wwPDB X-ray Structure Validation Summary Report (i)

Nov 22, 2023 – 06:04 PM EST

PDB ID	:	8STS
Title	:	Crystal Structure of HIV-1 Reverse Transcriptase (Y181C, V106A) varient in Complex with 5-(2-(2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy)-4-fluorophenoxy)-7-fluoro-2-naphthonitrile (JLJ636), a non-nucleoside inhibitor
Authors	:	Hollander, K.; Jorgensen, W.L.; Anderson, K.S.
Deposited on	:	2023-05-11
Resolution	:	3.02 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:

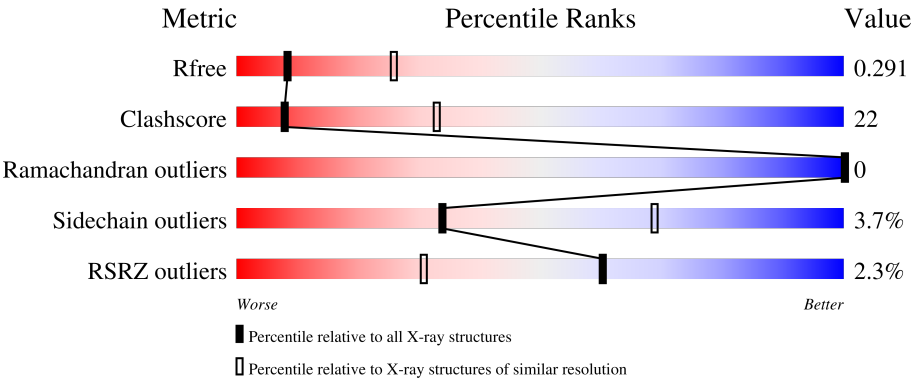
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.02 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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	558	3% 54% 38%
1	C	558	3% 58% 38%
2	B	428	2% 55% 37% 6%
2	D	428	% 58% 36% 5%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 14639 atoms, of which 30 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	C	546	Total	C	N	O	S	0	0	0
			4141	2658	688	788	7			
1	A	533	Total	C	N	O	S	0	0	0
			3972	2550	658	758	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	expression tag	UNP P03366
C	0	VAL	-	expression tag	UNP P03366
C	106	ALA	VAL	engineered mutation	UNP P03366
C	172	ALA	LYS	conflict	UNP P03366
C	173	ALA	LYS	conflict	UNP P03366
C	181	CYS	TYR	engineered mutation	UNP P03366
C	280	SER	CYS	engineered mutation	UNP P03366
A	-1	MET	-	expression tag	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	106	ALA	VAL	engineered mutation	UNP P03366
A	172	ALA	LYS	engineered mutation	UNP P03366
A	173	ALA	LYS	engineered mutation	UNP P03366
A	181	CYS	TYR	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366

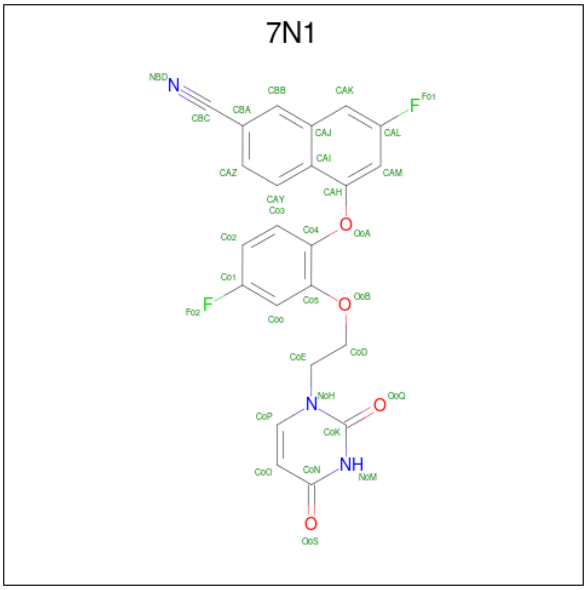
- Molecule 2 is a protein called p51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	402	Total	C	N	O	S	23	0	0
			3189	2071	522	591	5			
2	D	407	Total	C	N	O	S	0	0	0
			3229	2092	534	596	7			

There are 2 discrepancies between the modelled and reference sequences:

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

- Molecule 3 is 5-{2-[2-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)ethoxy]-4-fluorophenoxy}-7-fluoronaphthalene-2-carbonitrile (three-letter code: 7N1) (formula: $C_{23}H_{15}F_2N_3O_4$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	C	1	Total	C	F	H	N	O	0	0
			47	23	2	15	3	4		
3	A	1	Total	C	F	H	N	O	0	0
			47	23	2	15	3	4		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Mg	0	0
			1	1		
4	A	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	4	Total	O	0	0
			4	4		

Continued on next page...

