



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 18, 2024 – 10:28 AM JST

PDB ID : 5X2B  
Title : Crystal structure of mouse sulfotransferase SULT7A1 complexed with PAP  
Authors : Kanekiyo, M.; Teramoto, T.; Kakuta, Y.  
Deposited on : 2017-01-31  
Resolution : 2.08 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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

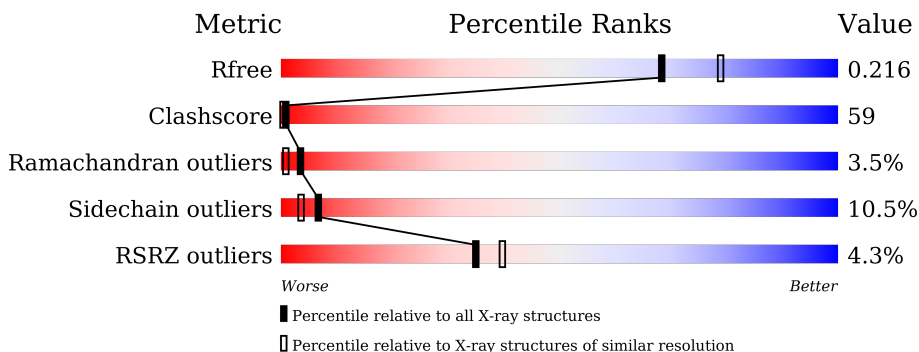
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.08 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	6189 (2.10-2.06)
Clashscore	141614	6738 (2.10-2.06)
Ramachandran outliers	138981	6663 (2.10-2.06)
Sidechain outliers	138945	6664 (2.10-2.06)
RSRZ outliers	127900	6057 (2.10-2.06)

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  $\geq 3$ , 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	283	 4% 29% 62% 7%
1	B	283	 2% 35% 54% 8%
1	C	283	 6% 28% 58% 12%
1	E	283	 4% 33% 54% 10%
1	I	283	 4% 35% 52% 11%
1	K	283	 4% 41% 49% 9%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	L	283	
2	D	281	
3	F	282	
3	G	282	
3	H	282	
3	J	282	

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
4	A3P	A	301	-	-	X	-
4	A3P	D	301	-	-	X	-
4	A3P	G	301	-	-	X	-
4	A3P	J	301	-	-	X	-
4	A3P	L	301	-	-	X	-

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 29402 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 Sulfotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	C	282	2313	1485	384	431	13	0	4	0
1	A	281	2293	1472	383	426	12	0	1	0
1	B	275	2272	1463	376	422	11	0	3	0
1	E	277	2264	1456	378	418	12	0	1	0
1	I	277	2261	1453	377	419	12	0	1	0
1	K	281	2316	1487	388	429	12	0	4	0
1	L	283	2327	1494	387	433	13	0	5	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	6	ALA	-	expression tag	UNP B7ZWN4
C	263	GLY	VAL	conflict	UNP B7ZWN4
C	288	ALA	-	expression tag	UNP B7ZWN4
A	6	ALA	-	expression tag	UNP B7ZWN4
A	263	GLY	VAL	conflict	UNP B7ZWN4
A	288	ALA	-	expression tag	UNP B7ZWN4
B	6	ALA	-	expression tag	UNP B7ZWN4
B	263	GLY	VAL	conflict	UNP B7ZWN4
B	288	ALA	-	expression tag	UNP B7ZWN4
E	6	ALA	-	expression tag	UNP B7ZWN4
E	263	GLY	VAL	conflict	UNP B7ZWN4
E	288	ALA	-	expression tag	UNP B7ZWN4
I	6	ALA	-	expression tag	UNP B7ZWN4
I	263	GLY	VAL	conflict	UNP B7ZWN4
I	288	ALA	-	expression tag	UNP B7ZWN4

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
K	6	ALA	-	expression tag	UNP B7ZWN4
K	263	GLY	VAL	conflict	UNP B7ZWN4
K	288	ALA	-	expression tag	UNP B7ZWN4
L	6	ALA	-	expression tag	UNP B7ZWN4
L	263	GLY	VAL	conflict	UNP B7ZWN4
L	288	ALA	-	expression tag	UNP B7ZWN4

- Molecule 2 is a protein called Sulfotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	276	2266	1455	378	421	12	0	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	8	ALA	-	expression tag	UNP B7ZWN4
D	9	ILE	-	expression tag	UNP B7ZWN4
D	263	GLY	VAL	conflict	UNP B7ZWN4
D	288	ALA	-	expression tag	UNP B7ZWN4

- Molecule 3 is a protein called Sulfotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	F	279	2277	1465	378	423	11	0	2	0
3	G	276	2264	1454	377	421	12	0	2	0
3	H	278	2265	1457	377	420	11	0	1	0
3	J	274	2249	1448	373	417	11	0	2	0

There are 8 discrepancies between the modelled and reference sequences:

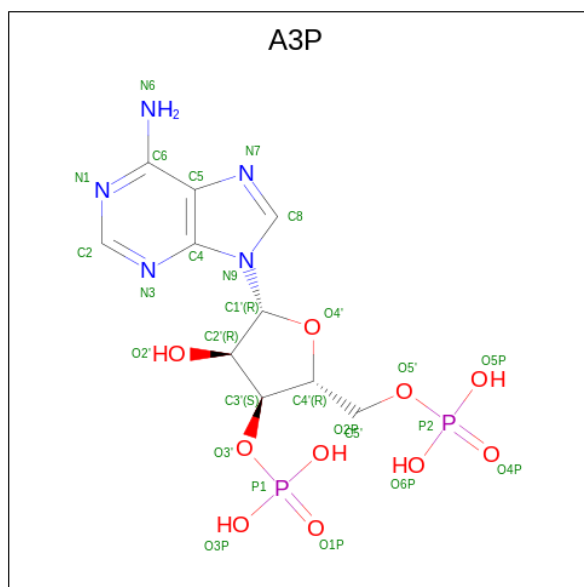
Chain	Residue	Modelled	Actual	Comment	Reference
F	263	GLY	VAL	conflict	UNP B7ZWN4
F	288	ALA	-	expression tag	UNP B7ZWN4
G	263	GLY	VAL	conflict	UNP B7ZWN4
G	288	ALA	-	expression tag	UNP B7ZWN4
H	263	GLY	VAL	conflict	UNP B7ZWN4
H	288	ALA	-	expression tag	UNP B7ZWN4

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
J	263	GLY	VAL	conflict	UNP B7ZWN4
J	288	ALA	-	expression tag	UNP B7ZWN4

- Molecule 4 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	C	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	D	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	E	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	F	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	G	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	H	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	I	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	J	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	K	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	L	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total	Ca	0	0
			1	1		
5	A	2	Total	Ca	0	0
			2	2		
5	B	3	Total	Ca	0	0
			3	3		
5	D	3	Total	Ca	0	0
			3	3		
5	E	1	Total	Ca	0	0
			1	1		
5	F	1	Total	Ca	0	0
			1	1		
5	G	2	Total	Ca	0	0
			2	2		
5	H	1	Total	Ca	0	0
			1	1		
5	I	1	Total	Ca	0	0
			1	1		
5	J	2	Total	Ca	0	0
			2	2		
5	K	2	Total	Ca	0	0
			2	2		
5	L	1	Total	Ca	0	0
			1	1		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	F	1	Total C O 6 3 3	0	0
6	J	1	Total C O 6 3 3	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	126	Total O 126 126	0	0
7	A	149	Total O 149 149	0	0
7	B	151	Total O 151 151	0	0
7	D	169	Total O 169 169	0	0
7	E	136	Total O 136 136	0	0
7	F	108	Total O 108 108	0	0
7	G	172	Total O 172 172	0	0
7	H	131	Total O 131 131	0	0
7	I	131	Total O 131 131	0	0
7	J	157	Total O 157 157	0	0

*Continued on next page...*



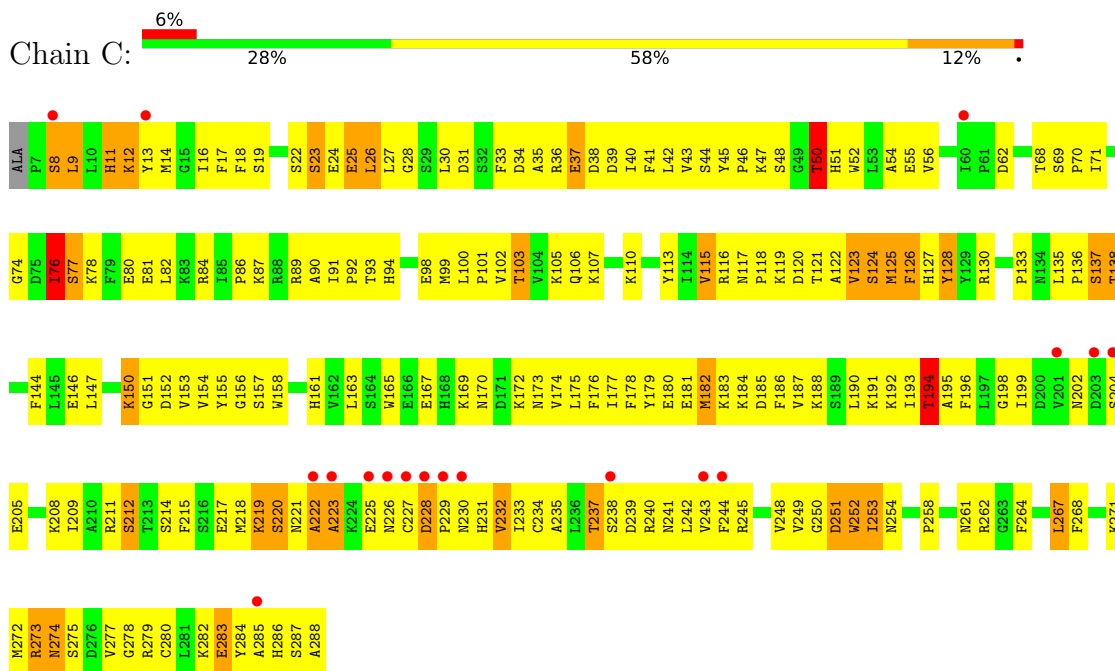
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
7	K	151	Total 151	O 151	0	0
7	L	98	Total 98	O 98	0	0

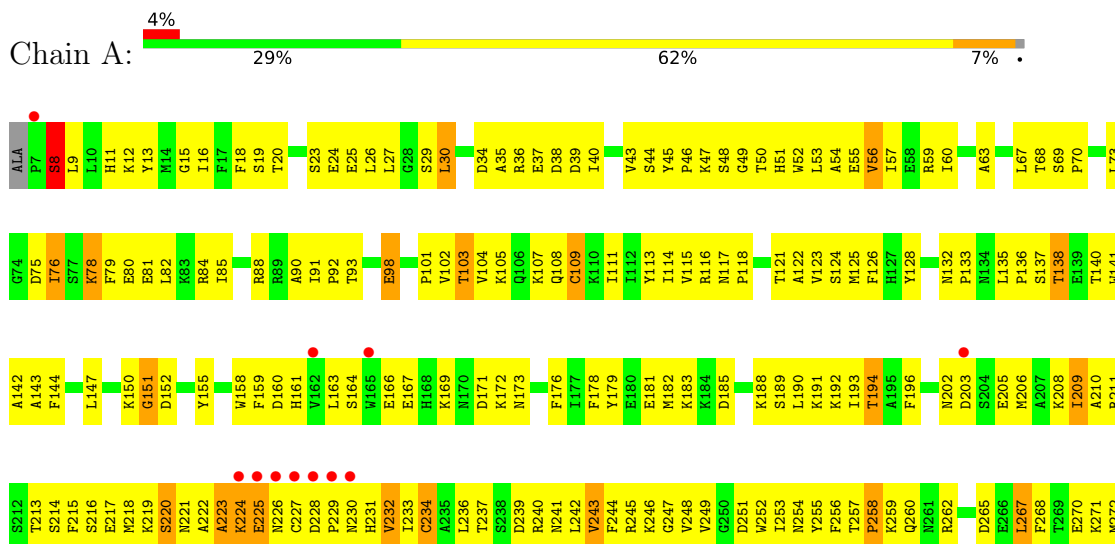
### 3 Residue-property plots

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: Sulfotransferase

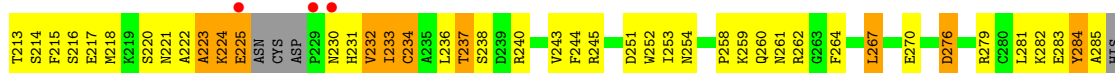
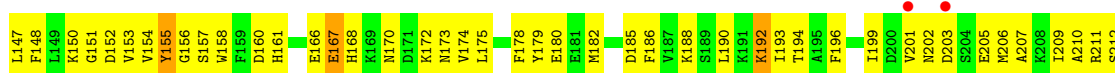
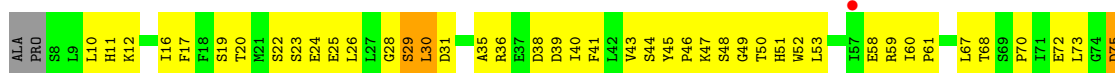


#### • Molecule 1: Sulfotransferase

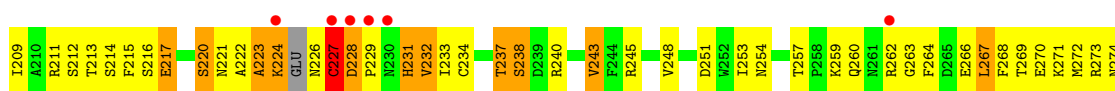
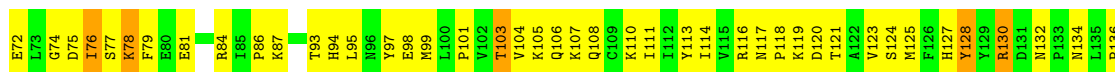
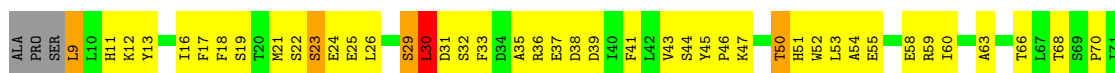




