wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 10:33 PM EDT

PDB ID	:	7MMG
Title	:	Crystal structure of HCV NS3/4A D168A protease in complex with NR02-58
Authors	:	Zephyr, J.; Schiffer, C.A.
Deposited on	:	2021-04-29
Resolution	:	1.95 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:

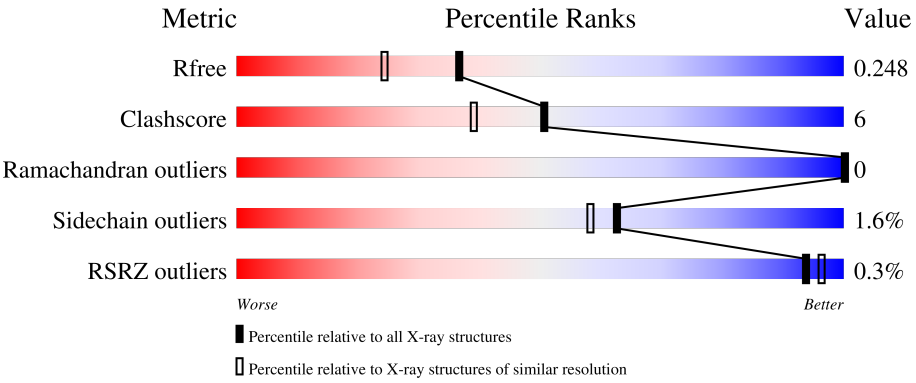
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	A	200	83% 10% • 6%
1	B	200	84% 9% 6%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5317 atoms, of which 2468 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 NS3/4A protease.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	188	Total	C	H	N	O	S	0	3	0
			2500	816	1184	239	255	6			
1	B	187	Total	C	H	N	O	S	0	1	0
			2493	810	1190	236	251	6			

There are 68 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...

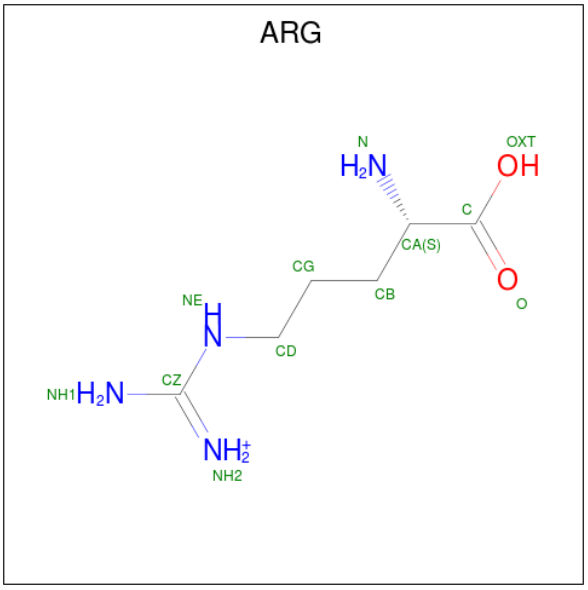
Chain	Residue	Modelled	Actual	Comment	Reference
A	1018	GLU	ILE	conflict	UNP A8DG50
A	1021	GLN	LEU	conflict	UNP A8DG50
A	1040	THR	ALA	conflict	UNP A8DG50
A	1047	SER	CYS	conflict	UNP A8DG50
A	1052	LEU	CYS	conflict	UNP A8DG50
A	1072	THR	ILE	conflict	UNP A8DG50
A	1086	GLN	PRO	conflict	UNP A8DG50
A	1159	SER	CYS	conflict	UNP A8DG50
A	1168	ALA	ASP	engineered mutation	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	1159	SER	CYS	conflict	UNP A8DG50

Continued on next page...

Continued from previous page...

