



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 06:44 pm BST

PDB ID : 4INT  
Title : Yeast 20S proteasome in complex with the vinyl sulfone LU122  
Authors : Geurink, P.P.; van der Linden, W.A.; Mirabella, A.C.; Gallastegui, N.; de Bruin, G.; Blom, A.E.M.; Voges, M.J.; Mock, E.D.; Florea, B.I.; van der Marel, G.A.; Driessen, C.; van der Stelt, M.; Groll, M.; Overkleeft, H.S.; Kisselev, A.F.  
Deposited on : 2013-01-06  
Resolution : 2.90 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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 : 2.11  
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) : 2.11

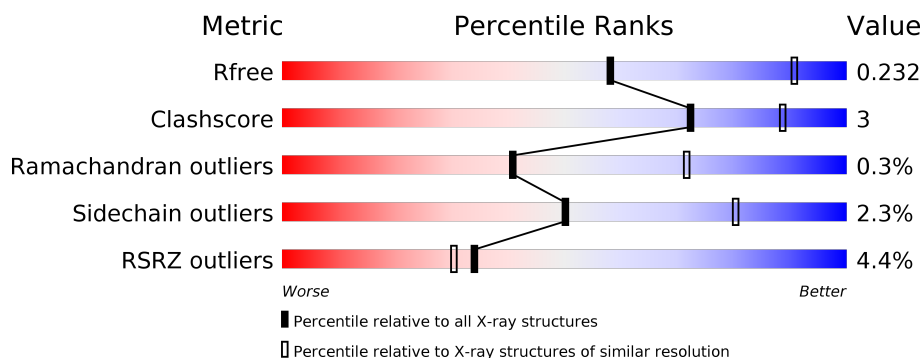
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	250	<div> <div>4%</div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>
1	O	250	<div> <div>3%</div> <div> <div></div> <div>91%</div> <div>8%</div> </div> </div>
2	B	258	<div> <div>7%</div> <div> <div></div> <div>81%</div> <div>14%</div> <div>5%</div> </div> </div>
2	P	258	<div> <div>8%</div> <div> <div></div> <div>81%</div> <div>13%</div> <div>5%</div> </div> </div>
3	C	254	<div> <div>6%</div> <div> <div></div> <div>82%</div> <div>11%</div> <div>5%</div> </div> </div>
3	Q	254	<div> <div>10%</div> <div> <div></div> <div>84%</div> <div>10%</div> <div>5%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	260	
4	R	260	
5	E	234	
5	S	234	
6	F	288	
6	T	288	
7	G	252	
7	U	252	
8	H	232	
8	V	232	
9	I	205	
9	W	205	
10	J	198	
10	X	198	
11	K	212	
11	Y	212	
12	L	222	
12	Z	222	
13	M	233	
13	a	233	
14	N	196	
14	b	196	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
15	1G5	Y	301	-	-	X	-

## 2 Entry composition

There are 16 unique types of molecules in this entry. The entry contains 51033 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 Proteasome component Y7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	0	0	0
			1915	1219	315	377	4			
1	O	250	Total	C	N	O	S	0	0	0
			1915	1219	315	377	4			

- Molecule 2 is a protein called Proteasome component Y13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	244	Total	C	N	O	S	0	0	0
			1904	1201	321	379	3			
2	P	244	Total	C	N	O	S	0	0	0
			1904	1201	321	379	3			

- Molecule 3 is a protein called Proteasome component PRE6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	241	Total	C	N	O	S	0	0	0
			1890	1181	331	374	4			
3	Q	241	Total	C	N	O	S	0	0	0
			1890	1181	331	374	4			

- Molecule 4 is a protein called Proteasome component PUP2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	242	Total	C	N	O	S	0	0	0
			1861	1162	314	378	7			
4	R	242	Total	C	N	O	S	0	0	0
			1861	1162	314	378	7			

- Molecule 5 is a protein called Proteasome component PRE5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	233	Total	C	N	O	S	0	0	0
			1795	1129	312	350	4			
5	S	233	Total	C	N	O	S	0	0	0
			1795	1129	312	350	4			

- Molecule 6 is a protein called Proteasome component C1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	244	Total	C	N	O	S	0	0	0
			1896	1205	330	357	4			
6	T	244	Total	C	N	O	S	0	0	0
			1896	1205	330	357	4			

- Molecule 7 is a protein called Proteasome component C7-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	243	Total	C	N	O	S	0	0	0
			1921	1221	322	370	8			
7	U	243	Total	C	N	O	S	0	0	0
			1921	1221	322	370	8			

- Molecule 8 is a protein called Proteasome component PUP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	222	Total	C	N	O	S	0	0	0
			1684	1061	293	323	7			
8	V	222	Total	C	N	O	S	0	0	0
			1684	1061	293	323	7			

- Molecule 9 is a protein called Proteasome component PUP3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	204	Total	C	N	O	S	0	0	0
			1581	1010	258	305	8			
9	W	204	Total	C	N	O	S	0	0	0
			1581	1010	258	305	8			

- Molecule 10 is a protein called Proteasome component C11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	198	Total	C	N	O	S	0	0	0
			1585	1005	269	305	6			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	X	198	Total	C	N	O	S	0	0	0
			1585	1005	269	305	6			

- Molecule 11 is a protein called Proteasome component PRE2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	212	Total	C	N	O	S	0	0	0
			1644	1045	280	312	7			
11	Y	212	Total	C	N	O	S	0	0	0
			1644	1045	280	312	7			

- Molecule 12 is a protein called Proteasome component C5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	222	Total	C	N	O	S	0	0	0
			1757	1115	303	335	4			
12	Z	222	Total	C	N	O	S	0	0	0
			1757	1115	303	335	4			

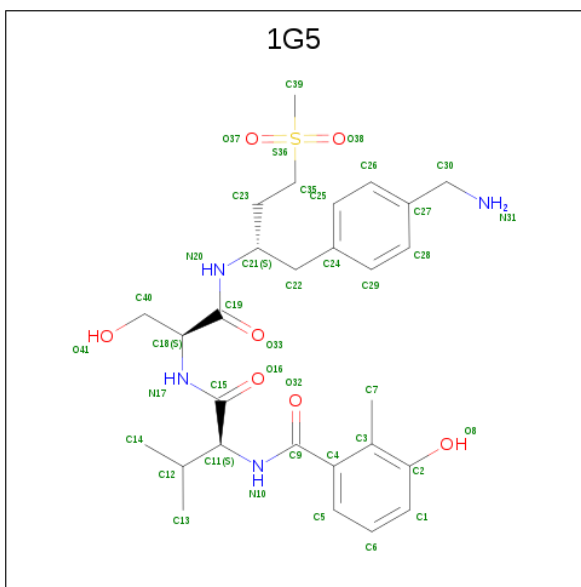
- Molecule 13 is a protein called Proteasome component PRE4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	M	233	Total	C	N	O	S	0	0	0
			1824	1154	312	351	7			
13	a	233	Total	C	N	O	S	0	0	0
			1824	1154	312	351	7			

- Molecule 14 is a protein called Proteasome component PRE3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	N	196	Total	C	N	O	S	0	0	0
			1512	955	250	300	7			
14	b	196	Total	C	N	O	S	0	0	0
			1512	955	250	300	7			

- Molecule 15 is HMB-Val-Ser-Phe(4-NH<sub>2</sub>CH<sub>2</sub>)-methyl vinyl sulfone, bound form (three-letter code: 1G5) (formula: C<sub>28</sub>H<sub>40</sub>N<sub>4</sub>O<sub>7</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
15	H	1	Total	C	N	O	S	0	0
			40	28	4	7	1		
15	K	1	Total	C	N	O	S	0	0
			40	28	4	7	1		
15	V	1	Total	C	N	O	S	0	0
			40	28	4	7	1		
15	Y	1	Total	C	N	O	S	0	0
			40	28	4	7	1		