• Molecule 1: Sulfotransferase

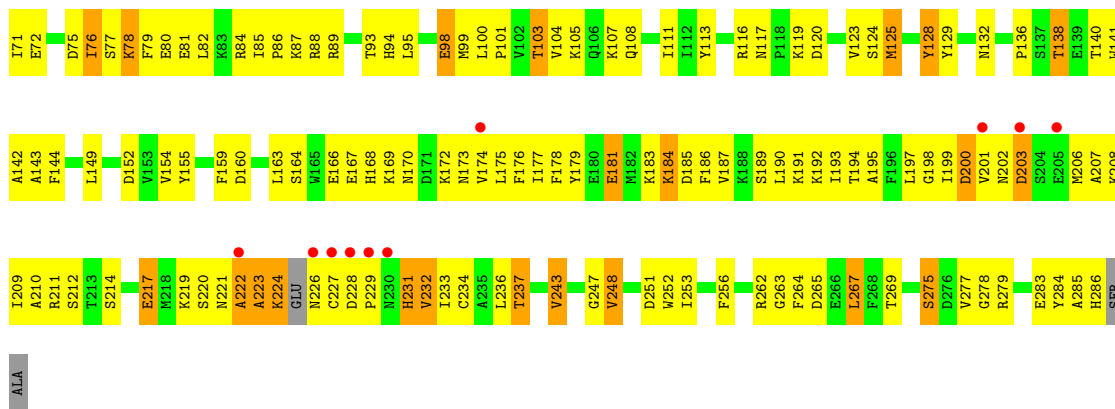


• Molecule 1: Sulfotransferase

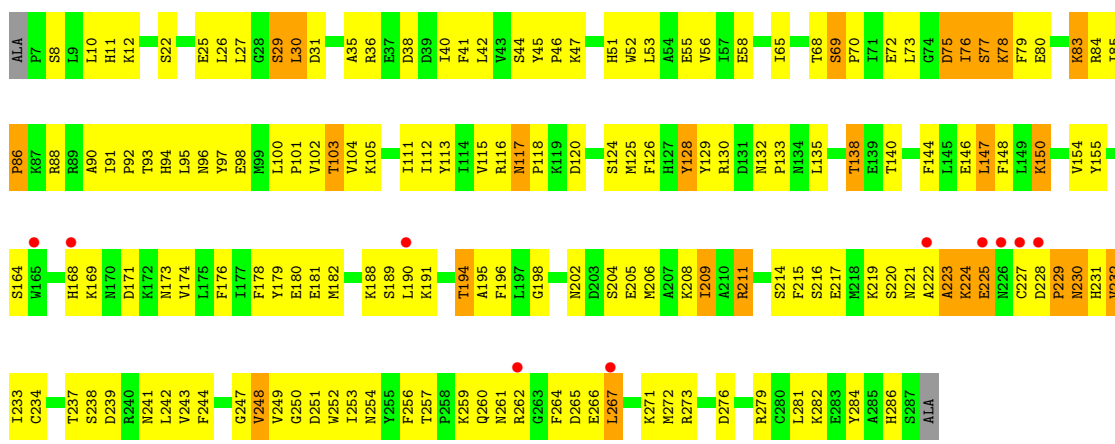


• Molecule 1: Sulfotransferase

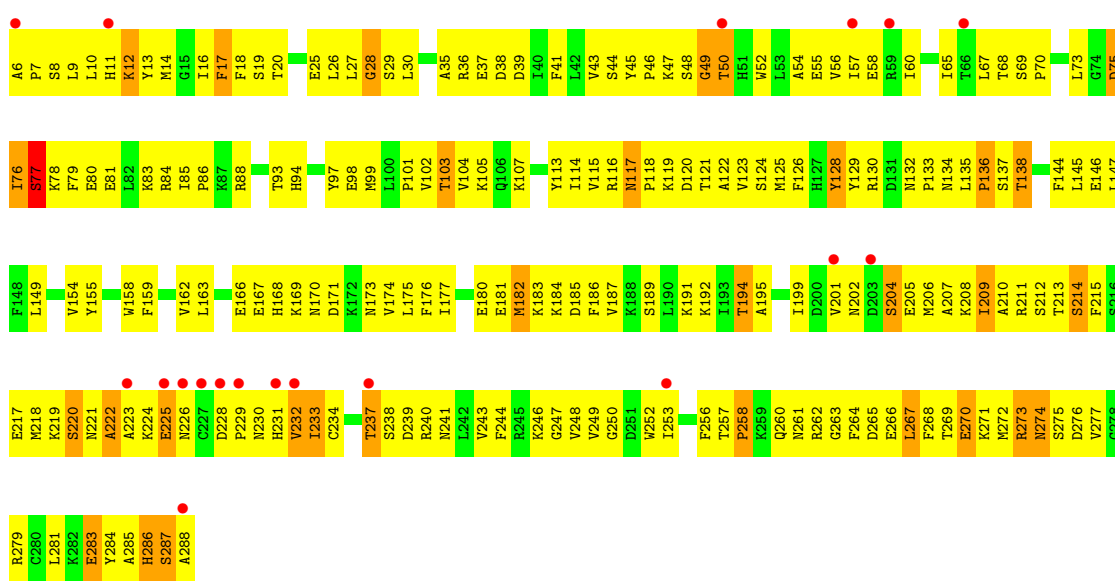




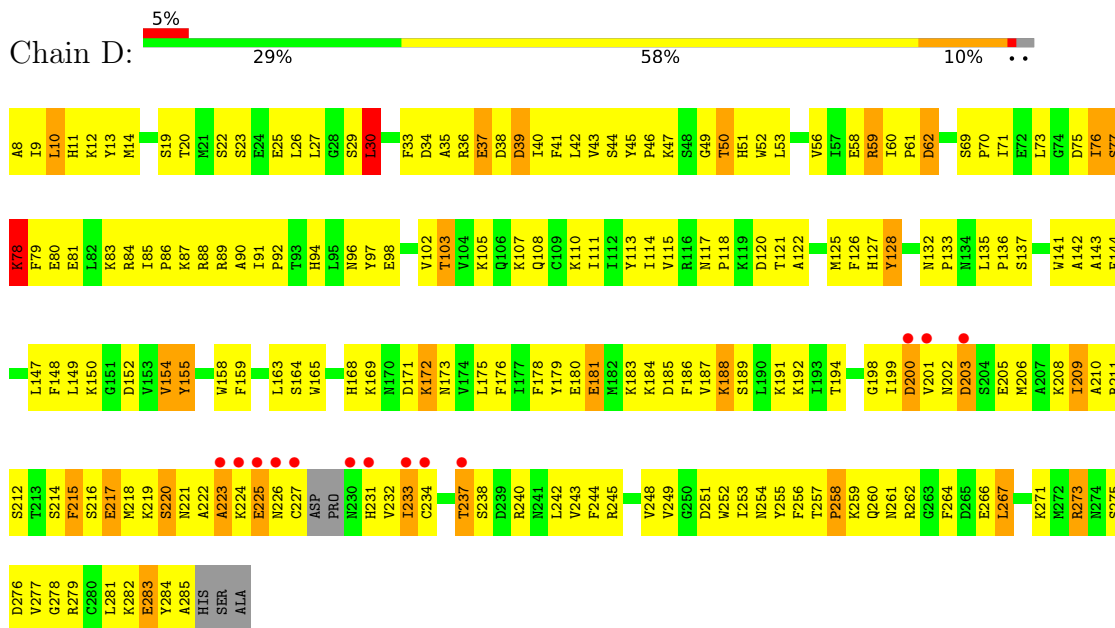
• Molecule 1: Sulfotransferase

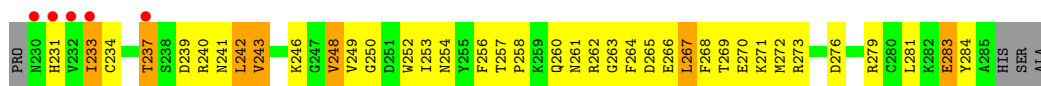


• Molecule 1: Sulfotransferase

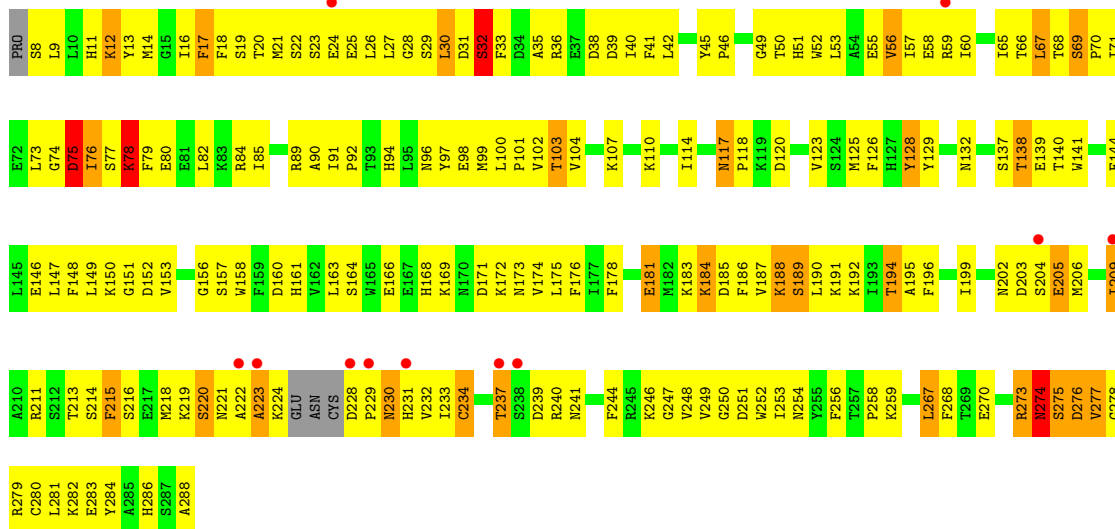


• Molecule 2: Sulfotransferase

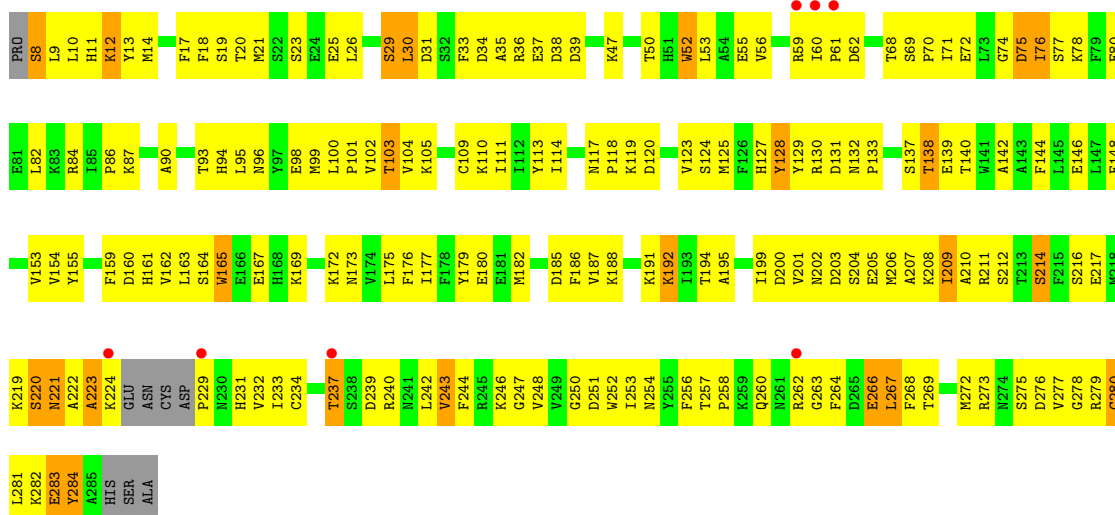




• Molecule 3: Sulfotransferase



• Molecule 3: Sulfotransferase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	166.17Å 81.13Å 166.18Å 90.00° 119.95° 90.00°	Depositor
Resolution (Å)	50.00 – 2.08 45.20 – 2.08	Depositor EDS
% Data completeness (in resolution range)	99.7 (50.00-2.08) 99.5 (45.20-2.08)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.24 (at 2.08Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.221 , 0.236 0.200 , 0.216	Depositor DCC
$R_{free}$ test set	11949 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.6	Xtrriage
Anisotropy	0.096	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 60.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.000 for -h-l,k,h 0.000 for l,k,-h-l 0.000 for h,-k,-h-l 0.248 for -h-l,-k,l 0.000 for l,-k,h	Xtrriage
Reported twinning fraction	0.907 for H, K, L 0.093 for H+L, -K, -L	Depositor
Outliers	11 of 229863 reflections (0.005%)	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	29402	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 82.82 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.5363e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: CA, A3P, GOL

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	0.91	2/2350 (0.1%)	0.95	0/3173
1	B	0.94	1/2327 (0.0%)	0.95	0/3138
1	C	0.94	0/2379	0.96	2/3211 (0.1%)
1	E	0.96	3/2318 (0.1%)	0.97	1/3127 (0.0%)
1	I	0.96	2/2315 (0.1%)	0.97	1/3124 (0.0%)
1	K	0.90	1/2379 (0.0%)	0.91	2/3210 (0.1%)
1	L	0.88	0/2393	0.90	1/3230 (0.0%)
2	D	0.93	1/2319 (0.0%)	0.94	2/3128 (0.1%)
3	F	0.89	0/2337	0.95	1/3153 (0.0%)
3	G	0.92	1/2320 (0.0%)	0.99	4/3129 (0.1%)
3	H	0.87	0/2319	0.92	2/3130 (0.1%)
3	J	1.01	3/2306 (0.1%)	0.98	1/3110 (0.0%)
All	All	0.93	14/28062 (0.0%)	0.95	17/37863 (0.0%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	109	CYS	CB-SG	-6.90	1.70	1.82
1	E	181	GLU	CG-CD	6.55	1.61	1.51
3	J	52	TRP	CB-CG	-6.45	1.38	1.50
2	D	39	ASP	CB-CG	-6.12	1.38	1.51
1	I	52	TRP	CB-CG	-5.94	1.39	1.50