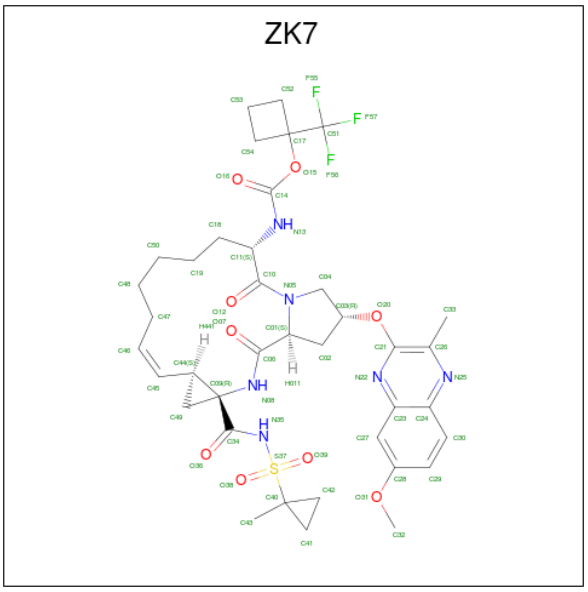
Chain	Residue	Modelled	Actual	Comment	Reference
B	1168	ALA	ASP	engineered mutation	UNP A8DG50

- Molecule 2 is ARGININE (three-letter code: ARG) (formula: $C_6H_{15}N_4O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	5	3	1	1	0	0

- Molecule 3 is 1-(trifluoromethyl)cyclobutyl {(2R,4S,6S,12Z,13aS,14aR,16aS)-2-[(7-methoxy-3-methylquinoxalin-2-yl)oxy]-14a-[(1-methylcyclopropane-1-sulfonyl)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: ZK7) (formula: $C_{38}H_{47}F_3N_6O_9S$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
3	A	1	Total	C	F	H	N	O	S	0	0
			104	38	3	47	6	9	1		
3	B	1	Total	C	F	H	N	O	S	0	0
			104	38	3	47	6	9	1		

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

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

- Molecule 5 is water.

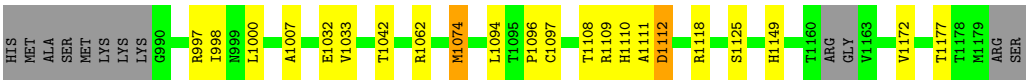
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	49	Total	O	0	0
			49	49		
5	B	60	Total	O	0	0
			60	60		

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.

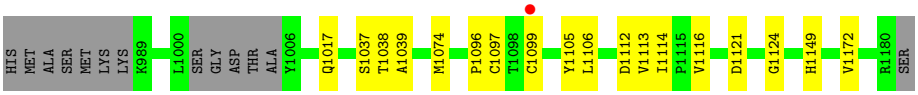
- Molecule 1: NS3/4A protease

Chain A:



- Molecule 1: NS3/4A protease

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	39.85Å 50.99Å 77.07Å 90.00° 92.84° 90.00°	Depositor
Resolution (Å)	42.51 – 1.95 42.51 – 1.95	Depositor EDS
% Data completeness (in resolution range)	100.0 (42.51-1.95) 99.9 (42.51-1.95)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.80 (at 1.95Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.199 , 0.248 0.198 , 0.248	Depositor DCC
R_{free} test set	1141 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	14.6	Xtriage
Anisotropy	1.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 36.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.136 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5317	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.05% 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: ZK7, ZN

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	0.52	0/1338	0.71	0/1832
1	B	0.57	0/1324	0.74	1/1812 (0.1%)
All	All	0.54	0/2662	0.72	1/3644 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1121	ASP	CB-CG-OD2	-5.10	113.71	118.30

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	A	1316	1184	1229	13	0
1	B	1303	1190	1245	13	0
2	A	5	0	1	1	0
3	A	57	47	0	4	0
3	B	57	47	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	49	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	60	0	0	2	0
All	All	2849	2468	2475	30	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 6.

The worst 5 of 30 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)
3:A:1202:ZK7:C01	3:A:1202:ZK7:N05	1.67	1.15
3:A:1202:ZK7:N05	3:A:1202:ZK7:C06	2.42	0.82
1:A:1097:CYS:HB2	1:A:1149:HIS:HB2	1.76	0.68
1:A:1177:THR:O	2:A:1201:ARG:N	2.32	0.62
1:B:1097:CYS:HB2	1:B:1149:HIS:HB2	1.82	0.62

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	187/200 (94%)	180 (96%)	7 (4%)	0	100	100
1	B	184/200 (92%)	180 (98%)	4 (2%)	0	100	100
All	All	371/400 (93%)	360 (97%)	11 (3%)	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	128/162 (79%)	124 (97%)	4 (3%)	40 28
1	B	131/162 (81%)	131 (100%)	0	100 100
All	All	259/324 (80%)	255 (98%)	4 (2%)	62 60

