wwPDB X-ray Structure Validation Summary Report (i)

Jan 6, 2021 – 02:04 PM EST

PDB ID	:	7KRC
Title	:	Crystal Structure of HIV-1 Reverse Transcriptase in Complex with (E)-4-(3-chloro-5-(2-cyanovinyl)phenoxy)-3-(2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy)phenyl sulfurofluoridate (JLJ709)
Authors	:	Bertoletti, N.; Ippolito, J.A.; Jorgensen, W.L.; Anderson, K.S.
Deposited on	:	2020-11-19
Resolution	:	2.65 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:

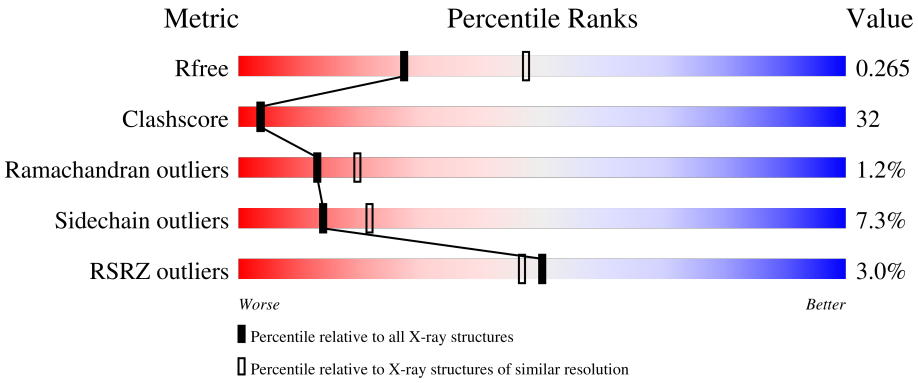
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	557	3%, 52%, 42%, 5%
2	B	428	3%, 50%, 42%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7805 atoms, of which 15 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 HIV-1 REVERSE TRANSCRIPTASE, P66 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	548	4416	2861	731	817	7	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	172	ALA	LYS	engineered mutation	UNP P03366
A	173	ALA	LYS	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366

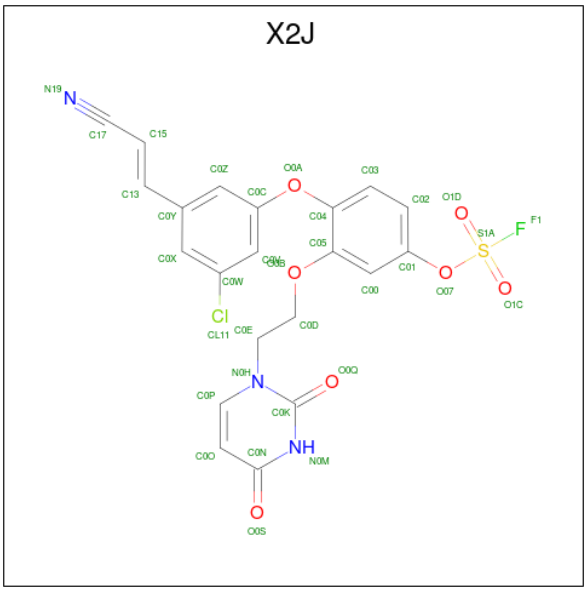
- Molecule 2 is a protein called HIV-1 REVERSE TRANSCRIPTASE, P51 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	409	3341	2179	548	608	6	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P03366

- Molecule 3 is 4-{3-chloro-5-[(E)-2-cyanoethenyl]phenoxy}-3-[2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy]phenyl sulfurofluoridate (three-letter code: X2J) (formula: $C_{21}H_{15}ClFN_3O_7S$) (labeled as "Ligand of Interest" by depositor).



