Full wwPDB X-ray Structure Validation Report (i)

Oct 23, 2023 – 05:38 PM EDT

PDB ID	:	8ETF
Title	:	Bile Salt Hydrolase B from Lactobacillus gasseri with covalent inhibitor bound
Authors	:	Walker, M.E.; Redinbo, M.R.
Deposited on	:	2022-10-17
Resolution	:	1.79 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

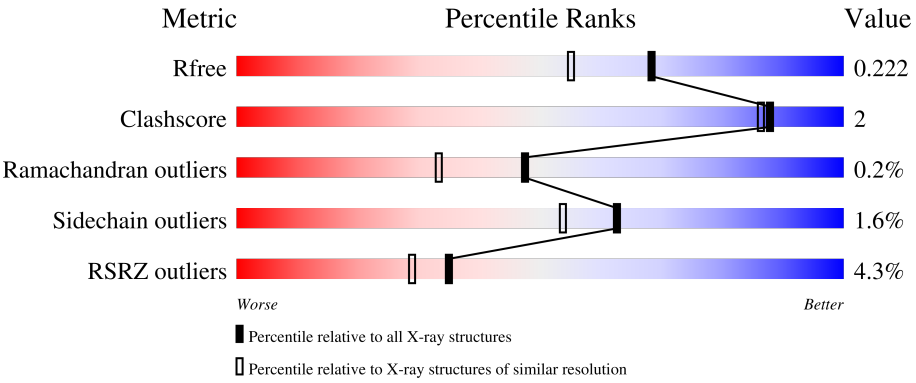
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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 1.79 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.



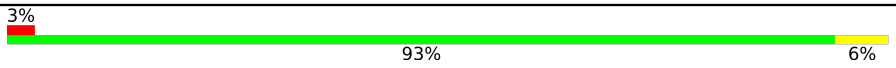
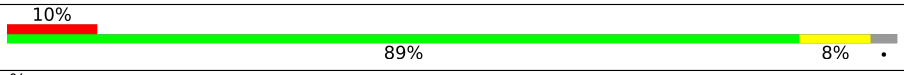
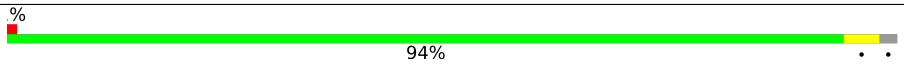
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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	331	4% 90% 8%
1	B	331	3% 95%
1	C	331	5% 94% 6%
1	D	331	5% 91% 7%
1	E	331	3% 92% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	331	
1	G	331	
1	H	331	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 21685 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 Choloylglycine hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	324	Total	C	N	O	S	0	0	0
			2496	1595	403	486	12			
1	A	324	Total	C	N	O	S	0	0	0
			2496	1595	403	486	12			
1	B	329	Total	C	N	O	S	0	0	0
			2544	1626	418	488	12			
1	C	330	Total	C	N	O	S	0	0	0
			2548	1627	418	491	12			
1	D	324	Total	C	N	O	S	0	1	0
			2487	1591	402	482	12			
1	E	329	Total	C	N	O	S	0	0	0
			2533	1617	414	490	12			
1	F	330	Total	C	N	O	S	0	0	0
			2547	1628	419	488	12			
1	G	320	Total	C	N	O	S	0	0	0
			2444	1567	393	472	12			

There are 48 discrepancies between the modelled and reference sequences:

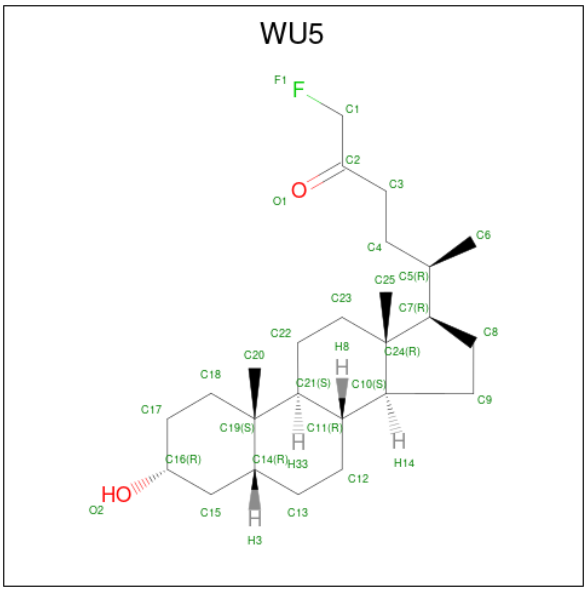
Chain	Residue	Modelled	Actual	Comment	Reference
H	326	HIS	-	expression tag	UNP A0A833FHE1
H	327	HIS	-	expression tag	UNP A0A833FHE1
H	328	HIS	-	expression tag	UNP A0A833FHE1
H	329	HIS	-	expression tag	UNP A0A833FHE1
H	330	HIS	-	expression tag	UNP A0A833FHE1
H	331	HIS	-	expression tag	UNP A0A833FHE1
A	326	HIS	-	expression tag	UNP A0A833FHE1
A	327	HIS	-	expression tag	UNP A0A833FHE1
A	328	HIS	-	expression tag	UNP A0A833FHE1
A	329	HIS	-	expression tag	UNP A0A833FHE1
A	330	HIS	-	expression tag	UNP A0A833FHE1
A	331	HIS	-	expression tag	UNP A0A833FHE1
B	326	HIS	-	expression tag	UNP A0A833FHE1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	327	HIS	-	expression tag	UNP A0A833FHE1
B	328	HIS	-	expression tag	UNP A0A833FHE1
B	329	HIS	-	expression tag	UNP A0A833FHE1
B	330	HIS	-	expression tag	UNP A0A833FHE1
B	331	HIS	-	expression tag	UNP A0A833FHE1
C	326	HIS	-	expression tag	UNP A0A833FHE1
C	327	HIS	-	expression tag	UNP A0A833FHE1
C	328	HIS	-	expression tag	UNP A0A833FHE1
C	329	HIS	-	expression tag	UNP A0A833FHE1
C	330	HIS	-	expression tag	UNP A0A833FHE1
C	331	HIS	-	expression tag	UNP A0A833FHE1
D	326	HIS	-	expression tag	UNP A0A833FHE1
D	327	HIS	-	expression tag	UNP A0A833FHE1
D	328	HIS	-	expression tag	UNP A0A833FHE1
D	329	HIS	-	expression tag	UNP A0A833FHE1
D	330	HIS	-	expression tag	UNP A0A833FHE1
D	331	HIS	-	expression tag	UNP A0A833FHE1
E	326	HIS	-	expression tag	UNP A0A833FHE1
E	327	HIS	-	expression tag	UNP A0A833FHE1
E	328	HIS	-	expression tag	UNP A0A833FHE1
E	329	HIS	-	expression tag	UNP A0A833FHE1
E	330	HIS	-	expression tag	UNP A0A833FHE1
E	331	HIS	-	expression tag	UNP A0A833FHE1
F	326	HIS	-	expression tag	UNP A0A833FHE1
F	327	HIS	-	expression tag	UNP A0A833FHE1
F	328	HIS	-	expression tag	UNP A0A833FHE1
F	329	HIS	-	expression tag	UNP A0A833FHE1
F	330	HIS	-	expression tag	UNP A0A833FHE1
F	331	HIS	-	expression tag	UNP A0A833FHE1
G	326	HIS	-	expression tag	UNP A0A833FHE1
G	327	HIS	-	expression tag	UNP A0A833FHE1
G	328	HIS	-	expression tag	UNP A0A833FHE1
G	329	HIS	-	expression tag	UNP A0A833FHE1
G	330	HIS	-	expression tag	UNP A0A833FHE1
G	331	HIS	-	expression tag	UNP A0A833FHE1

- Molecule 2 is (5R)-1-fluoro-5-[(1R,3aS,3bR,5aR,7R,9aS,9bS,11aR)-7-hydroxy-9a,11a-dimethylhexadecahydro-1H-cyclopenta[a]phenanthren-1-yl]hexan-2-one (non-preferred name) (three-letter code: WU5) (formula: $C_{25}H_{41}FO_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	H	1	Total C O 27 25 2	0	0
2	A	1	Total C O 27 25 2	0	0
2	B	1	Total C O 27 25 2	0	0
2	C	1	Total C O 27 25 2	0	0
2	D	1	Total C O 27 25 2	0	0
2	E	1	Total C O 27 25 2	0	0
2	F	1	Total C O 27 25 2	0	0
2	G	1	Total C O 27 25 2	0	0

- Molecule 3 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Ni 1 1	0	0
3	F	1	Total Ni 1 1	0	0

- Molecule 4 is water.

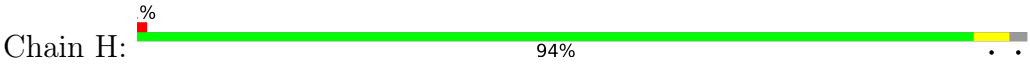
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	204	Total O 204 204	0	0
4	A	196	Total O 196 196	0	0
4	B	200	Total O 200 200	0	0
4	C	199	Total O 199 199	0	0
4	D	155	Total O 155 155	0	0
4	E	165	Total O 165 165	0	0
4	F	157	Total O 157 157	0	0
4	G	96	Total O 96 96	0	0

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.