- Molecule 16 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	A	61	Total	O	0	0
			61	61		
16	B	39	Total	O	0	0
			39	39		
16	C	42	Total	O	0	0
			42	42		
16	D	38	Total	O	0	0
			38	38		
16	E	22	Total	O	0	0
			22	22		
16	F	45	Total	O	0	0
			45	45		
16	G	59	Total	O	0	0
			59	59		
16	H	58	Total	O	0	0
			58	58		

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	I	63	Total 63	O 63	0	0
16	J	50	Total 50	O 50	0	0
16	K	47	Total 47	O 47	0	0
16	L	58	Total 58	O 58	0	0
16	M	74	Total 74	O 74	0	0
16	N	54	Total 54	O 54	0	0
16	O	33	Total 33	O 33	0	0
16	P	29	Total 29	O 29	0	0
16	Q	29	Total 29	O 29	0	0
16	R	33	Total 33	O 33	0	0
16	S	20	Total 20	O 20	0	0
16	T	41	Total 41	O 41	0	0
16	U	56	Total 56	O 56	0	0
16	V	50	Total 50	O 50	0	0
16	W	55	Total 55	O 55	0	0
16	X	47	Total 47	O 47	0	0
16	Y	47	Total 47	O 47	0	0
16	Z	49	Total 49	O 49	0	0
16	a	79	Total 79	O 79	0	0
16	b	57	Total 57	O 57	0	0

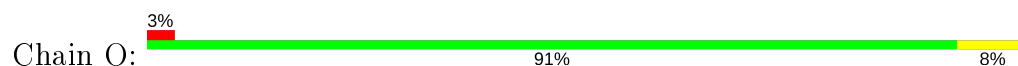
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

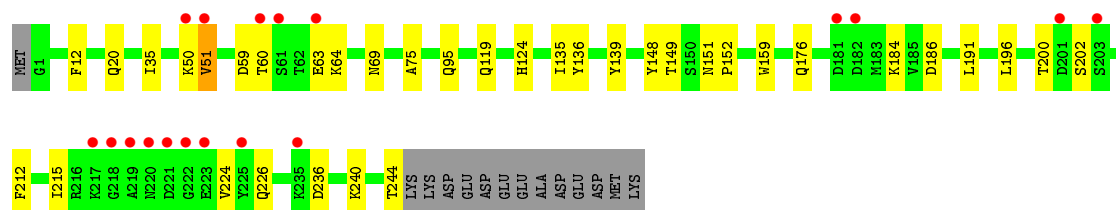
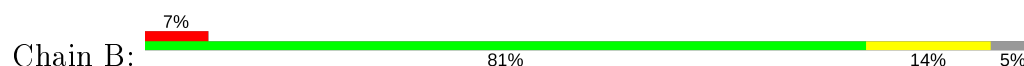
- Molecule 1: Proteasome component Y7



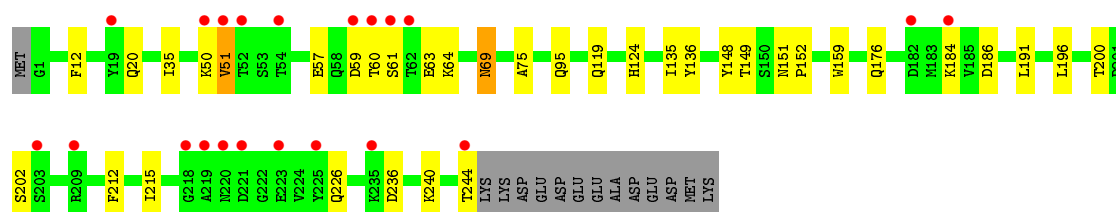
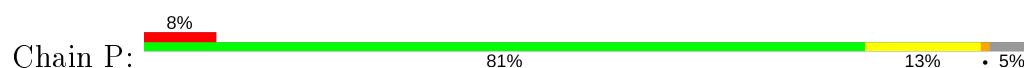
- Molecule 1: Proteasome component Y7



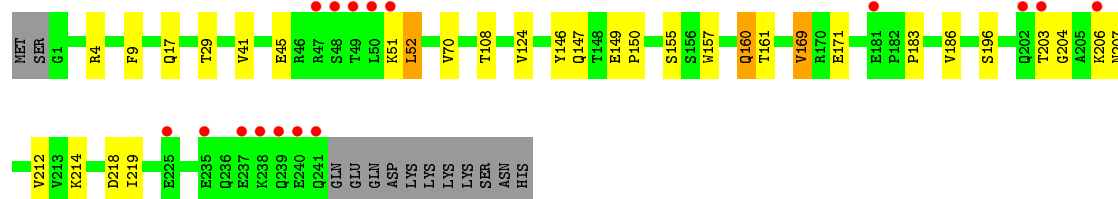
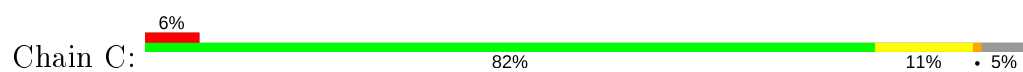
- Molecule 2: Proteasome component Y13



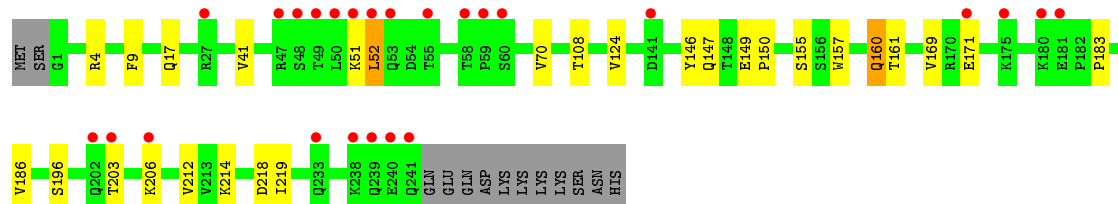
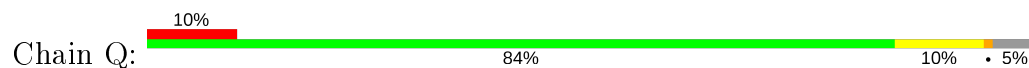
- Molecule 2: Proteasome component Y13



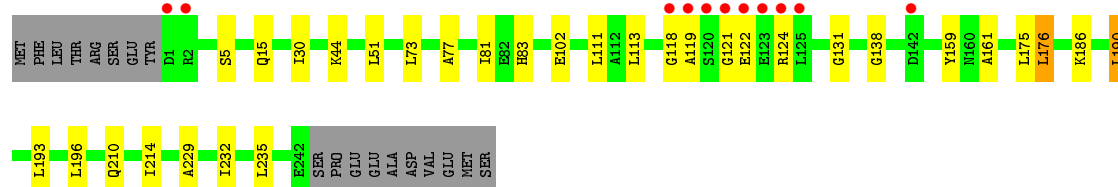
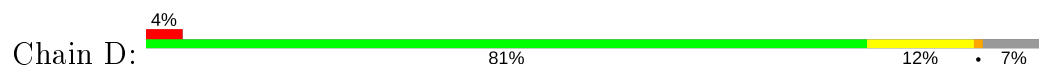
- Molecule 3: Proteasome component PRE6