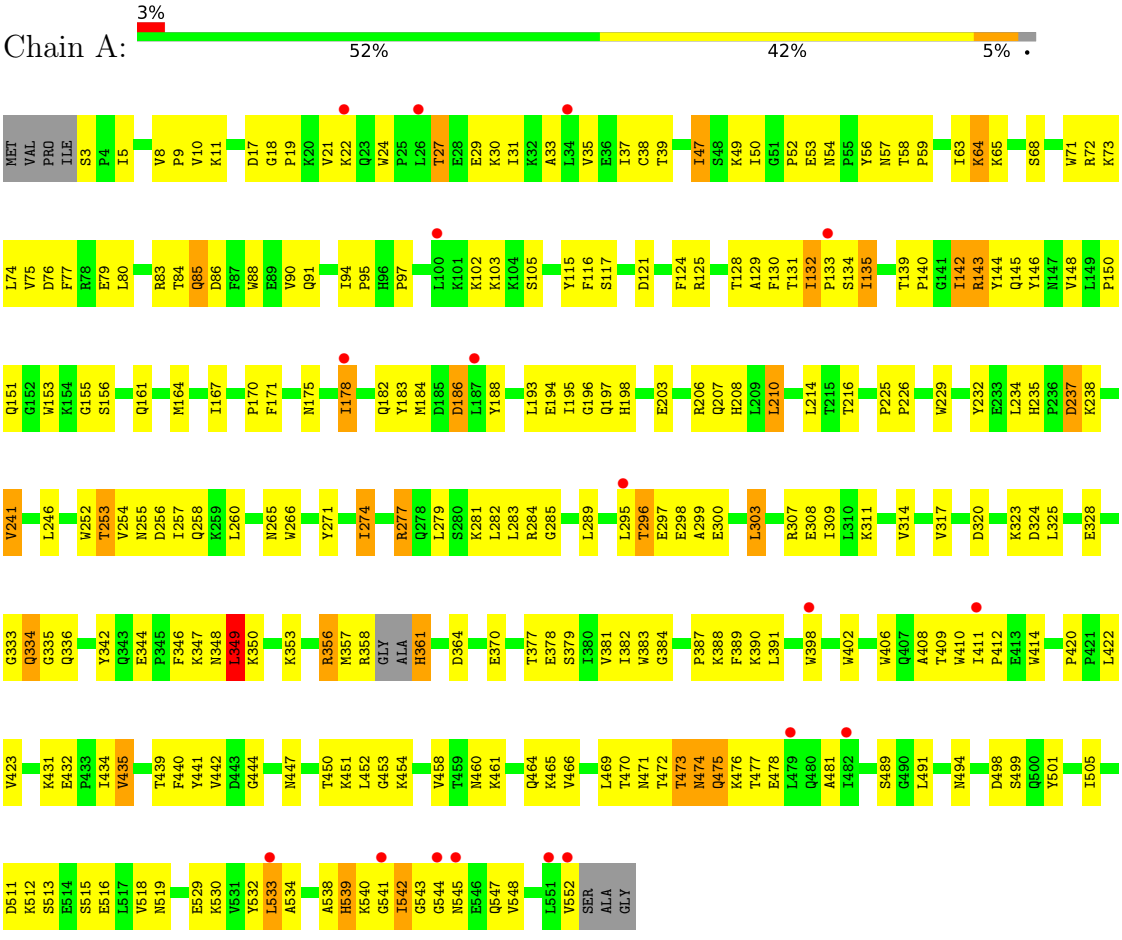
Mol	Chain	Residues	Atoms							ZeroOcc	AltConf
			Total	C	Cl	H	N	O	S		
3	A	1	48	21	1	15	3	7	1	0	0

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.