Continued from previous page...

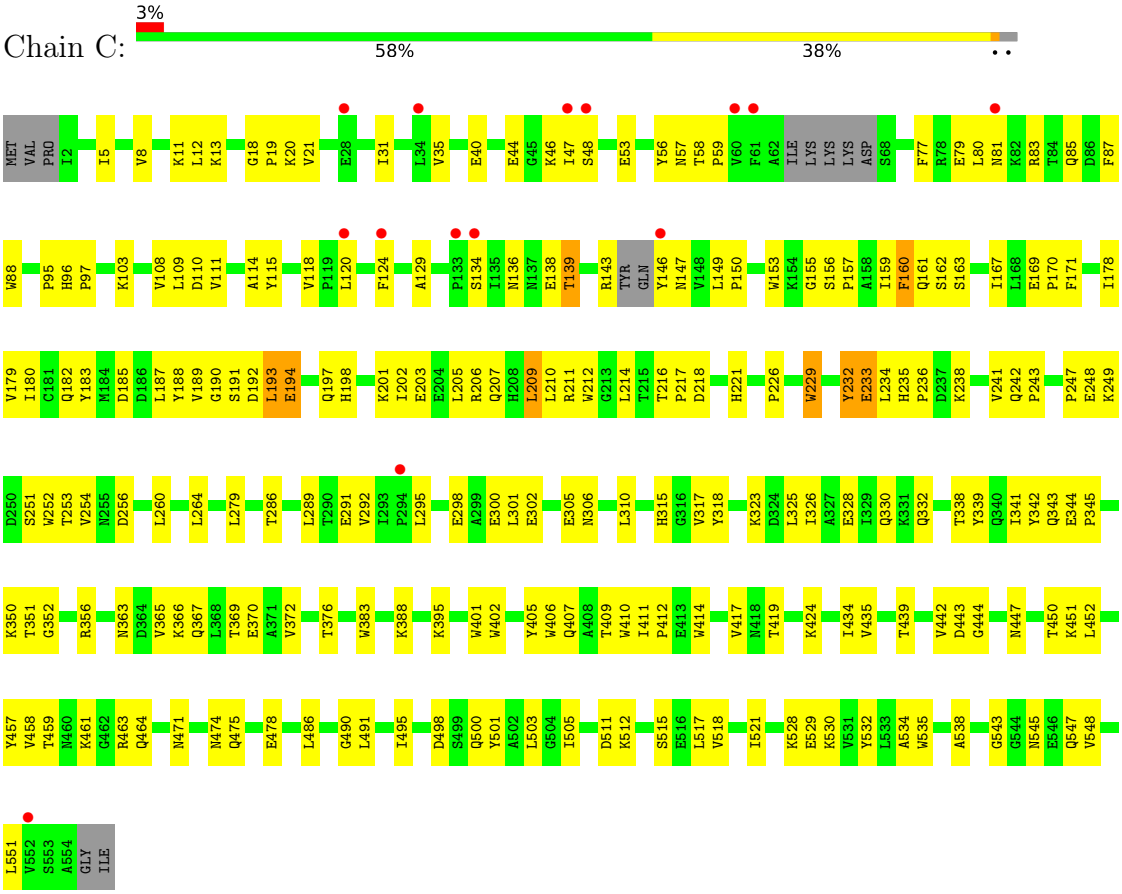
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total O 2 2	0	0
5	D	6	Total O 6 6	0	0

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.

