-	linker	UNP A0A140FB25
A	1029	GLY	ARG	engineered mutation	UNP A0A140DLX4
B	-4	GLY	-	expression tag	UNP A0A140FB25
B	-3	SER	-	expression tag	UNP A0A140FB25
B	-2	HIS	-	expression tag	UNP A0A140FB25
B	-1	MET	-	expression tag	UNP A0A140FB25
B	991	ALA	ARG	engineered mutation	UNP A0A140FB25
B	992	GLY	-	linker	UNP A0A140FB25
B	993	GLY	-	linker	UNP A0A140FB25
B	994	GLY	-	linker	UNP A0A140FB25
B	995	GLY	-	linker	UNP A0A140FB25
B	996	SER	-	linker	UNP A0A140FB25

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	997	GLY	-	linker	UNP A0A140FB25
B	998	GLY	-	linker	UNP A0A140FB25
B	999	GLY	-	linker	UNP A0A140FB25
B	1000	GLY	-	linker	UNP A0A140FB25
B	1029	GLY	ARG	engineered mutation	UNP A0A140DLX4

- Molecule 2 is N-((S)-3-(4-(aminomethyl)phenyl)-1-(((R)-4-guanidino-1-(5-hydroxy-1,3,2-dioxaborinan-2-yl)butyl)amino)-1-oxopropan-2-yl)benzamide (three-letter code: 6T8) (formula: C₂₅H₃₅BN₆O₅).

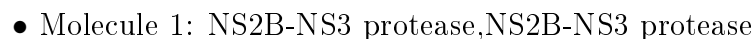


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	B	C	N	O	0	0
			37	1	25	6	5		
2	B	1	Total	B	C	N	O	0	0
			37	1	25	6	5		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	10	Total	O	0	0
			10	10		
3	B	5	Total	O	0	0
			5	5		

- Molecule 1: NS2B-NS3 protease, NS2B-NS3 protease



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	125.81 Å 125.81 Å 73.46 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	62.90 – 2.70 62.90 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (62.90-2.70) 100.0 (62.90-2.70)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.53 (at 2.69 Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.204 , 0.263 0.202 , 0.259	Depositor DCC
R_{free} test set	902 reflections (5.06%)	DCC
Wilson B-factor (Å ²)	69.0	Xtriage
Anisotropy	0.565	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 66.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.425 for -h,-k,l	Xtriage
Reported twinning fraction	0.435 for H, K, L 0.565 for -H, -K, L	Depositor
Outliers	0 of 18713 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	2981	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 3.24% 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: 6T8

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	0.62	0/1485	0.85	0/2014
1	B	0.50	0/1465	0.76	0/1988
All	All	0.56	0/2950	0.81	0/4002

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

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	1456	0	1425	26	0
1	B	1436	0	1400	26	0
2	A	37	0	0	2	0
2	B	37	0	0	1	0
3	A	10	0	0	0	2
3	B	5	0	0	0	0
All	All	2981	0	2825	46	2

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 8.