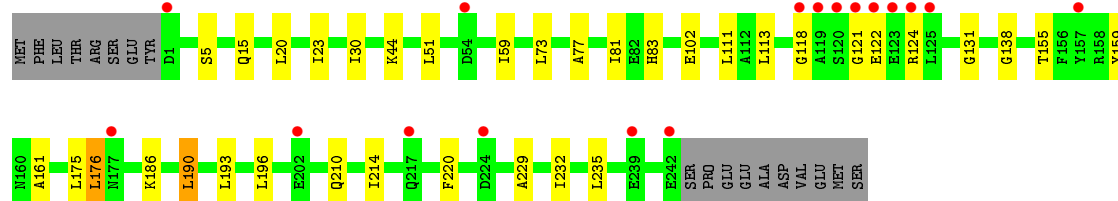
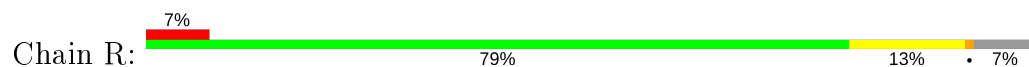
- Molecule 3: Proteasome component PRE6



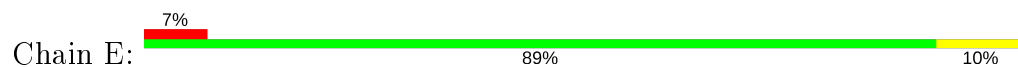
- Molecule 4: Proteasome component PUP2

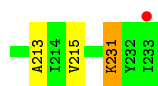


- Molecule 4: Proteasome component PUP2

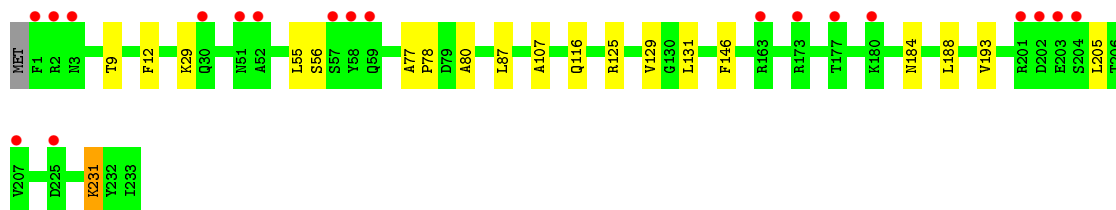
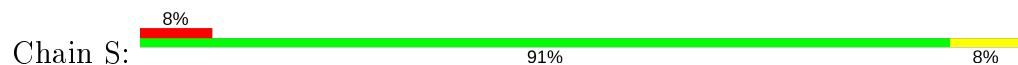


- Molecule 5: Proteasome component PRE5

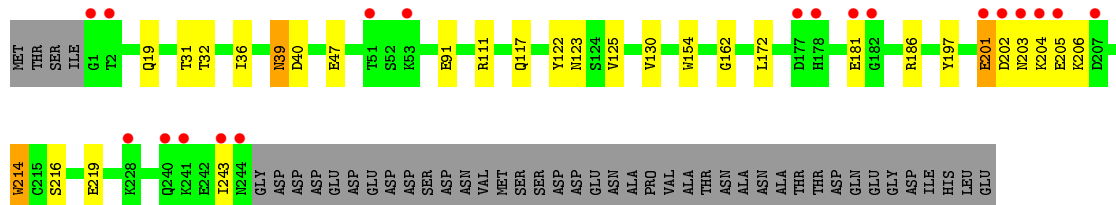
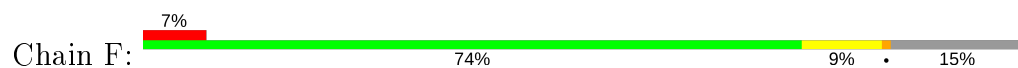




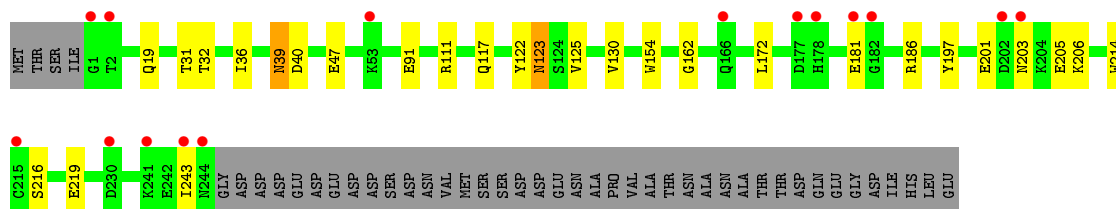
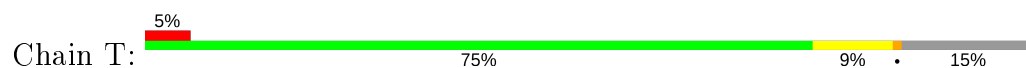
- Molecule 5: Proteasome component PRE5



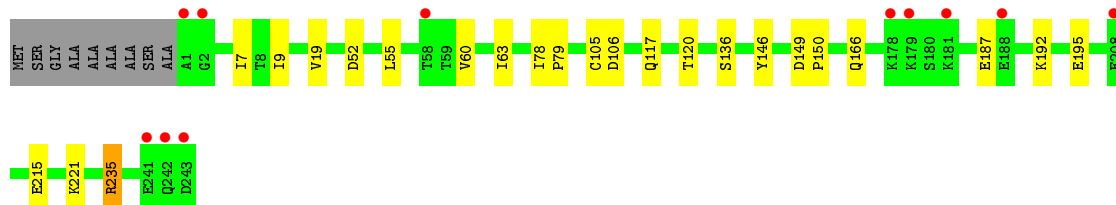
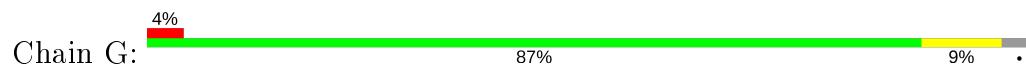
- Molecule 6: Proteasome component C1



