



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:00 PM GMT

PDB ID : 4WH8
Title : Crystal Structure of HCV NS3/4A protease in complex with an Asunaprevir P1-P3 macrocyclic analog.
Authors : Soumana, D.I.; Ali, A.; Schiffer, C.A.
Deposited on : 2014-09-20
Resolution : 2.70 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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

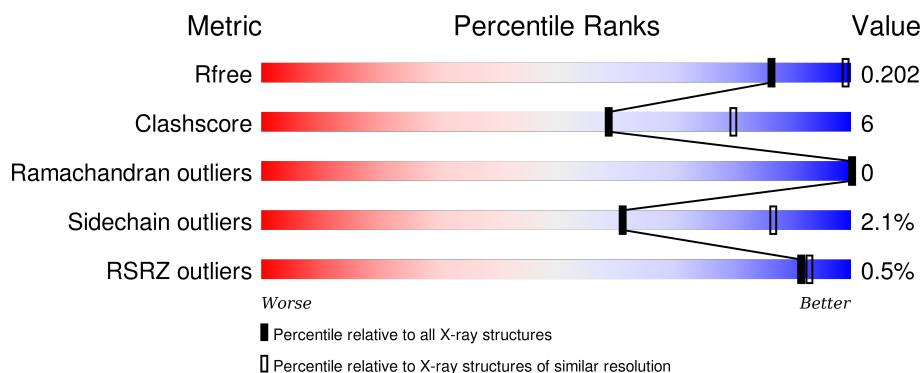
1 Overall quality at a glance

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	198	 91% 7% •
1	B	198	 86% 11% •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ZN	A	1204	-	-	-	X
4	ZN	B	1202	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5729 atoms, of which 2769 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 polypeptide.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	198	Total	C	H	N	O	S	0	0	0
			2727	861	1336	249	273	8			
1	B	198	Total	C	H	N	O	S	0	0	0
			2734	862	1341	251	271	9			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	982	HIS	-	expression tag	UNP A8DG50
A	983	MET	-	expression tag	UNP A8DG50
A	984	ALA	-	expression tag	UNP A8DG50
A	985	SER	-	expression tag	UNP A8DG50
A	986	MET	-	expression tag	UNP A8DG50
A	987	LYS	-	expression tag	UNP A8DG50
A	988	LYS	-	expression tag	UNP A8DG50
A	989	LYS	-	expression tag	UNP A8DG50
A	990	GLY	-	expression tag	UNP A8DG50
A	991	SER	-	expression tag	UNP A8DG50
A	992	VAL	-	expression tag	UNP A8DG50
A	993	VAL	-	expression tag	UNP A8DG50
A	994	ILE	-	expression tag	UNP A8DG50
A	995	VAL	-	expression tag	UNP A8DG50
A	996	GLY	-	expression tag	UNP A8DG50
A	997	ARG	-	expression tag	UNP A8DG50
A	998	ILE	-	expression tag	UNP A8DG50
A	999	ASN	-	expression tag	UNP A8DG50
A	1000	LEU	-	expression tag	UNP A8DG50
A	1001	SER	-	expression tag	UNP A8DG50
A	1002	GLY	-	expression tag	UNP A8DG50
A	1003	ASP	-	expression tag	UNP A8DG50
A	1013	GLU	LEU	conflict	UNP A8DG50
A	1014	GLU	LEU	conflict	UNP A8DG50
A	1017	GLN	ILE	conflict	UNP A8DG50