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	997	ARG
1	A	1074	MET
1	A	1112	ASP
1	A	1125	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 2 are monoatomic - leaving 3 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	ARG	A	1201	-	3,4,11	0.78	0	2,4,13	1.00	0
3	ZK7	B	1301	-	57,63,63	4.61	26 (45%)	79,98,98	2.53	28 (35%)
3	ZK7	A	1202	-	57,63,63	4.70	26 (45%)	79,98,98	2.67	25 (31%)

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	ARG	A	1201	-	-	0/0/2/11	-
3	ZK7	B	1301	-	-	7/61/103/103	0/6/7/7
3	ZK7	A	1202	-	-	4/61/103/103	0/6/7/7

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1202	ZK7	C02-C01	-13.23	1.28	1.53
3	B	1301	ZK7	C02-C01	-13.17	1.28	1.53
3	B	1301	ZK7	C26-N25	10.59	1.46	1.32
3	A	1202	ZK7	C26-N25	10.43	1.45	1.32
3	A	1202	ZK7	C01-N05	10.09	1.67	1.47

The worst 5 of 53 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	1301	ZK7	O39-S37-C40	9.23	115.01	107.60
3	A	1202	ZK7	O39-S37-C40	8.33	114.28	107.60
3	A	1202	ZK7	C34-C09-N08	-8.14	107.30	116.06
3	A	1202	ZK7	C49-C09-N08	7.21	127.48	117.80
3	B	1301	ZK7	O39-S37-O38	-6.39	107.59	120.57

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1202	ZK7	C34-N35-S37-C40
3	B	1301	ZK7	C34-N35-S37-C40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	1202	ZK7	C18-C19-C50-C48
3	B	1301	ZK7	C18-C19-C50-C48
3	B	1301	ZK7	C27-C28-O31-C32

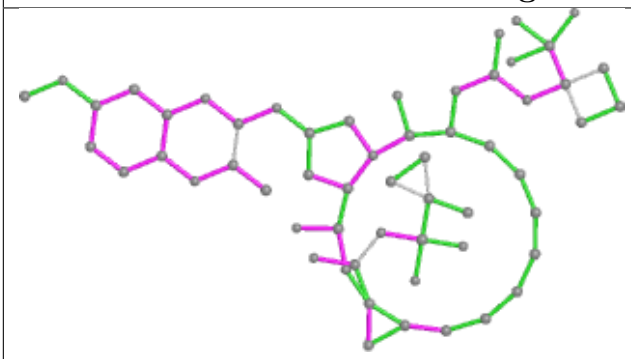
There are no ring outliers.

2 monomers are involved in 5 short contacts:

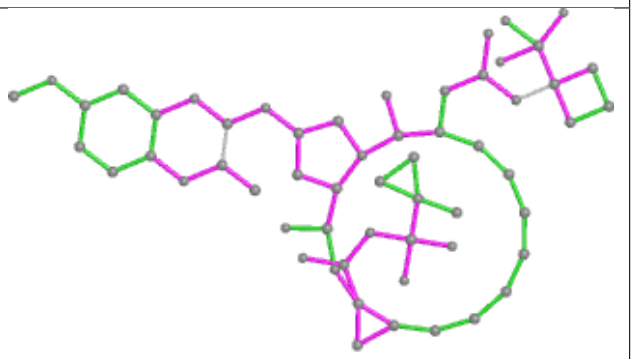
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1201	ARG	1	0
3	A	1202	ZK7	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