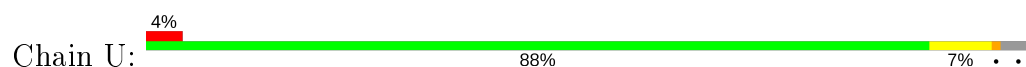
- Molecule 6: Proteasome component C1



- Molecule 7: Proteasome component C7-alpha

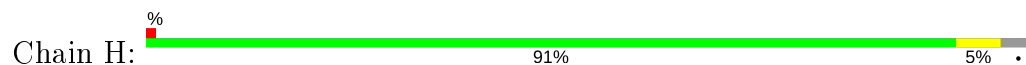


- Molecule 7: Proteasome component C7-alpha

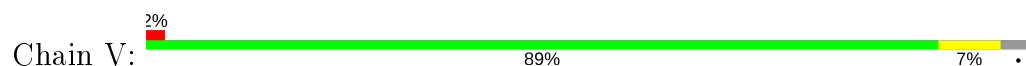




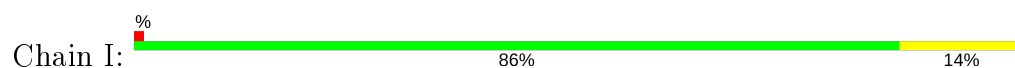
- Molecule 8: Proteasome component PUP1



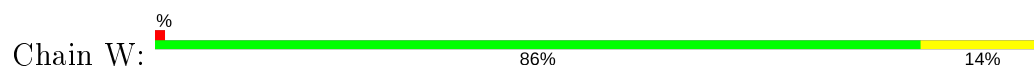
- Molecule 8: Proteasome component PUP1



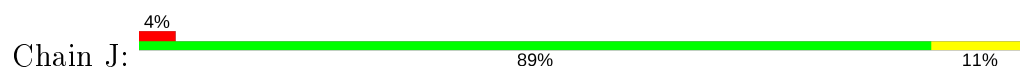
- Molecule 9: Proteasome component PUP3



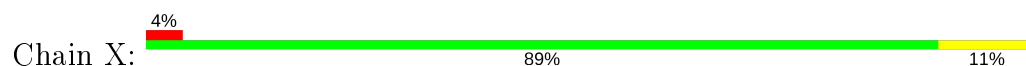
- Molecule 9: Proteasome component PUP3



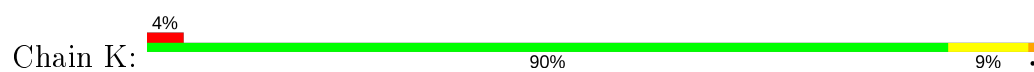
- Molecule 10: Proteasome component C11



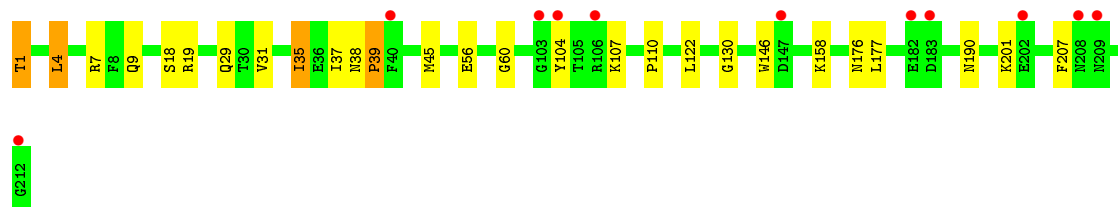
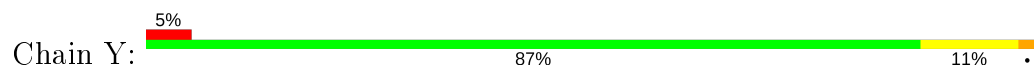
- Molecule 10: Proteasome component C11



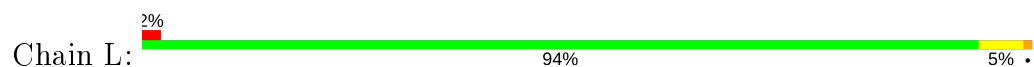
- Molecule 11: Proteasome component PRE2



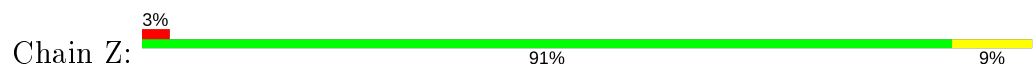
- Molecule 11: Proteasome component PRE2



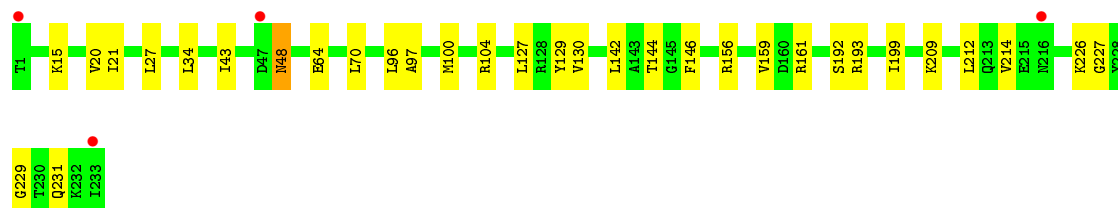
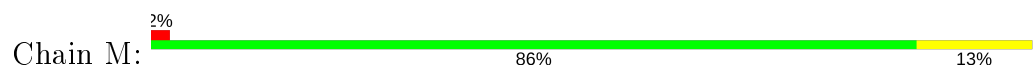
- Molecule 12: Proteasome component C5



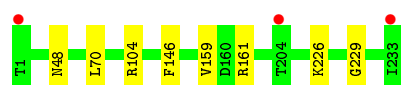
- Molecule 12: Proteasome component C5



- Molecule 13: Proteasome component PRE4

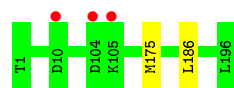


- Molecule 13: Proteasome component PRE4



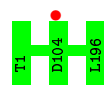
- Molecule 14: Proteasome component PRE3

Chain N:  99%



- Molecule 14: Proteasome component PRE3

Chain b:  100%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	137.01Å 299.51Å 145.04Å 90.00° 113.24° 90.00°	Depositor
Resolution (Å)	15.00 – 2.90 15.00 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (15.00-2.90) 98.9 (15.00-2.90)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	50.01 (at 2.91Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.222 , 0.230 0.223 , 0.232	Depositor DCC
$R_{free}$ test set	11610 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	64.9	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 33.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	51033	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.25% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: 1G5

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	0.36	0/1952	0.46	0/2642
1	O	0.36	0/1952	0.46	0/2642
2	B	0.33	0/1934	0.46	0/2618
2	P	0.33	0/1934	0.46	0/2618
3	C	0.34	0/1919	0.47	0/2598
3	Q	0.34	0/1919	0.47	0/2598
4	D	0.35	0/1886	0.48	0/2541
4	R	0.36	0/1886	0.48	0/2541
5	E	0.31	0/1823	0.45	0/2463
5	S	0.31	0/1823	0.45	0/2463
6	F	0.41	1/1936 (0.1%)	0.45	0/2614
6	T	0.41	0/1936	0.45	0/2614
7	G	0.35	0/1959	0.46	0/2652
7	U	0.34	0/1959	0.46	0/2652
8	H	0.44	0/1715	0.47	0/2326
8	V	0.44	0/1715	0.46	0/2326
9	I	0.34	0/1611	0.47	0/2174
9	W	0.34	0/1611	0.47	0/2174
10	J	0.31	0/1613	0.45	0/2173
10	X	0.31	0/1613	0.45	0/2173
11	K	0.50	1/1681 (0.1%)	0.49	1/2274 (0.0%)
11	Y	0.50	1/1681 (0.1%)	0.49	1/2274 (0.0%)
12	L	0.36	0/1795	0.45	0/2420
12	Z	0.36	0/1795	0.45	0/2420
13	M	0.36	0/1855	0.47	0/2514
13	a	0.36	0/1855	0.47	0/2514
14	N	0.39	0/1541	0.44	0/2087
14	b	0.39	0/1541	0.44	0/2087
All	All	0.37	3/50440 (0.0%)	0.46	2/68192 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	K	146	TRP	CD2-CE2	5.09	1.47	1.41
11	Y	146	TRP	CD2-CE2	5.05	1.47	1.41
6	F	214	TRP	CD2-CE2	5.02	1.47	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	Y	4	LEU	CA-CB-CG	5.21	127.28	115.30
11	K	4	LEU	CA-CB-CG	5.16	127.16	115.30

