wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2021 – 04:31 PM EDT

PDB ID	:	2HU0
Title	:	N1 neuraminidase in complex with oseltamivir 1
Authors	:	Russell, R.J.; Haire, L.F.; Stevens, D.J.; Collins, P.J.; Lin, Y.P.; Blackburn, G.M.; Hay, A.J.; Gamblin, S.J.; Skehel, J.J.
Deposited on	:	2006-07-26
Resolution	:	2.95 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 following versions of software and data (see references (i)) were used in the production of this report:

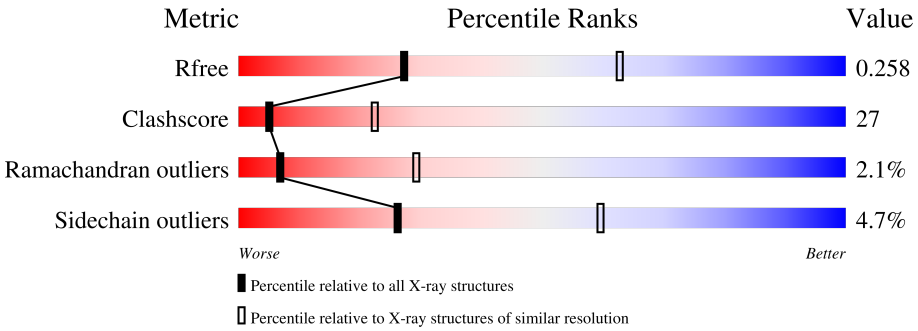
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)

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%

Mol	Chain	Length	Quality of chain
1	A	387	59% 33% 7% •
1	B	387	56% 37% 6% ••
1	C	387	56% 35% 8% •
1	D	387	53% 39% 7% ••
1	E	387	57% 37% 5% ••
1	F	387	58% 34% 7% ••
1	G	387	57% 36% 6% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	387	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 23716 atoms, of which 0 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.

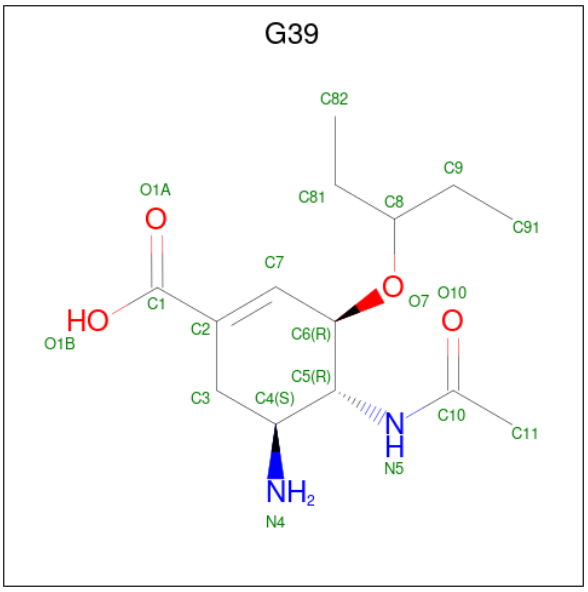
- Molecule 1 is a protein called Neuraminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	B	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	C	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	D	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	E	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	F	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	G	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			
1	H	385	Total	C	N	O	S	0	0	0
			2962	1858	510	573	21			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
B	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
C	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
D	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
E	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
F	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
G	169A	TYR	HIS	engineered mutation	UNP Q6DPL2
H	169A	TYR	HIS	engineered mutation	UNP Q6DPL2

- Molecule 2 is (3R,4R,5S)-4-(acetylamino)-5-amino-3-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid (three-letter code: G39) (formula: $C_{14}H_{24}N_2O_4$).