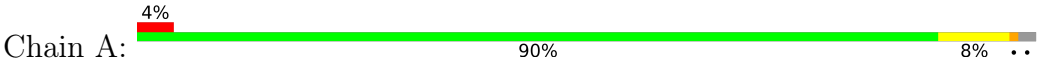
• Molecule 1: Cholоylglycine hydrolase

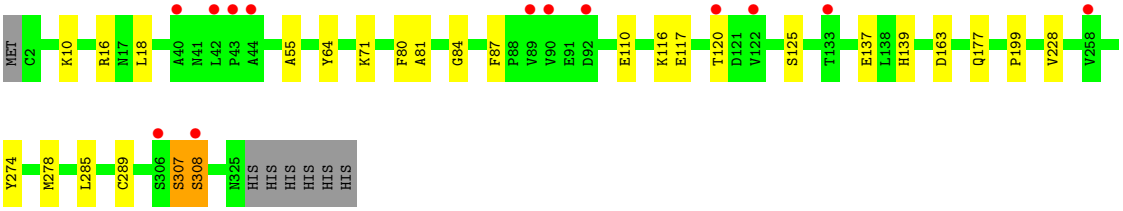
Chain H:



• Molecule 1: Cholоylglycine hydrolase

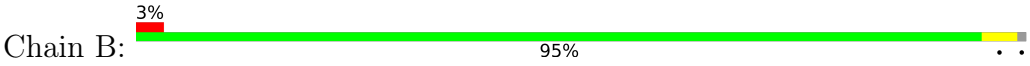
Chain A:

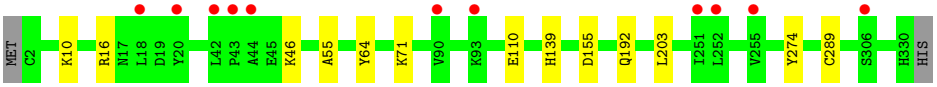




• Molecule 1: Cholоylglycine hydrolase

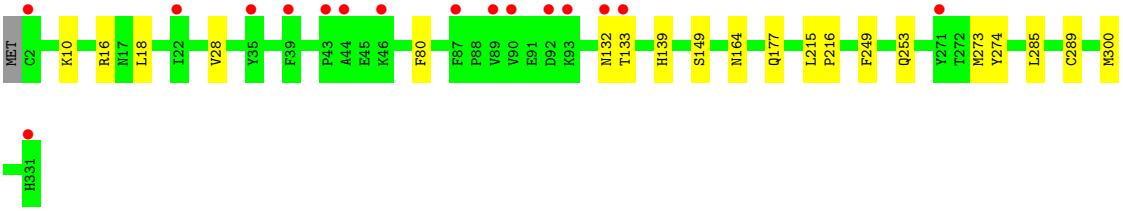
Chain B:





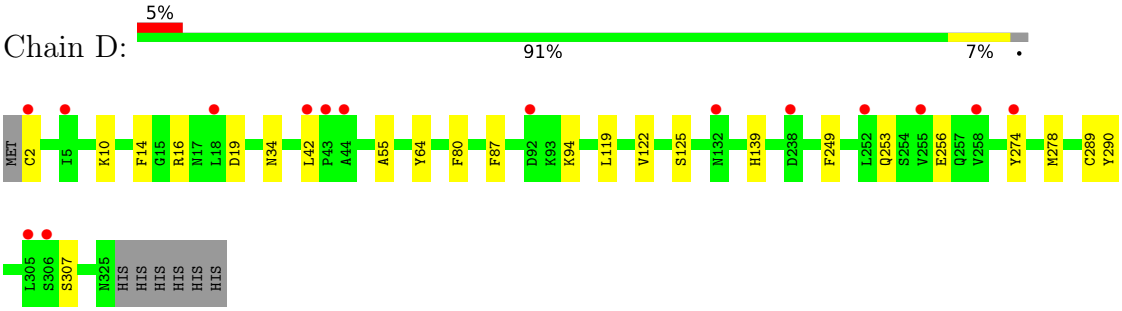
• Molecule 1: Cholоylglycine hydrolase

Chain C:



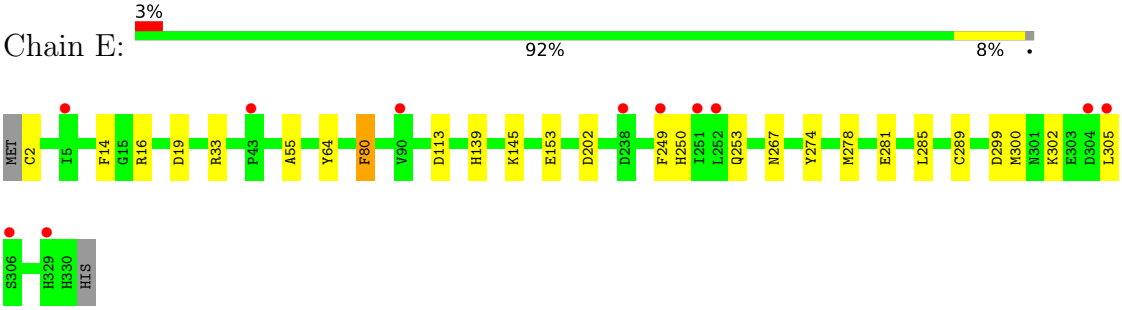
• Molecule 1: Cholоylglycine hydrolase

Chain D:



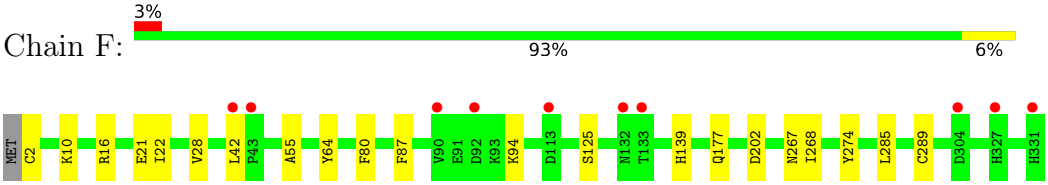
• Molecule 1: Choloylglycine hydrolase

Chain E:



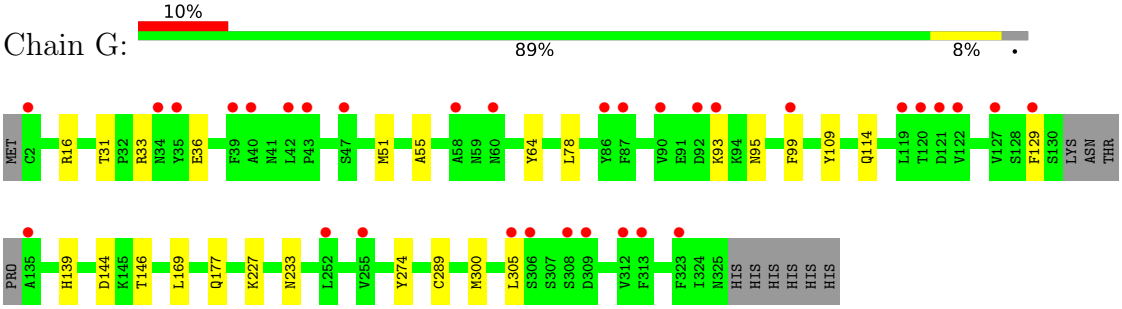
• Molecule 1: Choloylglycine hydrolase

Chain F:



• Molecule 1: Choloylglycine hydrolase

Chain G:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	104.52Å 147.99Å 104.56Å 90.00° 94.82° 90.00°	Depositor
Resolution (Å)	45.86 – 1.79 45.86 – 1.79	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.86-1.79) 99.5 (45.86-1.79)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.06 (at 1.79Å)	Xtriage
Refinement program	PHENIX 1.20_4459, PHENIX 1.20_4459	Depositor
R, R_{free}	0.186 , 0.222 0.186 , 0.222	Depositor DCC
R_{free} test set	2004 reflections (0.68%)	wwPDB-VP
Wilson B-factor (Å^2)	33.5	Xtriage
Anisotropy	0.241	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.014 for l,-k,h	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	21685	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.26% 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: WU5, NI

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	0.43	0/2556	0.59	0/3482
1	B	0.43	0/2609	0.58	0/3553
1	C	0.41	0/2613	0.59	0/3560
1	D	0.40	0/2550	0.57	0/3474
1	E	0.38	0/2598	0.56	0/3541
1	F	0.37	0/2613	0.55	0/3560
1	G	0.37	0/2502	0.57	0/3410
1	H	0.43	0/2556	0.62	0/3482
All	All	0.40	0/20597	0.58	0/28062

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	2496	0	2361	14	0
1	B	2544	0	2401	6	0
1	C	2548	0	2394	11	0
1	D	2487	0	2343	12	0
1	E	2533	0	2366	13	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	2547	0	2393	11	0
1	G	2444	0	2289	17	0
1	H	2496	0	2361	6	0
2	A	27	0	0	0	0
2	B	27	0	0	0	0
2	C	27	0	0	0	0
2	D	27	0	0	0	0
2	E	27	0	0	0	0
2	F	27	0	0	1	0
2	G	27	0	0	0	0
2	H	27	0	0	1	0
3	B	1	0	0	0	0
3	F	1	0	0	0	0
4	A	196	0	0	1	0
4	B	200	0	0	2	0
4	C	199	0	0	1	0
4	D	155	0	0	0	0
4	E	165	0	0	0	0
4	F	157	0	0	0	0
4	G	96	0	0	1	0
4	H	204	0	0	0	0
All	All	21685	0	18908	80	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 2.