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1915	0	1929	10	0
1	O	1915	0	1929	11	0
2	B	1904	0	1904	17	0
2	P	1904	0	1904	18	0
3	C	1890	0	1903	17	0
3	Q	1890	0	1903	15	0
4	D	1861	0	1839	19	0
4	R	1861	0	1839	21	0
5	E	1795	0	1800	15	0
5	S	1795	0	1800	12	0
6	F	1896	0	1889	14	0
6	T	1896	0	1889	14	0
7	G	1921	0	1913	11	0
7	U	1921	0	1913	10	0
8	H	1684	0	1687	7	0
8	V	1684	0	1687	10	0
9	I	1581	0	1574	17	0
9	W	1581	0	1574	15	0
10	J	1585	0	1590	13	0
10	X	1585	0	1590	14	0
11	K	1644	0	1594	9	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
11	Y	1644	0	1594	19	0
12	L	1757	0	1711	7	0
12	Z	1757	0	1711	10	0
13	M	1824	0	1832	16	0
13	a	1824	0	1832	0	0
14	N	1512	0	1481	1	0
14	b	1512	0	1481	0	0
15	H	40	0	39	1	0
15	K	40	0	39	0	0
15	V	40	0	39	3	0
15	Y	40	0	39	24	0
16	A	61	0	0	0	0
16	B	39	0	0	0	0
16	C	42	0	0	0	0
16	D	38	0	0	0	0
16	E	22	0	0	0	0
16	F	45	0	0	0	0
16	G	59	0	0	0	0
16	H	58	0	0	0	0
16	I	63	0	0	0	0
16	J	50	0	0	1	0
16	K	47	0	0	0	0
16	L	58	0	0	0	0
16	M	74	0	0	0	0
16	N	54	0	0	0	0
16	O	33	0	0	0	0
16	P	29	0	0	0	0
16	Q	29	0	0	0	0
16	R	33	0	0	0	0
16	S	20	0	0	0	0
16	T	41	0	0	0	0
16	U	56	0	0	0	0
16	V	50	0	0	0	0
16	W	55	0	0	0	0
16	X	47	0	0	1	0
16	Y	47	0	0	0	0
16	Z	49	0	0	0	0
16	a	79	0	0	0	0
16	b	57	0	0	0	0
All	All	51033	0	49448	318	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 3.

The worst 5 of 318 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)
15:Y:301:1G5:H20	15:Y:301:1G5:C39	1.68	1.22
15:Y:301:1G5:C40	15:Y:301:1G5:C39	2.23	1.16
15:Y:301:1G5:H20	15:Y:301:1G5:H30	1.36	1.05
15:Y:301:1G5:H20	15:Y:301:1G5:H29	1.52	0.89
15:Y:301:1G5:C39	15:Y:301:1G5:H19	2.11	0.79

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	248/250 (99%)	240 (97%)	7 (3%)	1 (0%)	34	66
1	O	248/250 (99%)	240 (97%)	7 (3%)	1 (0%)	34	66
2	B	242/258 (94%)	235 (97%)	6 (2%)	1 (0%)	34	66
2	P	242/258 (94%)	235 (97%)	6 (2%)	1 (0%)	34	66
3	C	239/254 (94%)	231 (97%)	5 (2%)	3 (1%)	12	37
3	Q	239/254 (94%)	230 (96%)	6 (2%)	3 (1%)	12	37
4	D	240/260 (92%)	233 (97%)	4 (2%)	3 (1%)	12	37
4	R	240/260 (92%)	233 (97%)	4 (2%)	3 (1%)	12	37
5	E	231/234 (99%)	225 (97%)	6 (3%)	0	100	100
5	S	231/234 (99%)	224 (97%)	7 (3%)	0	100	100
6	F	242/288 (84%)	231 (96%)	11 (4%)	0	100	100
6	T	242/288 (84%)	232 (96%)	10 (4%)	0	100	100
7	G	241/252 (96%)	235 (98%)	6 (2%)	0	100	100
7	U	241/252 (96%)	235 (98%)	6 (2%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
8	H	220/232 (95%)	213 (97%)	7 (3%)	0	100	100
8	V	220/232 (95%)	212 (96%)	8 (4%)	0	100	100
9	I	202/205 (98%)	194 (96%)	8 (4%)	0	100	100
9	W	202/205 (98%)	194 (96%)	8 (4%)	0	100	100
10	J	196/198 (99%)	188 (96%)	7 (4%)	1 (0%)	29	61
10	X	196/198 (99%)	188 (96%)	7 (4%)	1 (0%)	29	61
11	K	210/212 (99%)	204 (97%)	5 (2%)	1 (0%)	29	61
11	Y	210/212 (99%)	203 (97%)	6 (3%)	1 (0%)	29	61
12	L	220/222 (99%)	215 (98%)	5 (2%)	0	100	100
12	Z	220/222 (99%)	215 (98%)	5 (2%)	0	100	100
13	M	231/233 (99%)	219 (95%)	11 (5%)	1 (0%)	34	66
13	a	231/233 (99%)	221 (96%)	9 (4%)	1 (0%)	34	66
14	N	194/196 (99%)	188 (97%)	6 (3%)	0	100	100
14	b	194/196 (99%)	188 (97%)	6 (3%)	0	100	100
All	All	6312/6588 (96%)	6101 (97%)	189 (3%)	22 (0%)	41	71

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	52	LEU
4	D	121	GLY
3	Q	52	LEU
4	R	121	GLY
1	A	2	THR

### 5.3.2 Protein sidechains ⓘ

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	209/209 (100%)	206 (99%)	3 (1%)	67	89
1	O	209/209 (100%)	206 (99%)	3 (1%)	67	89

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	203/216 (94%)	194 (96%)	9 (4%)	28	61
2	P	203/216 (94%)	194 (96%)	9 (4%)	28	61
3	C	213/226 (94%)	206 (97%)	7 (3%)	38	72
3	Q	213/226 (94%)	207 (97%)	6 (3%)	43	76
4	D	198/215 (92%)	193 (98%)	5 (2%)	47	78
4	R	198/215 (92%)	193 (98%)	5 (2%)	47	78
5	E	192/193 (100%)	186 (97%)	6 (3%)	40	74
5	S	192/193 (100%)	186 (97%)	6 (3%)	40	74
6	F	201/239 (84%)	191 (95%)	10 (5%)	24	57
6	T	201/239 (84%)	192 (96%)	9 (4%)	27	61
7	G	207/210 (99%)	203 (98%)	4 (2%)	57	84
7	U	207/210 (99%)	203 (98%)	4 (2%)	57	84
8	H	181/190 (95%)	179 (99%)	2 (1%)	73	92
8	V	181/190 (95%)	178 (98%)	3 (2%)	60	86
9	I	172/173 (99%)	170 (99%)	2 (1%)	71	91
9	W	172/173 (99%)	169 (98%)	3 (2%)	60	86
10	J	175/175 (100%)	174 (99%)	1 (1%)	86	96
10	X	175/175 (100%)	174 (99%)	1 (1%)	86	96
11	K	169/169 (100%)	165 (98%)	4 (2%)	49	79
11	Y	169/169 (100%)	164 (97%)	5 (3%)	41	75
12	L	185/185 (100%)	183 (99%)	2 (1%)	73	92
12	Z	185/185 (100%)	184 (100%)	1 (0%)	88	96
13	M	199/199 (100%)	192 (96%)	7 (4%)	36	70
13	a	199/199 (100%)	192 (96%)	7 (4%)	36	70
14	N	162/162 (100%)	162 (100%)	0	100	100
14	b	162/162 (100%)	162 (100%)	0	100	100
All	All	5332/5522 (97%)	5208 (98%)	124 (2%)	50	80

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
13	M	70	LEU
2	P	184	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
11	Y	35	ILE
13	M	104	ARG
1	O	30	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 132 such sidechains are listed below:

Mol	Chain	Res	Type
13	M	18	ASN
2	P	155	ASN
12	Z	49	ASN
13	M	102	GLN
1	O	143	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

4 ligands are 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	$\# Z  > 2$
15	1G5	Y	301	11	41,41,41	3.37	6 (14%)	53,57,57	5.98	17 (32%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
15	1G5	V	301	8	41,41,41	1.24	3 (7%)	53,57,57	1.89	12 (22%)
15	1G5	K	301	11	41,41,41	1.22	3 (7%)	53,57,57	1.82	10 (18%)
15	1G5	H	301	8	41,41,41	1.26	3 (7%)	53,57,57	1.93	11 (20%)

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
15	1G5	Y	301	11	-	15/42/42/42	0/2/2/2
15	1G5	V	301	8	-	13/42/42/42	0/2/2/2
15	1G5	K	301	11	-	16/42/42/42	0/2/2/2
15	1G5	H	301	8	-	14/42/42/42	0/2/2/2

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	Y	301	1G5	O37-S36	15.96	1.79	1.44
15	Y	301	1G5	O38-S36	8.66	1.63	1.44
15	Y	301	1G5	C39-S36	-8.10	1.45	1.75
15	Y	301	1G5	C23-C35	4.48	1.57	1.52
15	H	301	1G5	C23-C35	4.44	1.57	1.52

The worst 5 of 50 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	Y	301	1G5	O38-S36-C39	26.58	135.62	108.91
15	Y	301	1G5	O38-S36-C35	-22.34	92.70	108.34
15	Y	301	1G5	O37-S36-O38	-16.56	81.68	117.09
15	Y	301	1G5	O37-S36-C35	11.56	116.44	108.34
15	Y	301	1G5	O37-S36-C39	-9.42	99.44	108.91

There are no chirality outliers.

