wwPDB X-ray Structure Validation Summary Report (i)

Nov 29, 2021 – 03:09 pm GMT

PDB ID	:	7OXQ
Title	:	Crystal structure of HIV-1 reverse transcriptase with a double stranded DNA in complex with fragment 048 at the transient P-pocket.
Authors	:	Martinez, S.E.; Singh, A.K.; Das, K.
Deposited on	:	2021-06-22
Resolution	:	3.30 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 following versions of software and data (see references (i)) were used in the production of this report:

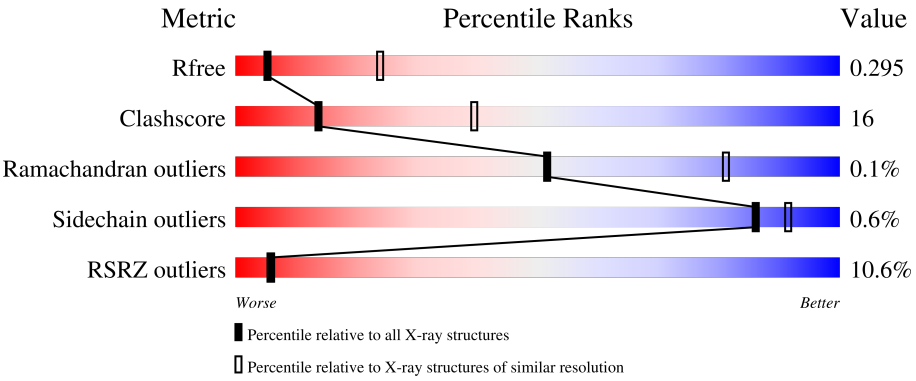
MolProbity	:	4.02b-467
Mogul	:	1.8.4 (270009), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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	556	14% 67% 33%
1	C	556	14% 64% 35%
2	B	444	6% 65% 29% 6%
2	D	444	8% 61% 30% 9%
3	E	28	46% 32% 21%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	T	28	
4	F	21	
4	P	21	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 17577 atoms, of which 12 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	555	Total	C	N	O	S	0	0	0
			4504	2913	749	834	8			
1	C	551	Total	C	N	O	S	0	0	0
			4486	2902	747	829	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	63	CYS	ILE	conflict	UNP P03366
A	280	SER	CYS	conflict	UNP P03366
C	-1	MET	-	initiating methionine	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	63	CYS	ILE	conflict	UNP P03366
C	280	SER	CYS	conflict	UNP P03366

- Molecule 2 is a protein called Gag-Pol polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	418	Total	C	N	O	S	0	0	0
			3431	2228	572	625	6			
2	D	406	Total	C	N	O	S	0	0	0
			3349	2180	552	611	6			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-15	MET	-	initiating methionine	UNP P03366
B	-14	ALA	-	expression tag	UNP P03366
B	-13	HIS	-	expression tag	UNP P03366
B	-12	HIS	-	expression tag	UNP P03366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-11	HIS	-	expression tag	UNP P03366
B	-10	HIS	-	expression tag	UNP P03366
B	-9	HIS	-	expression tag	UNP P03366
B	-8	HIS	-	expression tag	UNP P03366
B	-7	ALA	-	expression tag	UNP P03366
B	-6	LEU	-	expression tag	UNP P03366
B	-5	GLU	-	expression tag	UNP P03366
B	-4	VAL	-	expression tag	UNP P03366
B	-3	LEU	-	expression tag	UNP P03366
B	-2	PHE	-	expression tag	UNP P03366
B	-1	GLN	-	expression tag	UNP P03366
B	0	GLY	-	expression tag	UNP P03366
B	280	SER	CYS	engineered mutation	UNP P03366
D	-15	MET	-	initiating methionine	UNP P03366
D	-14	ALA	-	expression tag	UNP P03366
D	-13	HIS	-	expression tag	UNP P03366
D	-12	HIS	-	expression tag	UNP P03366
D	-11	HIS	-	expression tag	UNP P03366
D	-10	HIS	-	expression tag	UNP P03366
D	-9	HIS	-	expression tag	UNP P03366
D	-8	HIS	-	expression tag	UNP P03366
D	-7	ALA	-	expression tag	UNP P03366
D	-6	LEU	-	expression tag	UNP P03366
D	-5	GLU	-	expression tag	UNP P03366
D	-4	VAL	-	expression tag	UNP P03366
D	-3	LEU	-	expression tag	UNP P03366
D	-2	PHE	-	expression tag	UNP P03366
D	-1	GLN	-	expression tag	UNP P03366
D	0	GLY	-	expression tag	UNP P03366
D	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 3 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	23	Total	C	N	O	P	0	0	0
			474	223	92	136	23			
3	E	22	Total	C	N	O	P	0	0	0
			454	213	90	129	22			