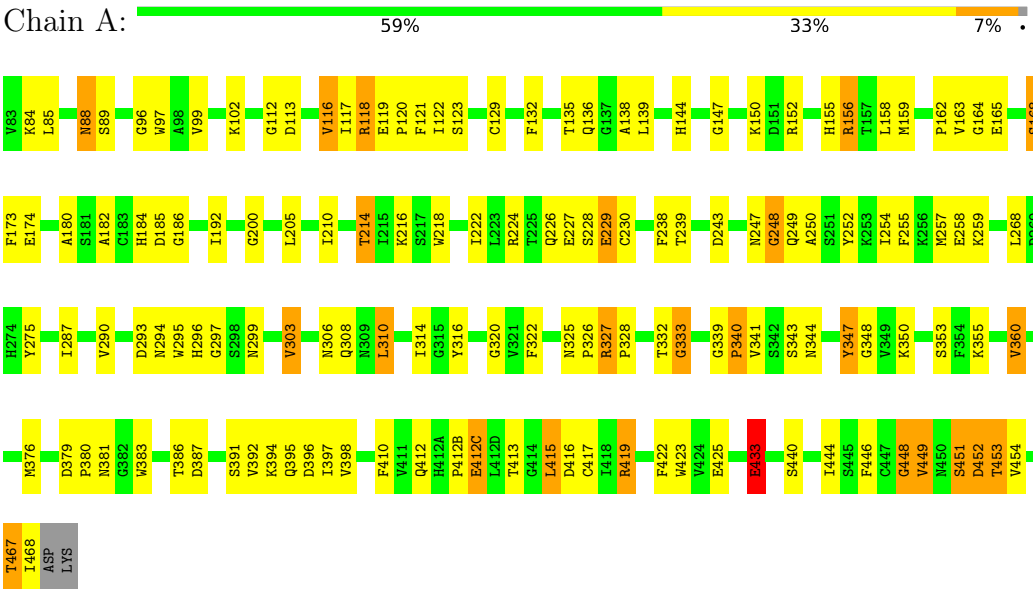
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	N	O	0	0
			20	14	2	4		

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. 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. 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.

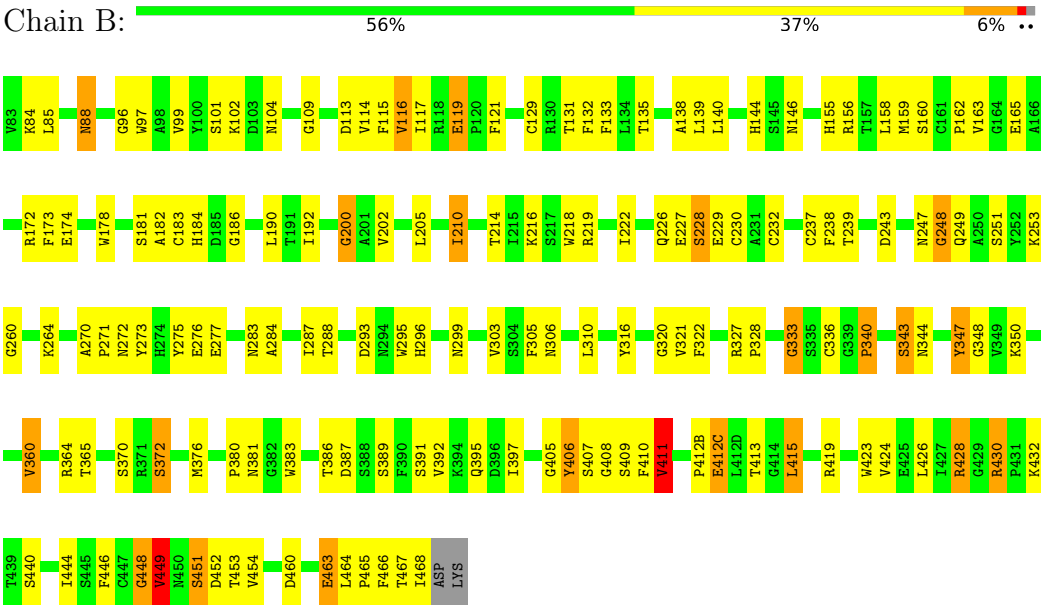
- Molecule 1: Neuraminidase

Chain A:



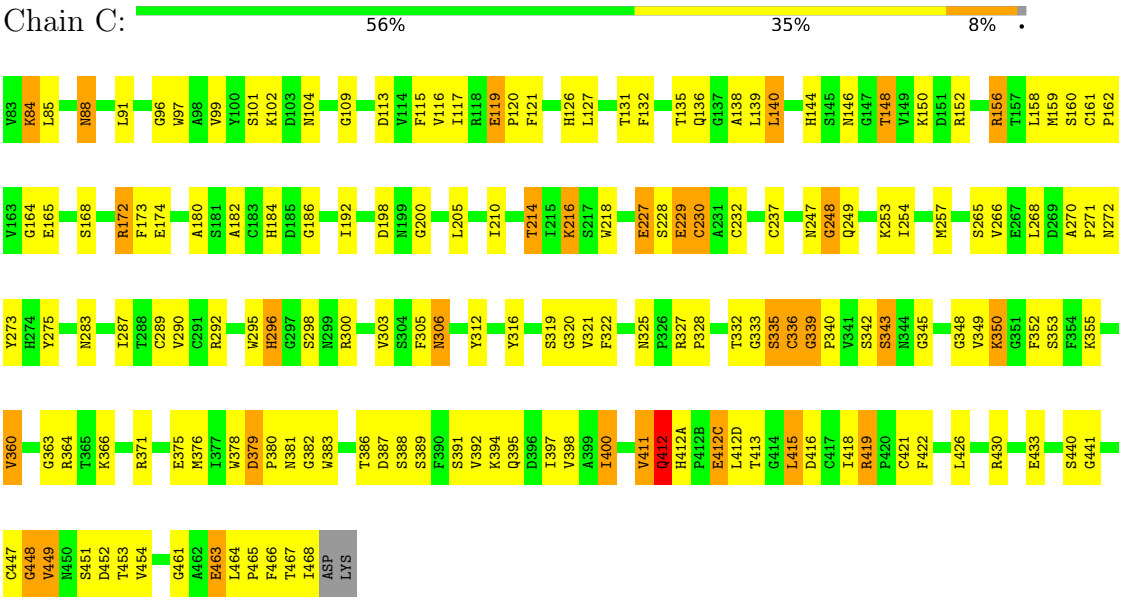
- Molecule 1: Neuraminidase

Chain B:



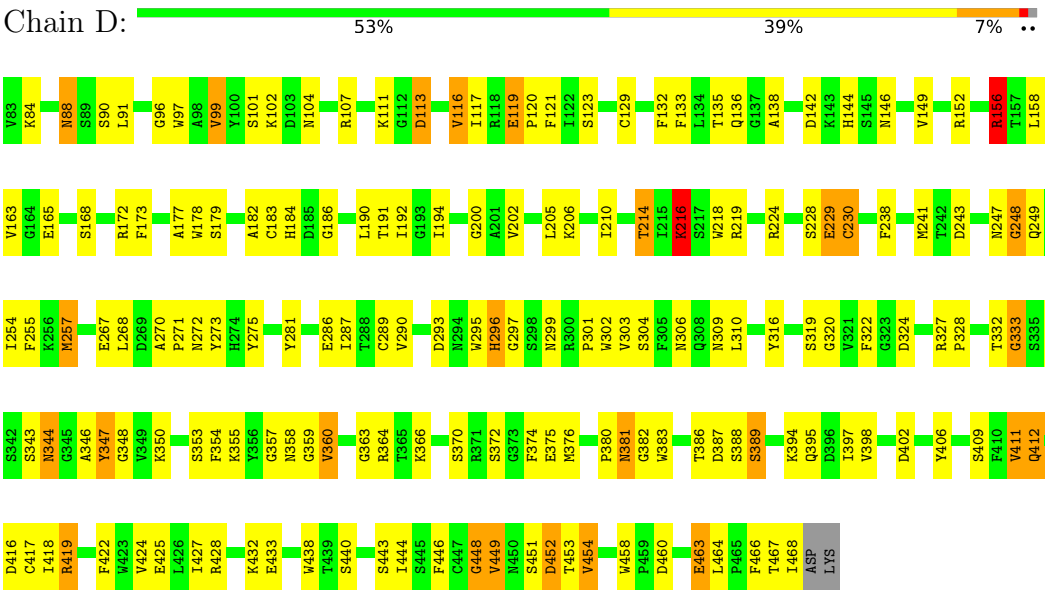
• Molecule 1: Neuraminidase

Chain C:



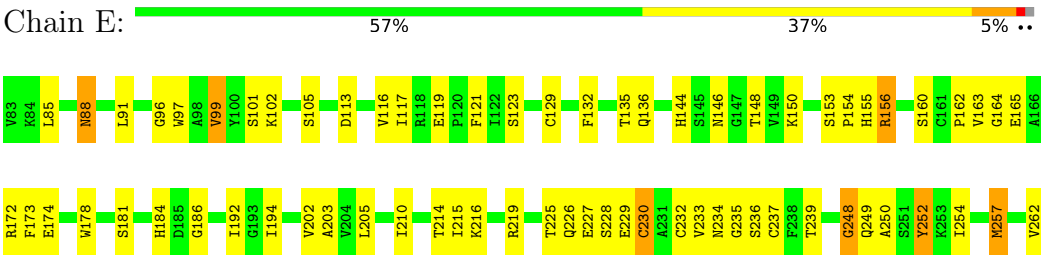
• Molecule 1: Neuraminidase

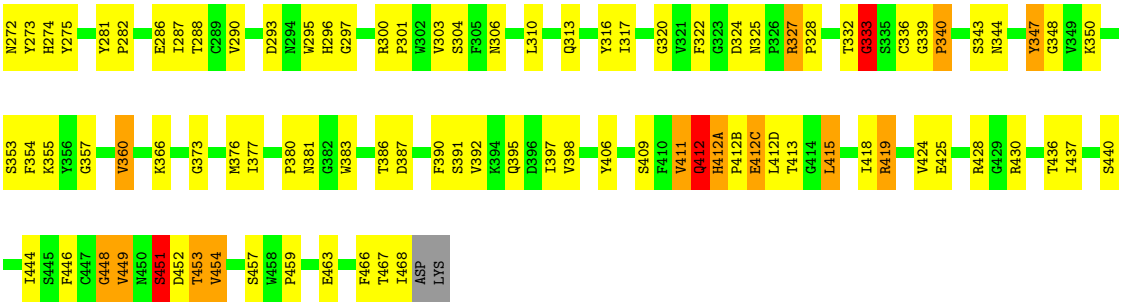
Chain D:



• Molecule 1: Neuraminidase

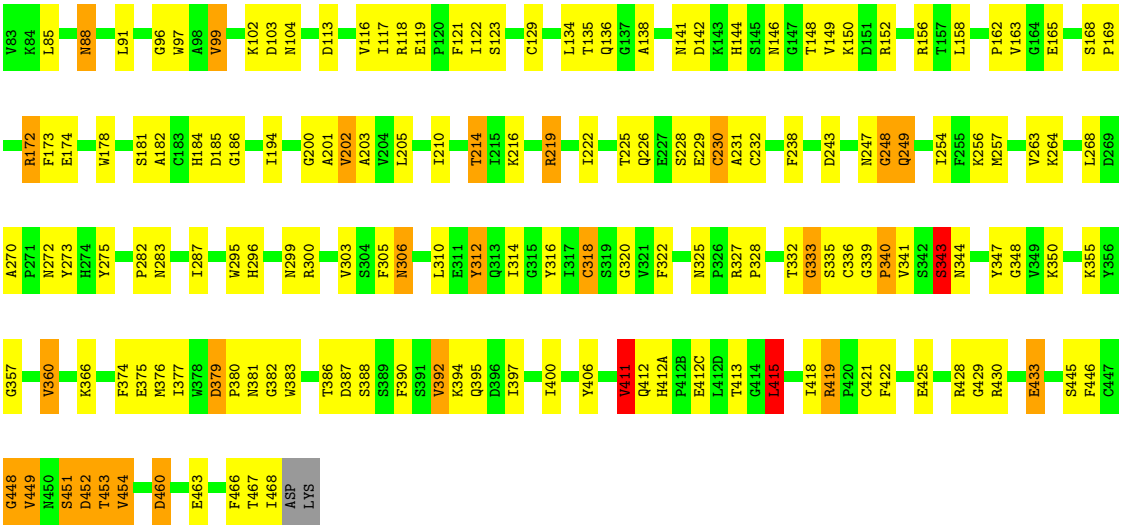
Chain E:





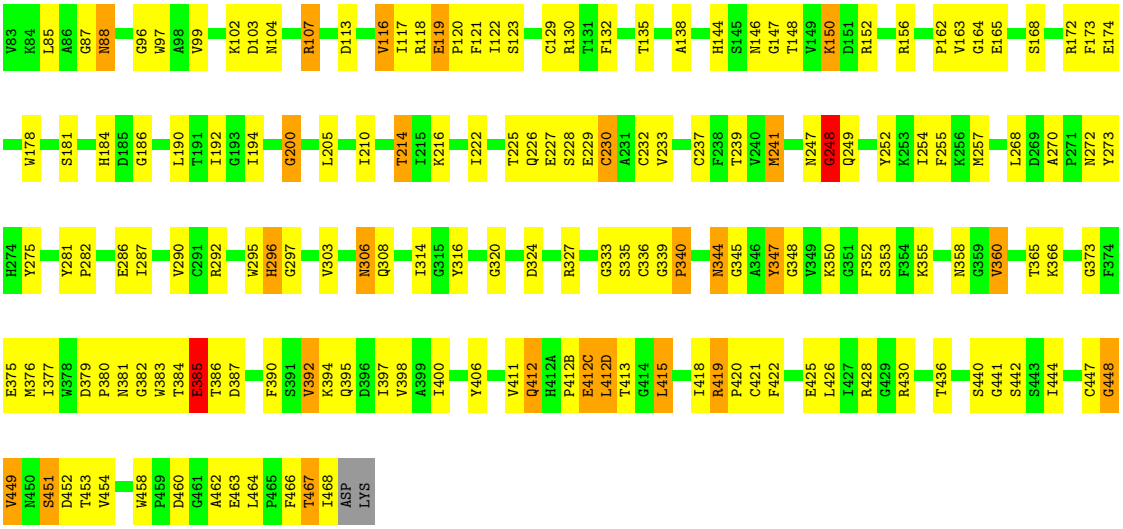
• Molecule 1: Neuraminidase

Chain F:



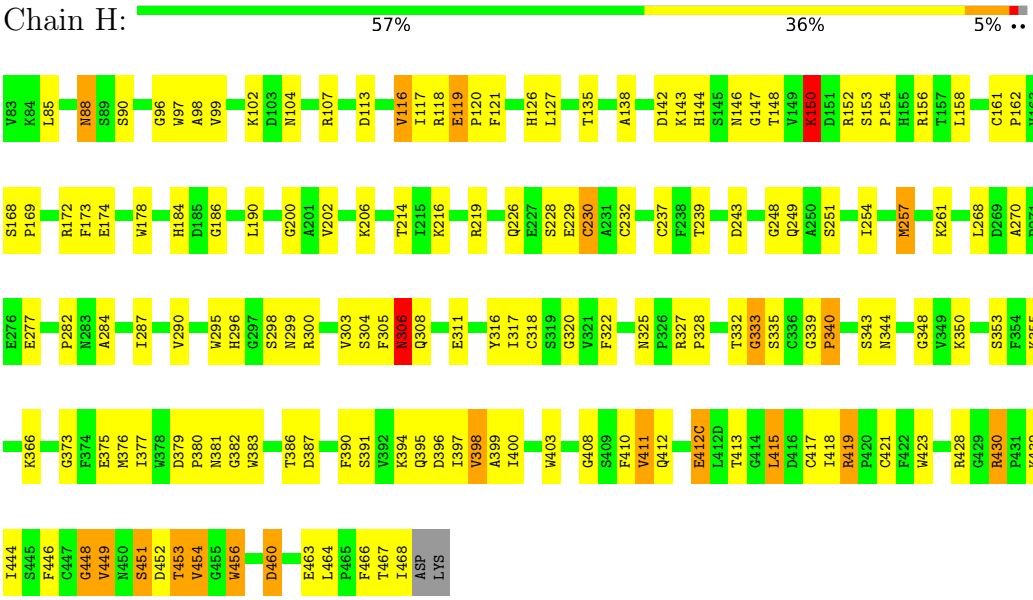
• Molecule 1: Neuraminidase

Chain G:



• Molecule 1: Neuraminidase

Chain H:


57%
36%
5%
V83 K84 L85 N88 S89 S90 G96 W97 A98 V99 K102 D103 N104 R107 D113 V116 I117 R118 E119 P120 F121 H126 L127 T135 A138 D142 K143 H144 S145 N146 G147 T148 V149 K150 D151 R152 S153 P154 H155 R156 T157 L158 C161 P162 V163 G164 E165 A166 P167
S168 P169 R172 F173 E174 W178 H184 D185 G186 L190 G200 A201 V202 K206 T214 I215 K216 R219 Q226 E227 S228 E229 C230 A231 C232 C237 F238 T239 D243 G248 Q249 A250 S251 I254 M257 K261 L268 D269 A270 P271 N272 Y273 H274 Y275
E276 E277 P282 N283 A284 I287 V290 W295 H296 G297 S298 N299 R300 V303 S304 F305 N306 Q308 E311 Y316 I317 C318 S319 G320 V321 F322 N325 P326 R327 P328 T332 G333 S335 C336 G339 P340 S343 N344 G348 V349 K350 S353 F354 K355 Y356 G357 V360
K366 G373 F374 E375 M376 I377 W378 D379 P380 N381 G382 W383 T386 D387 F390 S391 V392 K394 Q395 D396 I397 V398 A399 I400 W403 G408 S409 F410 V411 Q412 E412C L412D T413 G414 L415 D416 C417 I418 R419 P420 C421 F422 W423 R428 G429 R430 P431 K432 E433 S435 T436
I444 S445 F446 C447 G448 V449 N450 S451 D452 T453 V454 G455 W456 D460 E463 L464 P465 F466 T467 I468 ASP LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	198.08Å 200.58Å 210.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	142.86 – 2.95 19.70 – 2.86	Depositor EDS
% Data completeness (in resolution range)	79.1 (142.86-2.95) 77.7 (19.70-2.86)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.88Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.218 , 0.295 0.255 , 0.258	Depositor DCC
R_{free} test set	3705 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	32.1	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , -18.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.023 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	23716	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.80% 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: G39