All (80) 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:177:GLN:HG2	1:G:177:GLN:HG2	1.51	0.92
1:A:177:GLN:HG2	1:F:177:GLN:HG2	1.60	0.83
1:D:249:PHE:O	1:D:253:GLN:HG3	1.93	0.69
1:D:2:CYS:SG	1:D:19:ASP:HB2	2.35	0.67
1:H:278:MET:HG3	1:H:285:LEU:HD13	1.81	0.63
1:G:109:TYR:HD1	1:G:114:GLN:HE21	1.52	0.58
1:A:307:SER:OG	1:A:308:SER:N	2.40	0.54
1:D:256:GLU:OE2	1:E:250:HIS:HD2	1.91	0.54
1:D:14:PHE:O	1:D:278:MET:HB2	2.07	0.54
1:C:177:GLN:HG2	1:G:177:GLN:CG	2.32	0.54
1:A:177:GLN:CG	1:F:177:GLN:HG2	2.35	0.53
1:C:273:MET:HE1	4:C:601:HOH:O	2.09	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:177:GLN:HG2	1:F:177:GLN:CG	2.36	0.52
1:D:290:TYR:O	1:E:250:HIS:HE1	1.93	0.52
1:A:199:PRO:HD2	1:G:233:ASN:OD1	2.09	0.51
1:A:84:GLY:HA3	4:A:617:HOH:O	2.10	0.51
1:F:267:ASN:O	1:F:268:ILE:HD13	2.11	0.50
1:G:31:THR:HB	1:G:51:MET:HG2	1.94	0.50
1:B:71:LYS:HG3	1:B:110:GLU:HG2	1.94	0.49
1:F:42:LEU:HD21	1:F:94:LYS:HG3	1.95	0.49
1:E:55:ALA:HB2	1:E:64:TYR:CD1	2.48	0.49
1:F:274:TYR:HA	1:F:289:CYS:HA	1.95	0.49
1:G:78:LEU:HD13	1:G:169:LEU:HD23	1.94	0.49
1:C:249:PHE:O	1:C:253:GLN:HG3	2.12	0.49
1:C:177:GLN:CG	1:G:177:GLN:HG2	2.34	0.49
1:C:133:THR:O	1:C:133:THR:OG1	2.31	0.48
1:E:249:PHE:O	1:E:253:GLN:HG3	2.12	0.48
1:A:116:LYS:NZ	1:A:163:ASP:OD1	2.46	0.48
1:B:274:TYR:HA	1:B:289:CYS:HA	1.96	0.48
1:A:278:MET:HG2	1:A:285:LEU:HD13	1.97	0.47
1:A:71:LYS:HG3	1:A:110:GLU:HG2	1.95	0.47
1:A:81:ALA:HA	1:A:137:GLU:OE1	2.14	0.47
1:G:55:ALA:HB2	1:G:64:TYR:CD1	2.49	0.47
1:F:2:CYS:N	2:F:401:WU5:C1	2.78	0.47
1:G:144:ASP:OD1	1:G:146:THR:OG1	2.24	0.47
1:C:28:VAL:HG13	1:C:285:LEU:HD23	1.97	0.46
1:E:80:PHE:CE2	1:E:153:GLU:HG2	2.51	0.46
1:H:249:PHE:O	1:H:253:GLN:HG3	2.15	0.46
1:E:299:ASP:HB3	1:E:302:LYS:HD3	1.97	0.46
1:A:117:GLU:O	1:A:120:THR:HG23	2.15	0.46
1:G:227:LYS:NZ	4:G:504:HOH:O	2.48	0.46
1:E:274:TYR:HA	1:E:289:CYS:HA	1.98	0.46
1:F:55:ALA:HB2	1:F:64:TYR:CD1	2.51	0.46
1:G:93:LYS:O	1:G:95:ASN:ND2	2.49	0.46
1:H:33:ARG:HG3	1:H:305:LEU:O	2.16	0.45
1:G:33:ARG:HG3	1:G:305:LEU:O	2.17	0.44
1:G:99:PHE:CE2	1:G:129:PHE:HD1	2.35	0.44
1:G:274:TYR:HA	1:G:289:CYS:HA	1.99	0.44
1:D:34:ASN:HB2	1:D:307:SER:O	2.18	0.43
1:C:274:TYR:HA	1:C:289:CYS:HA	2.00	0.43
1:H:2:CYS:N	2:H:401:WU5:C1	2.82	0.43
1:E:2:CYS:SG	1:E:19:ASP:HB2	2.59	0.43
1:E:2:CYS:SG	1:E:19:ASP:O	2.76	0.43

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:119:LEU:HD23	1:D:122:VAL:HG21	2.00	0.43
1:E:14:PHE:O	1:E:278:MET:HB2	2.19	0.43
1:A:55:ALA:HB2	1:A:64:TYR:CD1	2.54	0.43
1:E:145:LYS:HD2	1:E:281:GLU:HG3	2.02	0.42
1:F:87:PHE:O	1:F:125:SER:HA	2.19	0.42
1:B:192:GLN:HB2	1:C:132:ASN:HB3	2.01	0.42
1:A:87:PHE:O	1:A:125:SER:HA	2.19	0.42
1:B:55:ALA:HB2	1:B:64:TYR:CD1	2.55	0.42
1:C:215:LEU:HA	1:C:216:PRO:HD3	1.93	0.42
1:D:87:PHE:O	1:D:125:SER:HA	2.20	0.42
1:F:28:VAL:HG13	1:F:285:LEU:HD23	2.01	0.41
1:G:99:PHE:H	1:G:99:PHE:HD1	1.67	0.41
1:H:274:TYR:HA	1:H:289:CYS:HA	2.03	0.41
1:A:274:TYR:HA	1:A:289:CYS:HA	2.02	0.41
1:E:33:ARG:HG3	1:E:305:LEU:O	2.20	0.41
1:G:36:GLU:O	1:G:36:GLU:HG3	2.19	0.41
1:D:42:LEU:HD21	1:D:94:LYS:HG3	2.03	0.41
1:D:55:ALA:HB2	1:D:64:TYR:CD1	2.56	0.41
1:F:21:GLU:HG2	1:F:22:ILE:HG13	2.02	0.41
1:B:46:LYS:N	4:B:502:HOH:O	2.37	0.41
1:D:274:TYR:HA	1:D:289:CYS:HA	2.03	0.40
1:G:51:MET:HE2	1:G:51:MET:HB3	1.96	0.40
1:H:87:PHE:O	1:H:125:SER:HA	2.20	0.40
1:B:155:ASP:HB2	4:B:527:HOH:O	2.21	0.40
1:C:149:SER:OG	1:C:164:ASN:HB3	2.22	0.40
1:D:2:CYS:SG	1:D:19:ASP:O	2.79	0.40
1:E:278:MET:HG3	1:E:285:LEU:HD13	2.04	0.40

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	322/331 (97%)	312 (97%)	9 (3%)	1 (0%)	41	27
1	B	327/331 (99%)	318 (97%)	8 (2%)	1 (0%)	41	27
1	C	328/331 (99%)	321 (98%)	6 (2%)	1 (0%)	41	27
1	D	323/331 (98%)	317 (98%)	5 (2%)	1 (0%)	41	27
1	E	327/331 (99%)	315 (96%)	12 (4%)	0	100	100
1	F	328/331 (99%)	320 (98%)	7 (2%)	1 (0%)	41	27
1	G	316/331 (96%)	305 (96%)	11 (4%)	0	100	100
1	H	322/331 (97%)	312 (97%)	9 (3%)	1 (0%)	41	27
All	All	2593/2648 (98%)	2520 (97%)	67 (3%)	6 (0%)	47	33

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	10	LYS
1	C	10	LYS
1	F	10	LYS
1	H	10	LYS
1	A	10	LYS
1	D	10	LYS

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	271/291 (93%)	264 (97%)	7 (3%)	46	32
1	B	275/291 (94%)	272 (99%)	3 (1%)	73	68
1	C	275/291 (94%)	270 (98%)	5 (2%)	59	48
1	D	267/291 (92%)	264 (99%)	3 (1%)	73	68
1	E	272/291 (94%)	265 (97%)	7 (3%)	46	32
1	F	274/291 (94%)	270 (98%)	4 (2%)	65	56
1	G	259/291 (89%)	256 (99%)	3 (1%)	71	65

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	H	271/291 (93%)	269 (99%)	2 (1%)	84 81
All	All	2164/2328 (93%)	2130 (98%)	34 (2%)	62 54

All (34) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	H	16	ARG
1	H	139	HIS
1	A	16	ARG
1	A	18	LEU
1	A	80	PHE
1	A	139	HIS
1	A	228	VAL
1	A	307	SER
1	A	308	SER
1	B	16	ARG
1	B	139	HIS
1	B	203	LEU
1	C	16	ARG
1	C	18	LEU
1	C	80	PHE
1	C	139	HIS
1	C	300	MET
1	D	16	ARG
1	D	80	PHE
1	D	139	HIS
1	E	16	ARG
1	E	80	PHE
1	E	113	ASP
1	E	139	HIS
1	E	202	ASP
1	E	267	ASN
1	E	300	MET
1	F	16	ARG
1	F	80	PHE
1	F	139	HIS
1	F	202	ASP
1	G	16	ARG
1	G	139	HIS
1	G	300	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such

sidechains are listed below:

Mol	Chain	Res	Type
1	H	177	GLN
1	H	181	ASN
1	H	253	GLN
1	A	253	GLN
1	D	253	GLN
1	E	250	HIS
1	E	253	GLN
1	F	253	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 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	WU5	H	401	1	30,30,31	0.46	0	47,47,48	1.33	1 (2%)
2	WU5	F	401	1	30,30,31	0.45	0	47,47,48	0.90	1 (2%)
2	WU5	D	401	1	30,30,31	0.40	0	47,47,48	0.66	1 (2%)
2	WU5	B	401	1	30,30,31	0.45	0	47,47,48	1.08	1 (2%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	WU5	C	401	1	30,30,31	0.40	0	47,47,48	0.60	1 (2%)
2	WU5	G	401	1	30,30,31	0.40	0	47,47,48	0.74	1 (2%)
2	WU5	A	401	1	30,30,31	0.43	0	47,47,48	1.10	1 (2%)
2	WU5	E	401	1	30,30,31	0.37	0	47,47,48	0.45	0

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	WU5	H	401	1	-	3/9/67/69	0/4/4/4
2	WU5	F	401	1	-	3/9/67/69	0/4/4/4
2	WU5	D	401	1	-	0/9/67/69	0/4/4/4
2	WU5	B	401	1	-	3/9/67/69	0/4/4/4
2	WU5	C	401	1	-	7/9/67/69	0/4/4/4
2	WU5	G	401	1	-	3/9/67/69	0/4/4/4
2	WU5	A	401	1	-	4/9/67/69	0/4/4/4
2	WU5	E	401	1	-	0/9/67/69	0/4/4/4

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	H	401	WU5	C4-C3-C2	-8.09	106.25	114.57
2	B	401	WU5	C4-C3-C2	-6.64	107.74	114.57
2	A	401	WU5	C4-C3-C2	-6.45	107.94	114.57
2	F	401	WU5	C4-C3-C2	-5.33	109.09	114.57
2	G	401	WU5	C4-C3-C2	-3.12	111.36	114.57
2	D	401	WU5	C4-C3-C2	-3.03	111.46	114.57
2	C	401	WU5	C4-C3-C2	-2.72	111.77	114.57

There are no chirality outliers.