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	I	38	ASP	CB-CG-OD1	5.85	123.56	118.30
3	H	209	ILE	CB-CA-C	-5.66	100.28	111.60
1	L	209	ILE	CB-CA-C	-5.54	100.53	111.60
1	E	147	LEU	CA-CB-CG	-5.47	102.71	115.30
1	K	209	ILE	CB-CA-C	-5.44	100.71	111.60



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	2293	0	2259	285	0
1	B	2272	0	2242	255	1
1	C	2313	0	2288	272	1
1	E	2264	0	2241	262	1
1	I	2261	0	2233	270	0
1	K	2316	0	2291	214	2
1	L	2327	0	2304	305	2
2	D	2266	0	2235	293	2
3	F	2277	0	2248	264	1
3	G	2264	0	2237	290	1
3	H	2265	0	2232	301	0
3	J	2249	0	2229	273	1
4	A	27	0	10	15	0
4	B	27	0	11	5	0
4	C	27	0	11	5	0
4	D	27	0	11	10	0
4	E	27	0	11	8	0
4	F	27	0	11	7	0
4	G	27	0	11	26	0
4	H	27	0	11	7	0
4	I	27	0	11	3	0
4	J	27	0	11	10	0
4	K	27	0	11	8	0
4	L	27	0	11	16	0
5	A	2	0	0	0	0
5	B	3	0	0	0	0
5	C	1	0	0	0	0
5	D	3	0	0	0	0
5	E	1	0	0	0	0
5	F	1	0	0	0	0
5	G	2	0	0	0	0
5	H	1	0	0	0	0
5	I	1	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	J	2	0	0	0	0
5	K	2	0	0	0	0
5	L	1	0	0	0	0
6	F	6	0	8	2	0
6	J	6	0	8	0	0
7	A	149	0	0	71	1
7	B	151	0	0	61	0
7	C	126	0	0	48	0
7	D	169	0	0	65	1
7	E	136	0	0	52	0
7	F	108	0	0	53	1
7	G	172	0	0	79	2
7	H	131	0	0	59	2
7	I	131	0	0	59	1
7	J	157	0	0	65	0
7	K	151	0	0	37	0
7	L	98	0	0	45	0
All	All	29402	0	27186	3236	10

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 59.

The worst 5 of 3236 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)
3:J:103:THR:HB	7:J:470:HOH:O	1.21	1.30
1:K:25:GLU:O	1:K:29:SER:OG	1.58	1.18
2:D:245:ARG:N	7:D:402:HOH:O	1.81	1.14
1:C:19:SER:N	7:C:401:HOH:O	1.83	1.09
1:L:218:MET:SD	7:L:406:HOH:O	2.10	1.07

The worst 5 of 10 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)
3:F:256:PHE:O	7:G:519:HOH:O[1_545]	1.43	0.77
2:D:135:LEU:O	7:H:426:HOH:O[2_646]	1.78	0.42
1:L:211:ARG:NH1	7:A:419:HOH:O[2_657]	1.99	0.21
1:E:130:ARG:O	1:K:211:ARG:NH2[2_747]	2.04	0.16
2:D:137:SER:N	7:H:426:HOH:O[2_646]	2.06	0.14

## 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	280/283 (99%)	235 (84%)	38 (14%)	7 (2%)	5	2
1	B	274/283 (97%)	242 (88%)	27 (10%)	5 (2%)	8	3
1	C	284/283 (100%)	231 (81%)	36 (13%)	17 (6%)	1	0
1	E	274/283 (97%)	240 (88%)	27 (10%)	7 (3%)	5	1
1	I	274/283 (97%)	230 (84%)	37 (14%)	7 (3%)	5	1
1	K	283/283 (100%)	247 (87%)	30 (11%)	6 (2%)	7	2
1	L	286/283 (101%)	236 (82%)	37 (13%)	13 (4%)	2	0
2	D	274/281 (98%)	234 (85%)	27 (10%)	13 (5%)	2	0
3	F	277/282 (98%)	231 (83%)	31 (11%)	15 (5%)	2	0
3	G	274/282 (97%)	240 (88%)	27 (10%)	7 (3%)	5	1
3	H	275/282 (98%)	237 (86%)	26 (10%)	12 (4%)	2	0
3	J	272/282 (96%)	230 (85%)	36 (13%)	6 (2%)	6	2
All	All	3327/3390 (98%)	2833 (85%)	379 (11%)	115 (4%)	3	1

5 of 115 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	50	THR
1	C	137	SER
1	C	223	ALA
1	C	228	ASP
1	A	8	SER

### 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	255/255 (100%)	234 (92%)	21 (8%)	11	7
1	B	252/255 (99%)	229 (91%)	23 (9%)	9	6
1	C	259/255 (102%)	227 (88%)	32 (12%)	4	2
1	E	251/255 (98%)	220 (88%)	31 (12%)	4	2
1	I	251/255 (98%)	222 (88%)	29 (12%)	5	2
1	K	258/255 (101%)	230 (89%)	28 (11%)	6	3
1	L	260/255 (102%)	233 (90%)	27 (10%)	7	4
2	D	251/253 (99%)	225 (90%)	26 (10%)	7	4
3	F	251/255 (98%)	227 (90%)	24 (10%)	8	5
3	G	252/255 (99%)	228 (90%)	24 (10%)	8	5
3	H	249/255 (98%)	219 (88%)	30 (12%)	5	2
3	J	250/255 (98%)	227 (91%)	23 (9%)	9	5
All	All	3039/3058 (99%)	2721 (90%)	318 (10%)	7	3

5 of 318 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	227	CYS
1	K	225	GLU
1	I	275	SER
3	J	266	GLU
1	L	117	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	K	11	HIS
1	K	51	HIS
1	L	11	HIS
3	F	11	HIS
1	E	274	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 34 ligands modelled in this entry, 20 are monoatomic - leaving 14 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
4	A3P	A	301	-	26,29,29	1.52	7 (26%)	31,45,45	1.66	7 (22%)
4	A3P	G	301	-	26,29,29	1.84	8 (30%)	31,45,45	1.91	8 (25%)
6	GOL	F	303	-	5,5,5	0.42	0	5,5,5	0.28	0
6	GOL	J	304	-	5,5,5	0.38	0	5,5,5	0.81	0
4	A3P	C	301	-	26,29,29	1.66	6 (23%)	31,45,45	1.55	5 (16%)
4	A3P	H	301	-	26,29,29	1.87	6 (23%)	31,45,45	1.48	5 (16%)
4	A3P	K	301	-	26,29,29	2.09	7 (26%)	31,45,45	1.47	3 (9%)
4	A3P	J	301	-	26,29,29	2.15	7 (26%)	31,45,45	1.64	7 (22%)
4	A3P	B	301	-	26,29,29	1.73	5 (19%)	31,45,45	1.83	6 (19%)
4	A3P	E	301	-	26,29,29	1.84	8 (30%)	31,45,45	1.65	5 (16%)
4	A3P	F	301	-	26,29,29	1.80	4 (15%)	31,45,45	1.54	3 (9%)
4	A3P	I	301	-	26,29,29	2.11	6 (23%)	31,45,45	1.47	5 (16%)
4	A3P	L	301	-	26,29,29	1.93	7 (26%)	31,45,45	1.84	6 (19%)
4	A3P	D	301	-	26,29,29	1.82	7 (26%)	31,45,45	1.43	4 (12%)

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
4	A3P	A	301	-	-	5/11/31/31	0/3/3/3
4	A3P	G	301	-	-	3/11/31/31	0/3/3/3
6	GOL	F	303	-	-	3/4/4/4	-
6	GOL	J	304	-	-	4/4/4/4	-
4	A3P	C	301	-	-	1/11/31/31	0/3/3/3
4	A3P	H	301	-	-	3/11/31/31	0/3/3/3
4	A3P	K	301	-	-	0/11/31/31	0/3/3/3
4	A3P	J	301	-	-	1/11/31/31	0/3/3/3
4	A3P	B	301	-	-	1/11/31/31	0/3/3/3
4	A3P	E	301	-	-	2/11/31/31	0/3/3/3
4	A3P	F	301	-	-	2/11/31/31	0/3/3/3
4	A3P	I	301	-	-	2/11/31/31	0/3/3/3
4	A3P	L	301	-	-	2/11/31/31	0/3/3/3
4	A3P	D	301	-	-	3/11/31/31	0/3/3/3

The worst 5 of 78 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	J	301	A3P	C2'-C1'	-6.51	1.43	1.53
4	I	301	A3P	C2'-C1'	-6.19	1.44	1.53
4	L	301	A3P	C2'-C1'	-5.08	1.46	1.53
4	K	301	A3P	C2'-C1'	-4.93	1.46	1.53
4	H	301	A3P	C2'-C1'	-4.91	1.46	1.53

The worst 5 of 64 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	301	A3P	O4'-C1'-C2'	-5.84	98.40	106.93
4	L	301	A3P	O4'-C1'-C2'	-5.39	99.04	106.93
4	B	301	A3P	O4'-C1'-C2'	-5.14	99.42	106.93
4	I	301	A3P	N3-C2-N1	-5.13	120.67	128.68
4	E	301	A3P	N3-C2-N1	-5.01	120.84	128.68

There are no chirality outliers.

5 of 32 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	301	A3P	C5'-O5'-P2-O4P
4	A	301	A3P	C5'-O5'-P2-O5P
4	B	301	A3P	C3'-O3'-P1-O1P

*Continued on next page...*

*Continued from previous page...*

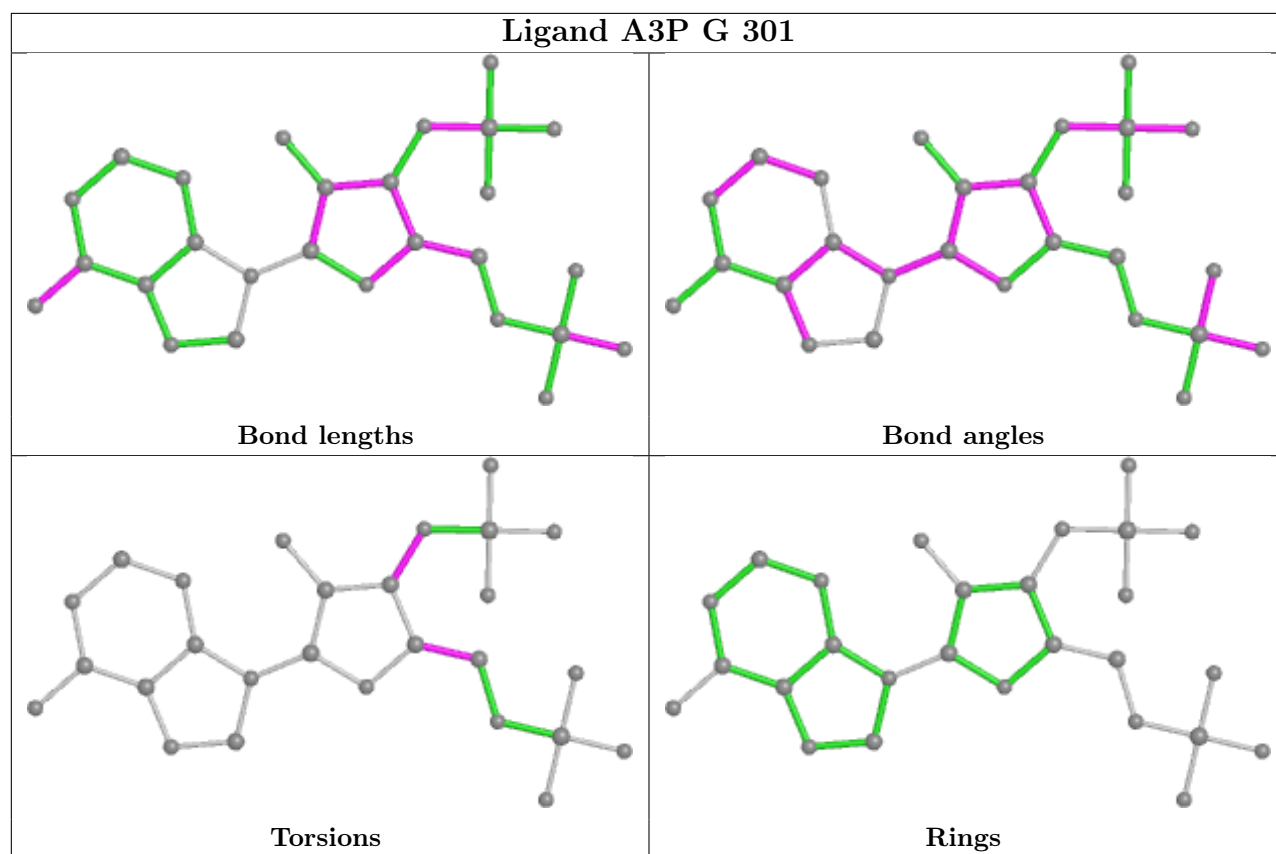
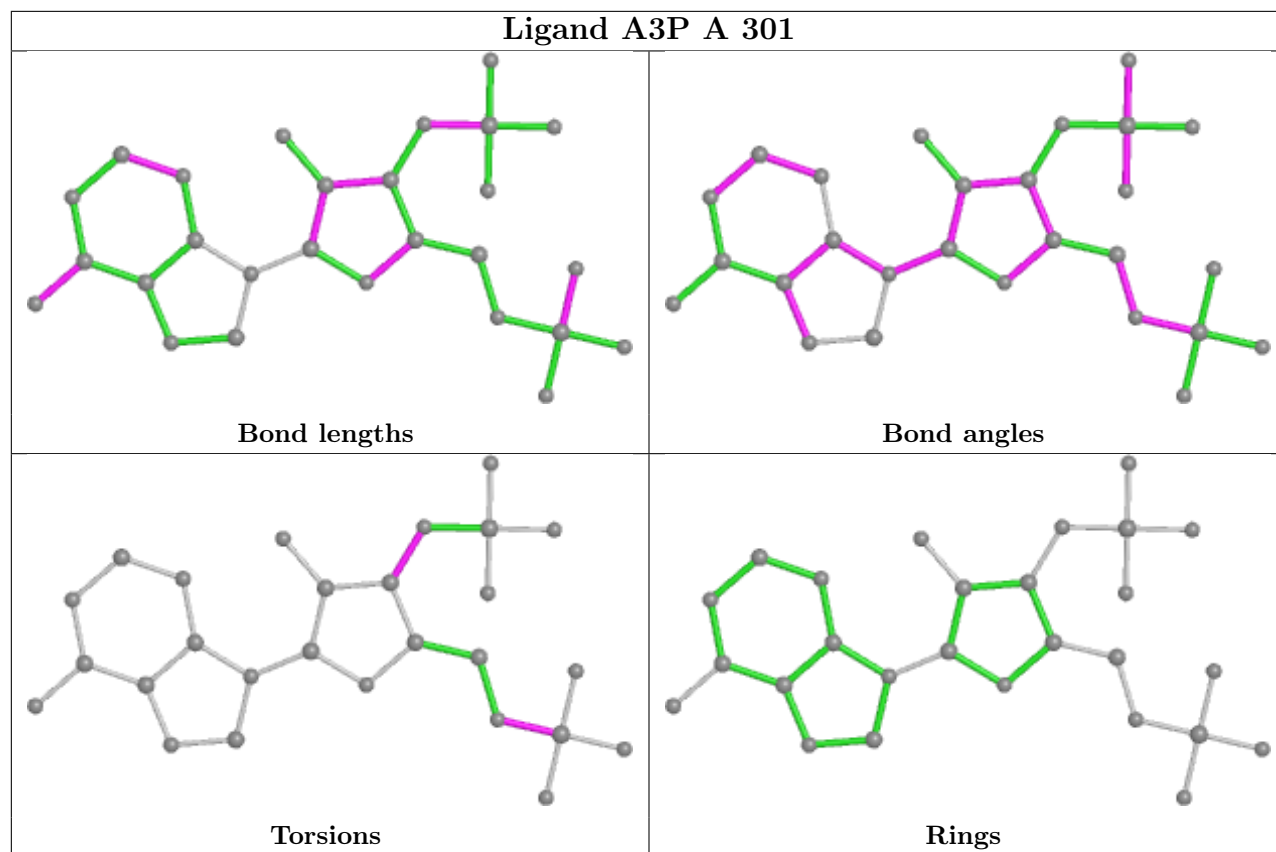
Mol	Chain	Res	Type	Atoms
4	E	301	A3P	C3'-O3'-P1-O3P
4	H	301	A3P	C5'-O5'-P2-O4P