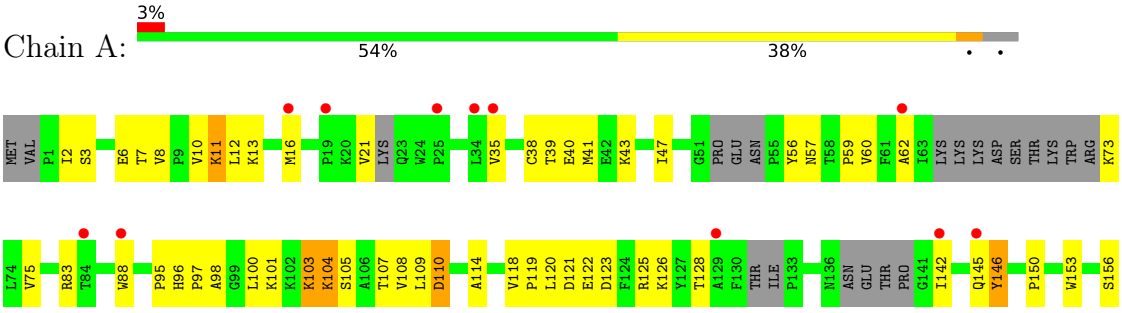
• Molecule 1: Reverse transcriptase/ribonuclease H

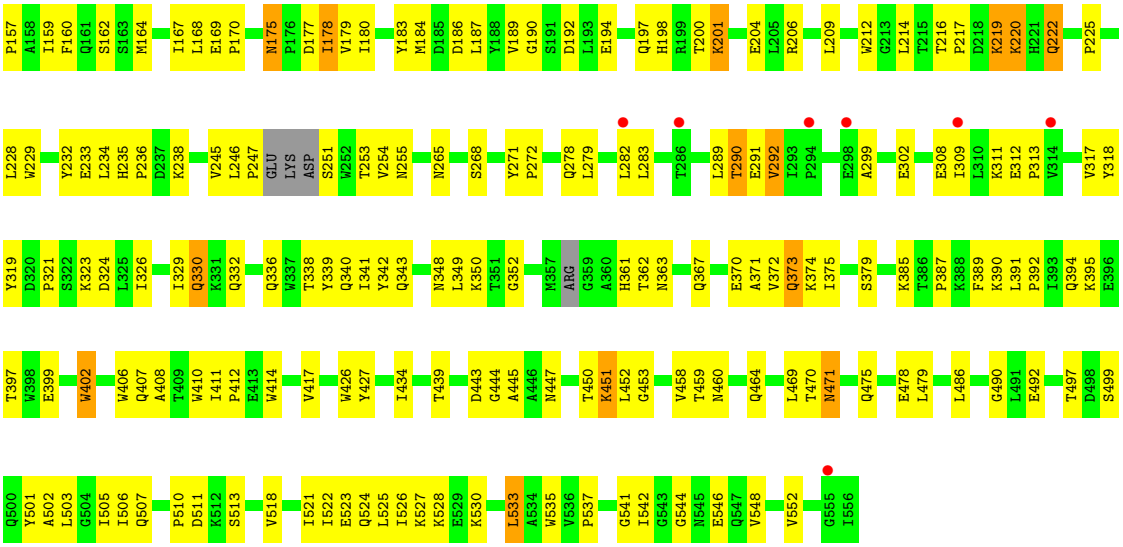
Chain C:



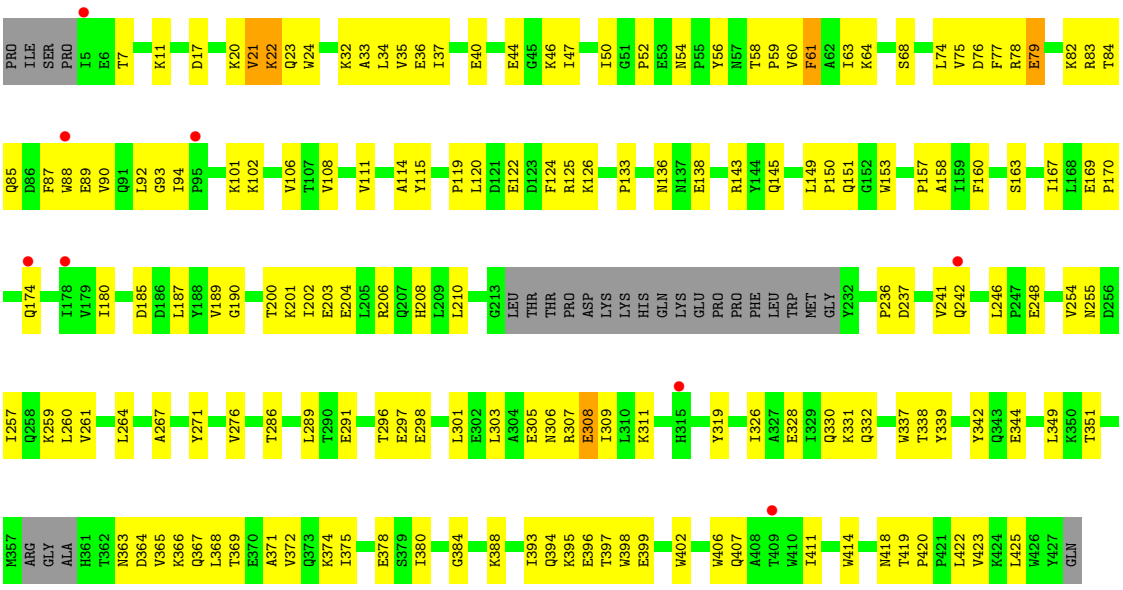
• Molecule 1: Reverse transcriptase/ribonuclease H