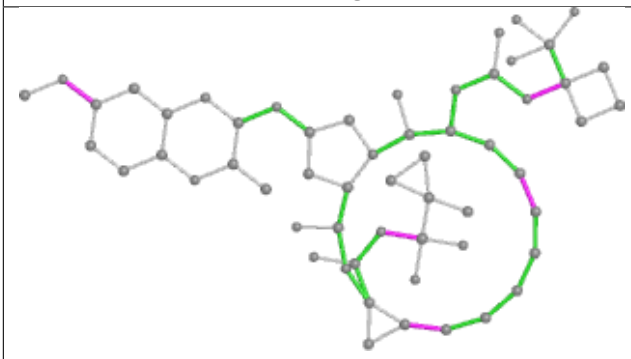
Ligand ZK7 B 1301



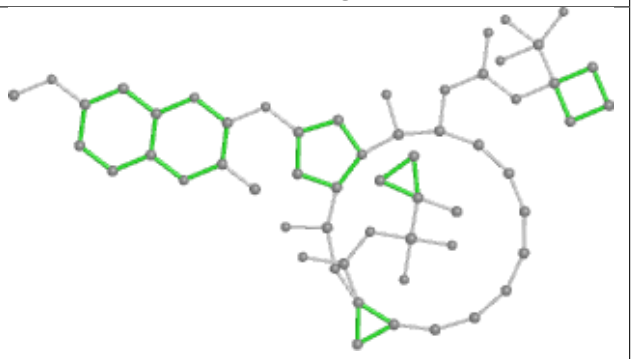
Bond lengths



Bond angles

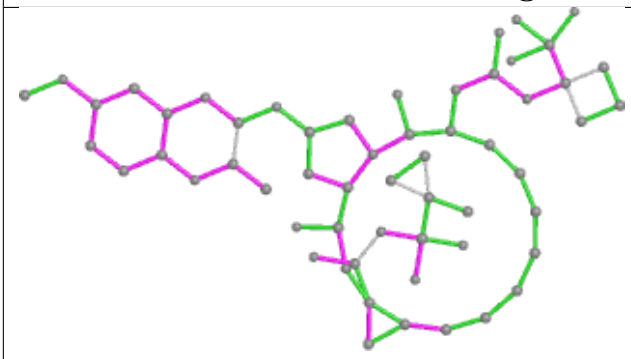


Torsions

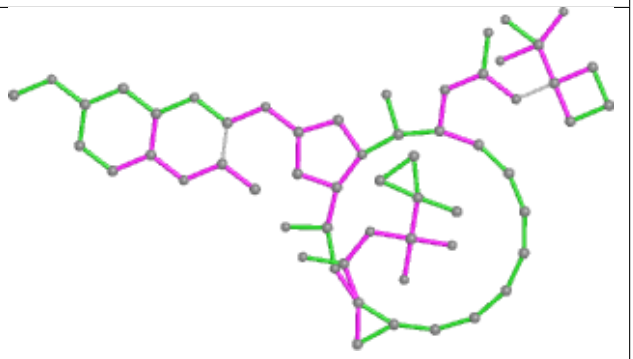


Rings

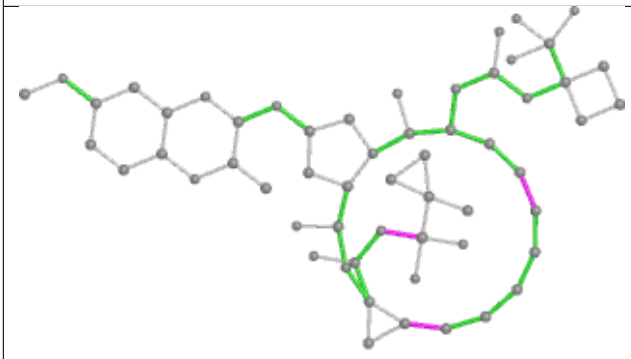
Ligand ZK7 A 1202



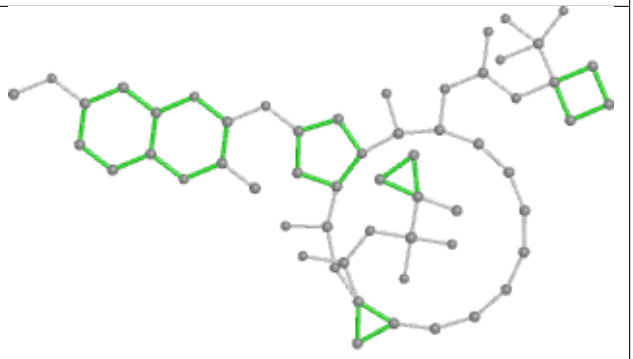
Bond lengths



Bond angles



Torsions



Rings

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, 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	A	188/200 (94%)	-0.09	0 100 100	11, 20, 34, 42	0
1	B	187/200 (93%)	-0.07	1 (0%) 91 94	11, 19, 33, 48	0
All	All	375/400 (93%)	-0.08	1 (0%) 94 96	11, 19, 34, 48	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1099	CYS	2.3

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

There are no non-standard protein/DNA/RNA residues in this entry.