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	1.26	7/3045 (0.2%)	1.18	13/4141 (0.3%)
1	B	1.35	13/3045 (0.4%)	1.15	11/4141 (0.3%)
1	C	1.38	19/3045 (0.6%)	1.27	19/4141 (0.5%)
1	D	1.36	16/3045 (0.5%)	1.20	18/4141 (0.4%)
1	E	1.19	6/3045 (0.2%)	1.11	11/4141 (0.3%)
1	F	1.22	9/3045 (0.3%)	1.09	8/4141 (0.2%)
1	G	1.30	12/3045 (0.4%)	1.12	6/4141 (0.1%)
1	H	1.25	8/3045 (0.3%)	1.09	7/4141 (0.2%)
All	All	1.29	90/24360 (0.4%)	1.15	93/33128 (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	2
1	B	0	1
1	C	0	4
1	D	0	2
1	E	0	3
1	F	0	3
1	G	0	3
1	H	0	3
All	All	0	21

The worst 5 of 90 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	411	VAL	CB-CG2	15.85	1.86	1.52
1	H	150	LYS	CE-NZ	13.79	1.83	1.49

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	411	VAL	CB-CG2	12.77	1.79	1.52
1	D	411	VAL	CB-CG2	12.26	1.78	1.52
1	C	336	CYS	C-N	11.22	1.53	1.33

The worst 5 of 93 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	412	GLN	O-C-N	-13.38	101.30	122.70
1	D	411	VAL	CA-CB-CG2	-12.09	92.77	110.90
1	H	306	ASN	O-C-N	-11.99	103.52	122.70
1	A	172	ARG	NE-CZ-NH1	-11.42	114.59	120.30
1	B	411	VAL	CA-CB-CG2	-11.21	94.08	110.90

There are no chirality outliers.

5 of 21 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	333	GLY	Mainchain
1	A	433	GLU	Mainchain
1	B	411	VAL	Mainchain
1	C	306	ASN	Mainchain
1	C	336	CYS	Mainchain

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	2962	0	2783	161	0
1	B	2962	0	2782	171	0
1	C	2962	0	2782	177	5
1	D	2962	0	2783	183	8
1	E	2962	0	2780	194	0
1	F	2962	0	2783	173	0
1	G	2962	0	2783	171	3
1	H	2962	0	2783	166	0
2	B	20	0	23	4	0
All	All	23716	0	22282	1245	8

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 27.

The worst 5 of 1245 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:E:411:VAL:CG2	1:E:411:VAL:CB	1.79	1.58
1:H:150:LYS:CE	1:H:150:LYS:CD	1.78	1.56
1:C:411:VAL:CG2	1:C:411:VAL:CB	1.76	1.56
1:D:411:VAL:CB	1:D:411:VAL:CG2	1.78	1.56
1:C:84:LYS:CD	1:C:84:LYS:CG	1.83	1.53

The worst 5 of 8 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:C:296:HIS:CD2	1:D:272:ASN:OD1[8_456]	1.92	0.28
1:D:381:ASN:OD1	1:G:384:THR:O[3_565]	1.93	0.27
1:D:416:ASP:OD2	1:G:344:ASN:N[3_565]	2.05	0.15
1:C:332:THR:OG1	1:D:332:THR:OG1[8_456]	2.12	0.08
1:C:272:ASN:OD1	1:D:296:HIS:CD2[8_456]	2.15	0.05

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	383/387 (99%)	339 (88%)	34 (9%)	10 (3%)	5 24
1	B	383/387 (99%)	345 (90%)	30 (8%)	8 (2%)	7 29
1	C	383/387 (99%)	339 (88%)	40 (10%)	4 (1%)	15 48
1	D	383/387 (99%)	341 (89%)	36 (9%)	6 (2%)	9 36
1	E	383/387 (99%)	344 (90%)	31 (8%)	8 (2%)	7 29
1	F	383/387 (99%)	339 (88%)	32 (8%)	12 (3%)	4 19

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	G	383/387 (99%)	345 (90%)	29 (8%)	9 (2%)	6 27
1	H	383/387 (99%)	339 (88%)	36 (9%)	8 (2%)	7 29
All	All	3064/3096 (99%)	2731 (89%)	268 (9%)	65 (2%)	7 29

5 of 65 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	412(A)	HIS
1	C	448	GLY
1	D	248	GLY
1	D	448	GLY
1	E	412(A)	HIS

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	329/331 (99%)	315 (96%)	14 (4%)	29 62
1	B	329/331 (99%)	315 (96%)	14 (4%)	29 62
1	C	329/331 (99%)	314 (95%)	15 (5%)	27 60
1	D	329/331 (99%)	311 (94%)	18 (6%)	21 53
1	E	329/331 (99%)	312 (95%)	17 (5%)	23 56
1	F	329/331 (99%)	312 (95%)	17 (5%)	23 56
1	G	329/331 (99%)	316 (96%)	13 (4%)	31 64
1	H	329/331 (99%)	312 (95%)	17 (5%)	23 56
All	All	2632/2648 (99%)	2507 (95%)	125 (5%)	26 59