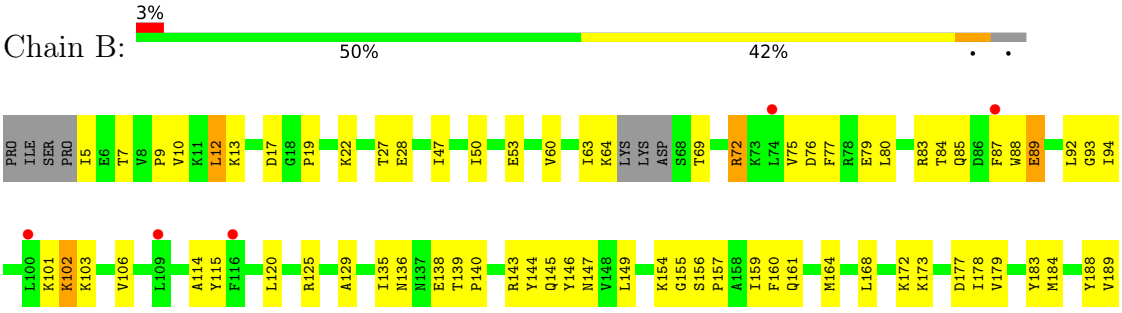
• Molecule 1: HIV-1 REVERSE TRANSCRIPTASE, P66 SUBUNIT

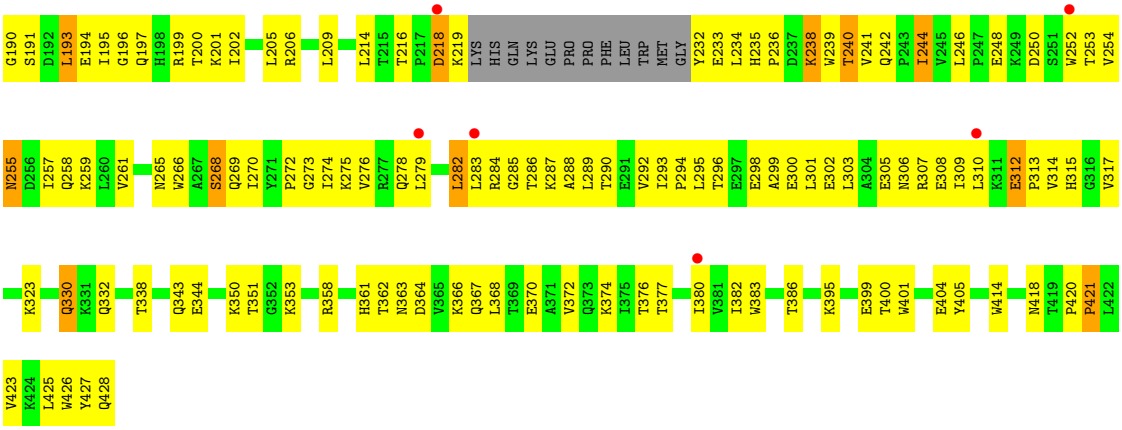
Chain A:



• Molecule 2: HIV-1 REVERSE TRANSCRIPTASE, P51 SUBUNIT

Chain B:




G190 S191 D192 L193 E194 I195 G196 Q197 H198 R199 T200 K201 I202 L205 R206 L209 L214 T215 T216 P217 D218 K219 LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET GLY Y232 E233 L234 H235 P236 D237 K238 W239 T240 V241 Q242 P243 I244 V245 L246 P247 E248 K249 D250 S251 W252 T253 V254
N255 D256 I257 Q258 K259 L260 V261 N265 W266 A267 S268 Q269 I270 Y271 P272 G273 I274 K275 V276 R277 Q278 L279 L282 L283 R284 G285 T286 K287 A288 L289 T290 E291 V292 I293 P294 L295 T296 E297 E298 A299 E300 L301 E302 L303 A304 E305 N306 R307 E308 I309 L310 K311 E312 P313 V314 H315 G316 V317
K323 Q330 K331 Q332 T338 Q343 E344 K350 T351 G352 K353 R358 H361 T362 N363 D364 V365 K366 Q367 L368 T369 E370 A371 V372 Q373 K374 I375 T376 T377 I380 V381 I382 W383 T386 K395 E399 T400 W401 E404 Y405 W414 N418 T419 P420 P421 L422
V423 K424 L425 W426 Y427 Q428