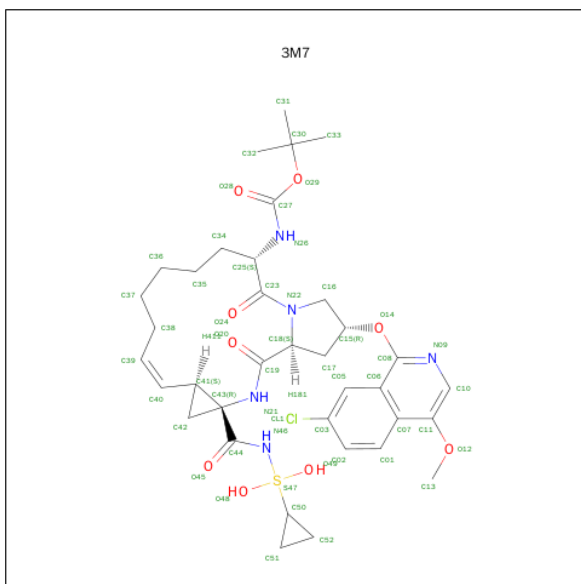
Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1018	GLU	ILE	conflict	UNP A8DG50
A	1021	GLN	LEU	conflict	UNP A8DG50
A	1040	THR	ALA	conflict	UNP A8DG50
A	1047	SER	CYS	conflict	UNP A8DG50
A	1052	LEU	CYS	conflict	UNP A8DG50
A	1072	THR	ILE	conflict	UNP A8DG50
A	1086	GLN	PRO	conflict	UNP A8DG50
A	1139	ALA	SER	conflict	UNP A8DG50
B	982	HIS	-	expression tag	UNP A8DG50
B	983	MET	-	expression tag	UNP A8DG50
B	984	ALA	-	expression tag	UNP A8DG50
B	985	SER	-	expression tag	UNP A8DG50
B	986	MET	-	expression tag	UNP A8DG50
B	987	LYS	-	expression tag	UNP A8DG50
B	988	LYS	-	expression tag	UNP A8DG50
B	989	LYS	-	expression tag	UNP A8DG50
B	990	GLY	-	expression tag	UNP A8DG50
B	991	SER	-	expression tag	UNP A8DG50
B	992	VAL	-	expression tag	UNP A8DG50
B	993	VAL	-	expression tag	UNP A8DG50
B	994	ILE	-	expression tag	UNP A8DG50
B	995	VAL	-	expression tag	UNP A8DG50
B	996	GLY	-	expression tag	UNP A8DG50
B	997	ARG	-	expression tag	UNP A8DG50
B	998	ILE	-	expression tag	UNP A8DG50
B	999	ASN	-	expression tag	UNP A8DG50
B	1000	LEU	-	expression tag	UNP A8DG50
B	1001	SER	-	expression tag	UNP A8DG50
B	1002	GLY	-	expression tag	UNP A8DG50
B	1003	ASP	-	expression tag	UNP A8DG50
B	1013	GLU	LEU	conflict	UNP A8DG50
B	1014	GLU	LEU	conflict	UNP A8DG50
B	1017	GLN	ILE	conflict	UNP A8DG50
B	1018	GLU	ILE	conflict	UNP A8DG50
B	1021	GLN	LEU	conflict	UNP A8DG50
B	1040	THR	ALA	conflict	UNP A8DG50
B	1047	SER	CYS	conflict	UNP A8DG50
B	1052	LEU	CYS	conflict	UNP A8DG50
B	1072	THR	ILE	conflict	UNP A8DG50
B	1086	GLN	PRO	conflict	UNP A8DG50
B	1139	ALA	SER	conflict	UNP A8DG50

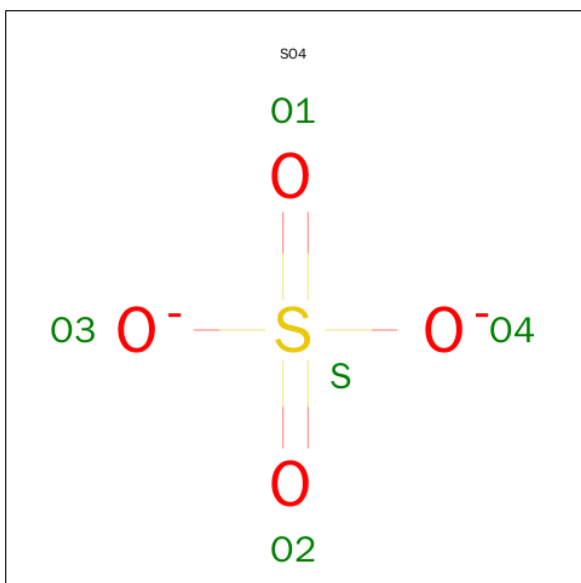
- Molecule 2 is tert-butyl [(2R,6S,12Z,13aS,14aR,16aS)-2-[(7-chloro-4-methoxyisoquinolin-1-yl