5 of 58 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
15	Y	301	1G5	N17-C18-C40-O41
15	Y	301	1G5	C23-C35-S36-O38
15	Y	301	1G5	C22-C21-C23-C35

*Continued on next page...*



*Continued from previous page...*

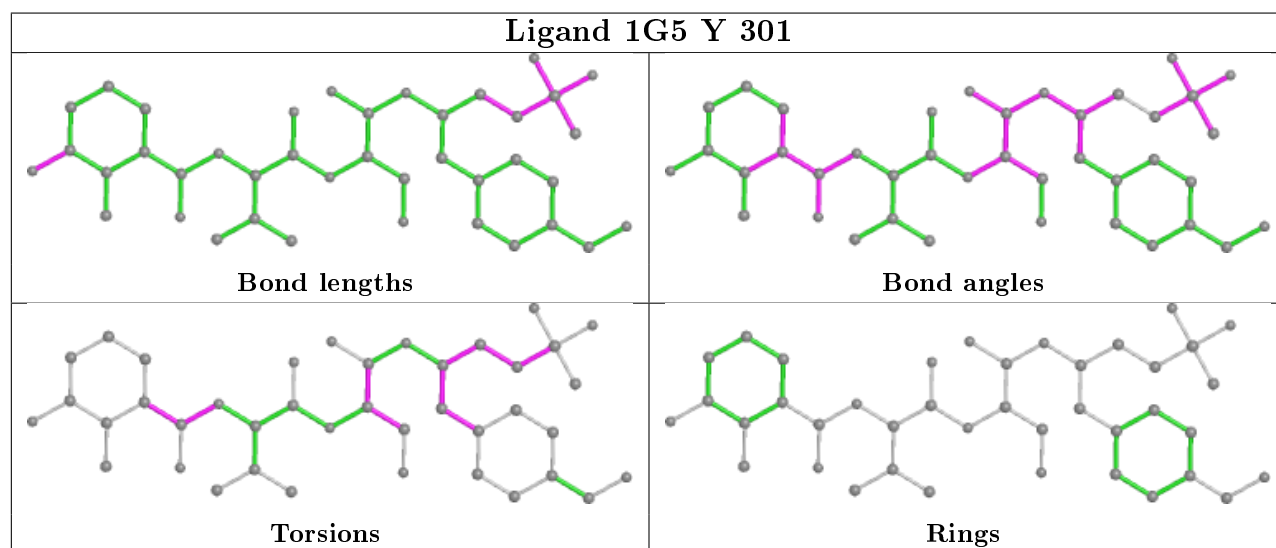
Mol	Chain	Res	Type	Atoms
15	Y	301	1G5	N20-C21-C23-C35
15	Y	301	1G5	C23-C21-C22-C24

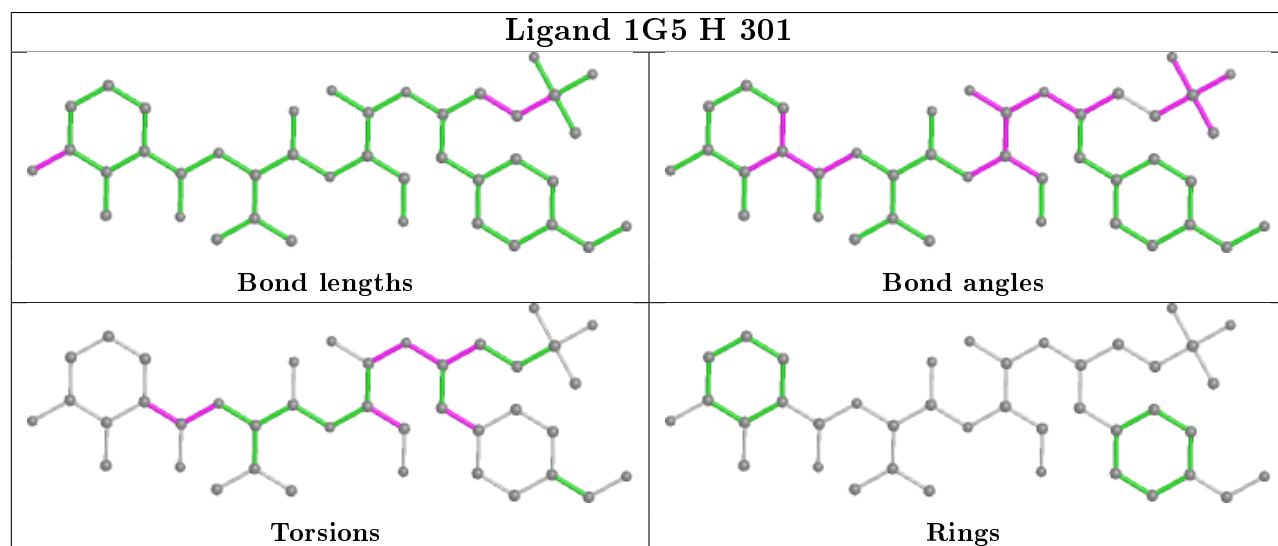
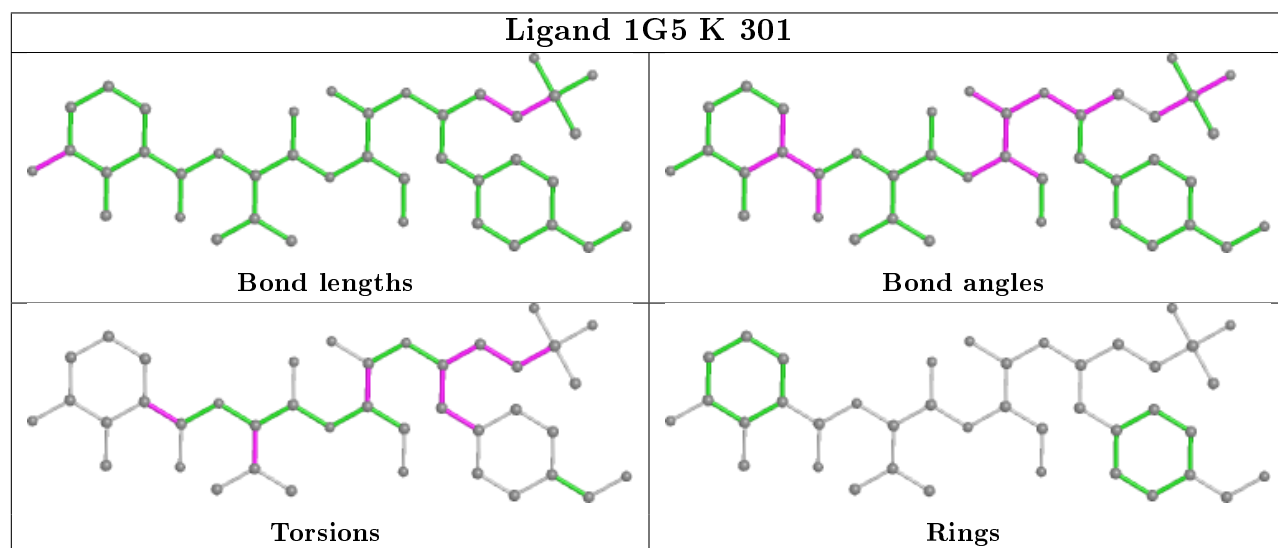
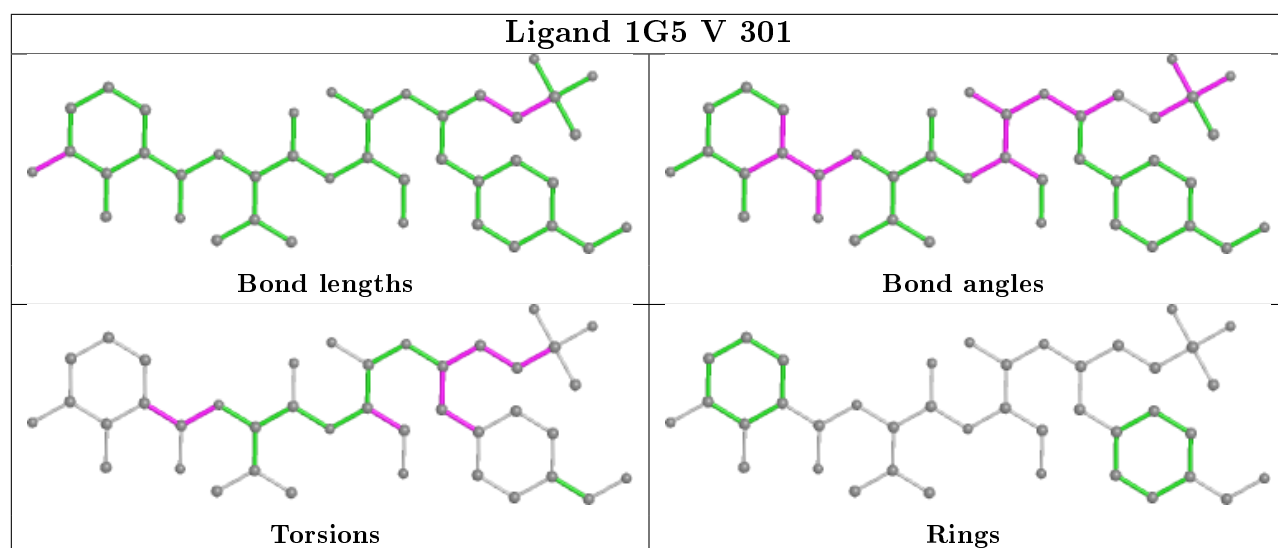
There are no ring outliers.