5 of 125 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	419	ARG
1	H	116	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	391	SER
1	H	99	VAL
1	H	412(C)	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 86 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	155	HIS
1	G	226	GLN
1	F	226	GLN
1	G	88	ASN
1	H	88	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)

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	$\#\|Z\| > 2$
2	G39	B	800	-	17,20,20	1.44	3 (17%)	15,27,27	1.12	2 (13%)

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	G39	B	800	-	-	4/12/32/32	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	800	G39	C3-C2	2.87	1.55	1.50
2	B	800	G39	C6-C5	2.80	1.56	1.53
2	B	800	G39	C6-C7	2.77	1.54	1.50

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	800	G39	C4-C3-C2	2.80	113.06	109.75
2	B	800	G39	C3-C4-N4	2.10	115.14	110.88

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	800	G39	C11-C10-N5-C5
2	B	800	G39	O10-C10-N5-C5
2	B	800	G39	O7-C8-C81-C82
2	B	800	G39	C9-C8-C81-C82

There are no ring outliers.

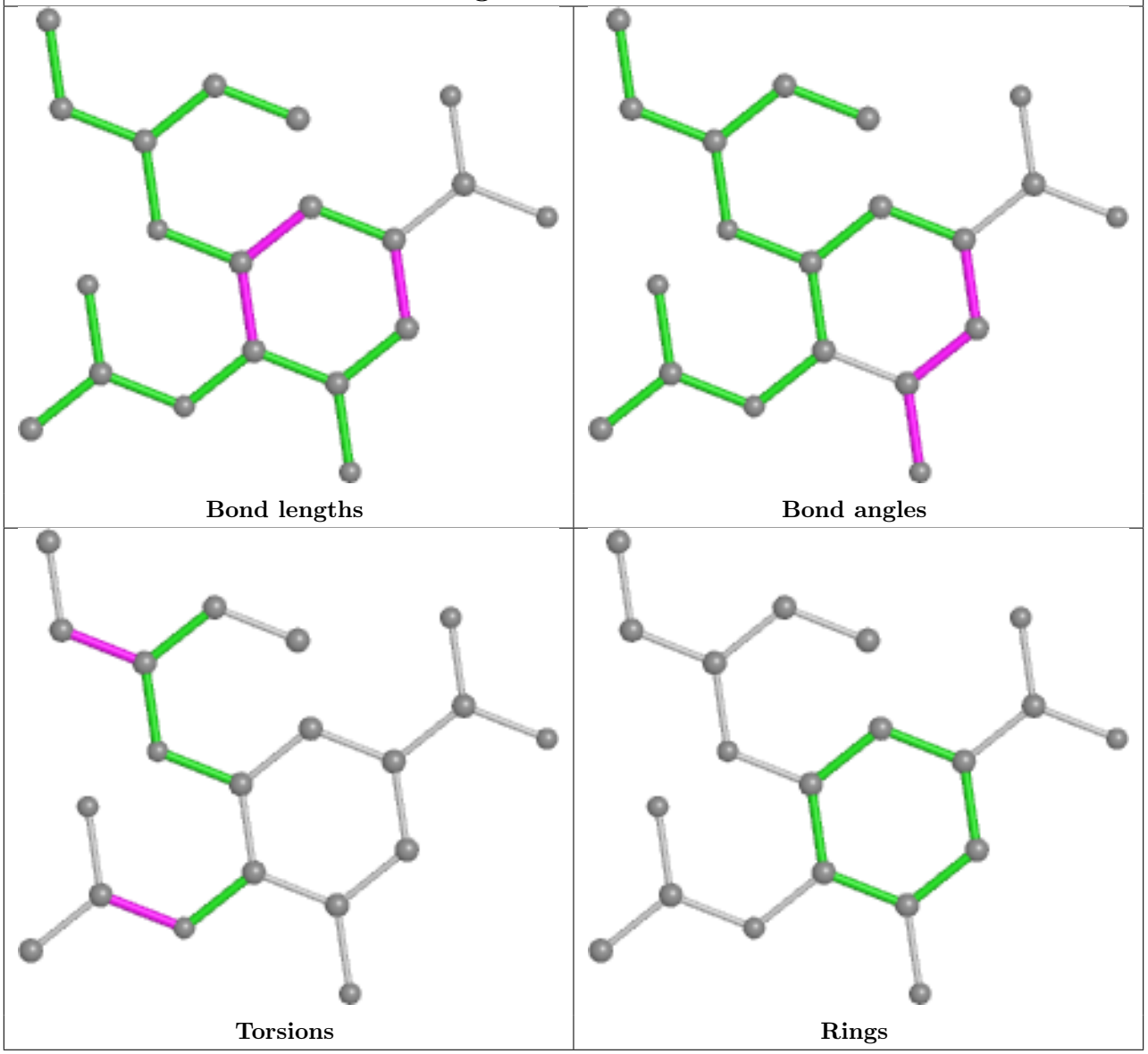
1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	800	G39	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 G39 B 800