Chain A:

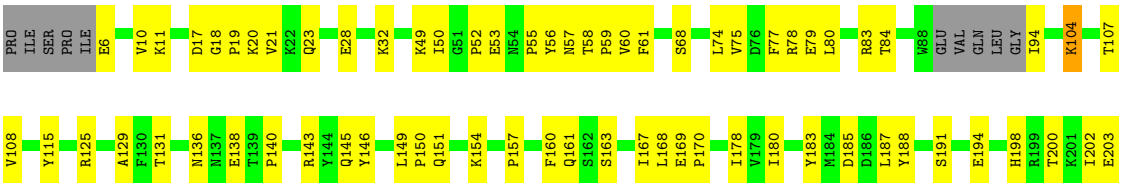


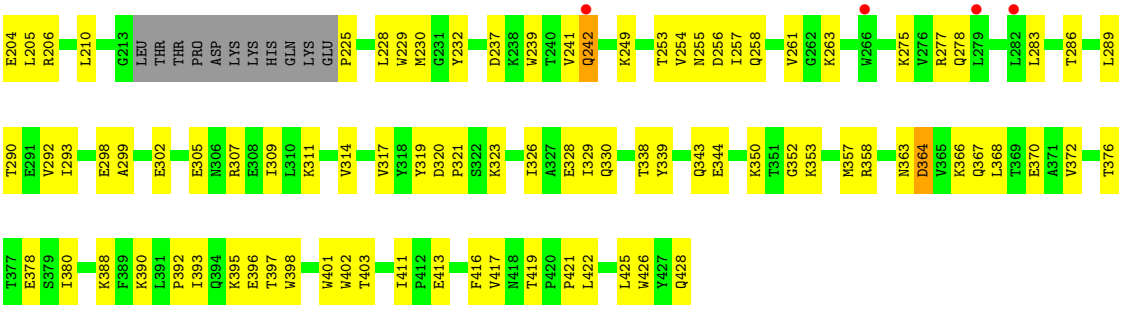


• Molecule 2: p51 RT



• Molecule 2: p51 RT




E204 L205 R206 L210 G213 LEU THR THR PRO ASP LYS LYS HIS GLN LYS GLU P225 L228 W229 M230 G231 Y232 D237 K238 W239 T240 V241 Q242 K249 T253 V254 N255 D256 I257 Q258 V261 G262 K263 W266 K275 V276 R277 Q278 L279 L282 L283 T286 L289
T290 E291 V292 I293 E298 A299 E302 E305 N306 R307 E308 I309 L310 K311 V314 V317 Y318 Y319 D320 P321 S322 K323 I326 A327 E328 I329 Q330 T338 Y339 Q343 E344 K350 T351 G352 K353 M357 R358 N363 D364 V365 K366 Q367 L368 T369 E370 A371 V372 T376
T377 E378 S379 I380 K388 F389 K390 L391 P392 I393 Q394 K395 E396 T397 W398 W401 W402 T403 I411 P412 E413 F416 V417 N418 T419 P420 P421 L422 L425 W426 Y427 Q428

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.86Å 128.26Å 133.04Å 90.00° 105.32° 90.00°	Depositor
Resolution (Å)	38.11 – 3.02 38.12 – 3.02	Depositor EDS
% Data completeness (in resolution range)	99.2 (38.11-3.02) 99.2 (38.12-3.02)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.240 , 0.285 0.247 , 0.291	Depositor DCC
R_{free} test set	2803 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	102.3	Xtriage
Anisotropy	0.371	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 83.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14639	wwPDB-VP
Average B, all atoms (Å^2)	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.27% 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: MG, 7N1