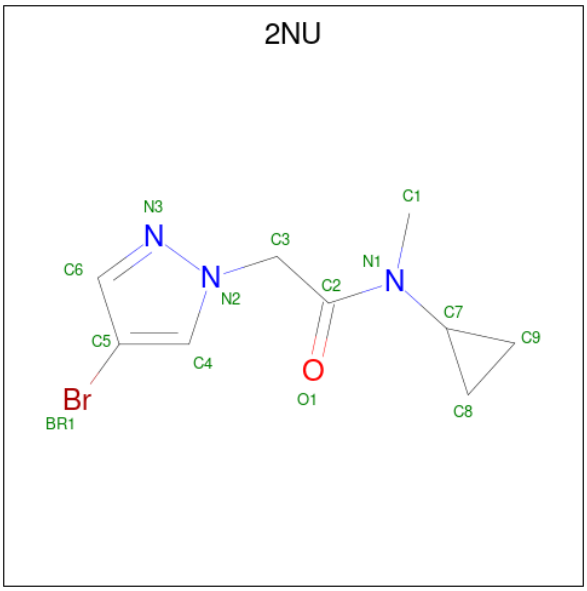
- Molecule 4 is a DNA chain called DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*GP*CP*GP*CP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	P	21	Total 425	C 202	N 77	O 126	P 20	0	0	0
4	F	20	Total 407	C 192	N 72	O 123	P 20	0	0	0

- Molecule 5 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	3	Total 3	Cd 3	0	0
5	B	3	Total 3	Cd 3	0	0
5	T	2	Total 2	Cd 2	0	0
5	P	1	Total 1	Cd 1	0	0
5	C	3	Total 3	Cd 3	0	0
5	D	6	Total 6	Cd 6	0	0
5	E	2	Total 2	Cd 2	0	0
5	F	1	Total 1	Cd 1	0	0

- Molecule 6 is 2-(4-bromanylpyrazol-1-yl)- {N}-cyclopropyl- {N}-methyl-ethanamide (three-letter code: 2NU) (formula: $C_9H_{12}BrN_3O$) (labeled as "Ligand of Interest" by depositor).