All (23) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	401	WU5	C3-C4-C5-C7
2	H	401	WU5	C3-C4-C5-C6
2	C	401	WU5	C3-C4-C5-C6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	H	401	WU5	C1-C2-C3-C4
2	A	401	WU5	C1-C2-C3-C4
2	B	401	WU5	C1-C2-C3-C4
2	C	401	WU5	C1-C2-C3-C4
2	F	401	WU5	C1-C2-C3-C4
2	F	401	WU5	C3-C4-C5-C6
2	H	401	WU5	O1-C2-C3-C4
2	G	401	WU5	C1-C2-C3-C4
2	B	401	WU5	C3-C4-C5-C6
2	A	401	WU5	O1-C2-C3-C4
2	B	401	WU5	O1-C2-C3-C4
2	C	401	WU5	O1-C2-C3-C4
2	F	401	WU5	O1-C2-C3-C4
2	A	401	WU5	C3-C4-C5-C6
2	G	401	WU5	O1-C2-C3-C4
2	C	401	WU5	C4-C5-C7-C24
2	C	401	WU5	C4-C5-C7-C8
2	A	401	WU5	C2-C3-C4-C5
2	C	401	WU5	C6-C5-C7-C24
2	G	401	WU5	C3-C4-C5-C6

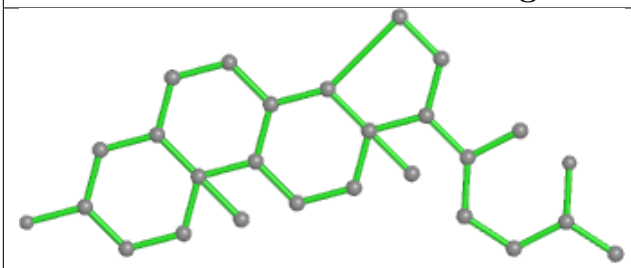
There are no ring outliers.

2 monomers are involved in 2 short contacts:

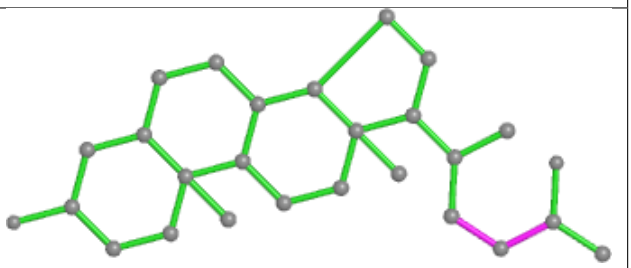
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	H	401	WU5	1	0
2	F	401	WU5	1	0

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.