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	0.26	0/4067	0.41	0/5563
1	C	0.23	0/4250	0.41	0/5826
2	B	0.23	0/3278	0.41	0/4477
2	D	0.23	0/3323	0.41	0/4540
All	All	0.24	0/14918	0.41	0/20406

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	3972	0	3686	183	0
1	C	4141	0	3858	182	0
2	B	3189	0	3079	135	0
2	D	3229	0	3089	134	0
3	A	32	15	0	2	0
3	C	32	15	0	2	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	A	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	4	0	0	1	0
5	D	6	0	0	0	0
All	All	14609	30	13712	608	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 22.

The worst 5 of 608 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:254:VAL:HG21	1:A:290:THR:HB	1.53	0.91
1:C:232:TYR:HB3	1:C:241:VAL:HA	1.54	0.90
2:B:59:PRO:HG2	2:B:76:ASP:HB3	1.63	0.81
1:C:253:THR:HG22	1:C:292:VAL:HG22	1.62	0.80
2:D:84:THR:HB	2:D:154:LYS:HD3	1.65	0.78

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	517/558 (93%)	511 (99%)	6 (1%)	0	100	100
1	C	540/558 (97%)	531 (98%)	9 (2%)	0	100	100
2	B	396/428 (92%)	388 (98%)	8 (2%)	0	100	100
2	D	401/428 (94%)	395 (98%)	6 (2%)	0	100	100
All	All	1854/1972 (94%)	1825 (98%)	29 (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	387/495 (78%)	359 (93%)	28 (7%)	14 44
1	C	412/495 (83%)	400 (97%)	12 (3%)	42 75
2	B	334/390 (86%)	323 (97%)	11 (3%)	38 72
2	D	335/390 (86%)	331 (99%)	4 (1%)	71 89
All	All	1468/1770 (83%)	1413 (96%)	55 (4%)	34 69

5 of 55 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	220	LYS
1	A	373	GLN
2	D	419	THR
2	B	364	ASP
1	A	222	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	330	GLN
1	A	373	GLN
2	B	348	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
3	7N1	A	601	-	35,35,35	1.60	7 (20%)	48,49,49	1.85	8 (16%)
3	7N1	C	601	-	35,35,35	1.61	7 (20%)	48,49,49	1.84	9 (18%)

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	7N1	A	601	-	-	0/12/12/12	0/4/4/4
3	7N1	C	601	-	-	0/12/12/12	0/4/4/4

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	601	7N1	CBA-CBC	3.85	1.53	1.44
3	A	601	7N1	CBA-CBC	3.83	1.53	1.44
3	A	601	7N1	C0O-C0N	-3.50	1.35	1.43
3	C	601	7N1	C0O-C0N	-3.49	1.35	1.43
3	C	601	7N1	C0N-N0M	-3.23	1.32	1.38

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	601	7N1	N0M-C0K-N0H	5.60	119.98	114.86
3	C	601	7N1	N0M-C0K-N0H	5.53	119.92	114.86

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	601	7N1	C0N-N0M-C0K	-4.92	120.09	126.58
3	C	601	7N1	C0N-N0M-C0K	-4.90	120.11	126.58
3	A	601	7N1	CAK-CAL-CAM	-4.86	120.25	124.09

There are no chirality outliers.

There are no torsion outliers.

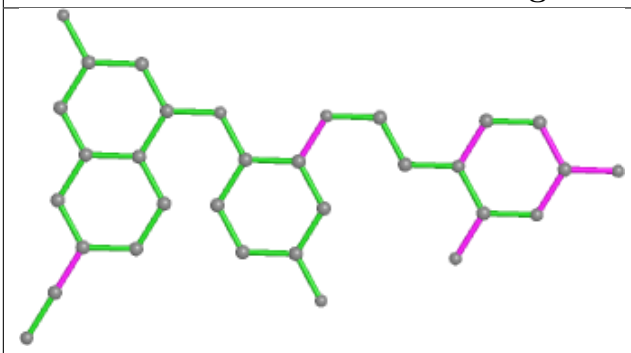
There are no ring outliers.