All (46) 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:A:-1:MET:C	1:B:1158:ASN:OD1	2.14	0.85
1:B:1139:ILE:HG21	1:B:1165:ILE:CD1	2.22	0.68
1:B:53:ILE:HD12	1:B:1023:TYR:CD2	2.31	0.65
1:B:1090:ASP:OD1	1:B:1090:ASP:N	2.29	0.65
1:A:-1:MET:N	1:B:1158:ASN:HD21	1.95	0.64
1:B:68:THR:CG2	1:B:1109:ILE:HD11	2.28	0.63
1:A:-1:MET:H3	1:B:1158:ASN:HD21	1.49	0.61
1:A:1098:LEU:O	1:A:1137:SER:OG	2.19	0.60
1:A:1142:LYS:H	1:A:1142:LYS:HD3	1.66	0.59
1:A:-1:MET:O	1:A:49:VAL:HG23	2.06	0.56
1:A:1089:TRP:HA	1:A:1147:ILE:HD11	1.87	0.56
1:A:78:LEU:HD12	1:A:79:ASP:O	2.06	0.55
1:A:-1:MET:O	1:A:49:VAL:CG2	2.54	0.55
2:A:1201:6T8:C28	1:B:1054:LYS:HD2	2.37	0.54
1:A:1034:THR:HG21	1:B:1131:PRO:HB3	1.89	0.54
1:B:1045:VAL:HG21	1:B:1084:LYS:HE3	1.91	0.53
1:A:72:PRO:HB2	1:A:74:LEU:HD13	1.89	0.53
1:B:55:ARG:HA	1:B:1023:TYR:HA	1.90	0.53
1:B:1139:ILE:HG21	1:B:1165:ILE:HD12	1.89	0.52
1:A:1038:VAL:O	1:A:1136:GLY:HA3	2.12	0.49
1:A:1141:ASP:O	1:A:1144:GLY:N	2.45	0.49
1:B:1050:TRP:HA	1:B:1077:VAL:CG2	2.43	0.48
1:A:1142:LYS:H	1:A:1142:LYS:CD	2.24	0.48
1:A:61:TRP:CD2	1:A:1142:LYS:HA	2.48	0.48
1:B:1089:TRP:CZ2	1:B:1113:PRO:HG2	2.48	0.48
1:A:-1:MET:O	1:B:1158:ASN:OD1	2.31	0.47
1:A:1047:HIS:CG	1:A:1085:LEU:HD11	2.50	0.46
1:B:59:ILE:N	1:B:59:ILE:HD12	2.32	0.45
1:B:83:ASP:OD1	2:B:1201:6T8:N38	2.50	0.45
1:A:1129:ASP:O	1:B:1030:LEU:HB2	2.18	0.44
1:A:1156:ILE:HG13	1:A:1158:ASN:HB3	1.98	0.44
1:A:-1:MET:C	1:A:49:VAL:HG23	2.38	0.44
1:B:1047:HIS:HB3	1:B:1083:TRP:CZ3	2.53	0.43
1:A:68:THR:HG21	1:A:1109:ILE:HD11	2.01	0.43
1:B:1139:ILE:HG13	1:B:1165:ILE:HD13	2.02	0.42
1:A:1129:ASP:OD1	2:A:1201:6T8:N15	2.53	0.42
1:A:68:THR:CG2	1:A:1109:ILE:HD11	2.50	0.42
1:B:1031:LEU:HD12	1:B:1031:LEU:HA	1.66	0.42
1:A:59:ILE:HD13	1:A:1022:VAL:HB	2.02	0.41
1:B:68:THR:HG23	1:B:1109:ILE:HD11	2.00	0.41
1:B:1114:GLY:N	1:B:1125:ALA:O	2.54	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:60:THR:HG22	1:B:61:TRP:N	2.35	0.41
1:A:1039:GLY:HA3	1:A:1046:PHE:CZ	2.56	0.41
1:B:1030:LEU:HD23	1:B:1030:LEU:HA	1.82	0.41
1:B:1051:HIS:CD2	1:B:1051:HIS:C	2.94	0.40
1:A:-1:MET:HG3	1:A:49:VAL:N	2.36	0.40

All (2) 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)
3:A:1305:HOH:O	3:A:1305:HOH:O[6_769]	0.88	1.32
3:A:1308:HOH:O	3:A:1308:HOH:O[6_768]	1.96	0.24

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	189/230 (82%)	177 (94%)	10 (5%)	2 (1%)	17	42
1	B	186/230 (81%)	163 (88%)	23 (12%)	0	100	100
All	All	375/460 (82%)	340 (91%)	33 (9%)	2 (0%)	34	63

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	VAL
1	A	1075	ASP

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	154/176 (88%)	139 (90%)	15 (10%)	10	23
1	B	152/176 (86%)	138 (91%)	14 (9%)	11	25
All	All	306/352 (87%)	277 (90%)	29 (10%)	11	24