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	224.63Å 69.11Å 105.86Å 90.00° 105.27° 90.00°	Depositor
Resolution (Å)	57.47 – 2.65 108.35 – 2.47	Depositor EDS
% Data completeness (in resolution range)	99.2 (57.47-2.65) 98.7 (108.35-2.47)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.67 (at 2.45Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.220 , 0.265 0.222 , 0.265	Depositor DCC
R_{free} test set	2000 reflections (3.57%)	wwPDB-VP
Wilson B-factor (Å^2)	103.5	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 77.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7805	wwPDB-VP
Average B, all atoms (Å^2)	109.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.28% 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: X2J

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	0.45	0/4532	0.58	2/6170 (0.0%)
2	B	0.49	0/3436	0.64	0/4676
All	All	0.47	0/7968	0.60	2/10846 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	186	ASP	CB-CG-OD1	5.91	123.62	118.30
1	A	349	LEU	CA-CB-CG	5.43	127.79	115.30

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	4416	0	4411	296	0
2	B	3341	0	3346	228	0
3	A	33	15	0	0	0
All	All	7790	15	7757	492	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 32.

The worst 5 of 492 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:54:ASN:HD21	1:A:129:ALA:HB2	1.19	1.02
2:B:254:VAL:HG23	2:B:283:LEU:HD22	1.41	1.02
2:B:85:GLN:HG3	2:B:154:LYS:HB3	1.42	0.97
2:B:282:LEU:HD21	2:B:295:LEU:HA	1.46	0.95
2:B:244:ILE:HB	2:B:310:LEU:HD21	1.50	0.94

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	544/557 (98%)	478 (88%)	59 (11%)	7 (1%)	12	18
2	B	403/428 (94%)	354 (88%)	45 (11%)	4 (1%)	15	23
All	All	947/985 (96%)	832 (88%)	104 (11%)	11 (1%)	13	19

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	85	GLN
1	A	542	ILE
2	B	93	GLY
1	A	539	HIS
1	A	135	ILE

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	477/495 (96%)	444 (93%)	33 (7%)	15 24
2	B	364/390 (93%)	336 (92%)	28 (8%)	13 20
All	All	841/885 (95%)	780 (93%)	61 (7%)	14 21

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	473	THR
2	B	12	LEU
2	B	353	LYS
1	A	475	GLN
1	A	511	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	336	GLN
2	B	242	GLN
1	A	545	ASN
1	A	182	GLN
1	A	373	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)

1 ligand is 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	$\#\|Z\| > 2$
3	X2J	A	601	1	29,35,36	2.21	3 (10%)	34,47,50	1.76	14 (41%)

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
3	X2J	A	601	1	-	0/15/18/19	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	X2J	C0P-N0H	8.48	1.50	1.35
3	A	601	X2J	C0N-N0M	5.46	1.42	1.33
3	A	601	X2J	O07-C01	2.27	1.46	1.42

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	601	X2J	C02-C01-C00	-3.61	115.58	120.53
3	A	601	X2J	C0C-C0V-C0W	3.07	121.73	118.14
3	A	601	X2J	C0O-C0P-N0H	-3.00	118.89	121.28
3	A	601	X2J	C03-C02-C01	2.97	123.37	119.73
3	A	601	X2J	C0Z-C0C-C0V	-2.93	116.31	120.98

There are no chirality outliers.