2 monomers are involved in 4 short contacts:

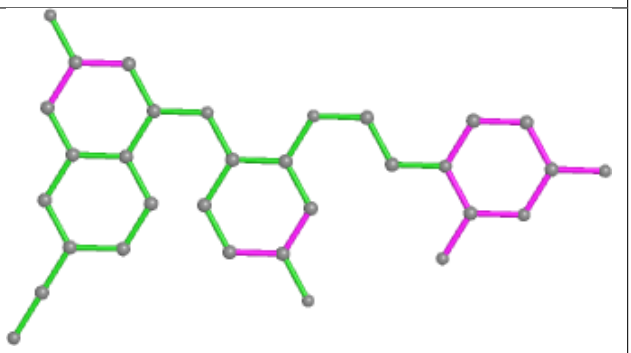
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	601	7N1	2	0
3	C	601	7N1	2	0

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.

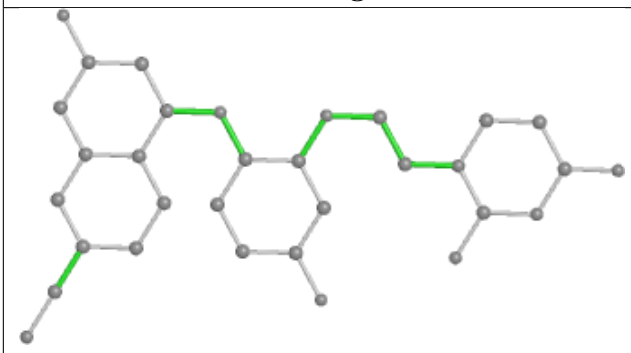
Ligand 7N1 A 601



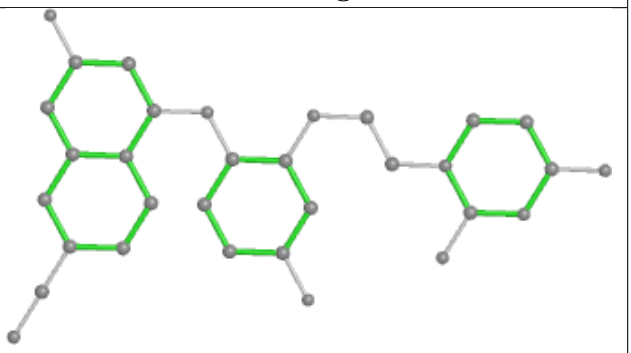
Bond lengths



Bond angles

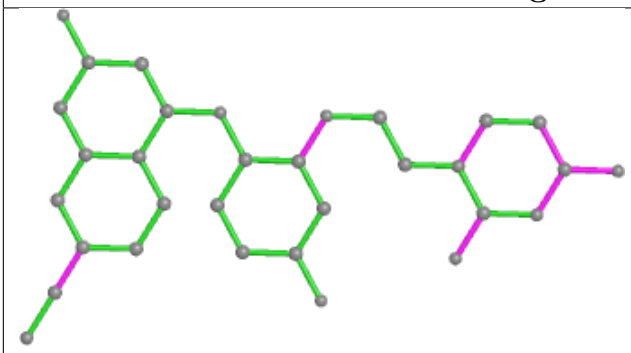


Torsions

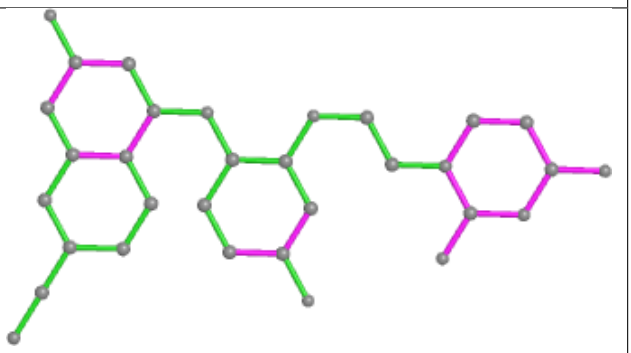


Rings

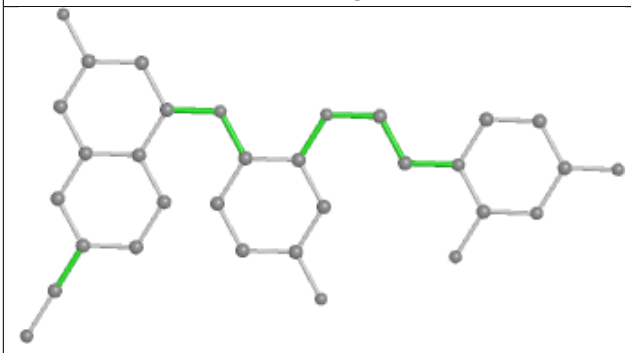
Ligand 7N1 C 601



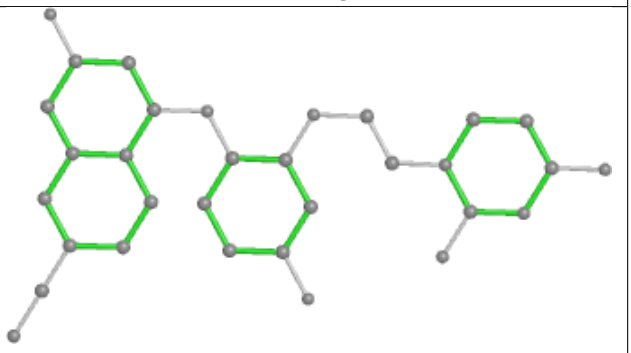
Bond lengths



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	533/558 (95%)	0.01	18 (3%)	45	19	55, 123, 191, 216	0
1	C	546/558 (97%)	-0.09	14 (2%)	56	27	58, 120, 182, 239	0