3 monomers are involved in 28 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
15	Y	301	1G5	24	0
15	V	301	1G5	3	0
15	H	301	1G5	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 than 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.





## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	250/250 (100%)	0.00	9 (3%) 42 37	59, 74, 97, 114	0
1	O	250/250 (100%)	-0.03	8 (3%) 47 43	56, 72, 99, 116	0
2	B	244/258 (94%)	0.22	18 (7%) 14 11	54, 73, 113, 122	0
2	P	244/258 (94%)	0.28	21 (8%) 10 8	57, 76, 109, 124	0
3	C	241/254 (94%)	0.14	16 (6%) 18 14	55, 73, 112, 151	0
3	Q	241/254 (94%)	0.36	25 (10%) 6 5	65, 91, 137, 179	0
4	D	242/260 (93%)	0.12	11 (4%) 33 29	59, 78, 106, 127	0
4	R	242/260 (93%)	0.21	17 (7%) 16 12	62, 83, 113, 135	0
5	E	233/234 (99%)	0.15	16 (6%) 16 13	67, 86, 110, 119	0
5	S	233/234 (99%)	0.21	19 (8%) 11 9	66, 89, 120, 133	0
6	F	244/288 (84%)	0.13	19 (7%) 13 10	61, 79, 110, 130	0
6	T	244/288 (84%)	0.10	15 (6%) 21 17	58, 78, 116, 142	0
7	G	243/252 (96%)	0.11	11 (4%) 33 29	60, 77, 104, 137	0
7	U	243/252 (96%)	0.01	9 (3%) 41 37	55, 70, 95, 122	0
8	H	222/232 (95%)	-0.09	3 (1%) 75 75	57, 70, 85, 98	0
8	V	222/232 (95%)	-0.12	5 (2%) 60 58	54, 67, 84, 102	0
9	I	204/205 (99%)	-0.32	2 (0%) 82 82	52, 63, 80, 86	0
9	W	204/205 (99%)	-0.28	2 (0%) 82 82	54, 63, 81, 88	0
10	J	198/198 (100%)	-0.04	8 (4%) 38 33	51, 63, 81, 119	0
10	X	198/198 (100%)	-0.04	8 (4%) 38 33	55, 66, 84, 120	0
11	K	212/212 (100%)	-0.08	8 (3%) 40 36	50, 64, 83, 90	0
11	Y	212/212 (100%)	-0.01	11 (5%) 27 23	53, 65, 86, 96	0
12	L	222/222 (100%)	-0.16	5 (2%) 60 58	52, 68, 91, 99	0
12	Z	222/222 (100%)	-0.12	6 (2%) 54 50	55, 69, 92, 103	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
13	M	233/233 (100%)	-0.21	4 (1%) 70 69	54, 67, 81, 86	0
13	a	233/233 (100%)	-0.22	3 (1%) 77 77	52, 67, 83, 85	0
14	N	196/196 (100%)	-0.20	3 (1%) 73 73	58, 65, 81, 91	0
14	b	196/196 (100%)	-0.21	1 (0%) 91 91	54, 63, 80, 91	0
All	All	6368/6588 (96%)	0.01	283 (4%) 34 30	50, 72, 108, 179	0

The worst 5 of 283 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	R	119	ALA	10.8
2	P	220	ASN	10.5
10	J	198	GLN	10.5
2	B	220	ASN	9.0
4	D	119	ALA	8.9

## 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 carbohydrates 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<sup>th</sup> 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.

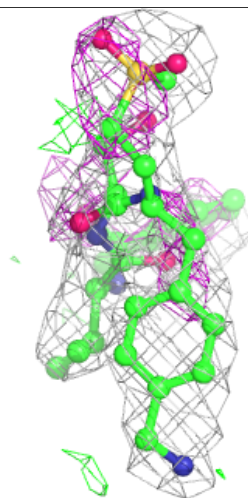
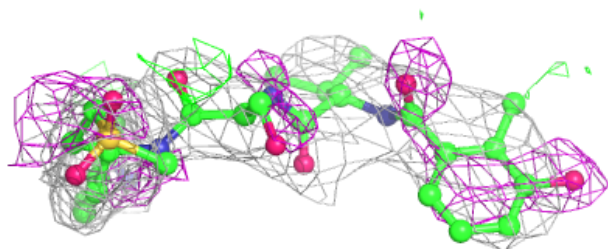
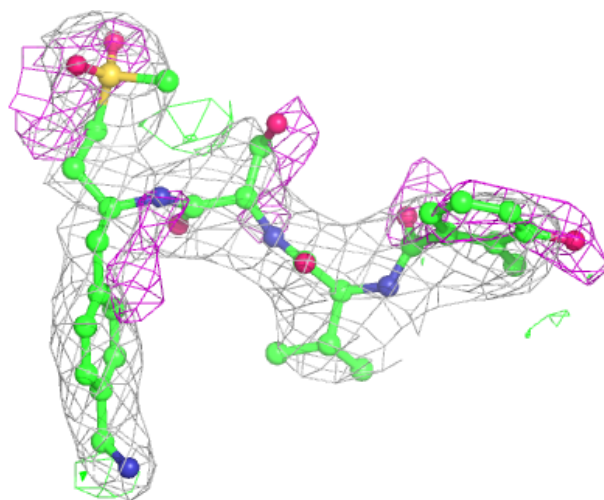
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
15	1G5	V	301	40/40	0.87	0.28	54,59,67,68	0
15	1G5	Y	301	40/40	0.88	0.25	48,50,53,53	0
15	1G5	K	301	40/40	0.89	0.25	47,48,52,53	0
15	1G5	H	301	40/40	0.89	0.26	57,61,66,66	0

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.