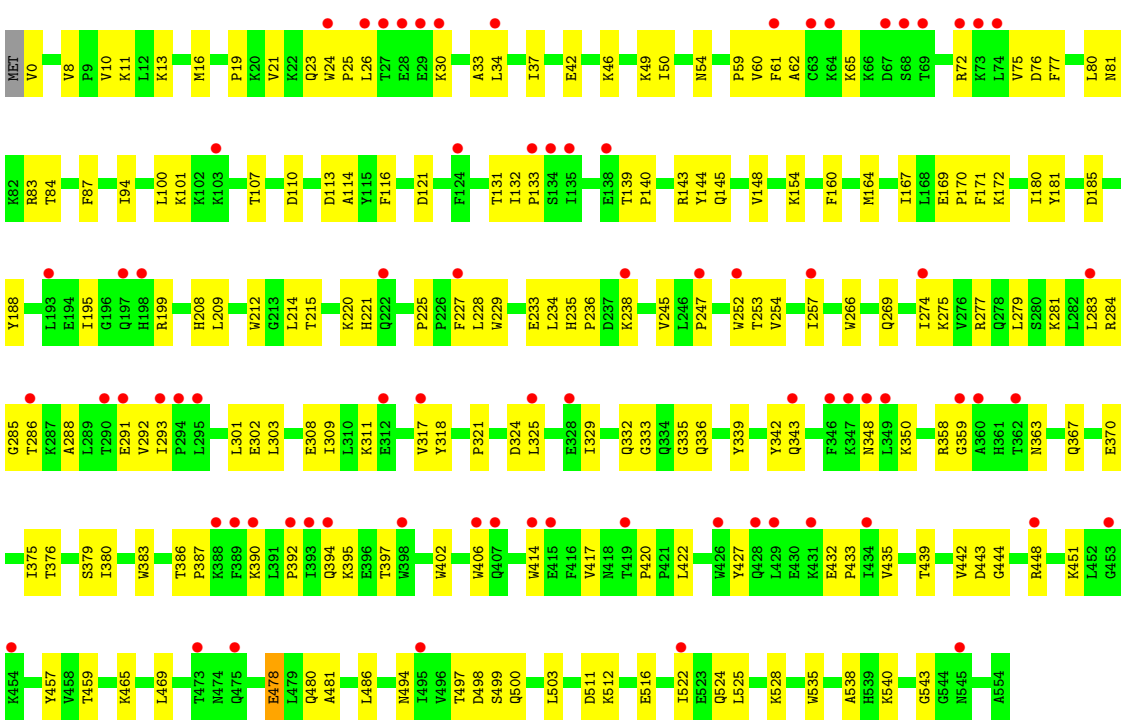
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
6	C	1	Total	Br	C	H	N	O	0	0
			26	1	9	12	3	1		

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.

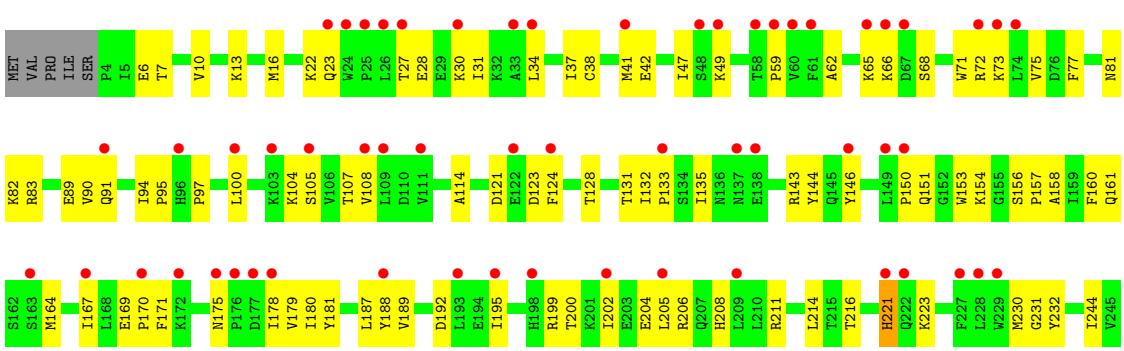
• Molecule 1: Reverse transcriptase/ribonuclease H