There are no ring outliers.

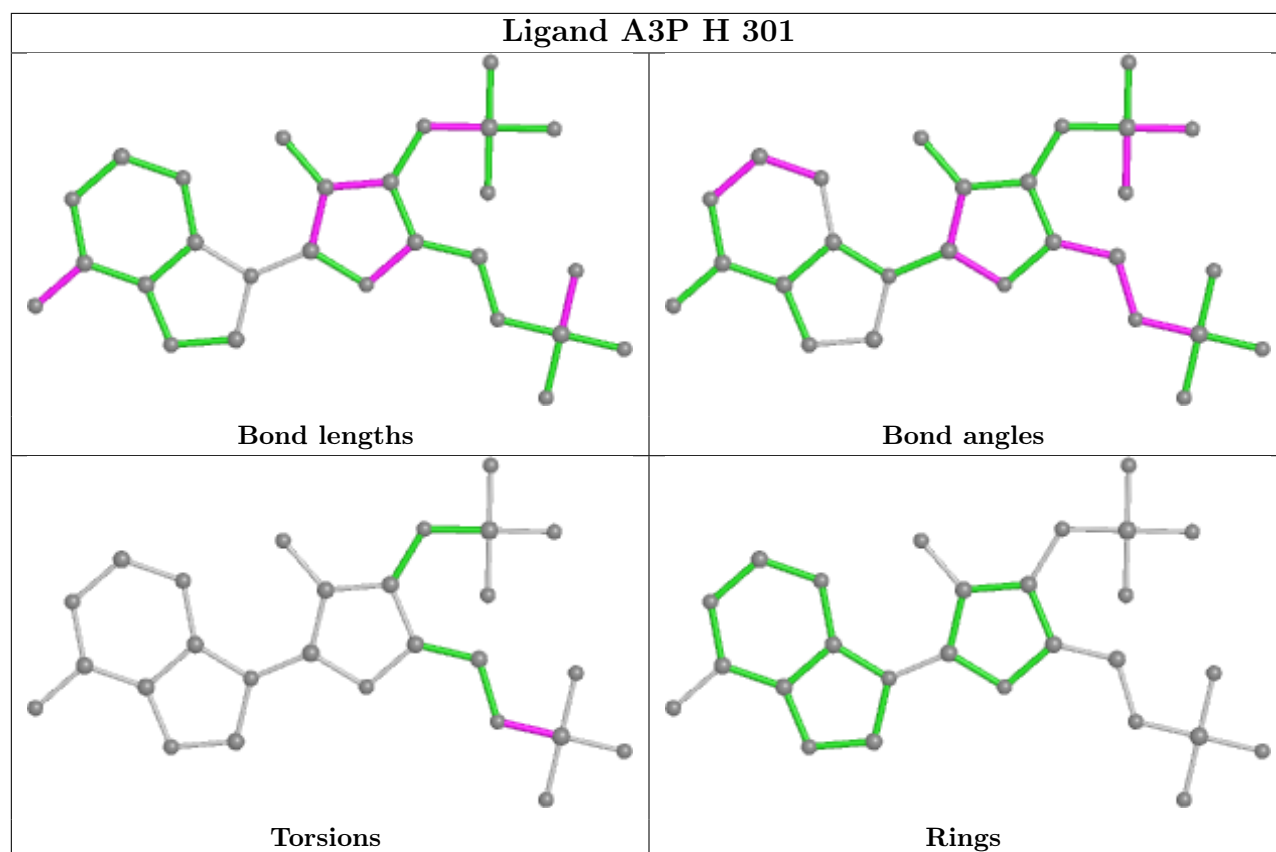
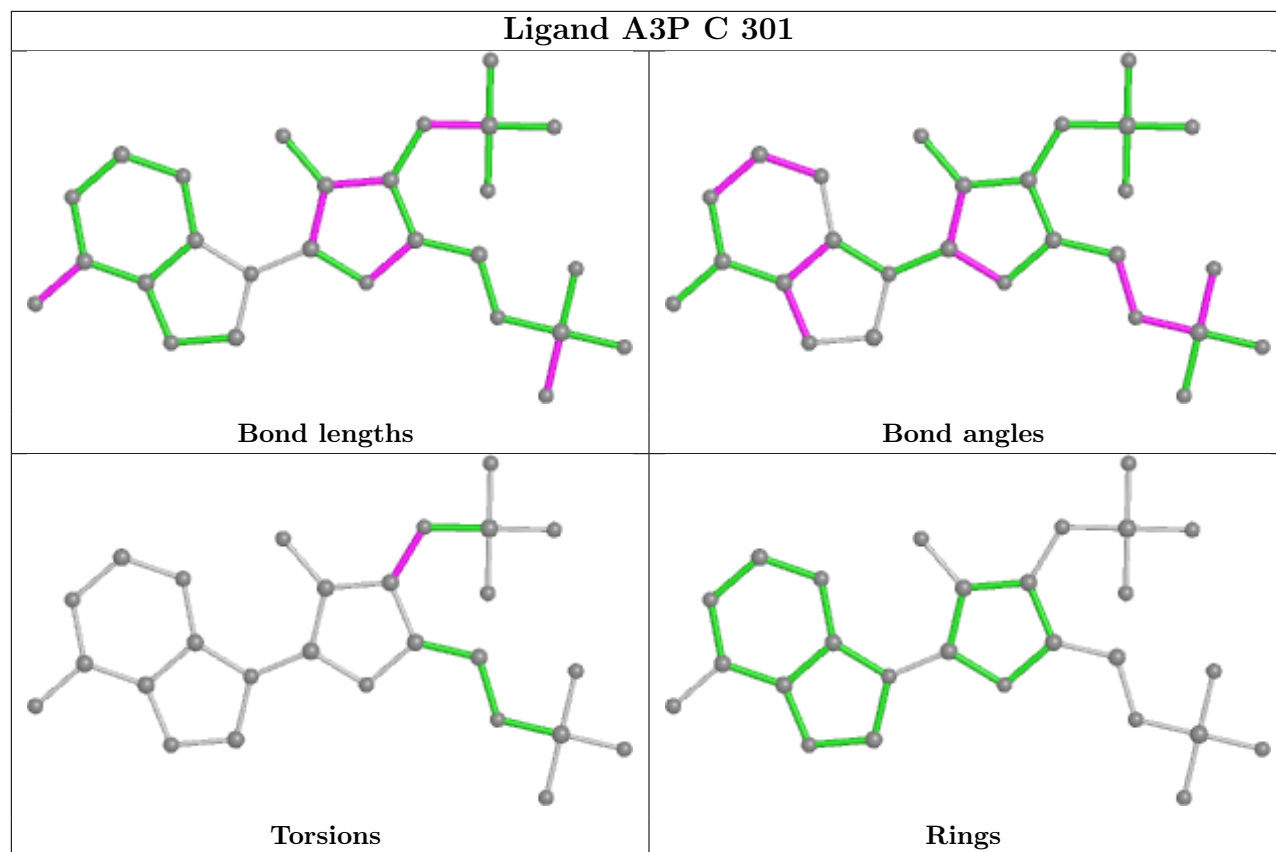
13 monomers are involved in 122 short contacts:

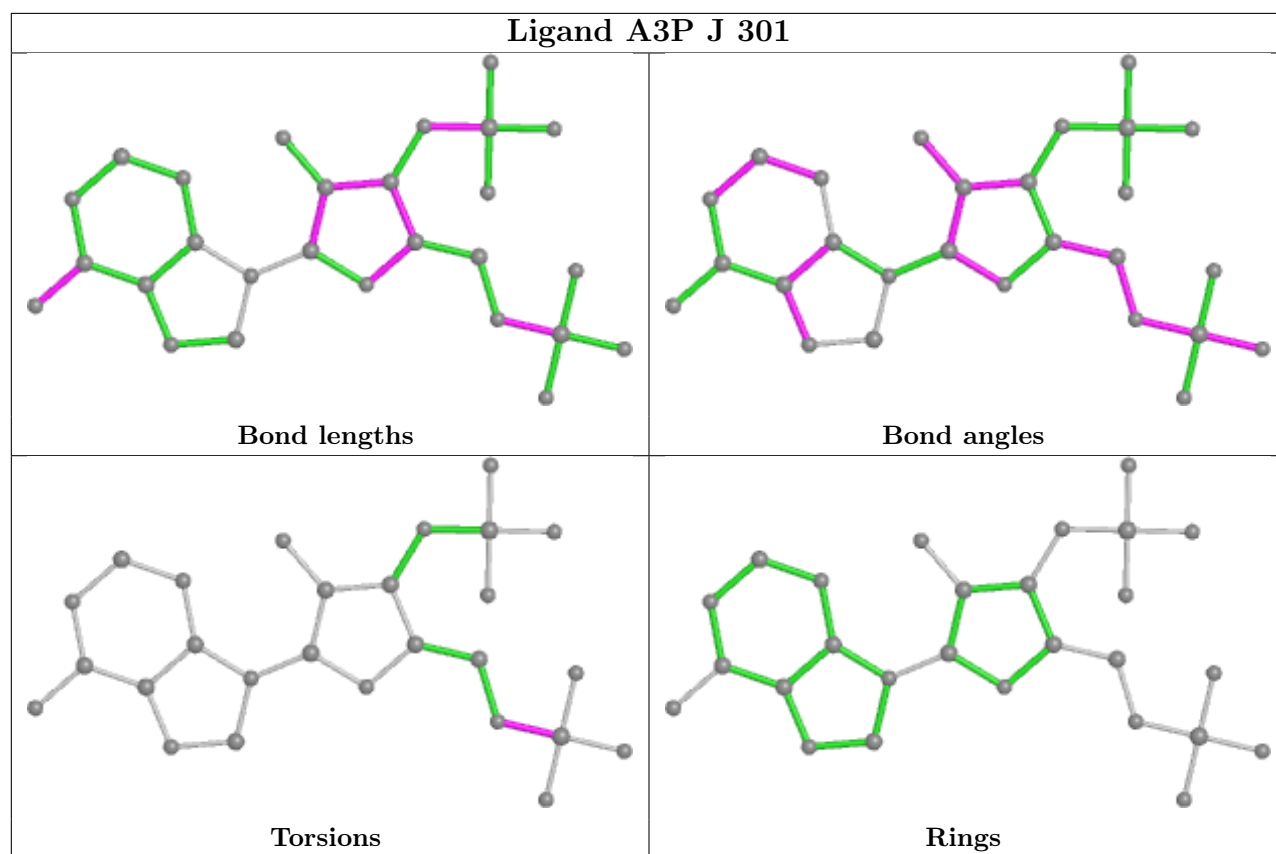
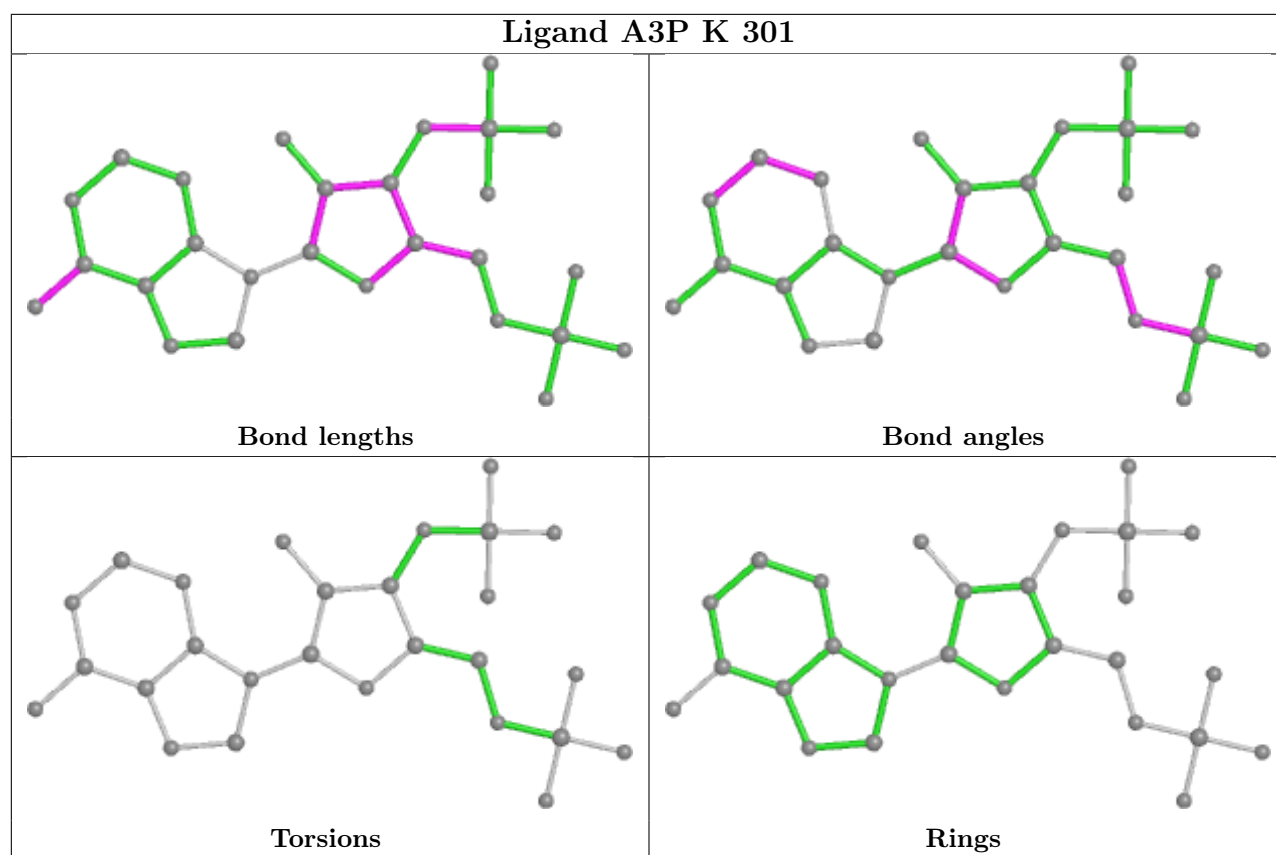
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	301	A3P	15	0
4	G	301	A3P	26	0
6	F	303	GOL	2	0
4	C	301	A3P	5	0
4	H	301	A3P	7	0
4	K	301	A3P	8	0
4	J	301	A3P	10	0
4	B	301	A3P	5	0
4	E	301	A3P	8	0
4	F	301	A3P	7	0
4	I	301	A3P	3	0
4	L	301	A3P	16	0
4	D	301	A3P	10	0