l)oxy]-14a-{[cyclopropyl(dihydroxy)-lambda 4 -sulfanyl]carbamoyl}-5,16-dioxo-1,2,3,5,6,7,8,9,10,11,13a,14,14a,15,16,16a-hexadecahydrocyclopropa[e]pyrrolo[1,2-a][1,4]diazacyclopentadecin-6-yl]carbamate (three-letter code: 3M7) (formula: C₃₆H₄₈ClN₅O₉S).



Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
			Total	C	Cl	H	N	O	S		
2	A	1	98	36	1	46	5	9	1	0	0
2	B	1	98	36	1	46	5	9	1	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO₄) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Zn	0	0
			1	1		
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	32	Total	O	0	0
			32	32		
5	B	28	Total	O	0	0
			28	28		

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 ($\text{RSRZ} > 2$). 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

Chain A:  91% 7% .



- Molecule 1: Genome polyprotein

Chain B:  86% 11% .



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	58.41Å 54.98Å 59.86Å 90.00° 90.09° 90.00°	Depositor
Resolution (Å)	24.87 – 2.70 24.87 – 2.70	Depositor EDS
% Data completeness (in resolution range)	97.7 (24.87-2.70) 93.3 (24.87-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	8.57 (at 2.72Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.171 , 0.239 0.185 , 0.202	Depositor DCC
R_{free} test set	466 reflections (4.95%)	DCC
Wilson B-factor (Å ²)	22.3	Xtriage
Anisotropy	0.539	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.46 , 15.4	EDS
Estimated twinning fraction	0.006 for l,k,-h 0.477 for h,-k,-l 0.017 for l,-k,h	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 10299 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	5729	wwPDB-VP
Average B, all atoms (Å ²)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.85% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 3M7, ZN, SO4

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	0.77	3/1415 (0.2%)	0.95	6/1935 (0.3%)
1	B	0.95	7/1417 (0.5%)	0.92	4/1937 (0.2%)
All	All	0.87	10/2832 (0.4%)	0.94	10/3872 (0.3%)

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	A	0	1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1145	CYS	CB-SG	-10.90	1.63	1.82
1	B	1097	CYS	CB-SG	7.66	1.95	1.82
1	B	1146	PRO	N-CD	6.82	1.57	1.47
1	B	1133	SER	CB-OG	-6.45	1.33	1.42
1	B	1159	CYS	CB-SG	-6.40	1.71	1.82

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1148	GLY	C-N-CA	12.04	151.79	121.70
1	A	1148	GLY	O-C-N	-9.37	107.70	122.70
1	B	1121	ASP	CB-CG-OD1	7.56	125.10	118.30
1	A	1148	GLY	CA-C-N	6.53	131.57	117.20
1	A	1062	ARG	NE-CZ-NH2	-6.13	117.23	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1149	HIS	Mainchain

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	1391	1336	1332	8	3
1	B	1393	1341	1337	20	1
2	A	52	46	46	8	0
2	B	52	46	46	5	0
3	A	10	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	32	0	0	2	0
5	B	28	0	0	1	0
All	All	2960	2769	2761	33	4

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 6.