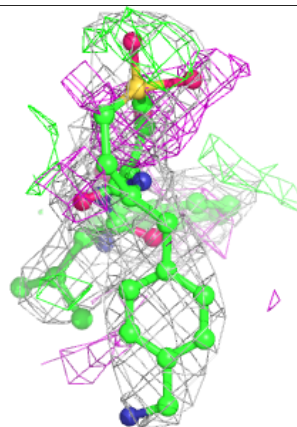
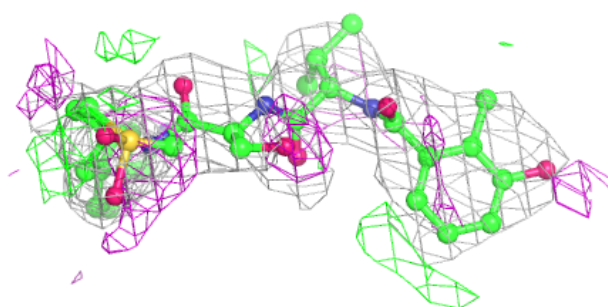
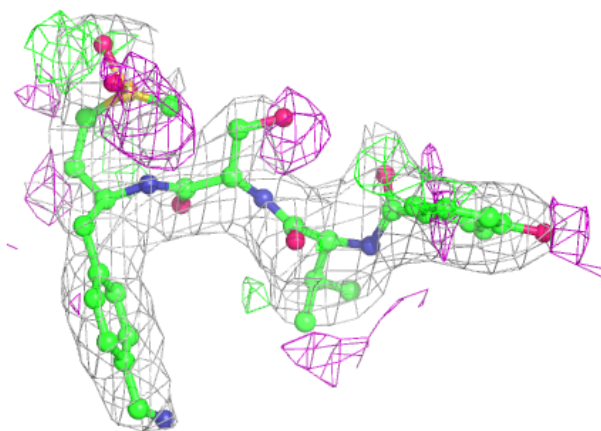
**Electron density around 1G5 V 301:**

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



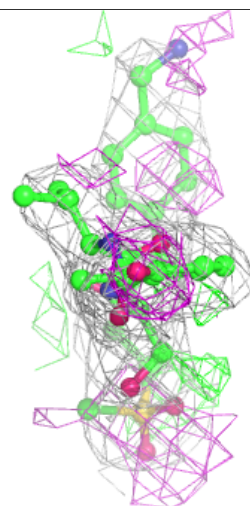
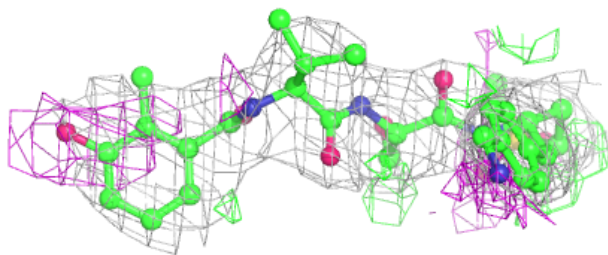
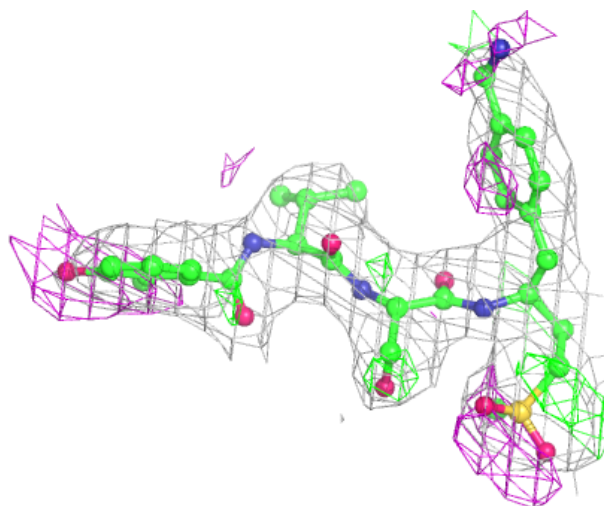
**Electron density around 1G5 Y 301:**

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



**Electron density around 1G5 K 301:**

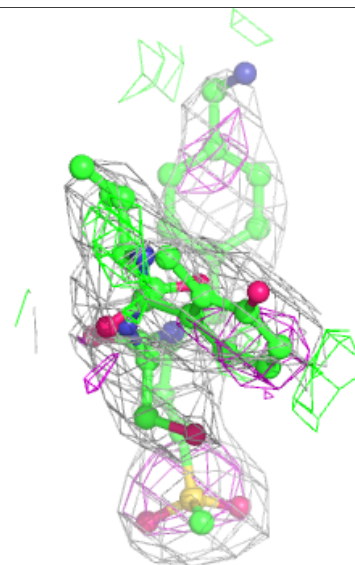
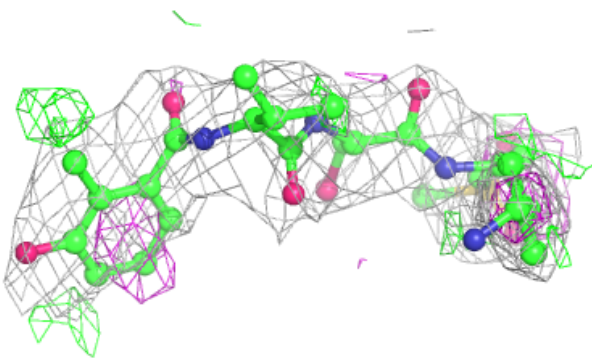
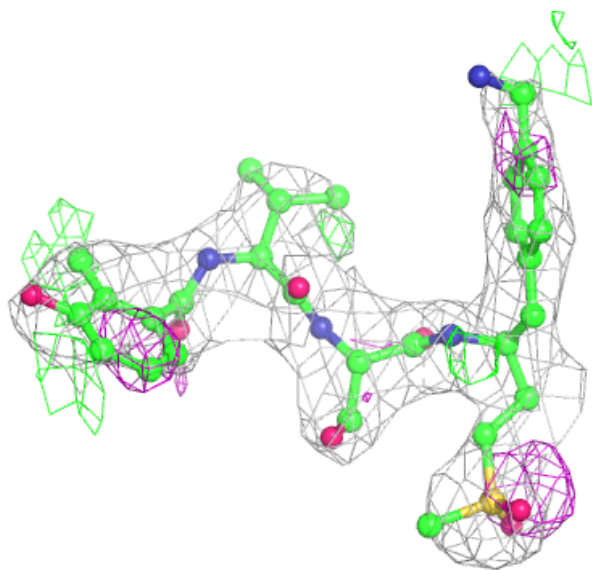
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)





**Electron density around 1G5 H 301:**

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



## 6.5 Other polymers ⓘ

There are no such residues in this entry.