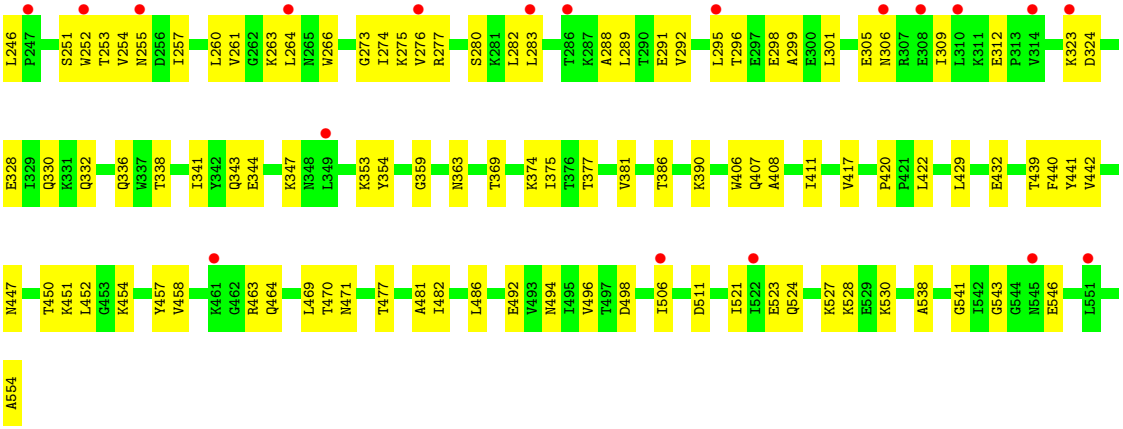
Chain A:



• Molecule 1: Reverse transcriptase/ribonuclease H

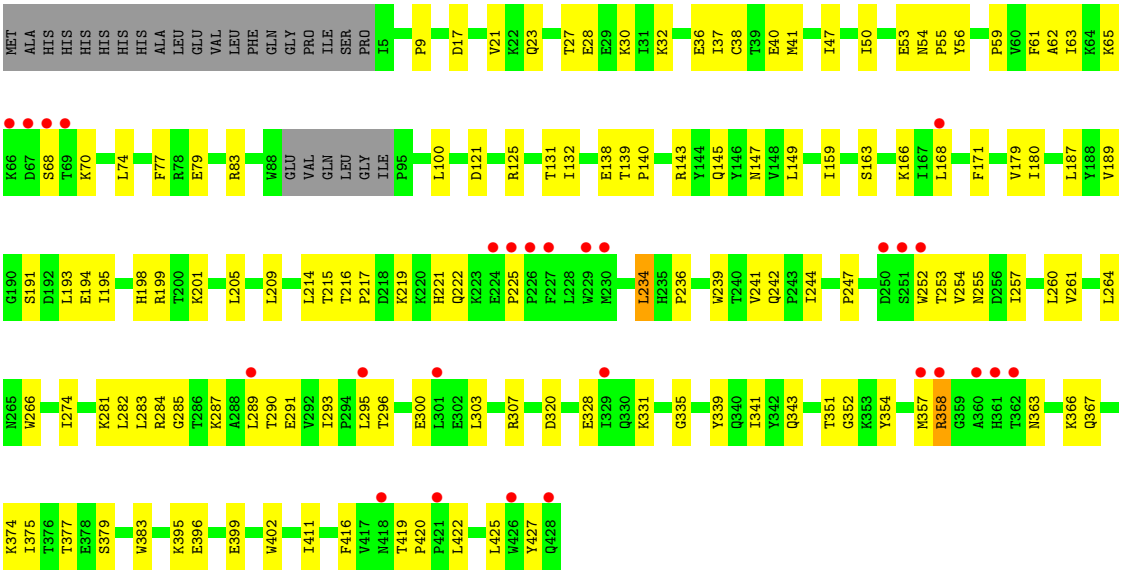
Chain C:





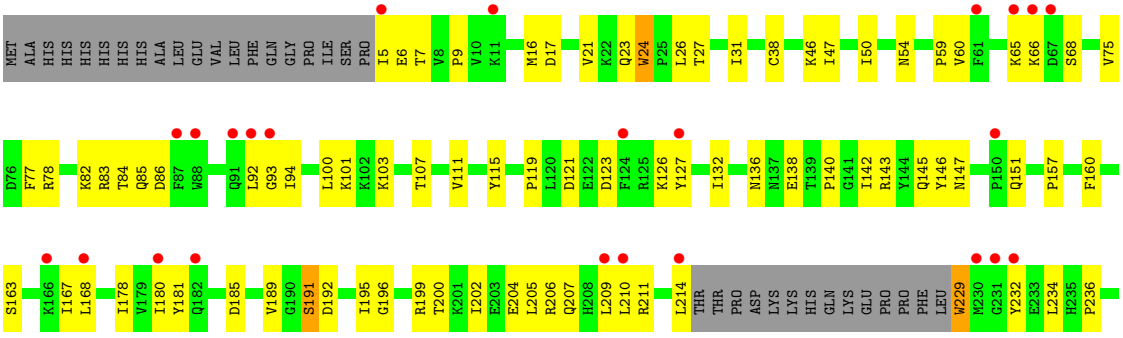
• Molecule 2: Gag-Pol polyprotein

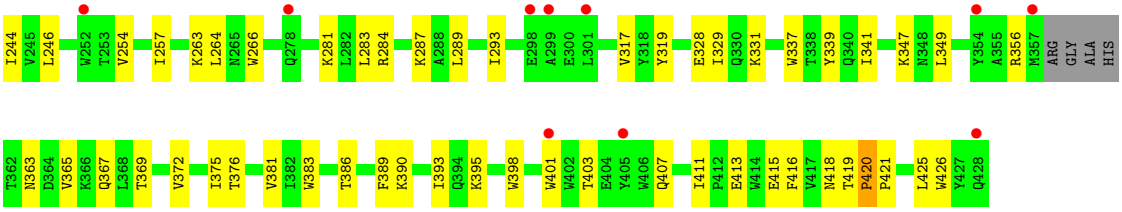
Chain B:



• Molecule 2: Gag-Pol polyprotein

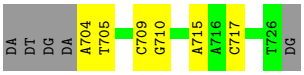
Chain D:





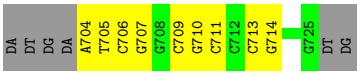
• Molecule 3: DNA (28-MER)

Chain T:



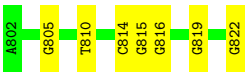
• Molecule 3: DNA (28-MER)

Chain E:



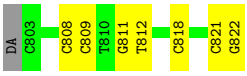
• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*GP *CP*GP*CP*CP*G)-3')

Chain P:



• Molecule 4: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*GP *CP*GP*CP*CP*G)-3')

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	309.56Å 61.90Å 168.78Å 90.00° 104.55° 90.00°	Depositor
Resolution (Å)	81.68 – 3.30 81.68 – 3.30	Depositor EDS
% Data completeness (in resolution range)	93.8 (81.68-3.30) 93.8 (81.68-3.30)	Depositor EDS
R_{merge}	0.32	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 3.33Å)	Xtriage
Refinement program	PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.262 , 0.296 0.261 , 0.295	Depositor DCC
R_{free} test set	2249 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	62.7	Xtriage
Anisotropy	0.516	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.81	EDS
Total number of atoms	17577	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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: 2NU, CD

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	0.30	0/4622	0.54	0/6283
1	C	0.31	0/4603	0.59	0/6252
2	B	0.34	0/3530	0.59	3/4797 (0.1%)
2	D	0.34	0/3442	0.58	1/4674 (0.0%)
3	E	0.74	0/510	0.88	0/785
3	T	0.58	0/532	0.80	0/819
4	F	0.67	1/454 (0.2%)	0.90	1/698 (0.1%)
4	P	0.58	0/475	0.81	0/731
All	All	0.37	1/18168 (0.0%)	0.61	5/25039 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	F	811	DG	C4'-O4'	5.63	1.50	1.45

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	358	ARG	N-CA-CB	7.10	123.37	110.60
4	F	818	DC	O4'-C4'-C3'	-5.60	102.26	104.50
2	B	357	MET	N-CA-C	5.44	125.69	111.00
2	B	234	LEU	CA-CB-CG	5.36	127.62	115.30
2	D	93	GLY	N-CA-C	5.15	125.96	113.10

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	4504	0	4543	169	0
1	C	4486	0	4538	185	1
2	B	3431	0	3431	101	1
2	D	3349	0	3370	105	1
3	E	454	0	245	6	0
3	T	474	0	257	5	1
4	F	407	0	225	8	0
4	P	425	0	237	10	0
5	A	3	0	0	0	0
5	B	3	0	0	0	0
5	C	3	0	0	0	0
5	D	6	0	0	0	0
5	E	2	0	0	0	0
5	F	1	0	0	0	0
5	P	1	0	0	0	0
5	T	2	0	0	0	0
6	C	14	12	0	2	0
All	All	17565	12	16846	546	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 16.