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
2	ARG	A	1201	5/12	0.88	0.16	35,37,39,46	0
3	ZK7	B	1301	57/57	0.94	0.13	13,17,28,34	0
3	ZK7	A	1202	57/57	0.95	0.11	12,19,28,30	0
4	ZN	A	1203	1/1	0.97	0.06	35,35,35,35	0

Continued on next page...

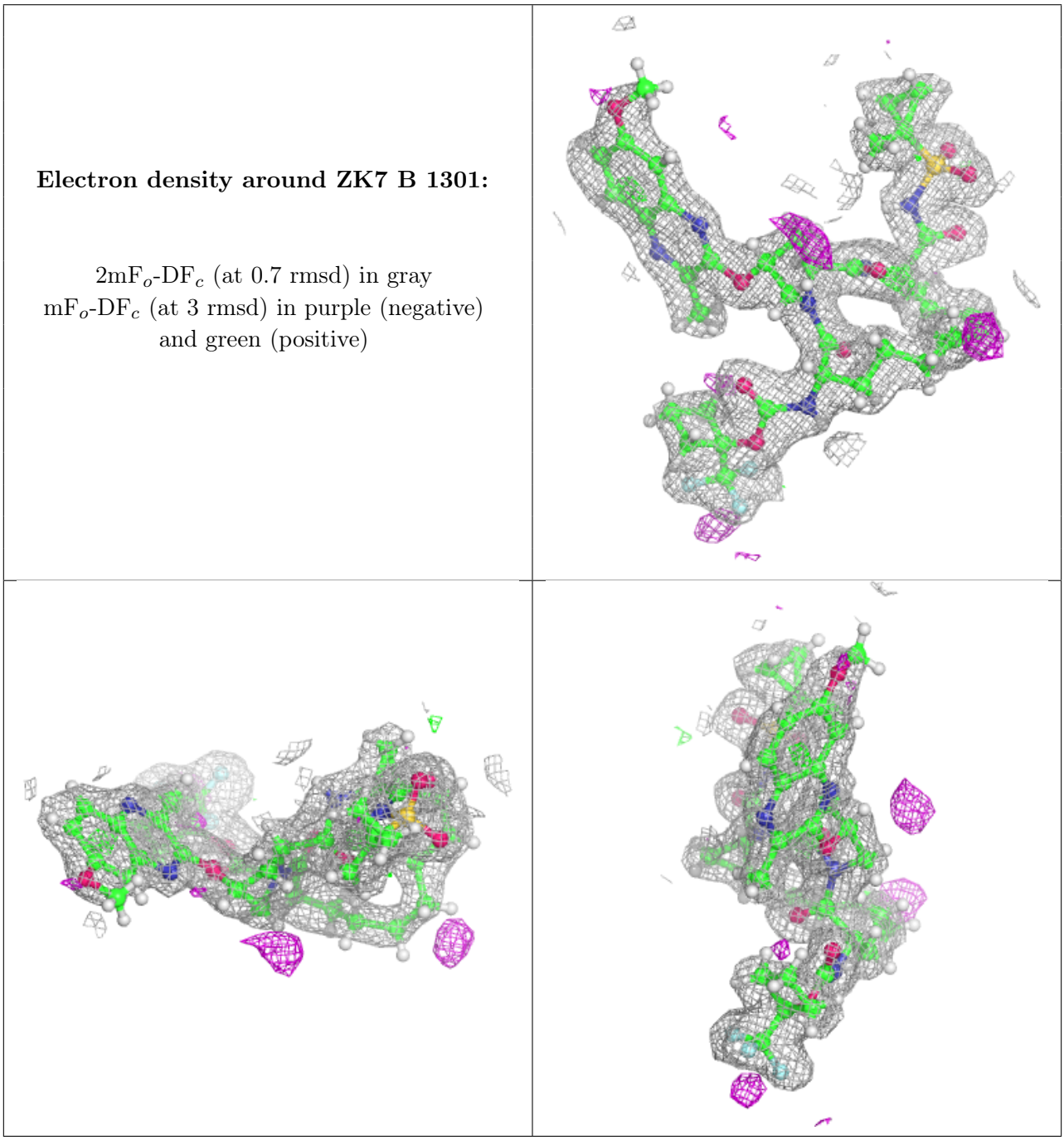
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	ZN	B	1302	1/1	0.99	0.05	22,22,22,22	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