The worst 5 of 33 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:A:1201:3M7:C43	2:A:1201:3M7:C44	1.75	1.60
1:B:1147:ALA:HB3	1:B:1149:HIS:CE1	2.12	0.85
1:A:1115:PRO:HB2	1:A:1127:LEU:HD22	1.60	0.81
1:B:1147:ALA:CB	1:B:1149:HIS:CE1	2.63	0.81
1:B:1115:PRO:HB2	1:B:1127:LEU:HD12	1.69	0.74

All (4) 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)
1:A:1068:LYS:NZ	1:A:1176:GLU:OE1[2_645]	1.67	0.53
1:A:1068:LYS:HZ3	1:A:1176:GLU:OE1[2_645]	1.51	0.09
1:B:1068:LYS:HZ1	1:B:1097:CYS:O[2_756]	1.57	0.03
1:A:1068:LYS:HZ2	1:A:1176:GLU:OE1[2_645]	1.59	0.01

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	196/198 (99%)	192 (98%)	4 (2%)	0	100	100
1	B	196/198 (99%)	193 (98%)	3 (2%)	0	100	100
All	All	392/396 (99%)	385 (98%)	7 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	143/160 (89%)	141 (99%)	2 (1%)	74	92
1	B	143/160 (89%)	139 (97%)	4 (3%)	51	81
All	All	286/320 (89%)	280 (98%)	6 (2%)	61	87

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1047	SER
1	B	1133	SER
1	B	1098	THR
1	A	1133	SER
1	B	1122	SER

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

Mol	Chain	Res	Type
1	B	982	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](#)

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 for Mogul analysis.

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	3M7	A	1201	-	51,57,57	3.29	22 (43%)	66,86,86	3.04	22 (33%)
3	SO4	A	1202	-	4,4,4	0.23	0	6,6,6	0.21	0
3	SO4	A	1203	-	4,4,4	2.01	3 (75%)	6,6,6	0.36	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	3M7	B	1201	-	51,57,57	3.24	21 (41%)	66,86,86	3.25	30 (45%)

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	3M7	A	1201	-	-	0/50/81/81	0/3/6/6
3	SO4	A	1202	-	-	0/0/0/0	0/0/0/0
3	SO4	A	1203	-	-	0/0/0/0	0/0/0/0
2	3M7	B	1201	-	-	0/50/81/81	0/3/6/6

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1201	3M7	C16-N22	-9.96	1.30	1.47
2	A	1201	3M7	C16-N22	-9.52	1.31	1.47
2	A	1201	3M7	C17-C18	-6.10	1.41	1.53
2	B	1201	3M7	C17-C18	-6.01	1.41	1.53
2	B	1201	3M7	O45-C44	-5.42	1.13	1.22

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1201	3M7	O45-C44-C43	-8.51	110.71	120.76
2	A	1201	3M7	C42-C43-N21	-8.08	106.60	117.84
2	B	1201	3M7	C17-C15-C16	-8.05	96.39	103.64
2	B	1201	3M7	C42-C43-N21	-7.73	107.09	117.84
2	A	1201	3M7	O45-C44-C43	-7.70	111.68	120.76

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1201	3M7	8	0
2	B	1201	3M7	5	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	198/198 (100%)	-0.53	0 100 100	2, 13, 35, 57	0
1	B	198/198 (100%)	-0.50	2 (1%) 84 85	2, 13, 33, 65	0
All	All	396/396 (100%)	-0.52	2 (0%) 91 93	2, 13, 35, 65	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	984	ALA	3.7
1	B	982	HIS	2.3

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(Å ²)	Q<0.9
4	ZN	A	1204	1/1	0.24	0.56	12.43	159,159,159,159	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
4	ZN	B	1202	1/1	0.89	0.24	3.26	30,30,30,30	0
3	SO4	A	1202	5/5	0.97	0.13	1.36	28,28,41,47	0
2	3M7	A	1201	52/52	0.94	0.18	1.19	20,20,20,20	0
3	SO4	A	1203	5/5	0.96	0.15	0.76	29,39,48,49	0
2	3M7	B	1201	52/52	0.95	0.16	0.07	20,20,20,20	0

6.5 Other polymers [i](#)

There are no such residues in this entry.