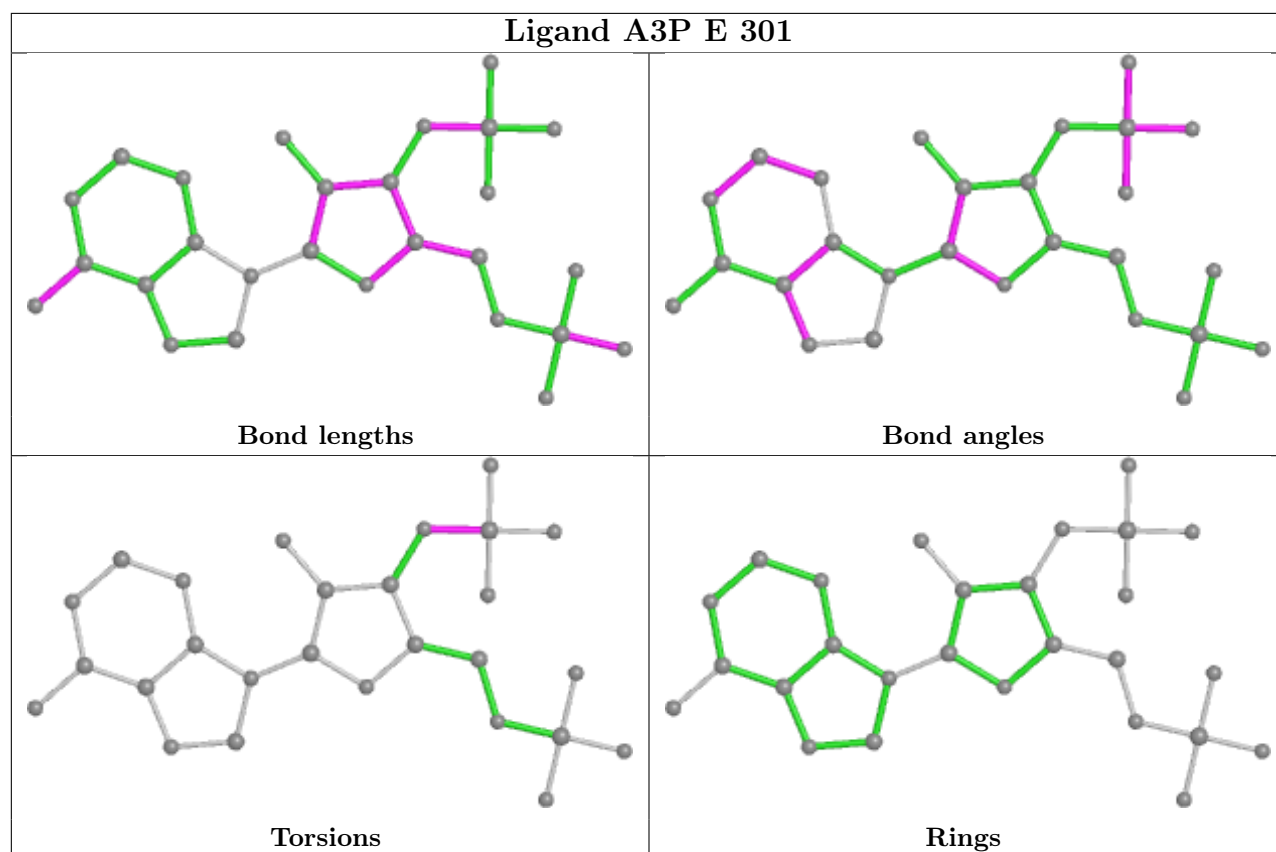
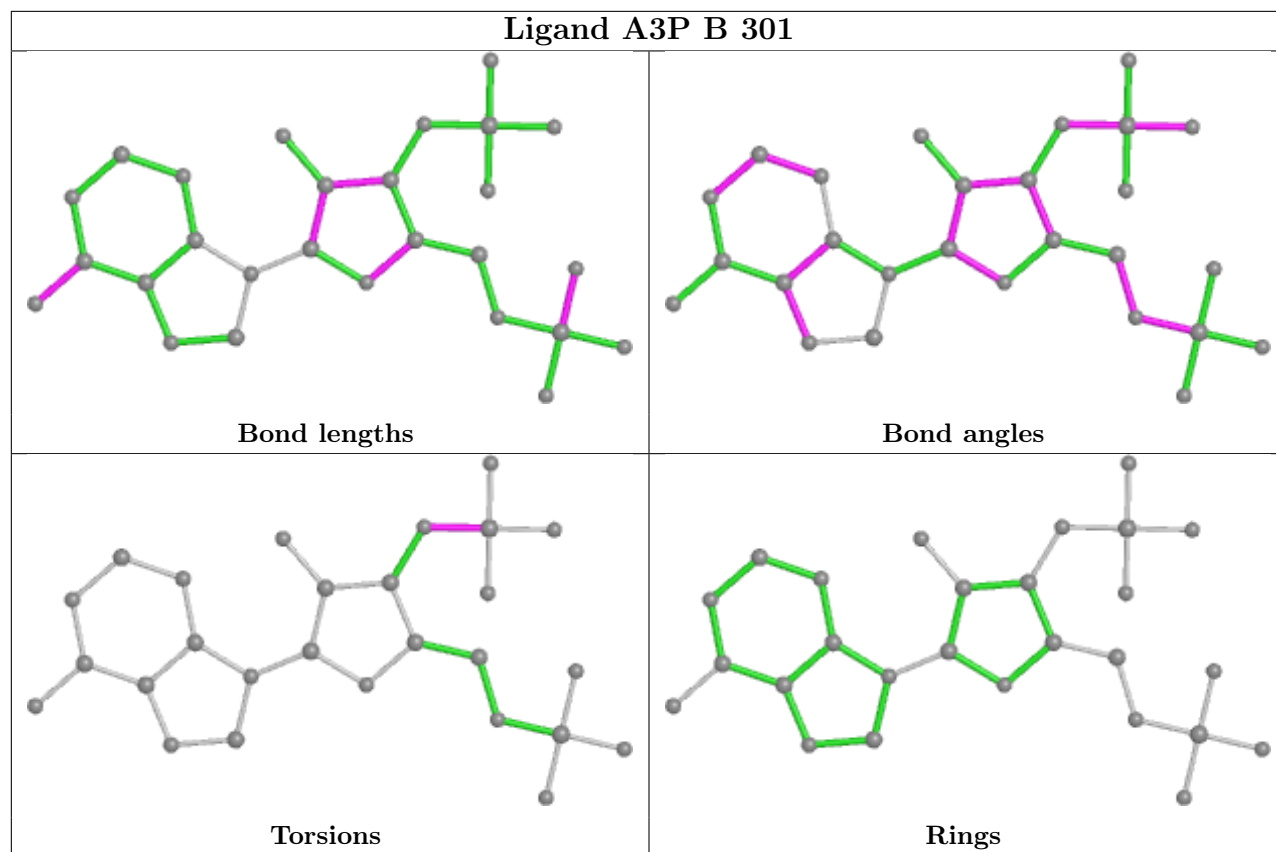
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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.

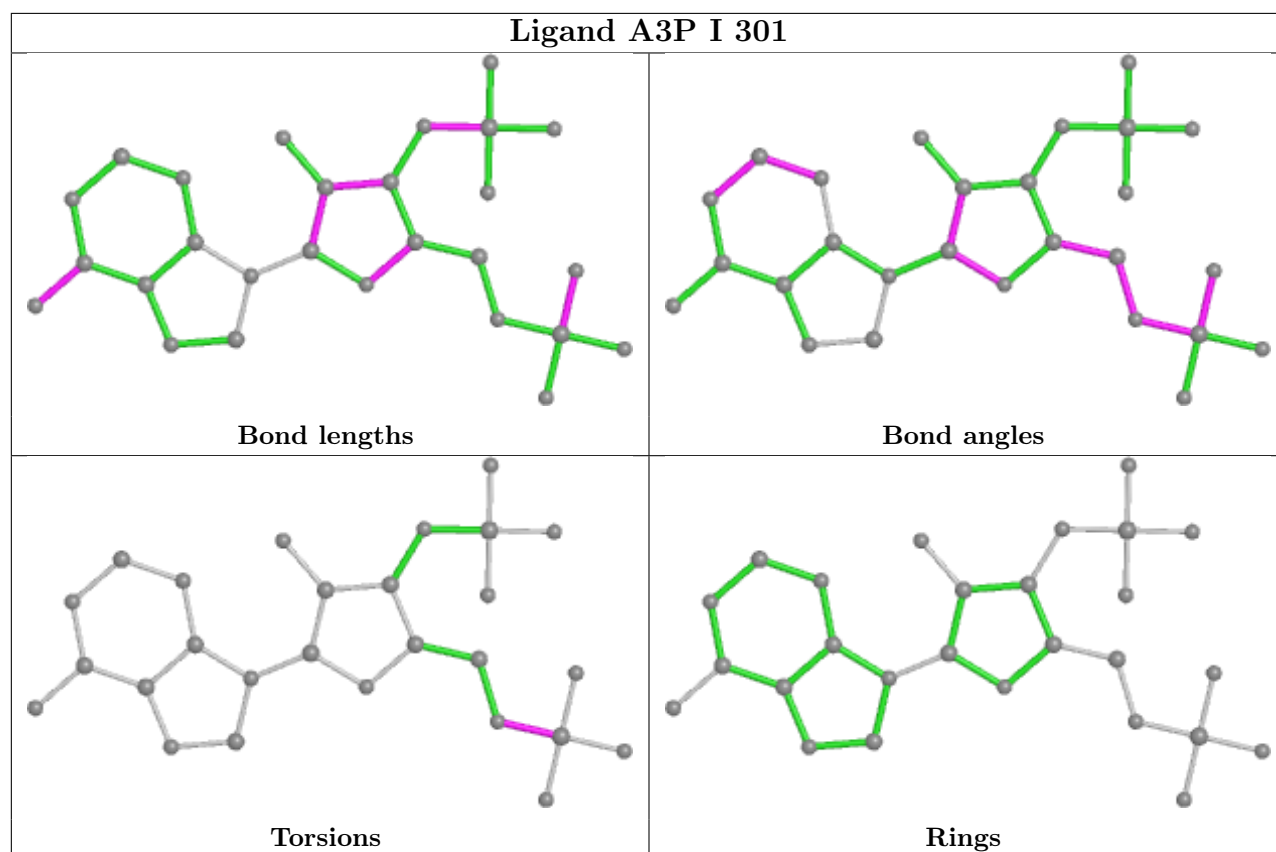
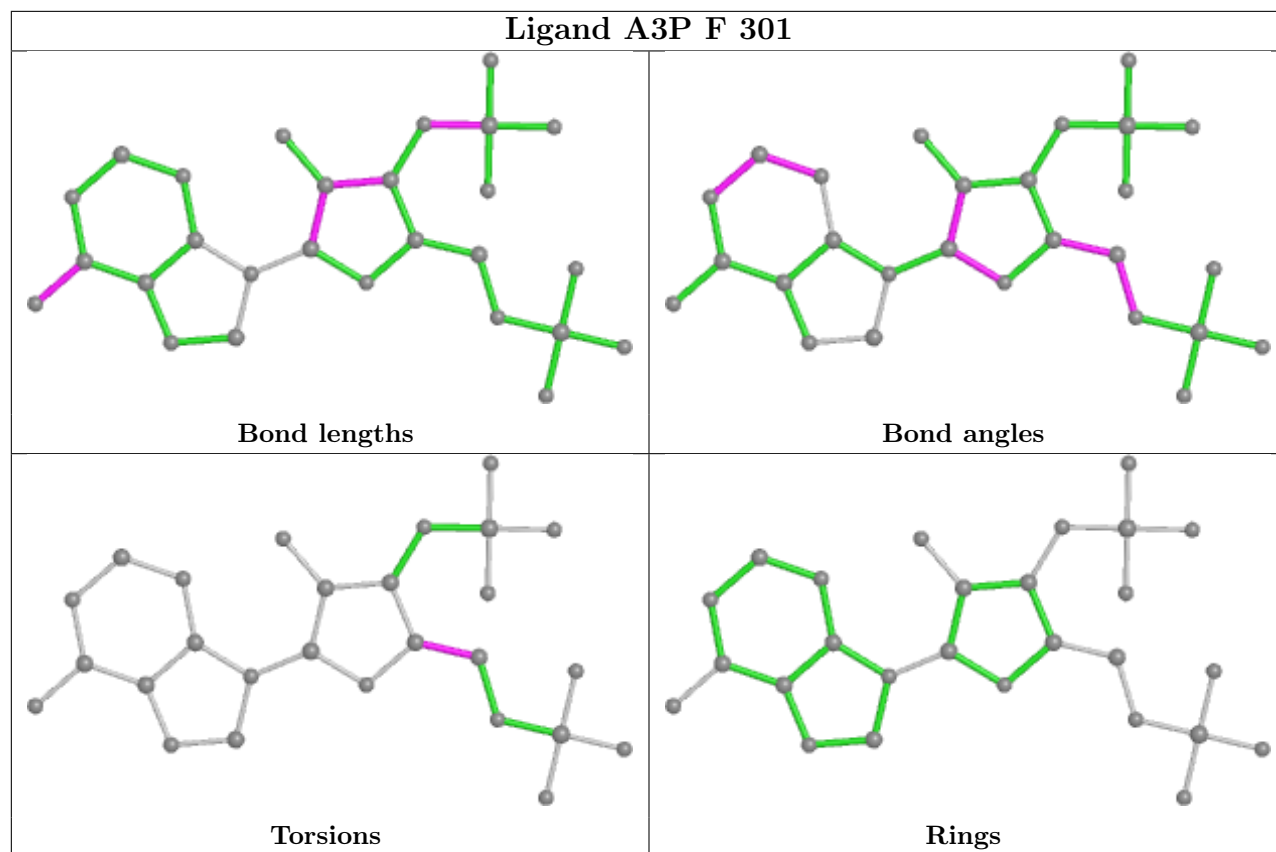