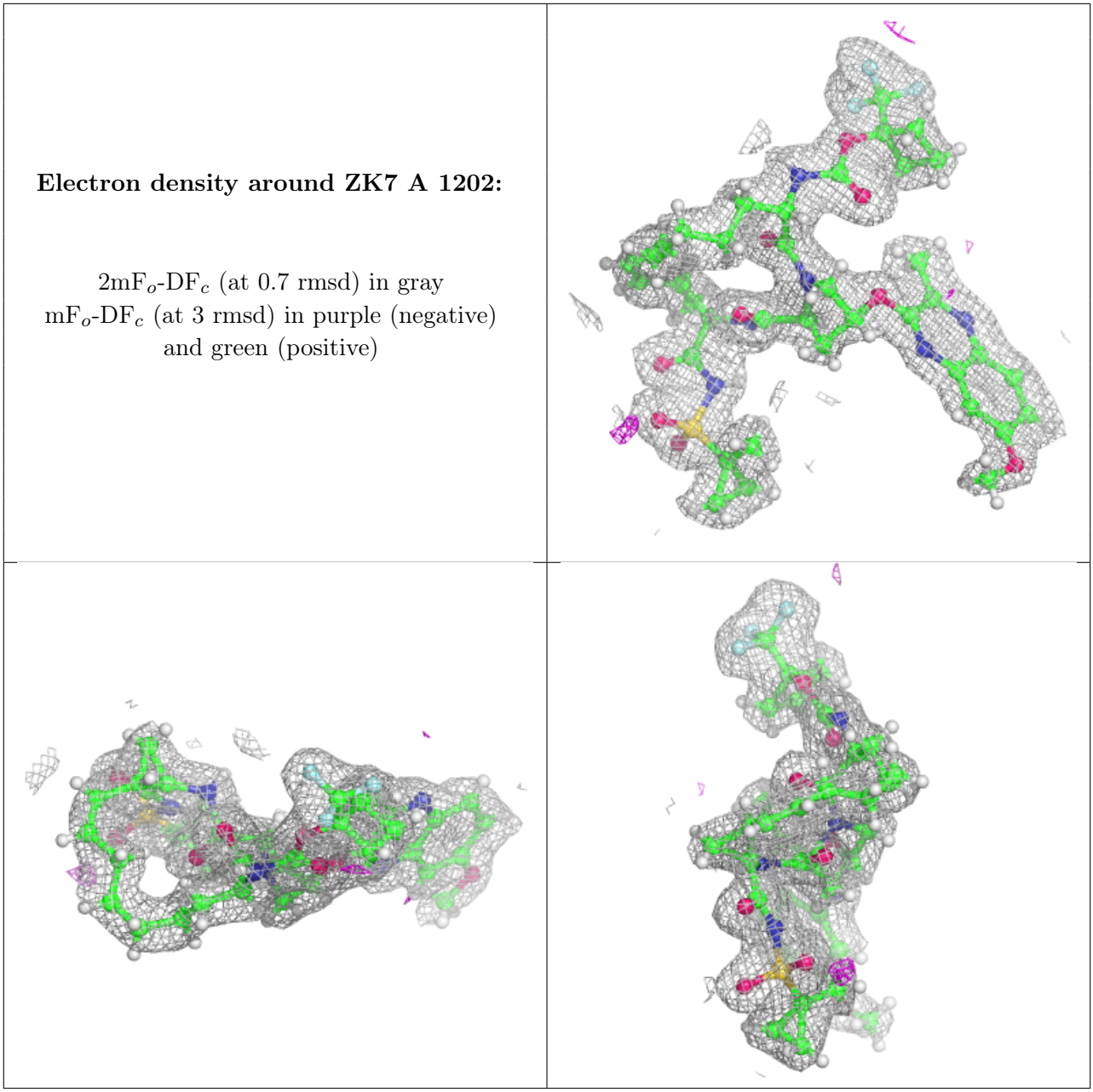
Electron density around ZK7 B 1301:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around ZK7 A 1202:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.