The worst 5 of 546 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:C:131:THR:HG22	1:C:143:ARG:HG2	1.18	1.11
1:C:202:ILE:HG21	1:C:223:LYS:HZ2	1.18	1.02
2:B:281:LYS:HG2	2:B:284:ARG:CZ	1.89	1.02
1:C:202:ILE:HG21	1:C:223:LYS:NZ	1.75	1.01
2:B:281:LYS:HG2	2:B:284:ARG:NH1	1.80	0.96

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)
2:B:199:ARG:NH2	3:T:717:DC:OP1[1_565]	1.98	0.22
1:C:6:GLU:OE1	2:D:121:ASP:OD2[2_455]	2.06	0.14

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	553/556 (100%)	539 (98%)	14 (2%)	0	100 100
1	C	549/556 (99%)	534 (97%)	14 (3%)	1 (0%)	47 77
2	B	414/444 (93%)	399 (96%)	15 (4%)	0	100 100
2	D	400/444 (90%)	385 (96%)	14 (4%)	1 (0%)	41 71
All	All	1916/2000 (96%)	1857 (97%)	57 (3%)	2 (0%)	51 81

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	543	GLY
2	D	420	PRO

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	493/497 (99%)	492 (100%)	1 (0%)	93 97
1	C	492/497 (99%)	490 (100%)	2 (0%)	91 95

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	373/403 (93%)	370 (99%)	3 (1%)	81 89
2	D	367/403 (91%)	362 (99%)	5 (1%)	67 82
All	All	1725/1800 (96%)	1714 (99%)	11 (1%)	86 91

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	86	ASP
2	D	92	LEU
2	D	229	TRP
2	D	191	SER
1	C	22	LYS

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

Mol	Chain	Res	Type
2	D	91	GLN
2	D	428	GLN

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 22 ligands modelled in this entry, 21 are monoatomic - leaving 1 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

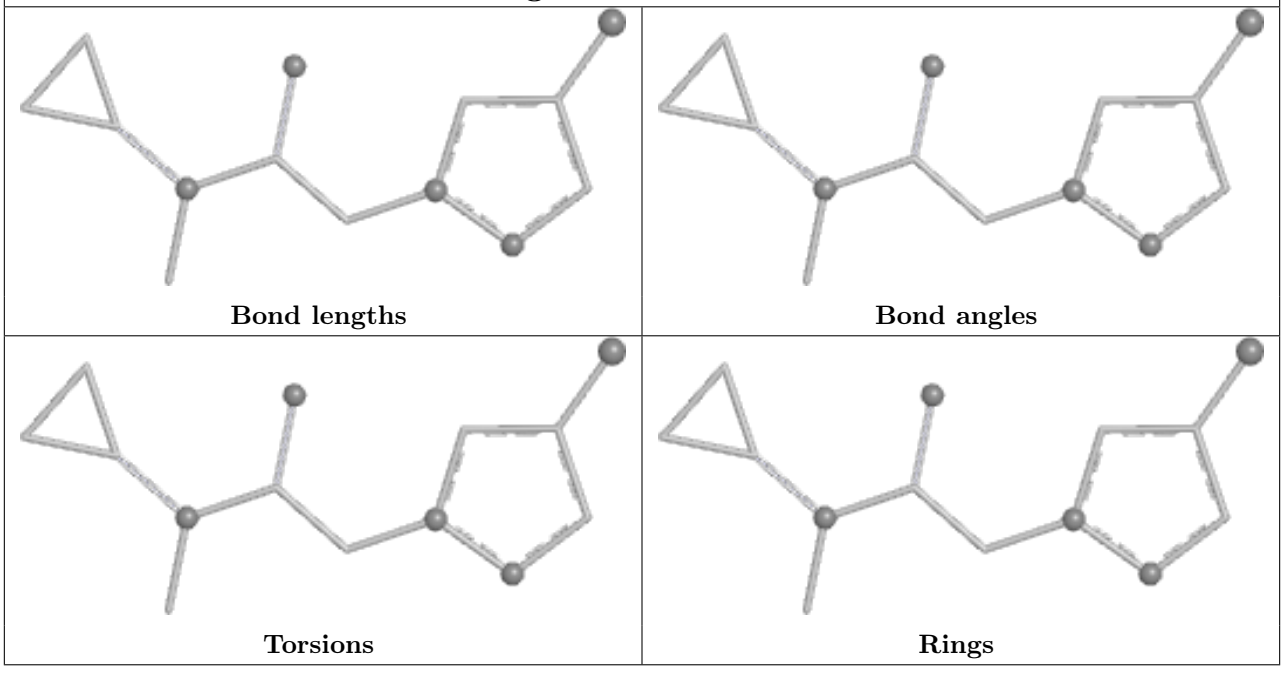
There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 2NU C 601