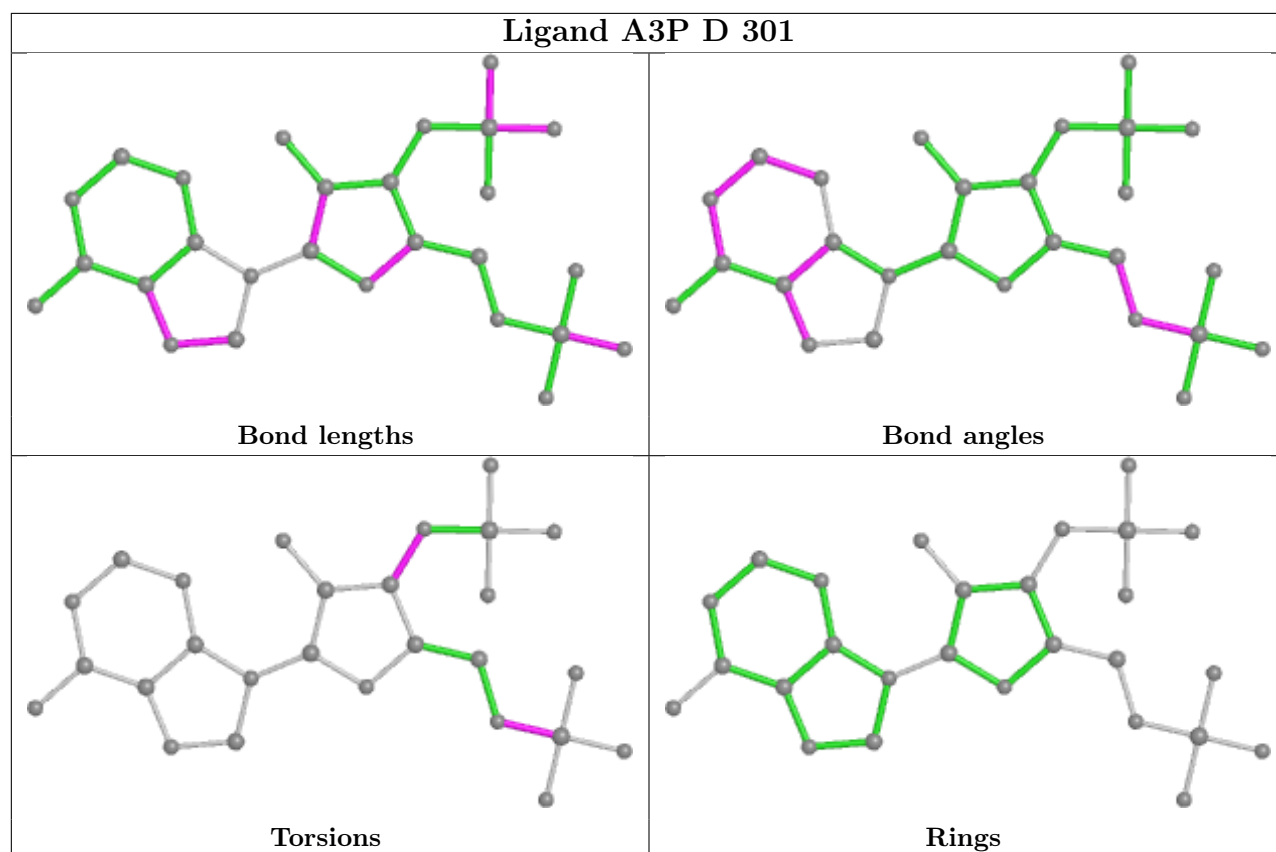
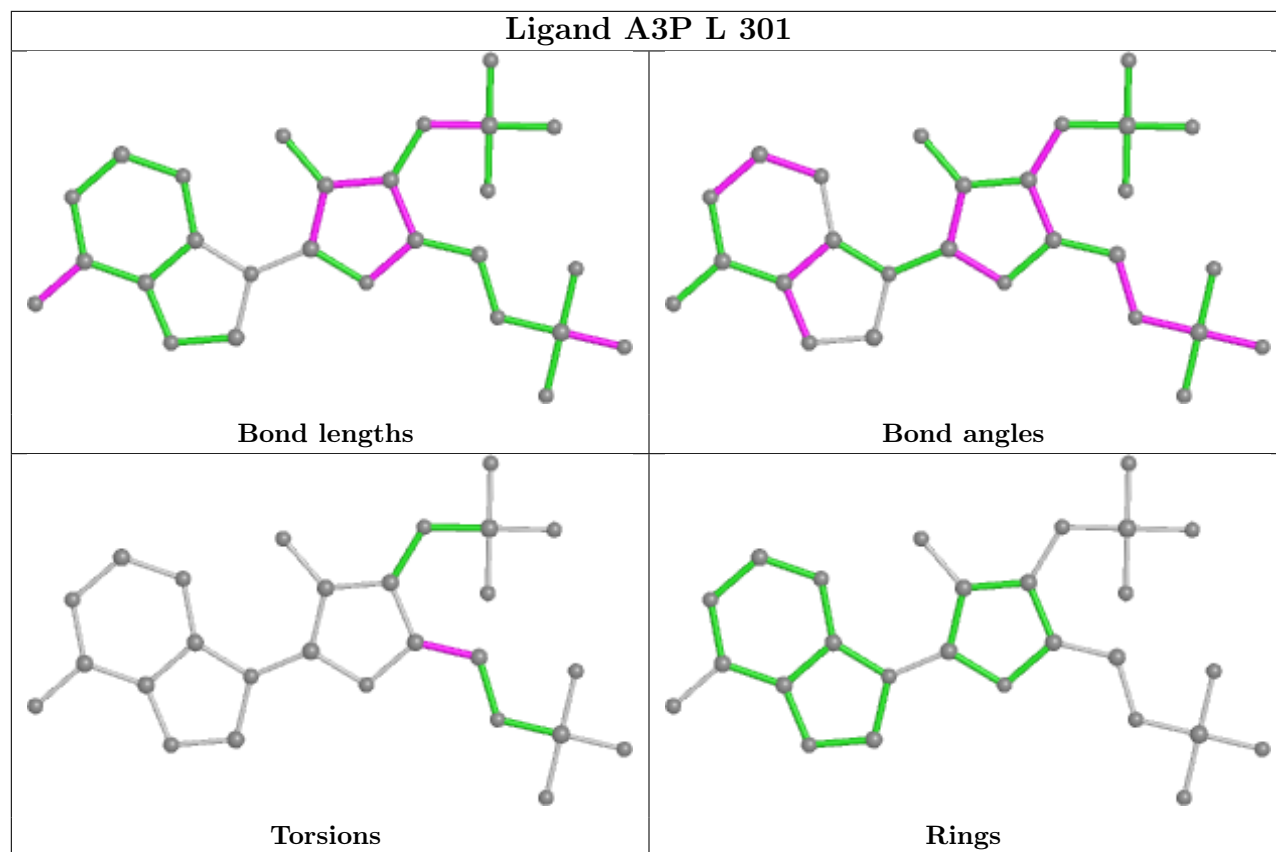












## 5.7 Other polymers [i](#)

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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	281/283 (99%)	0.29	12 (4%) 35 40	24, 39, 61, 105	0
1	B	275/283 (97%)	0.21	6 (2%) 62 66	23, 37, 61, 98	0
1	C	282/283 (99%)	0.35	18 (6%) 19 23	23, 41, 66, 107	0
1	E	277/283 (97%)	0.25	10 (3%) 42 48	22, 37, 62, 99	0
1	I	277/283 (97%)	0.31	10 (3%) 42 48	23, 37, 63, 115	0
1	K	281/283 (99%)	0.24	10 (3%) 42 48	23, 38, 62, 107	0
1	L	283/283 (100%)	0.45	19 (6%) 17 21	24, 42, 65, 116	0
2	D	276/281 (98%)	0.33	13 (4%) 31 36	24, 38, 61, 100	0
3	F	279/282 (98%)	0.32	11 (3%) 39 44	26, 41, 65, 99	0
3	G	276/282 (97%)	0.43	15 (5%) 25 30	22, 38, 61, 108	0
3	H	278/282 (98%)	0.30	11 (3%) 38 43	25, 43, 64, 109	0
3	J	274/282 (97%)	0.16	7 (2%) 56 60	23, 37, 61, 104	0
All	All	3339/3390 (98%)	0.30	142 (4%) 35 40	22, 39, 64, 116	0

The worst 5 of 142 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	227	CYS	22.6
1	I	227	CYS	21.4
3	G	227	CYS	20.3
3	F	229	PRO	12.1
1	I	226	ASN	11.4

### 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<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
5	CA	L	302	1/1	0.57	0.14	80,80,80,80	0
6	GOL	J	304	6/6	0.67	0.22	42,47,59,67	0
5	CA	B	304	1/1	0.76	0.09	69,69,69,69	0
5	CA	K	303	1/1	0.80	0.10	58,58,58,58	0
6	GOL	F	303	6/6	0.83	0.49	47,54,66,79	0
5	CA	J	302	1/1	0.83	0.18	58,58,58,58	0
5	CA	F	302	1/1	0.90	0.11	64,64,64,64	0
5	CA	B	303	1/1	0.90	0.07	44,44,44,44	0
4	A3P	L	301	27/27	0.90	0.13	31,44,62,64	0
5	CA	A	302	1/1	0.93	0.06	46,46,46,46	0
5	CA	D	304	1/1	0.93	0.14	58,58,58,58	0
5	CA	J	303	1/1	0.94	0.08	63,63,63,63	0
5	CA	K	302	1/1	0.94	0.09	46,46,46,46	0
5	CA	G	302	1/1	0.94	0.10	53,53,53,53	0
5	CA	G	303	1/1	0.95	0.13	45,45,45,45	0
5	CA	E	302	1/1	0.95	0.12	43,43,43,43	0
5	CA	D	302	1/1	0.95	0.08	47,47,47,47	0
4	A3P	B	301	27/27	0.95	0.11	20,37,49,53	0
4	A3P	K	301	27/27	0.96	0.11	29,39,52,58	0
4	A3P	A	301	27/27	0.96	0.12	28,40,54,60	0
4	A3P	D	301	27/27	0.96	0.11	28,40,62,91	0
4	A3P	F	301	27/27	0.96	0.11	28,44,53,57	0
4	A3P	H	301	27/27	0.96	0.12	36,48,63,65	0
4	A3P	I	301	27/27	0.96	0.11	27,37,44,47	0
4	A3P	J	301	27/27	0.97	0.11	24,38,50,55	0
4	A3P	G	301	27/27	0.97	0.10	20,43,54,63	0
4	A3P	E	301	27/27	0.97	0.11	28,36,51,67	0
4	A3P	C	301	27/27	0.97	0.11	33,42,49,54	0
5	CA	A	303	1/1	0.97	0.07	53,53,53,53	0
5	CA	I	302	1/1	0.98	0.22	39,39,39,39	0
5	CA	D	303	1/1	0.98	0.17	43,43,43,43	0
5	CA	B	302	1/1	0.98	0.13	40,40,40,40	0

