wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2024 – 04:09 AM EDT

PDB ID	:	1M8V
Title	:	Structure of Pyrococcus abyssii Sm Protein in Complex with a Uridine Heptamer
Authors	:	Thore, S.; Mayer, C.; Sauter, C.; Weeks, S.; Suck, D.
Deposited on	:	2002-07-26
Resolution	:	2.60 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

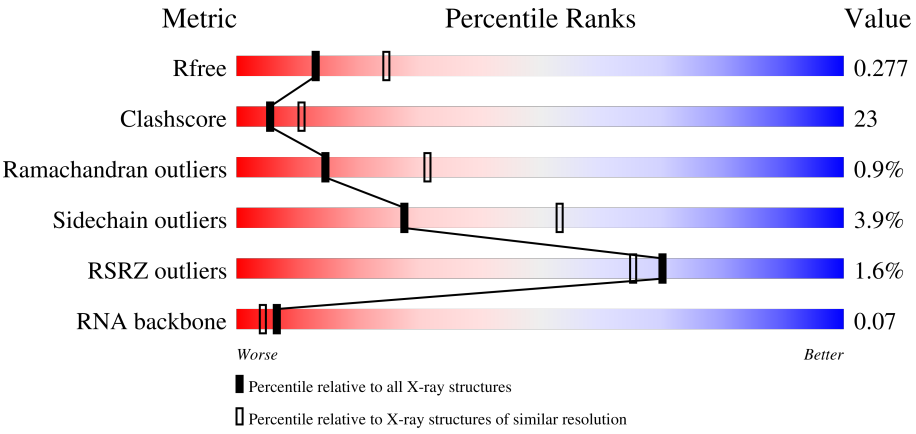
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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}	164625	3775 (2.60-2.60)
Clashscore	180529	4181 (2.60-2.60)
Ramachandran outliers	177936	4129 (2.60-2.60)
Sidechain outliers	177891	4129 (2.60-2.60)
RSRZ outliers	164620	3775 (2.60-2.60)
RNA backbone	3690	1025 (2.88-2.32)

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	O	7	29% 14% 57% 14% 14%
1	P	7	29% 29% 29% 29% 14%
1	Q	7	29% 29% 57% 14%
1	R	7	14% 71% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	S	7	29% 29% 57% 14%
1	T	7	29% 14% 57% 29%
1	U	7	29% 14% 29% 29% 14% 14%
2	A	77	% 56% 35% 6%
2	B	77	52% 39% 8%
2	C	77	60% 29% 8%
2	D	77	57% 32% 8%
2	E	77	52% 36% 8%
2	F	77	% 52% 36% 8%
2	G	77	45% 47% 8%
2	H	77	65% 26% 8%
2	I	77	% 64% 23% 5% 8%
2	J	77	61% 29% 8%
2	K	77	62% 27% 8%
2	L	77	61% 31% 8%
2	M	77	68% 25% 8%
2	N	77	3% 58% 31% 8%

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	U5P	B	411	-	-	X	-
4	U5P	J	419	-	-	X	-
4	U5P	K	476	-	-	X	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 9227 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 RNA chain called 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	O	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
1	P	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
1	Q	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
1	R	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
1	S	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			
1	T	5	Total	C	N	O	P	0	0	0
			100	45	10	40	5			
1	U	6	Total	C	N	O	P	0	0	0
			120	54	12	48	6			

- Molecule 2 is a protein called PUTATIVE SNRNP SM-LIKE PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	72	Total	C	N	O	S	0	0	0
			570	365	102	102	1			
2	B	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	C	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	D	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	E	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	F	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	G	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	I	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	J	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	K	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	L	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	M	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			
2	N	71	Total	C	N	O	S	0	0	0
			565	362	101	101	1			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	99	GLY	-	cloning artifact	UNP Q9V0Y8
A	102	ALA	-	cloning artifact	UNP Q9V0Y8
B	199	GLY	-	cloning artifact	UNP Q9V0Y8
B	202	ALA	-	cloning artifact	UNP Q9V0Y8
C	299	GLY	-	cloning artifact	UNP Q9V0Y8
C	302	ALA	-	cloning artifact	UNP Q9V0Y8
D	399	GLY	-	cloning artifact	UNP Q9V0Y8
D	402	ALA	-	cloning artifact	UNP Q9V0Y8
E	499	GLY	-	cloning artifact	UNP Q9V0Y8
E	502	ALA	-	cloning artifact	UNP Q9V0Y8
F	599	GLY	-	cloning artifact	UNP Q9V0Y8
F	602	ALA	-	cloning artifact	UNP Q9V0Y8
G	699	GLY	-	cloning artifact	UNP Q9V0Y8
G	702	ALA	-	cloning artifact	UNP Q9V0Y8
H	99	GLY	-	cloning artifact	UNP Q9V0Y8
H	102	ALA	-	cloning artifact	UNP Q9V0Y8
I	199	GLY	-	cloning artifact	UNP Q9V0Y8
I	202	ALA	-	cloning artifact	UNP Q9V0Y8
J	299	GLY	-	cloning artifact	UNP Q9V0Y8
J	302	ALA	-	cloning artifact	UNP Q9V0Y8
K	399	GLY	-	cloning artifact	UNP Q9V0Y8
K	402	ALA	-	cloning artifact	UNP Q9V0Y8
L	499	GLY	-	cloning artifact	UNP Q9V0Y8
L	502	ALA	-	cloning artifact	UNP Q9V0Y8
M	599	GLY	-	cloning artifact	UNP Q9V0Y8