2	B	402/428 (93%)	-0.13	8 (1%)	65	36	60, 108, 158, 211	7 (1%)
2	D	407/428 (95%)	-0.15	4 (0%)	82	58	53, 109, 156, 194	0
All	All	1888/1972 (95%)	-0.08	44 (2%)	60	31	53, 115, 182, 239	7 (0%)

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	286	THR	4.4
1	A	34	LEU	4.2
1	C	146	TYR	4.1
1	A	62	ALA	3.8
2	B	242	GLN	3.6

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

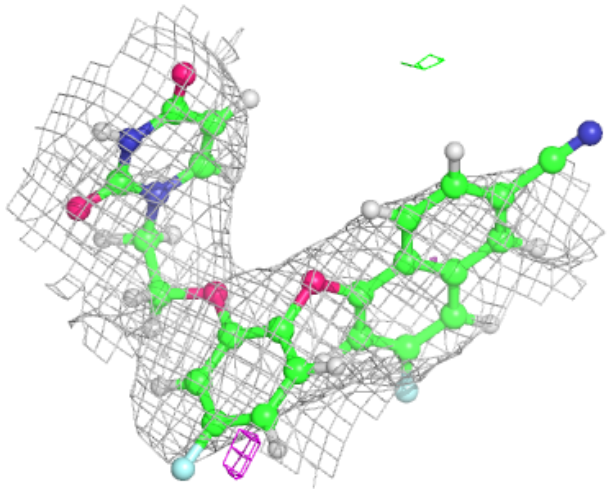
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

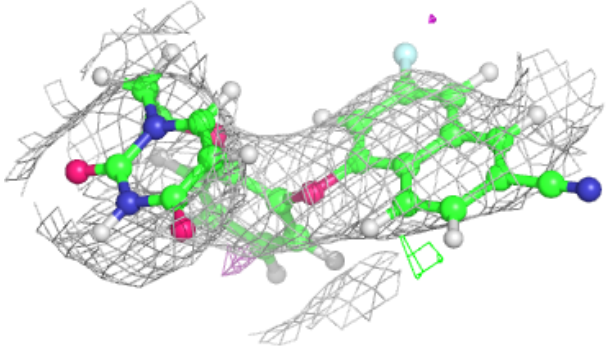
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	MG	A	602	1/1	0.76	0.36	86,86,86,86	0
3	7N1	A	601	32/32	0.91	0.24	111,130,156,160	0
3	7N1	C	601	32/32	0.91	0.30	88,107,132,134	0
4	MG	C	602	1/1	0.97	0.14	87,87,87,87	0