*Continued on next page...*



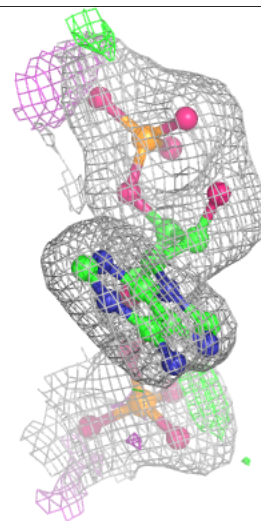
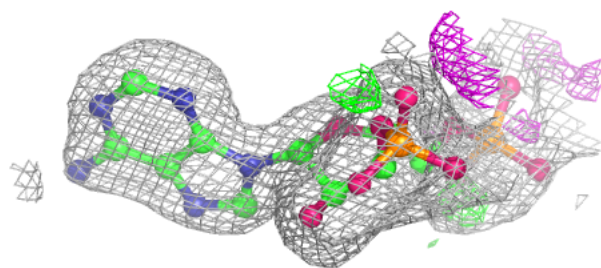
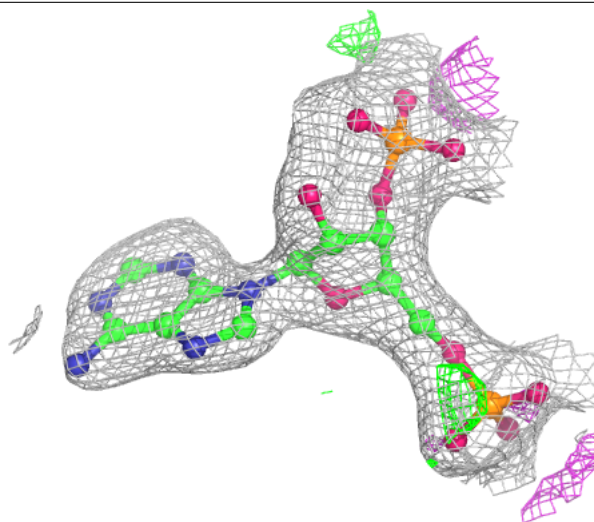
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	CA	H	302	1/1	0.98	0.05	49,49,49,49	0
5	CA	C	302	1/1	0.99	0.09	44,44,44,44	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.

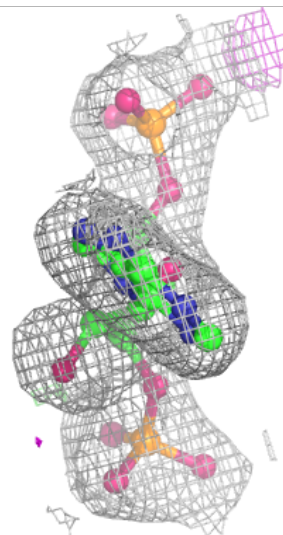
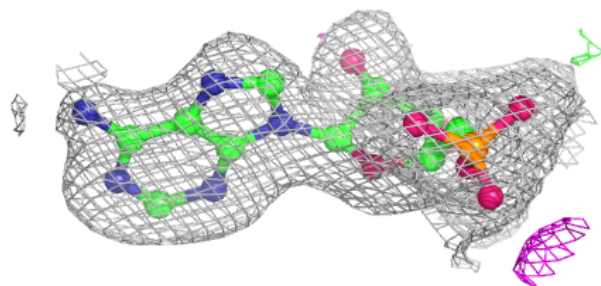
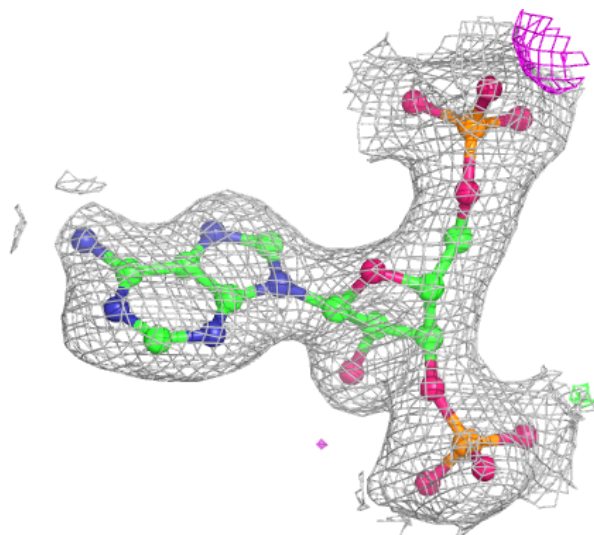
**Electron density around A3P L 301:**

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



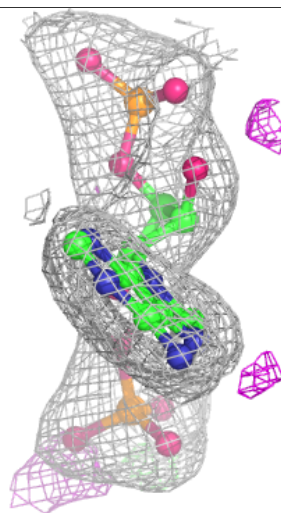
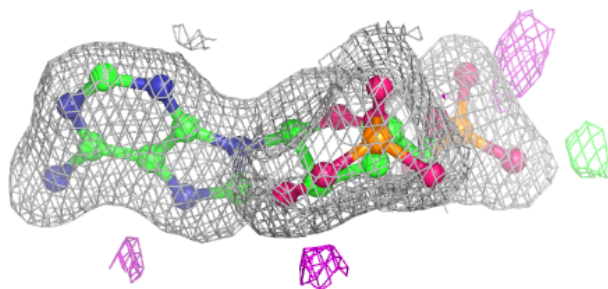
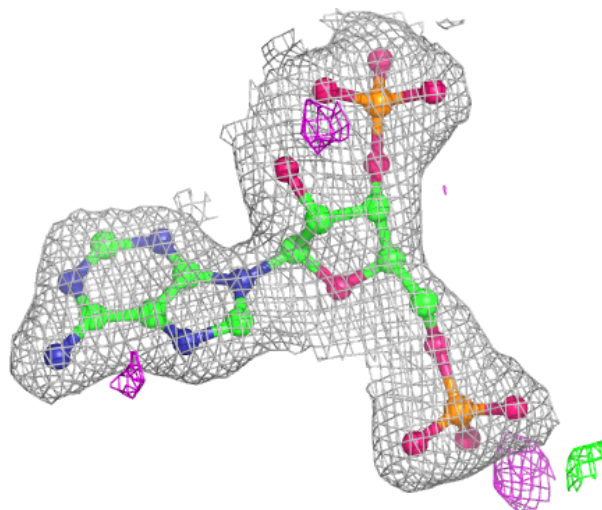
**Electron density around A3P B 301:**

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



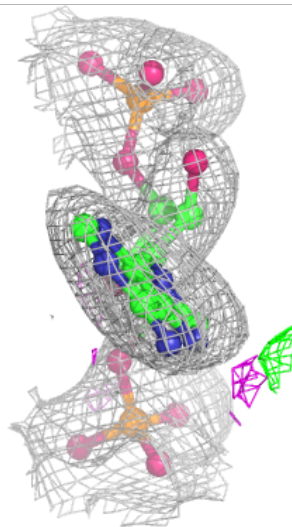
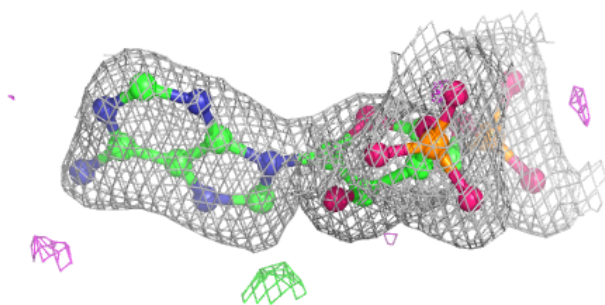
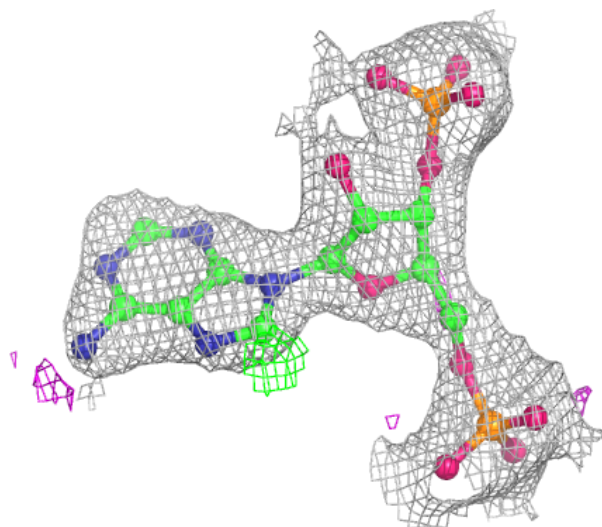
**Electron density around A3P K 301:**

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



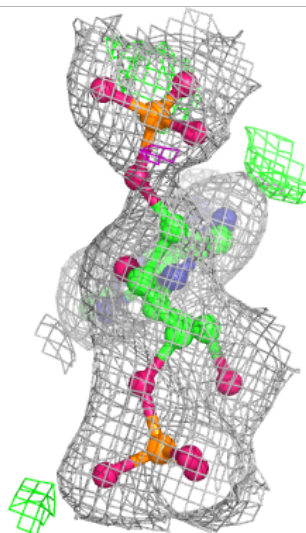
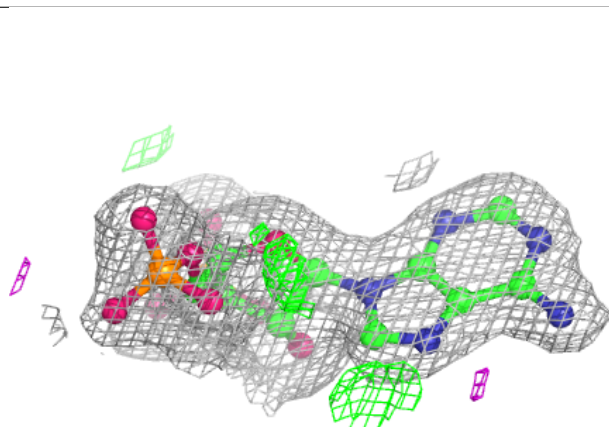
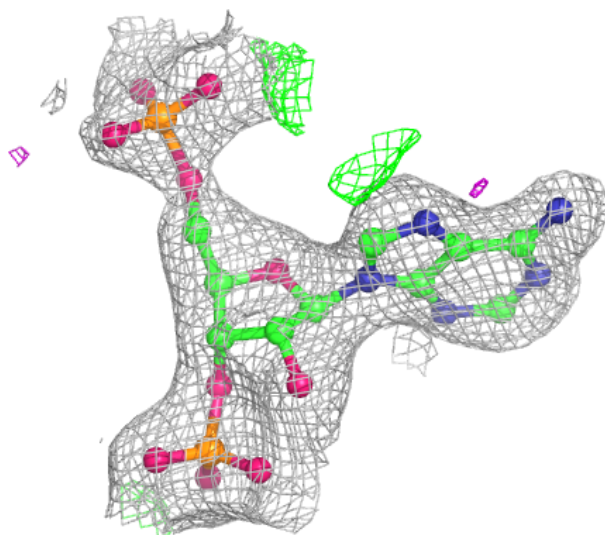
**Electron density around A3P A 301:**

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



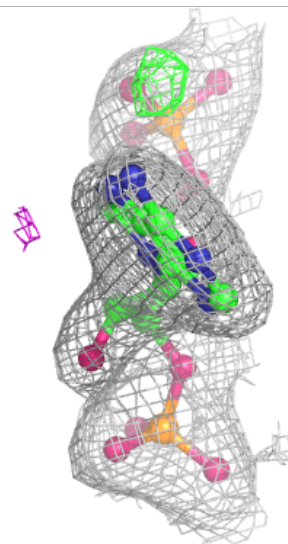
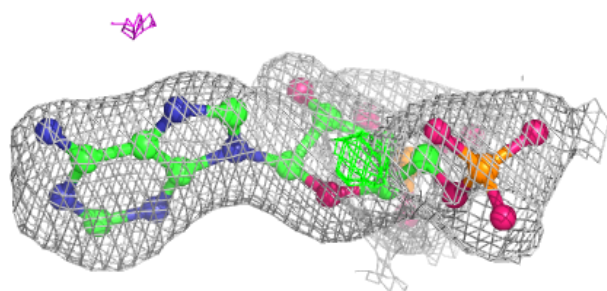
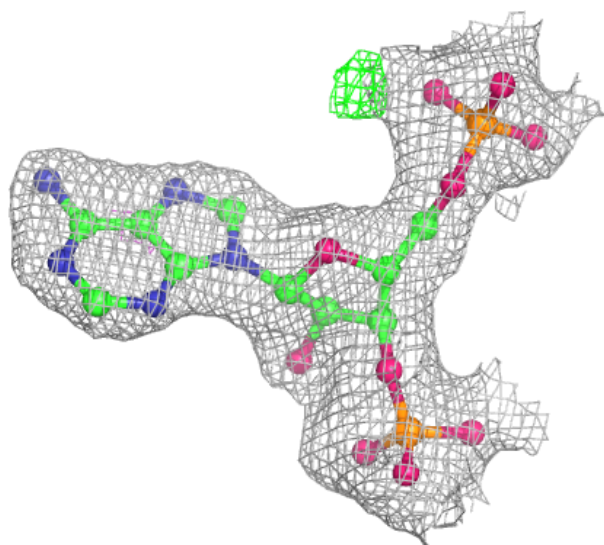
**Electron density around A3P D 301:**