Continued on next page...

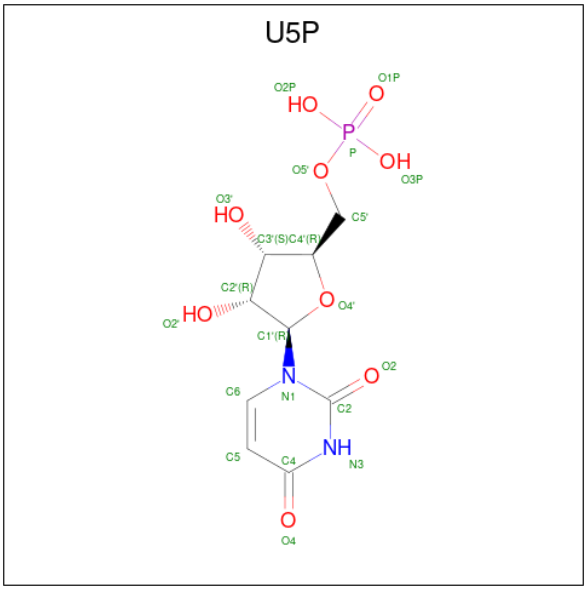
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	602	ALA	-	cloning artifact	UNP Q9V0Y8
N	699	GLY	-	cloning artifact	UNP Q9V0Y8
N	702	ALA	-	cloning artifact	UNP Q9V0Y8

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	O	1	Total Ca 1 1	0	0
3	P	1	Total Ca 1 1	0	0
3	Q	1	Total Ca 1 1	0	0
3	R	1	Total Ca 1 1	0	0
3	S	1	Total Ca 1 1	0	0
3	T	1	Total Ca 1 1	0	0
3	U	1	Total Ca 1 1	0	0

- Molecule 4 is URIDINE-5'-MONOPHOSPHATE (three-letter code: U5P) (formula: $C_9H_{13}N_2O_9P$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	B	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	C	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	D	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	E	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	F	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	G	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	H	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	I	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	J	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	K	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	L	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	M	1	Total	C	N	O	P	0	0
			20	9	2	8	1		
4	M	1	Total	C	N	O	P	0	0
			20	9	2	8	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	O	6	Total	O	0	0
			6	6		
5	P	7	Total	O	0	0
			7	7		
5	Q	9	Total	O	0	0
			9	9		
5	R	8	Total	O	0	0
			8	8		
5	S	6	Total	O	0	0
			6	6		
5	T	8	Total	O	0	0
			8	8		

Continued on next page...

Continued from previous page...

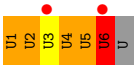
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	U	3	Total O 3 3	0	0
5	A	9	Total O 9 9	0	0
5	B	7	Total O 7 7	0	0
5	C	9	Total O 9 9	0	0
5	D	16	Total O 16 16	0	0
5	E	13	Total O 13 13	0	0
5	F	7	Total O 7 7	0	0
5	G	11	Total O 11 11	0	0
5	H	14	Total O 14 14	0	0
5	I	11	Total O 11 11	0	0
5	J	11	Total O 11 11	0	0
5	K	11	Total O 11 11	0	0
5	L	7	Total O 7 7	0	0
5	M	17	Total O 17 17	0	0
5	N	15	Total O 15 15	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.