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	412:GLN	C	412(A):HIS	N	1.18

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

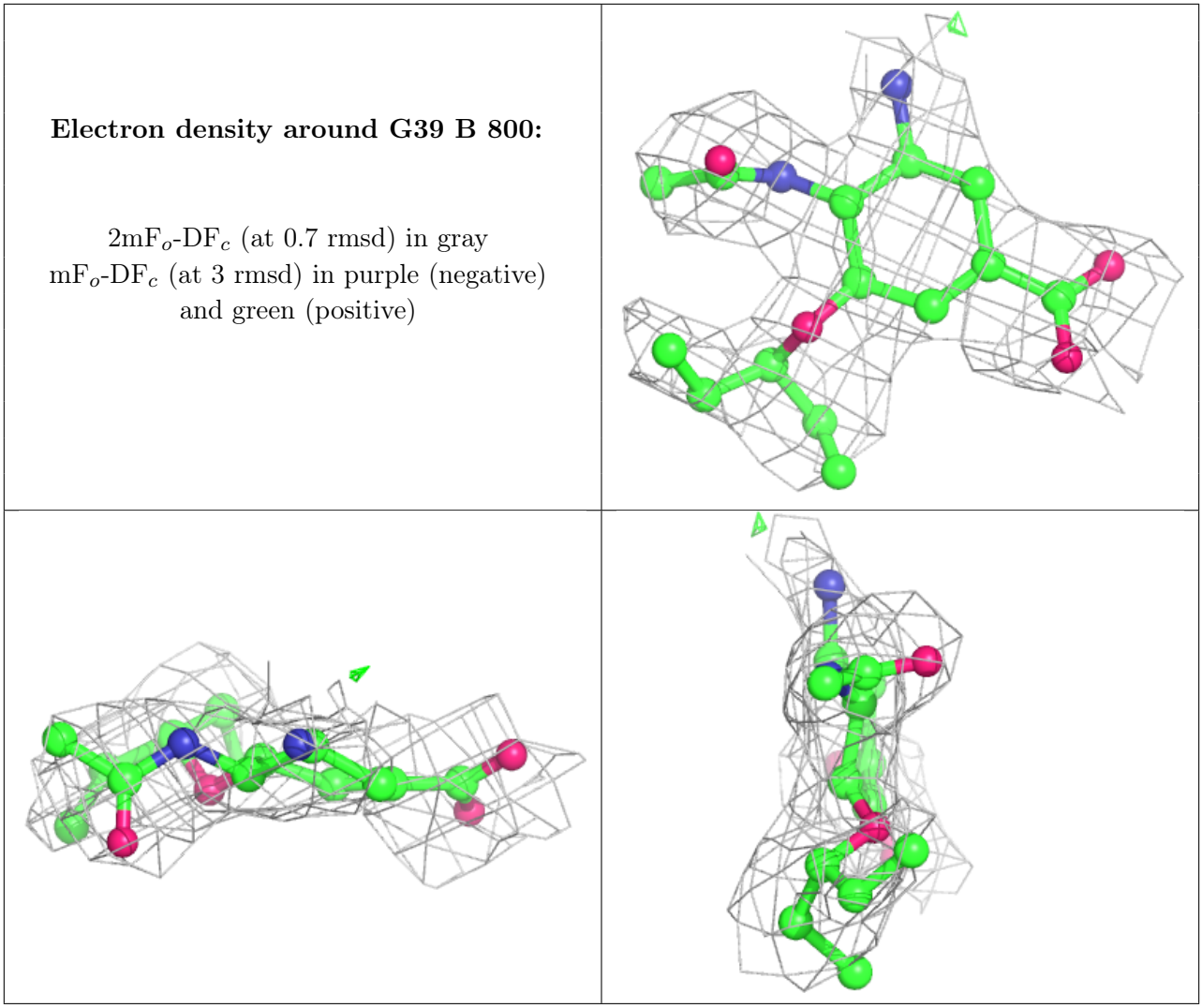
6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

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 G39 B 800:

$2mF_o$-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)

Unable to reproduce the depositors R factor - this section is therefore empty.