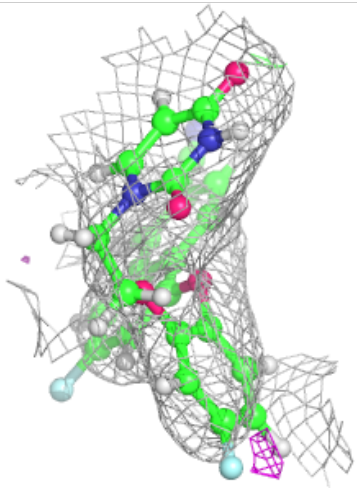
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around 7N1 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)

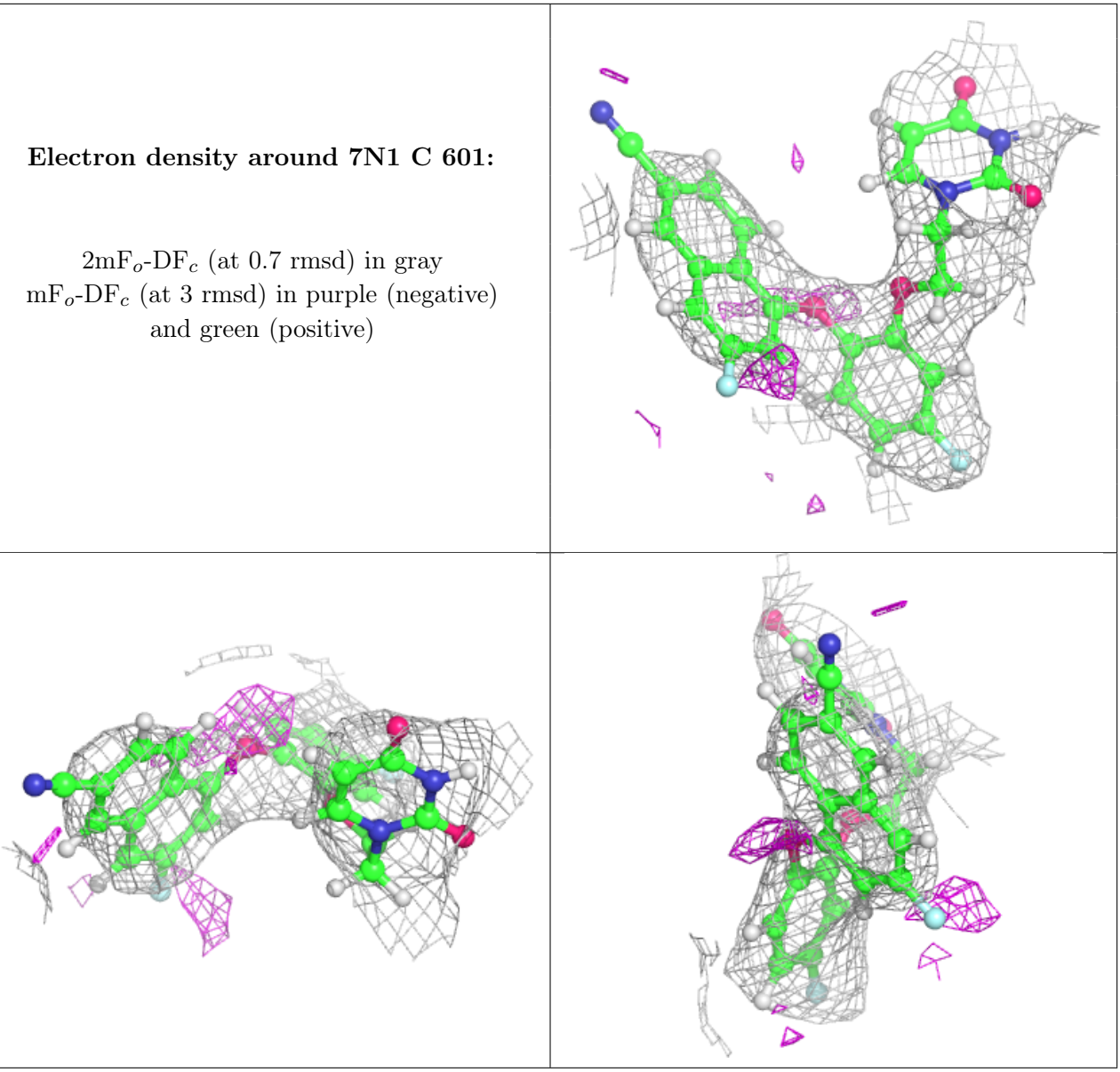






Electron density around 7N1 C 601:

$2mF_o$-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.