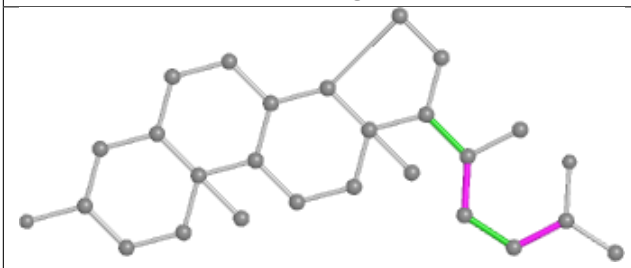
Ligand WU5 H 401



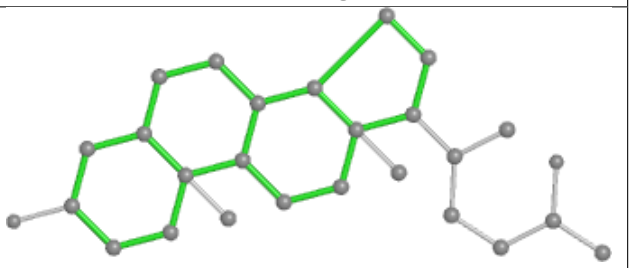
Bond lengths



Bond angles

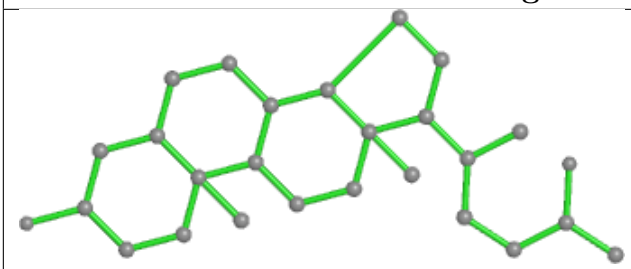


Torsions

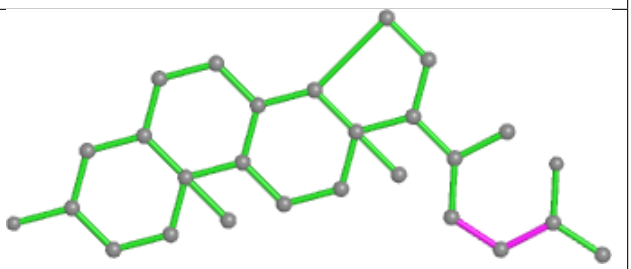


Rings

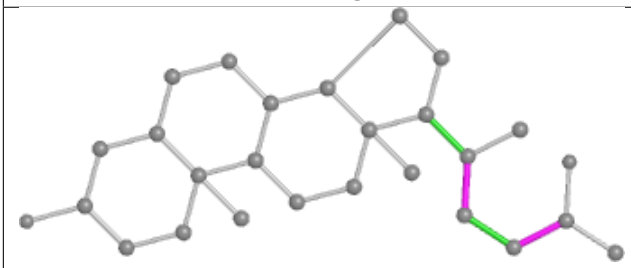
Ligand WU5 F 401



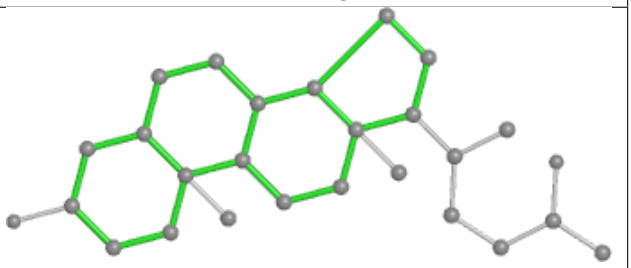
Bond lengths



Bond angles

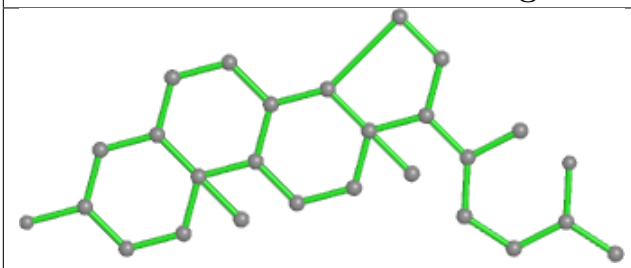


Torsions

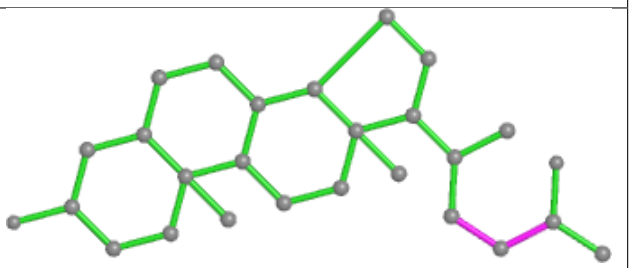


Rings

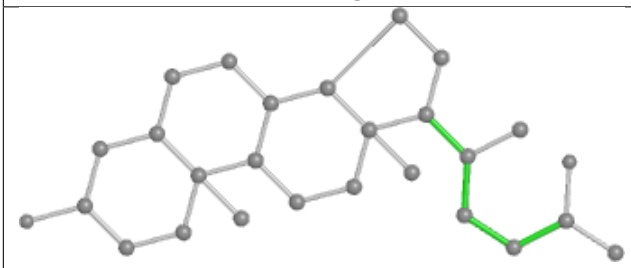
Ligand WU5 D 401



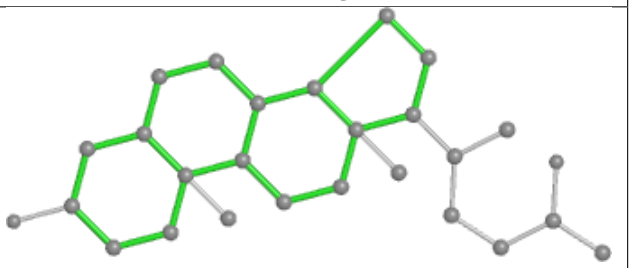
Bond lengths



Bond angles

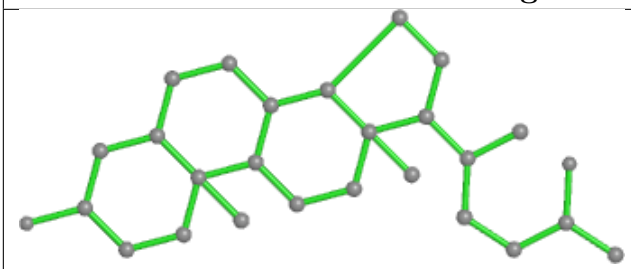


Torsions

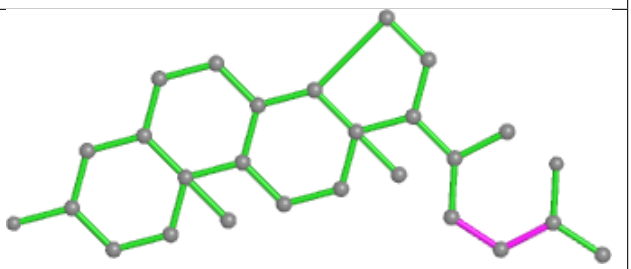


Rings

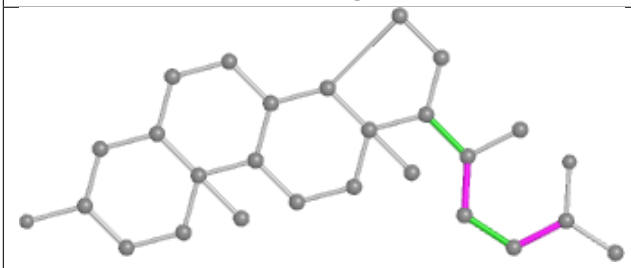
Ligand WU5 B 401



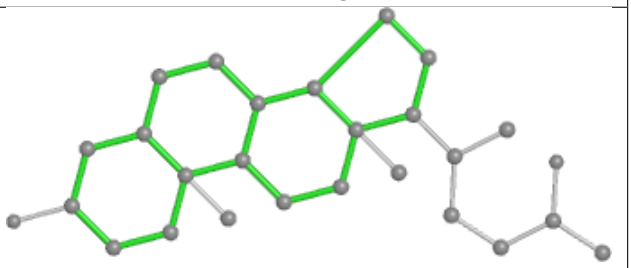
Bond lengths



Bond angles

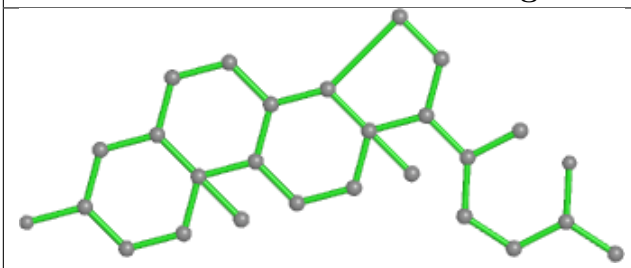


Torsions

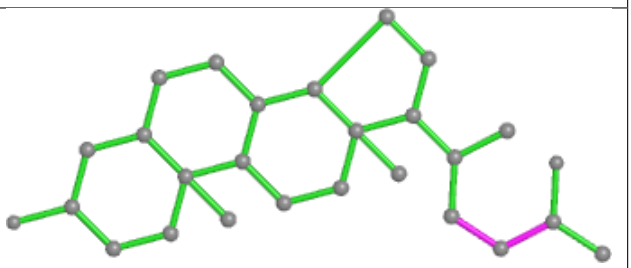


Rings

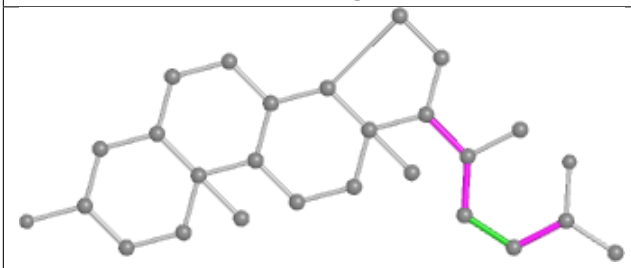
Ligand WU5 C 401



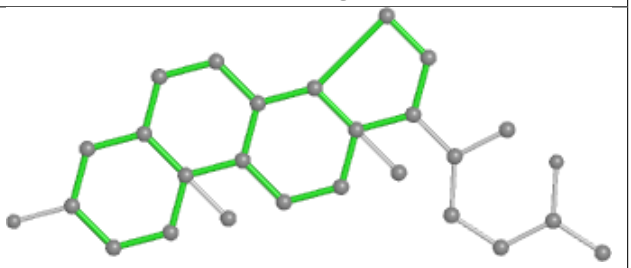
Bond lengths



Bond angles

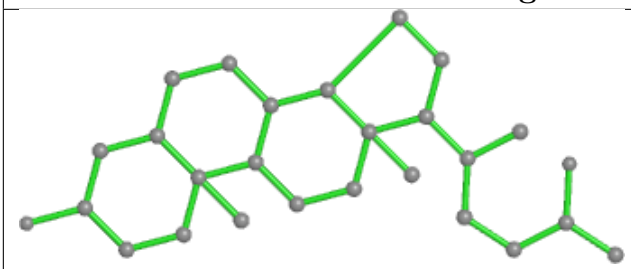


Torsions

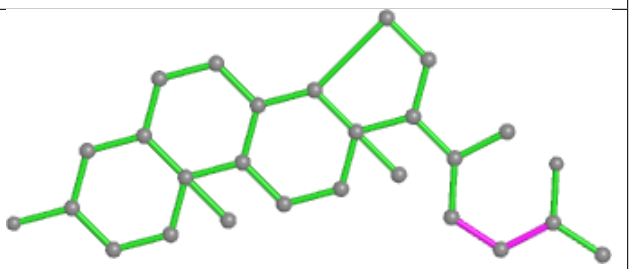


Rings

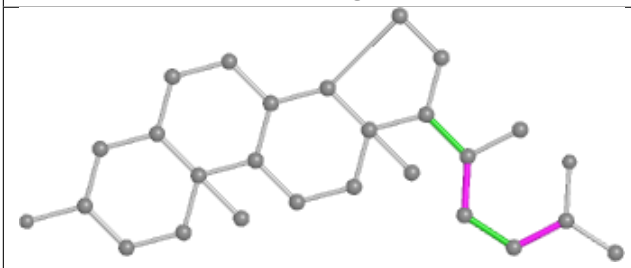
Ligand WU5 G 401



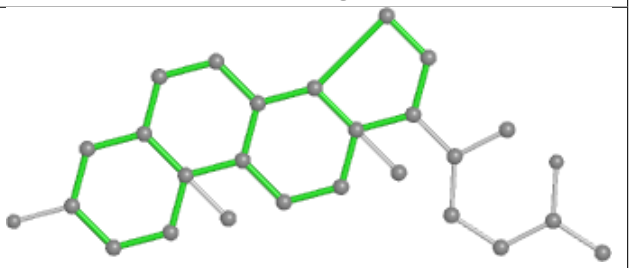
Bond lengths



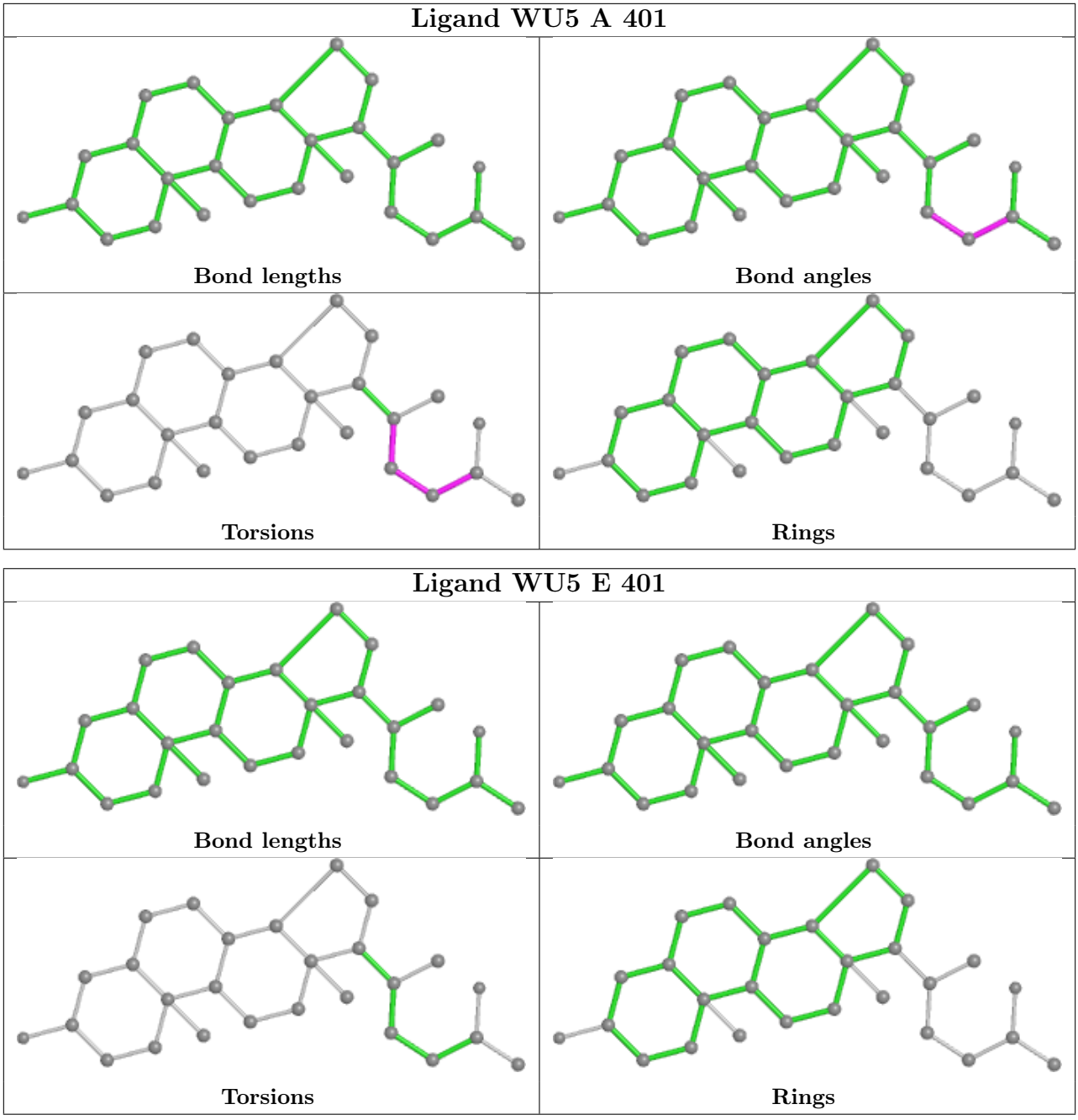
Bond angles



Torsions



Rings



5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	324/331 (97%)	0.15	13 (4%) 38 32	25, 34, 52, 61	0
1	B	329/331 (99%)	0.21	11 (3%) 46 40	24, 33, 49, 57	0
1	C	330/331 (99%)	0.23	16 (4%) 30 25	25, 34, 52, 63	0
1	D	324/331 (97%)	0.22	15 (4%) 32 26	25, 38, 53, 62	0
1	E	329/331 (99%)	0.11	11 (3%) 46 40	25, 39, 56, 63	0
1	F	330/331 (99%)	-0.02	10 (3%) 50 44	26, 39, 54, 67	1 (0%)
1	G	320/331 (96%)	0.53	32 (10%) 7 5	28, 44, 63, 73	1 (0%)
1	H	324/331 (97%)	0.01	3 (0%) 84 82	21, 33, 48, 60	0
All	All	2610/2648 (98%)	0.18	111 (4%) 35 29	21, 37, 55, 73	2 (0%)

All (111) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	40	ALA	4.7
1	G	305	LEU	4.7
1	F	133	THR	4.4
1	G	313	PHE	4.3
1	D	44	ALA	3.8
1	G	42	LEU	3.7
1	C	90	VAL	3.6
1	A	90	VAL	3.4
1	E	305	LEU	3.4
1	D	252	LEU	3.4
1	A	122	VAL	3.3
1	G	47	SER	3.3
1	F	331	HIS	3.3
1	B	255	VAL	3.3
1	G	122	VAL	3.3
1	A	43	PRO	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	F	42	LEU	3.2
1	G	127	VAL	3.2
1	G	43	PRO	3.2
1	H	90	VAL	3.2
1	C	39	PHE	3.2