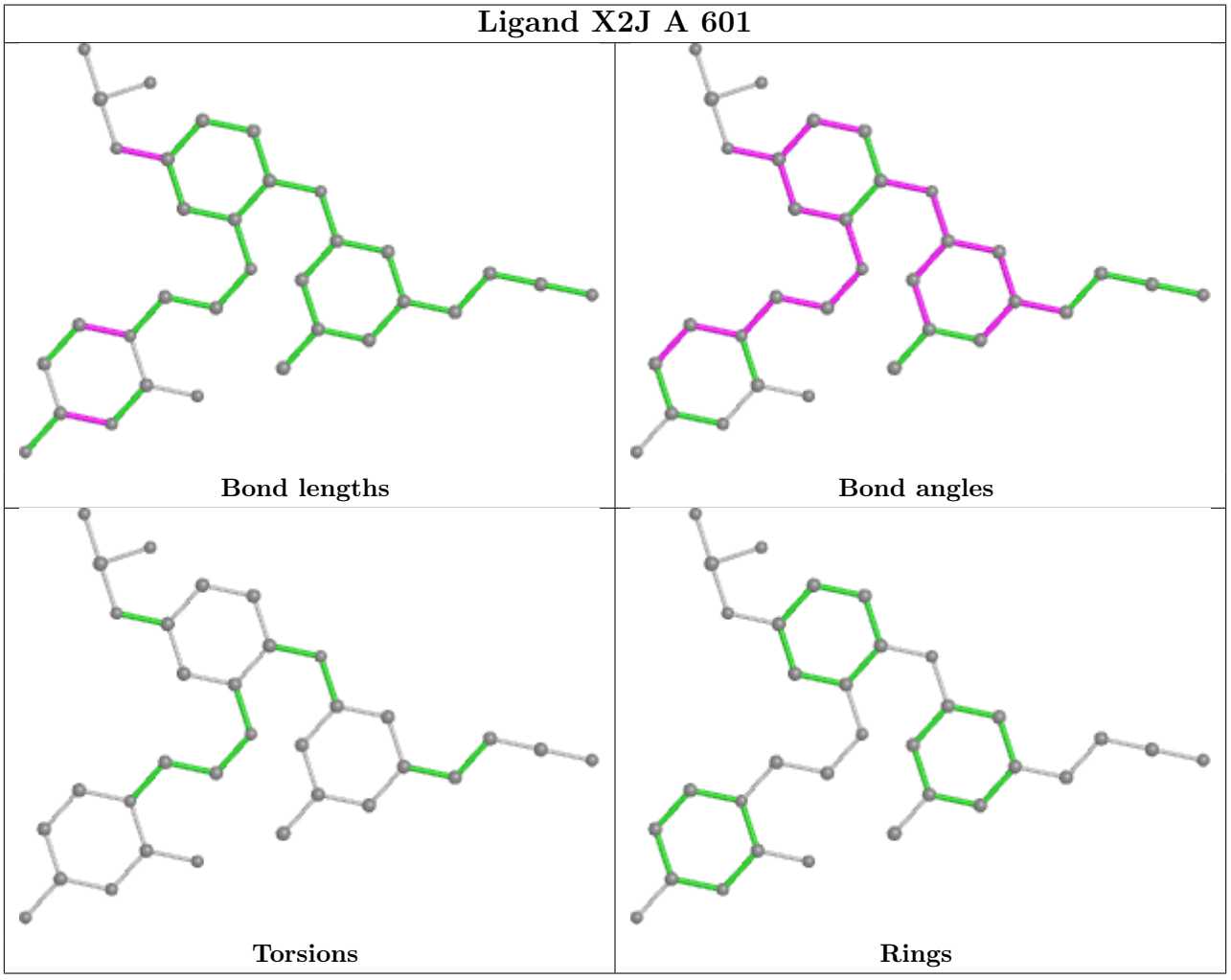
There are no torsion outliers.