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



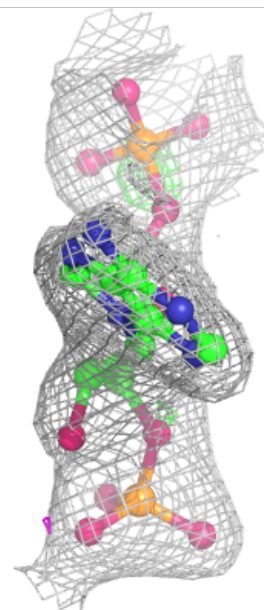
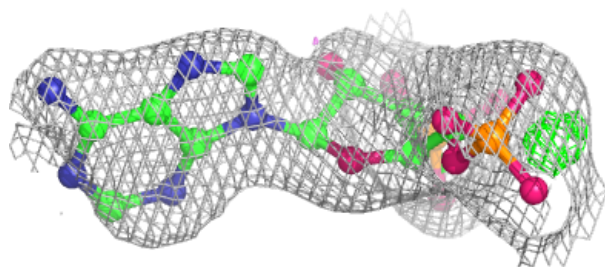
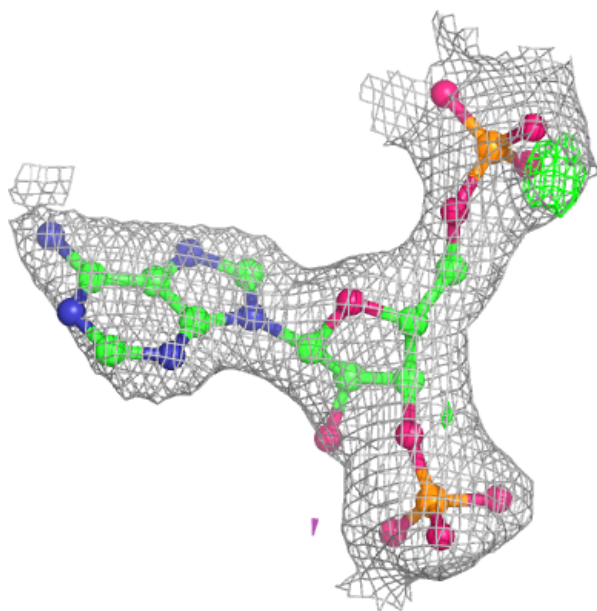
**Electron density around A3P F 301:**

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



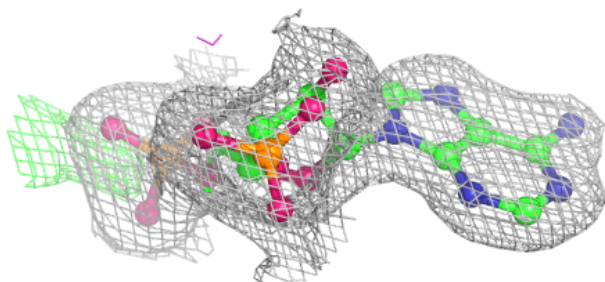
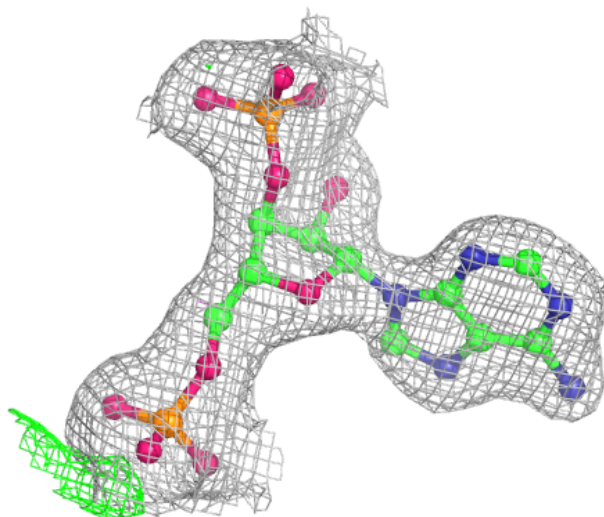
**Electron density around A3P H 301:**

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



**Electron density around A3P I 301:**

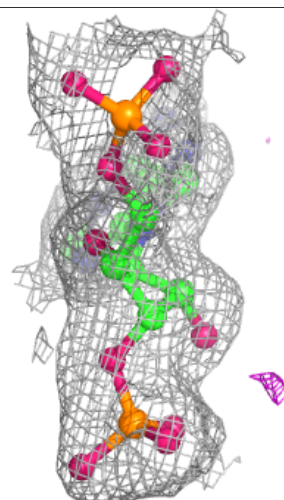
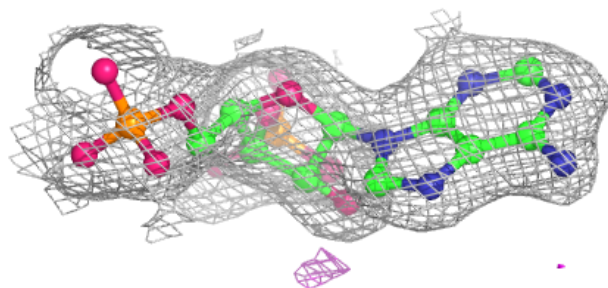
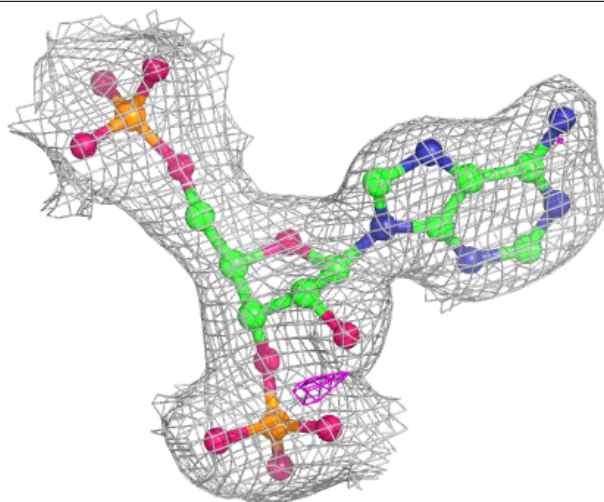
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)





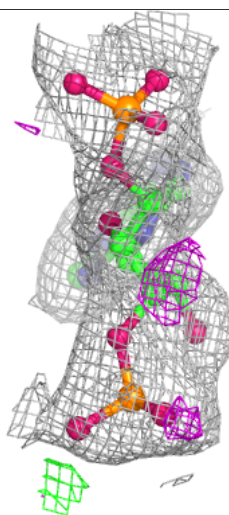
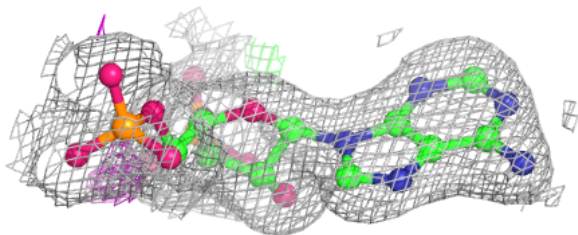
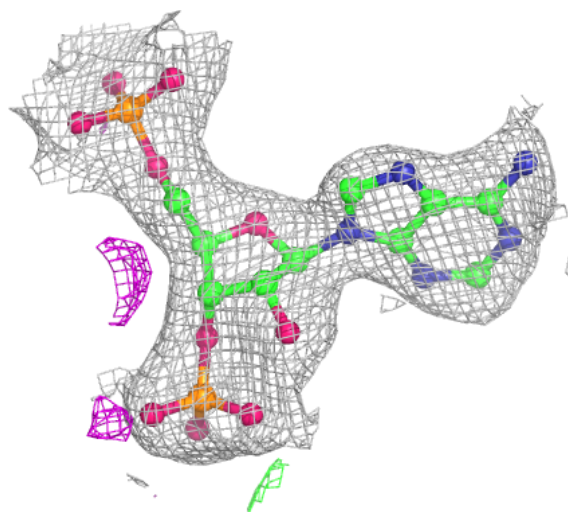
**Electron density around A3P J 301:**

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



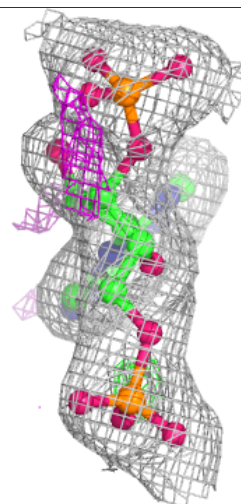
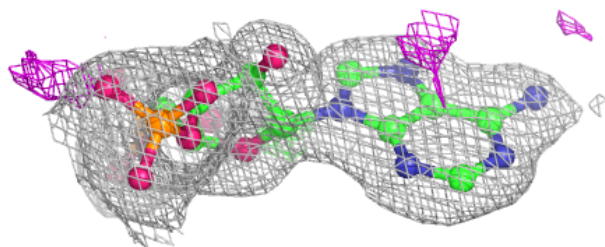
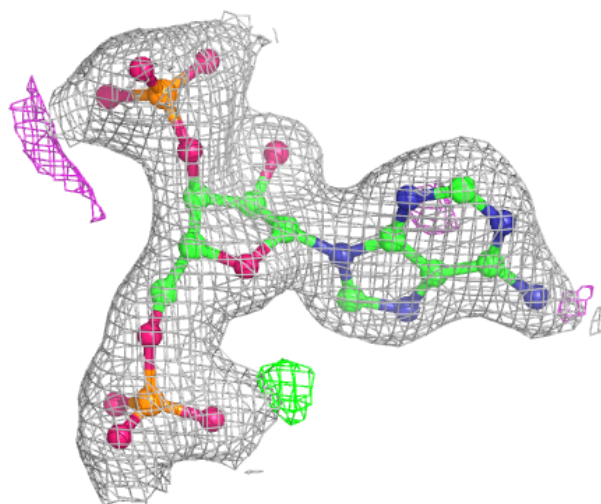
**Electron density around A3P G 301:**

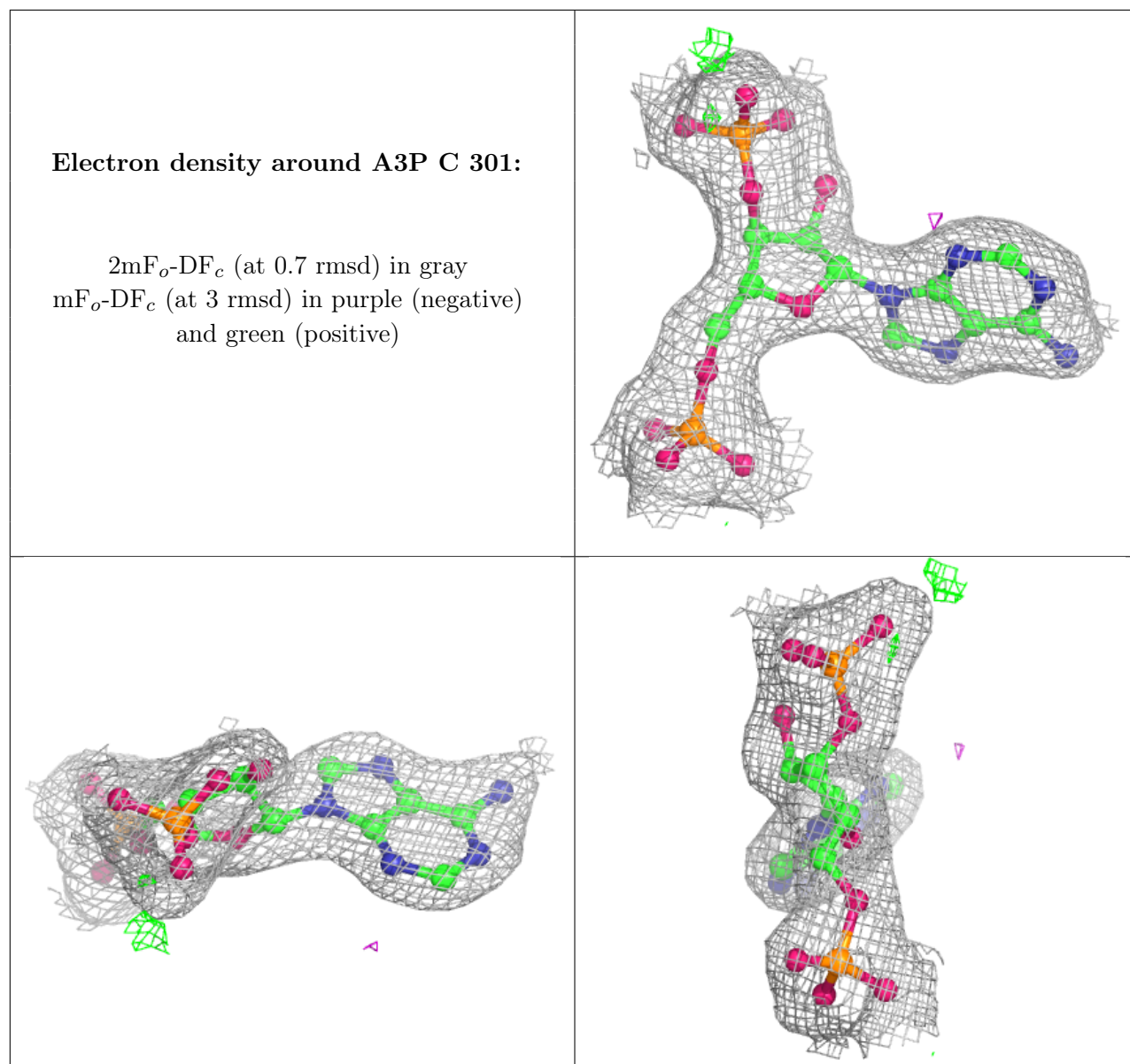
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)



**Electron density around A3P E 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 [i](#)

There are no such residues in this entry.