1	G	87	PHE	3.1
1	D	92	ASP	3.1
1	C	133	THR	3.1
1	G	121	ASP	3.0
1	F	327	HIS	3.0
1	B	90	VAL	3.0
1	G	93	LYS	3.0
1	A	44	ALA	2.9
1	G	92	ASP	2.9
1	H	133	THR	2.9
1	B	18	LEU	2.8
1	G	252	LEU	2.8
1	G	2	CYS	2.8
1	A	306	SER	2.8
1	G	129	PHE	2.8
1	E	5	ILE	2.7
1	A	40	ALA	2.7
1	G	58	ALA	2.7
1	A	89	VAL	2.7
1	A	133	THR	2.7
1	E	252	LEU	2.7
1	C	87	PHE	2.7
1	G	90	VAL	2.7
1	E	43	PRO	2.7
1	G	99	PHE	2.7
1	D	43	PRO	2.6
1	E	329	HIS	2.6
1	D	2	CYS	2.6
1	G	308	SER	2.6
1	D	132	ASN	2.6
1	G	135	ALA	2.6
1	C	2	CYS	2.6
1	B	306	SER	2.6
1	D	42	LEU	2.6
1	B	251	ILE	2.5
1	G	119	LEU	2.5
1	G	323	PHE	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	238	ASP	2.5
1	B	43	PRO	2.5
1	C	132	ASN	2.4
1	H	22	ILE	2.4
1	B	252	LEU	2.4
1	D	305	LEU	2.4
1	D	274	TYR	2.4
1	G	306	SER	2.4
1	D	5	ILE	2.4
1	E	251	ILE	2.3
1	F	43	PRO	2.3
1	F	92	ASP	2.3
1	E	249	PHE	2.3
1	G	312	VAL	2.3
1	B	42	LEU	2.3
1	G	120	THR	2.3
1	C	35	TYR	2.3
1	C	92	ASP	2.3
1	G	255	VAL	2.2
1	A	120	THR	2.2
1	D	255	VAL	2.2
1	F	90	VAL	2.2
1	C	44	ALA	2.2
1	G	34	ASN	2.2
1	D	18	LEU	2.2
1	C	331	HIS	2.2
1	G	86	TYR	2.2
1	G	309	ASP	2.2
1	C	46	LYS	2.2
1	F	304	ASP	2.2
1	C	22	ILE	2.1
1	F	113	ASP	2.1
1	B	44	ALA	2.1
1	E	304	ASP	2.1
1	B	20	TYR	2.1
1	A	308	SER	2.1
1	D	306	SER	2.1
1	G	60	ASN	2.1
1	E	306	SER	2.1
1	A	42	LEU	2.1
1	A	92	ASP	2.1
1	C	93	LYS	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	G	35	TYR	2.0
1	G	39	PHE	2.0
1	B	93	LYS	2.0
1	F	132	ASN	2.0
1	C	271	TYR	2.0
1	A	258	VAL	2.0
1	C	89	VAL	2.0
1	D	258	VAL	2.0
1	E	90	VAL	2.0
1	E	238	ASP	2.0
1	C	43	PRO	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

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.

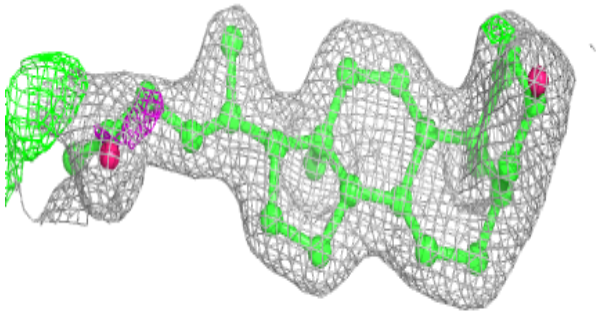
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	WU5	B	401	27/28	0.93	0.13	29,32,37,38	1
2	WU5	G	401	27/28	0.93	0.10	41,46,49,51	1
2	WU5	E	401	27/28	0.94	0.07	31,35,40,40	0
2	WU5	F	401	27/28	0.94	0.09	34,37,43,49	1
2	WU5	D	401	27/28	0.94	0.08	34,37,40,40	1
2	WU5	C	401	27/28	0.95	0.15	30,33,39,40	1
2	WU5	H	401	27/28	0.96	0.09	28,31,37,37	1
2	WU5	A	401	27/28	0.96	0.11	29,32,36,36	1
3	NI	B	402	1/1	0.99	0.05	41,41,41,41	0
3	NI	F	402	1/1	0.99	0.05	42,42,42,42	0

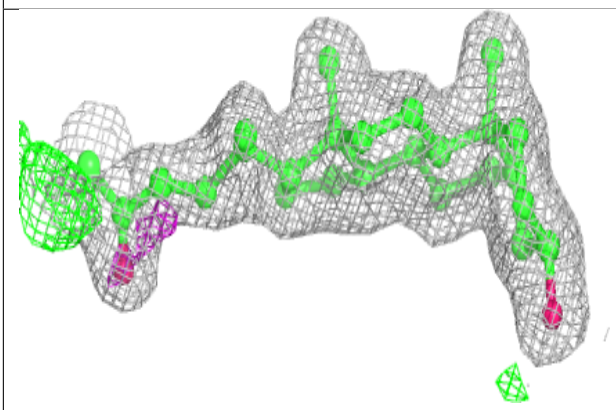
The following is a graphical depiction of the model fit to experimental electron density of all

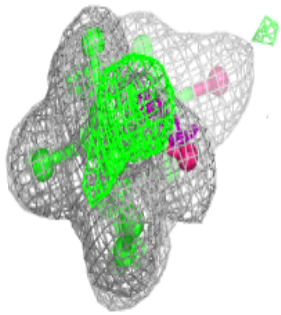
instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around WU5 B 401:

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

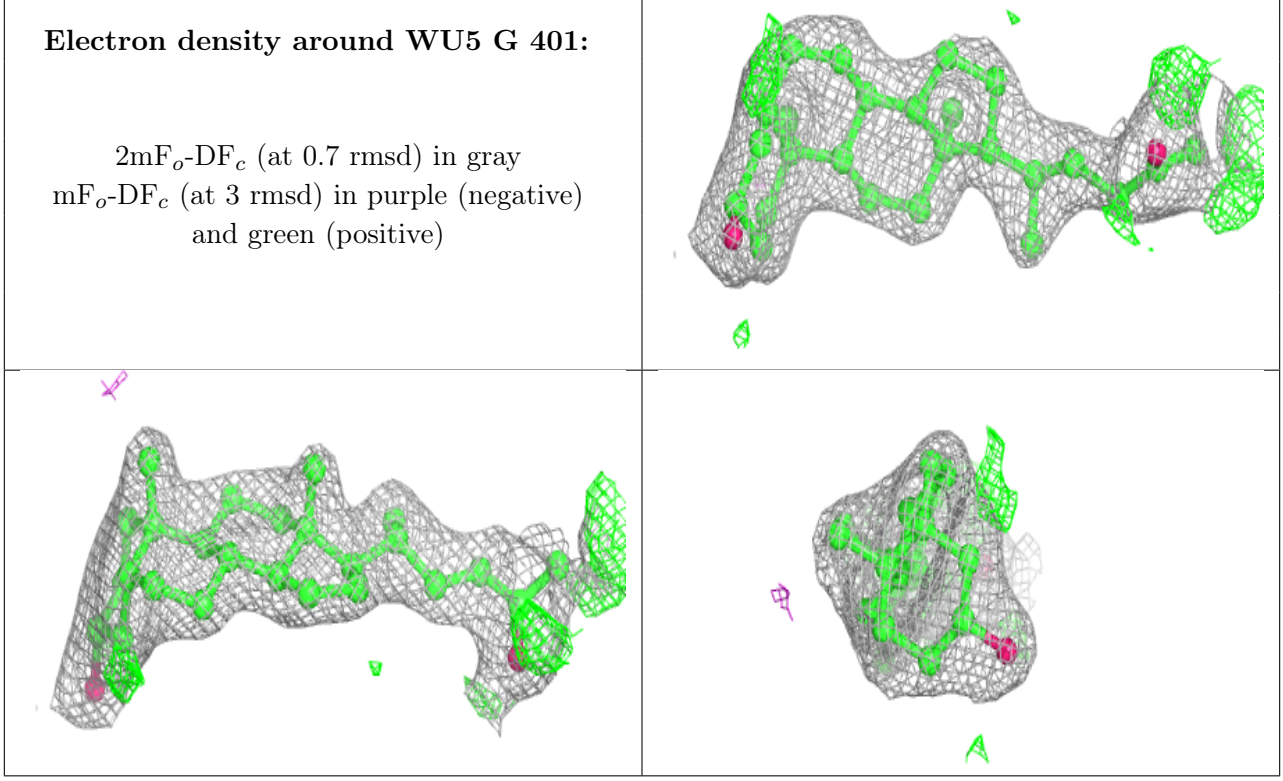






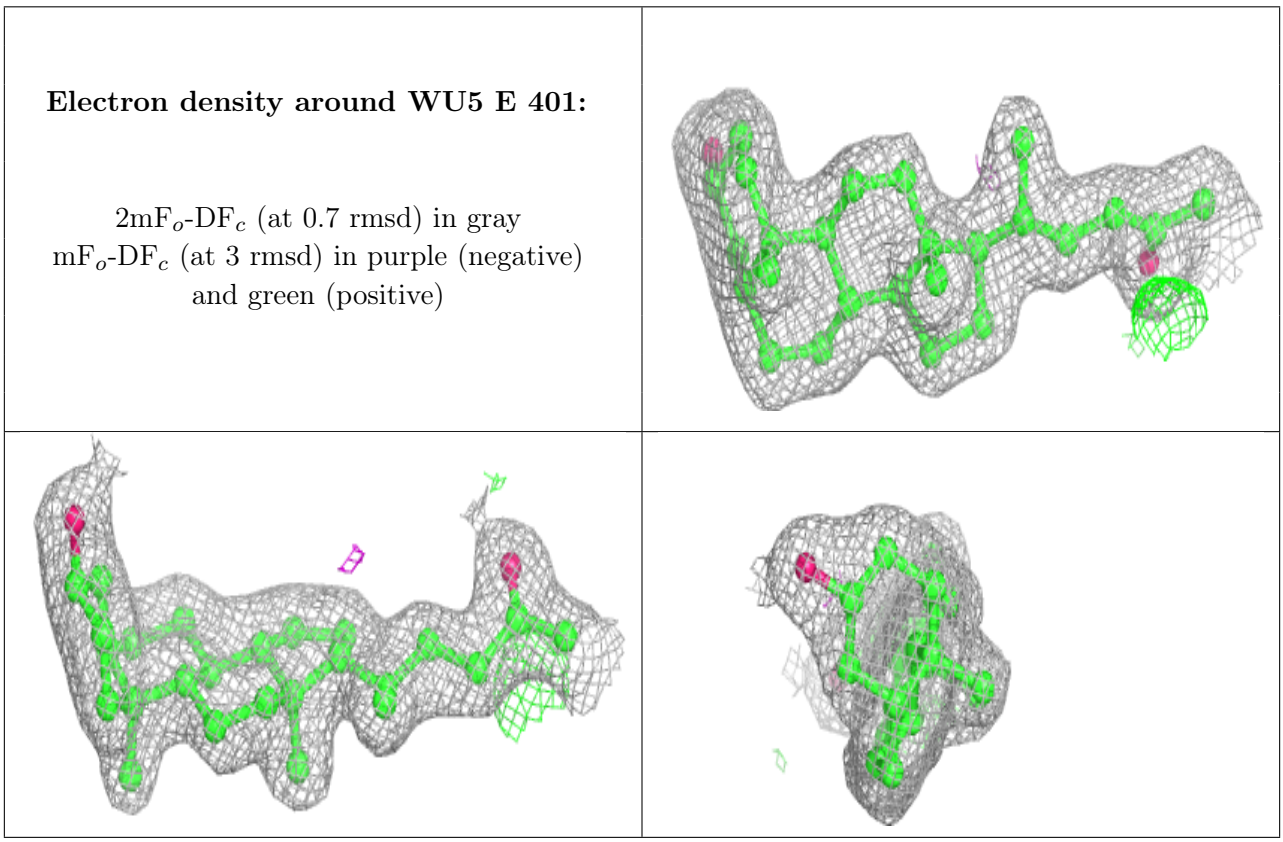
Electron density around WU5 G 401:

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



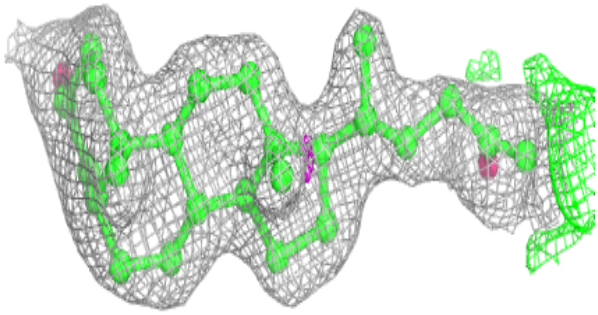
Electron density around WU5 E 401:

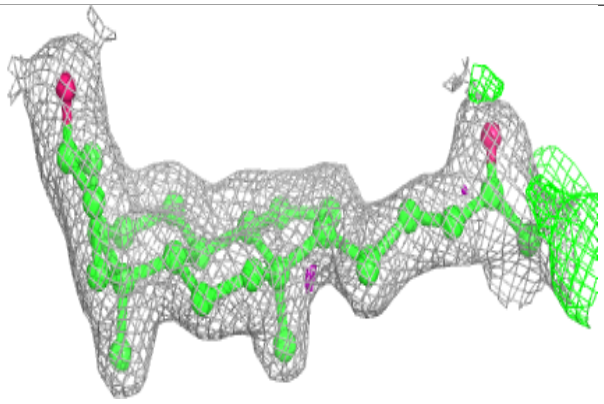
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

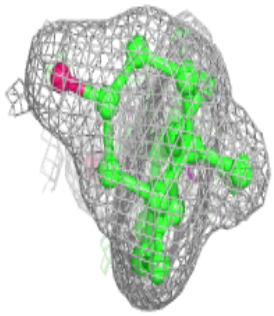


Electron density around WU5 F 401:

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

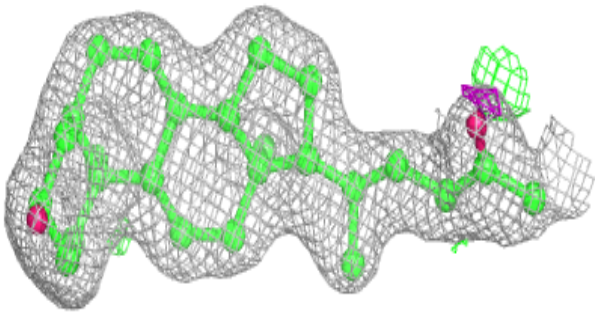


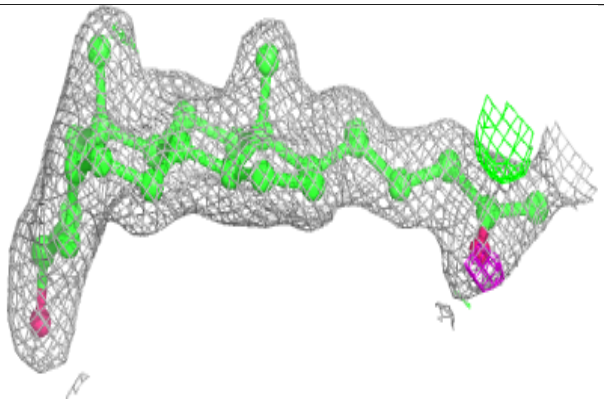


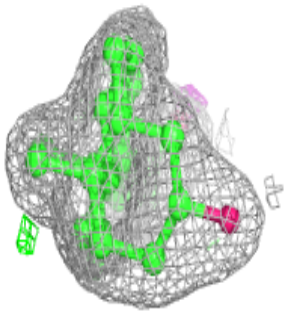


Electron density around WU5 D 401:

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

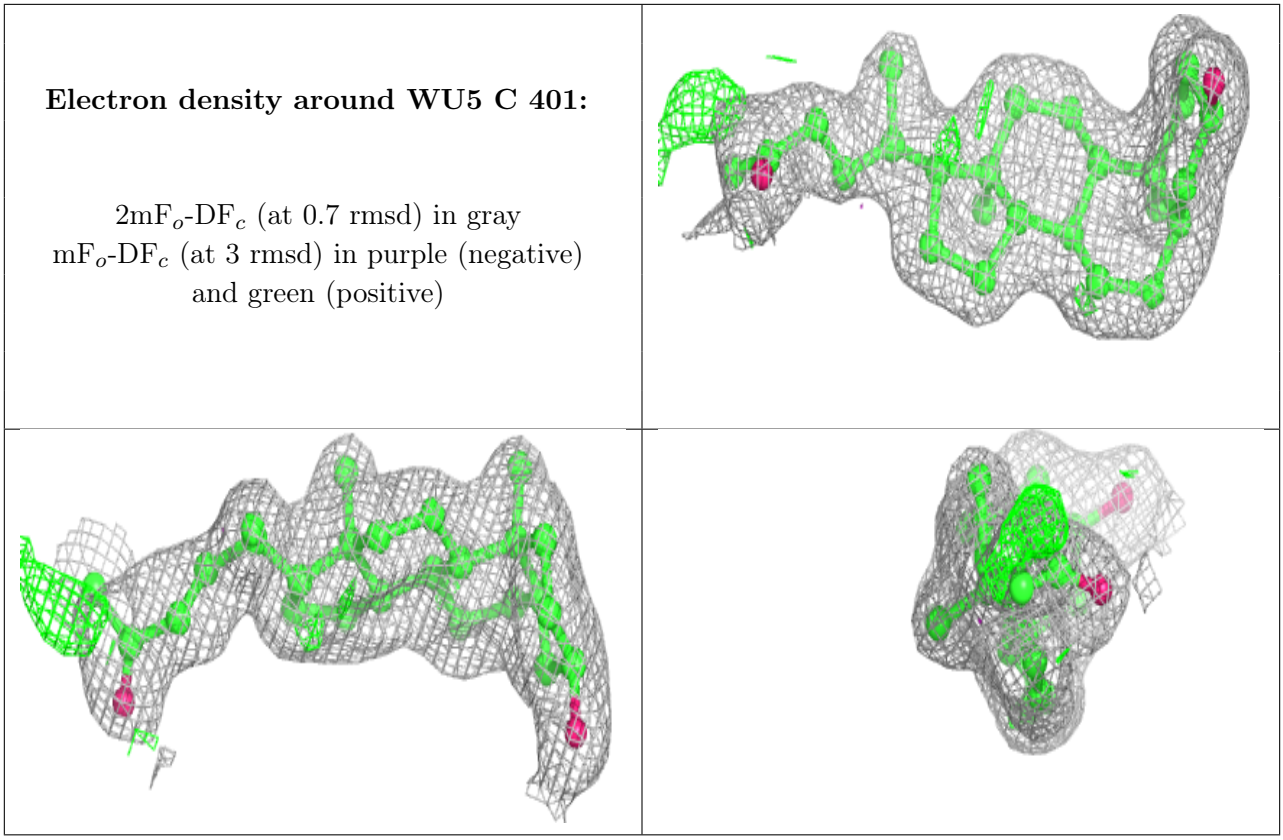






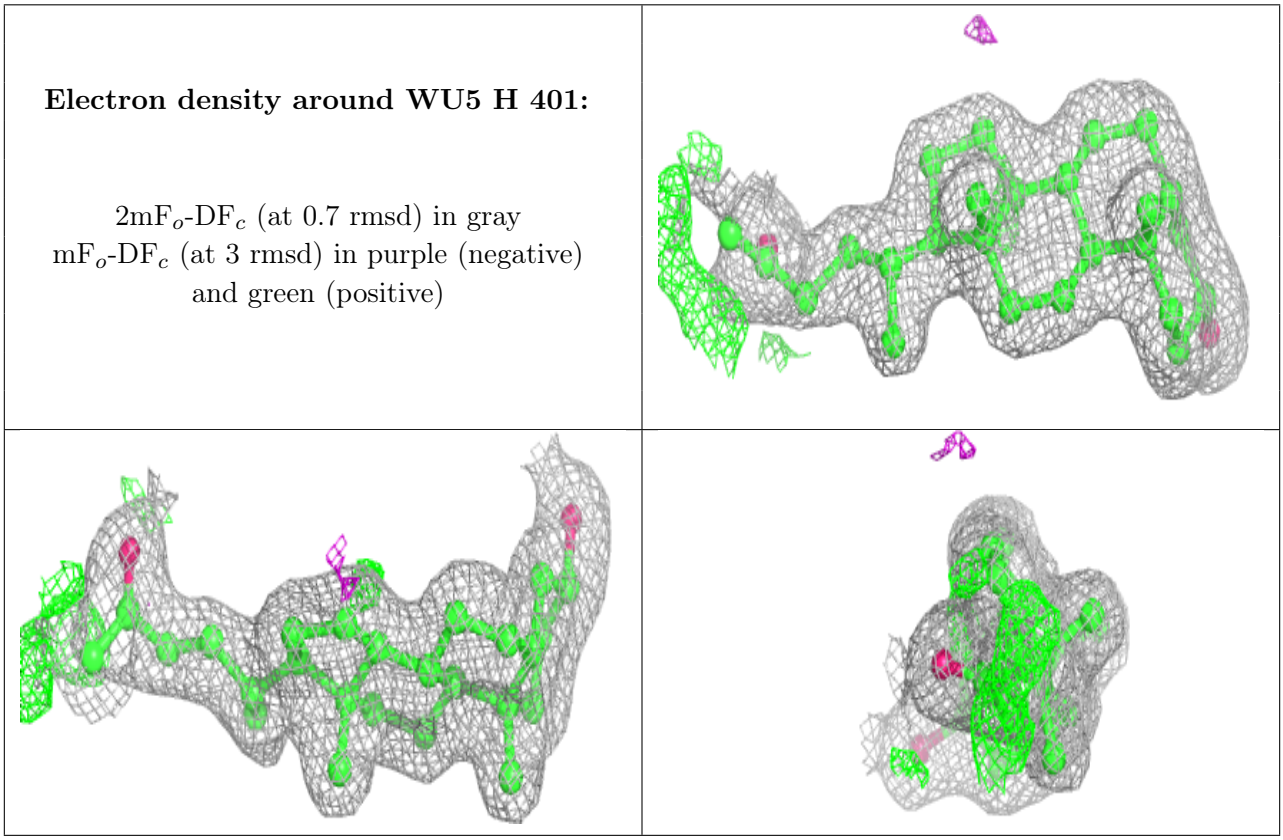
Electron density around WU5 C 401:

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



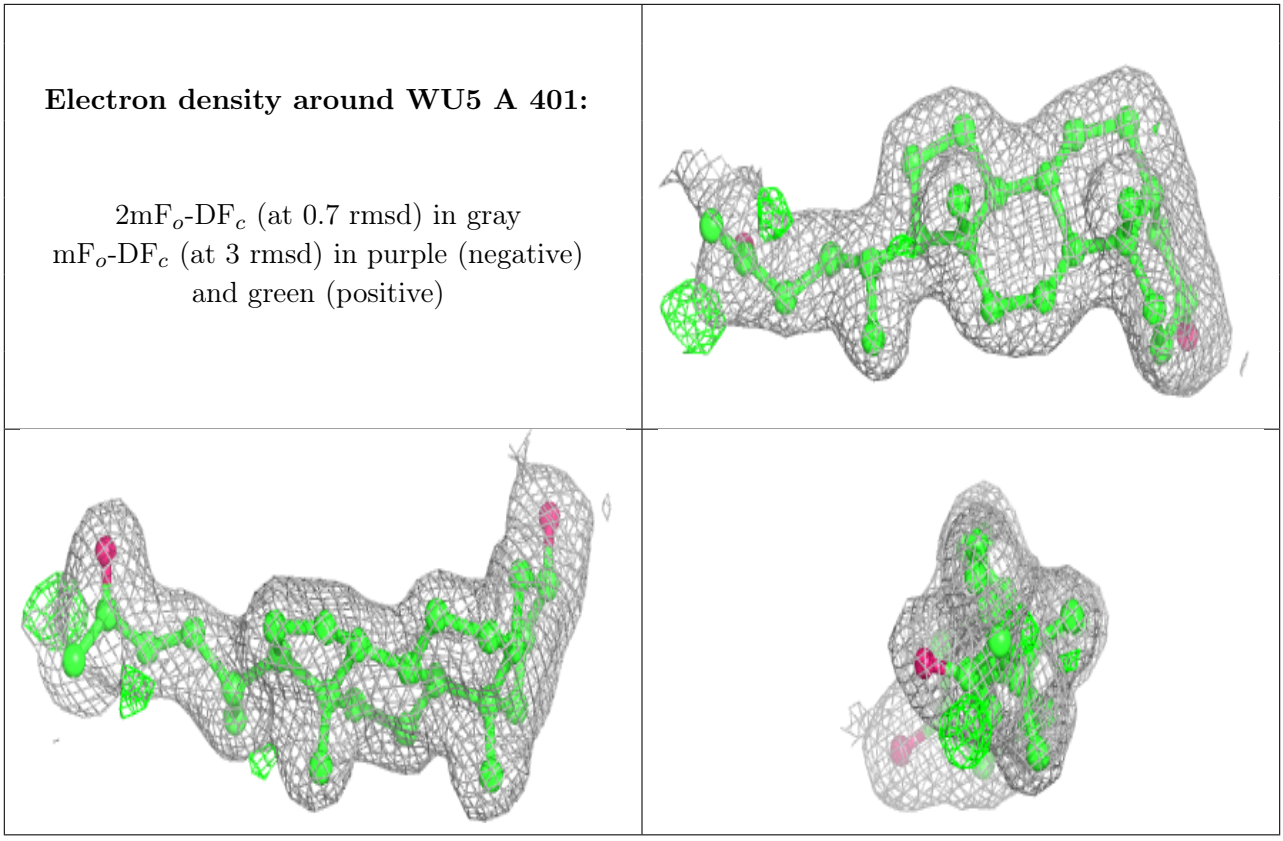
Electron density around WU5 H 401:

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



Electron density around WU5 A 401:

$2mF_o$-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.