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	555/556 (99%)	0.90	76 (13%) 3 2	22, 56, 124, 165	0
1	C	551/556 (99%)	0.80	76 (13%) 2 2	17, 63, 120, 174	0
2	B	418/444 (94%)	0.43	27 (6%) 18 18	18, 49, 101, 138	0
2	D	406/444 (91%)	0.51	34 (8%) 11 10	26, 53, 101, 180	0
3	E	22/28 (78%)	-0.34	0 100 100	56, 85, 117, 122	0
3	T	23/28 (82%)	-0.10	0 100 100	28, 73, 113, 152	0
4	F	20/21 (95%)	-0.38	0 100 100	66, 85, 120, 131	0
4	P	21/21 (100%)	-0.25	0 100 100	41, 57, 108, 130	0
All	All	2016/2098 (96%)	0.65	213 (10%) 6 6	17, 57, 117, 180	0

The worst 5 of 213 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	63	CYS	9.5
1	A	30	LYS	7.5
1	A	27	THR	7.4
1	C	24	TRP	7.4
2	B	225	PRO	7.2

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 (i)

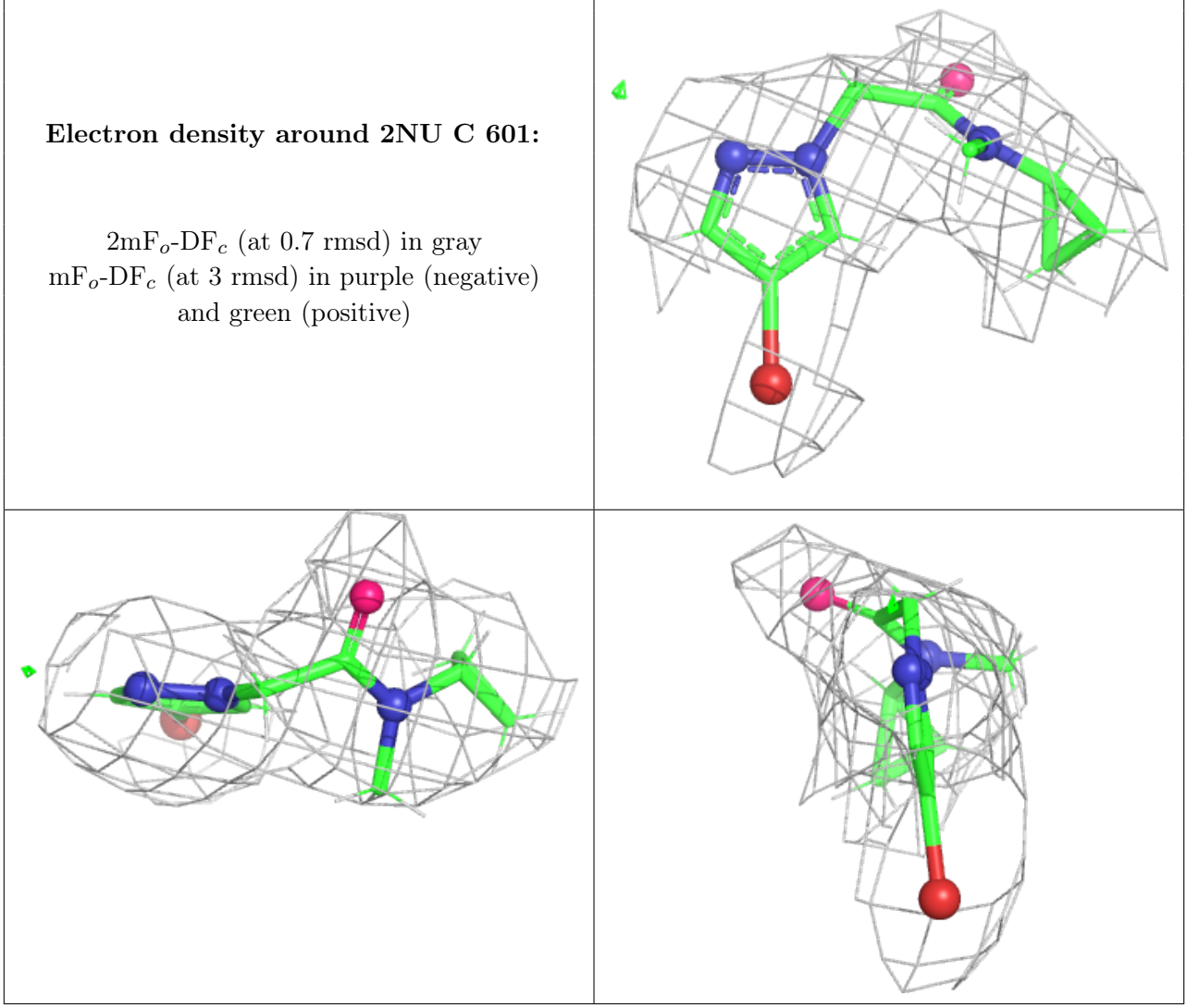
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
5	CD	B	502	1/1	0.62	0.11	203,203,203,203	0
5	CD	E	801	1/1	0.77	0.17	144,144,144,144	0