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	50	ASP
1	A	75	ASP
1	A	78	LEU
1	A	85	SER
1	A	1033	SER
1	A	1049	MET
1	A	1085	LEU
1	A	1093	SER
1	A	1107	ARG
1	A	1112	LEU
1	A	1120	ASP
1	A	1140	LEU
1	A	1142	LYS
1	A	1160	SER
1	A	1166	THR
1	B	-1	MET
1	B	68	THR
1	B	74	LEU
1	B	78	LEU
1	B	1018	THR
1	B	1024	ARG
1	B	1031	LEU
1	B	1064	ARG
1	B	1090	ASP
1	B	1094	GLU
1	B	1098	LEU
1	B	1126	VAL
1	B	1135	SER
1	B	1157	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	1158	ASN

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](#)

2 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	6T8	A	1201	1	35,39,39	2.05	3 (8%)	39,51,51	1.29	6 (15%)
2	6T8	B	1201	1	35,39,39	1.96	2 (5%)	39,51,51	1.13	5 (12%)

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	6T8	A	1201	1	-	0/27/43/43	0/3/3/3
2	6T8	B	1201	1	-	0/27/43/43	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1201	6T8	C2-C5	-2.74	1.49	1.51
2	B	1201	6T8	B4-O1	7.64	1.42	1.36
2	A	1201	6T8	B4-O1	7.81	1.42	1.36
2	A	1201	6T8	B4-O3	8.17	1.42	1.36
2	B	1201	6T8	B4-O3	8.35	1.42	1.36

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1201	6T8	C20-C31-C36	-3.29	114.26	120.91
2	B	1201	6T8	C20-C18-C17	-2.58	102.64	110.19
2	B	1201	6T8	C20-C31-C36	-2.50	115.85	120.91
2	B	1201	6T8	C11-C3-C1	-2.49	107.80	114.74
2	A	1201	6T8	O24-C22-C25	-2.44	116.77	120.95
2	A	1201	6T8	N13-C14-N16	-2.43	116.29	120.40
2	B	1201	6T8	O24-C22-C25	-2.12	117.32	120.95
2	A	1201	6T8	C18-C17-N2	2.08	121.64	116.80
2	A	1201	6T8	C25-C22-N21	2.18	120.79	116.98
2	A	1201	6T8	C20-C31-C32	2.59	126.13	120.91
2	B	1201	6T8	C25-C22-N21	2.62	121.56	116.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1201	6T8	2	0
2	B	1201	6T8	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

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, 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	193/230 (83%)	0.64	11 (5%) 27 26	54, 87, 117, 137	0
1	B	190/230 (82%)	0.58	10 (5%) 30 28	63, 101, 124, 143	0
All	All	383/460 (83%)	0.61	21 (5%) 29 27	54, 95, 121, 143	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1165	ILE	3.4
1	A	1017	GLU	3.1
1	B	1105	ARG	2.8
1	B	1112	LEU	2.8
1	A	1036	VAL	2.7
1	B	1147	ILE	2.6
1	A	1130	TYR	2.6
1	B	1077	VAL	2.6
1	B	1095	VAL	2.4
1	B	1145	ARG	2.4
1	B	1100	VAL	2.4
1	B	73	ARG	2.3
1	B	1074	GLN	2.3
1	A	1022	VAL	2.3
1	A	78	LEU	2.3
1	A	1040	VAL	2.3
1	A	1046	PHE	2.2
1	B	1165	ILE	2.1
1	A	53	ILE	2.1
1	A	1100	VAL	2.1
1	A	56	ALA	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 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
2	6T8	B	1201	37/37	0.94	0.23	0.24	69,85,100,104	0
2	6T8	A	1201	37/37	0.93	0.24	0.09	37,75,99,117	0

6.5 Other polymers [i](#)

There are no such residues in this entry.