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 X2J A 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	548/557 (98%)	0.21	18 (3%) 46 43	71, 108, 158, 192	0
2	B	409/428 (95%)	0.27	11 (2%) 54 50	71, 98, 159, 174	0
All	All	957/985 (97%)	0.24	29 (3%) 50 47	71, 105, 159, 192	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	26	LEU	6.1
1	A	551	LEU	4.7
1	A	133	PRO	4.2
2	B	218	ASP	3.9
2	B	252	TRP	3.4

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

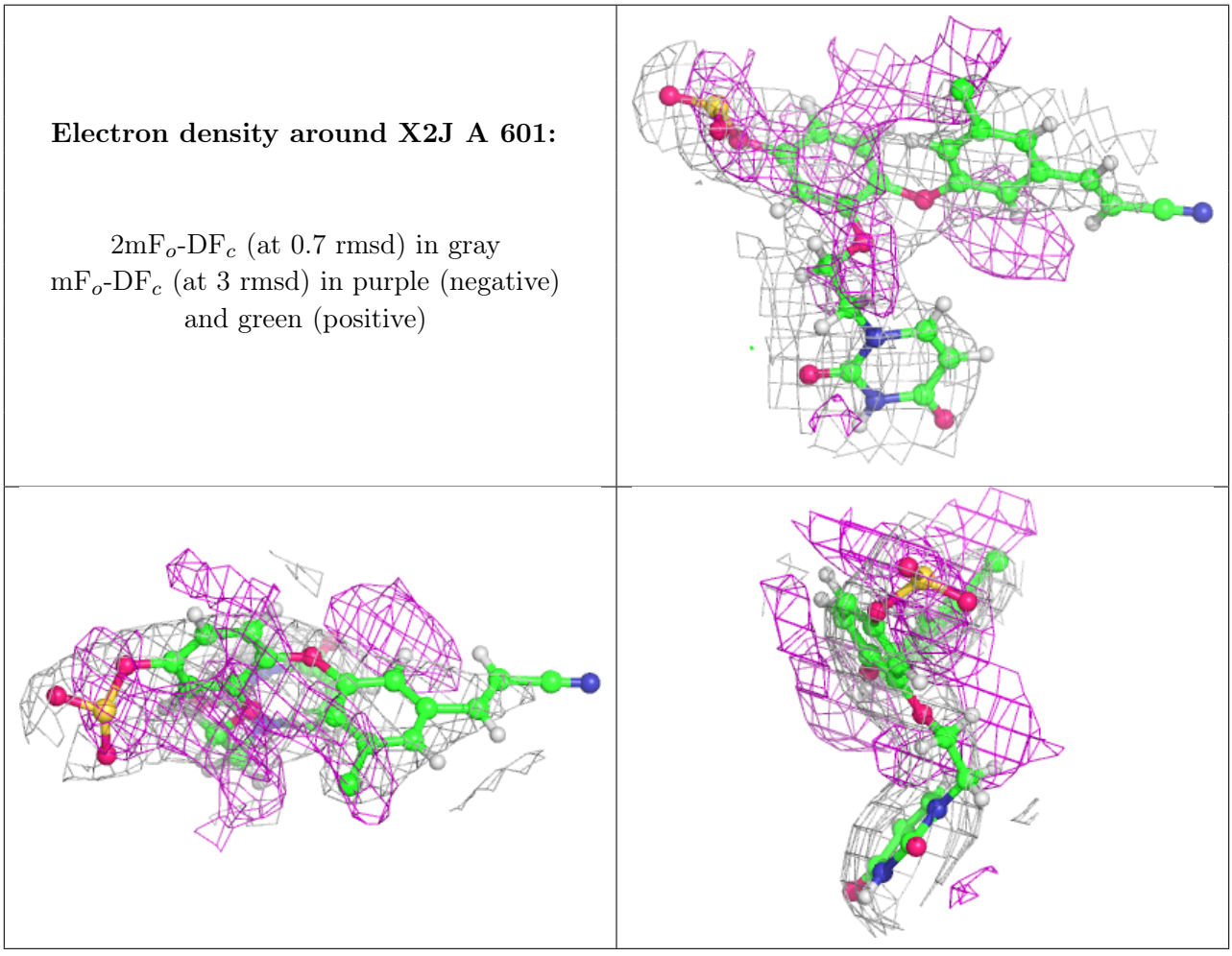
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
3	X2J	A	601	33/34	0.92	0.26	83,99,116,119	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 X2J A 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.