5	CD	D	506	1/1	0.78	0.19	190,190,190,190	0
5	CD	E	802	1/1	0.79	0.10	150,150,150,150	0
5	CD	T	801	1/1	0.80	0.09	157,157,157,157	0
5	CD	C	604	1/1	0.82	0.06	178,178,178,178	0
5	CD	B	503	1/1	0.84	0.09	143,143,143,143	0
6	2NU	C	601	14/14	0.84	0.16	57,76,100,128	0
5	CD	T	802	1/1	0.85	0.22	154,154,154,154	0
5	CD	D	503	1/1	0.87	0.11	118,118,118,118	0
5	CD	C	603	1/1	0.89	0.12	101,101,101,101	0
5	CD	B	501	1/1	0.92	0.12	102,102,102,102	0
5	CD	D	504	1/1	0.92	0.11	109,109,109,109	0
5	CD	D	505	1/1	0.93	0.05	134,134,134,134	0
5	CD	A	602	1/1	0.94	0.19	76,76,76,76	0
5	CD	D	502	1/1	0.94	0.17	93,93,93,93	0
5	CD	A	603	1/1	0.94	0.17	99,99,99,99	0
5	CD	F	901	1/1	0.97	0.14	83,83,83,83	0
5	CD	P	901	1/1	0.97	0.12	103,103,103,103	0
5	CD	D	501	1/1	0.98	0.18	51,51,51,51	0
5	CD	A	601	1/1	0.98	0.23	52,52,52,52	0
5	CD	C	602	1/1	0.99	0.20	40,40,40,40	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 2NU C 601:

$2mF_o$-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.