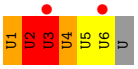
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

Chain O:



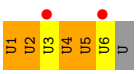
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

Chain P:



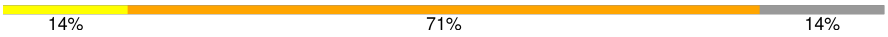
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

Chain Q:



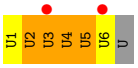
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

Chain R:



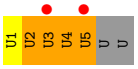
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

Chain S:

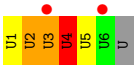


• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'

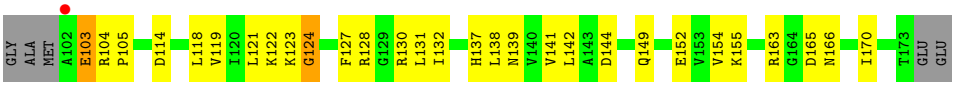
Chain T:



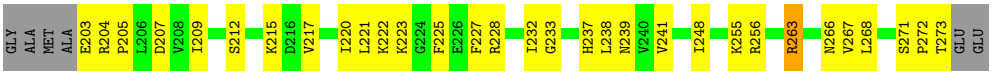
• Molecule 1: 5'-R(P*UP*UP*UP*UP*UP*UP*U)-3'



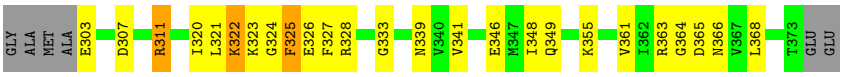
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN



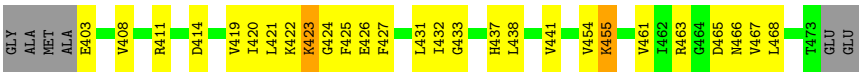
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN



• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

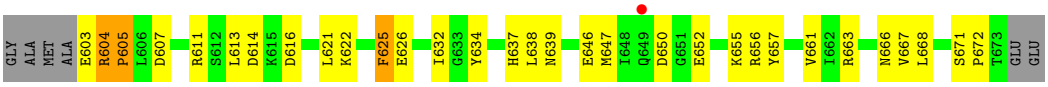


• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN



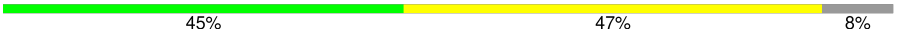
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN



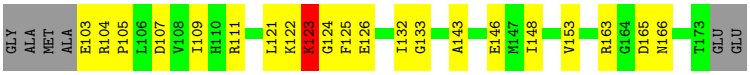
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain G:



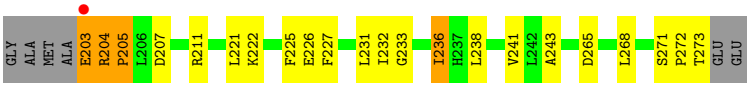
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain H:



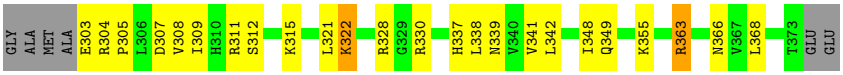
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain I:



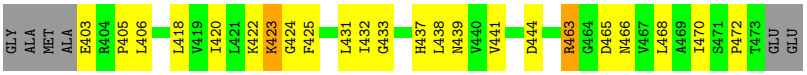
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain J:



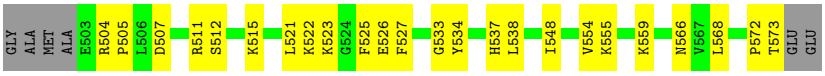
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain K:



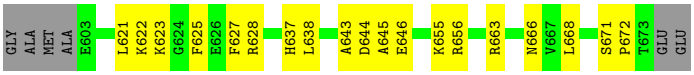
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain L:



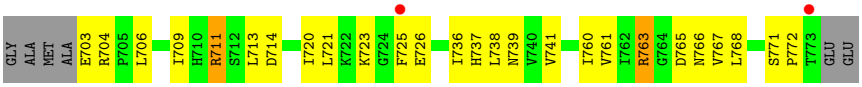
• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain M:



• Molecule 2: PUTATIVE SNRNP SM-LIKE PROTEIN

Chain N:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	68.00Å 68.00Å 84.80Å 105.00° 108.80° 100.00°	Depositor
Resolution (Å)	30.00 – 2.60 30.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	90.5 (30.00-2.60) 90.3 (30.00-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	4.19 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.212 , 0.282 0.206 , 0.277	Depositor DCC
R_{free} test set	1862 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	35.8	Xtriage
Anisotropy	0.429	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 63.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.064 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9227	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 9.17% 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: U5P, CA

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	O	0.60	0/131	1.04	0/200
1	P	1.01	0/131	2.93	8/200 (4.0%)
1	Q	0.61	0/131	0.75	0/200
1	R	0.57	0/131	0.87	0/200
1	S	0.61	0/131	0.95	0/200
1	T	0.69	0/109	0.99	0/166
1	U	0.74	0/131	0.88	2/200 (1.0%)
2	A	0.43	0/577	0.71	0/777
2	B	0.34	0/572	0.64	0/770
2	C	0.43	0/572	0.65	0/770
2	D	0.37	0/572	0.66	0/770
2	E	0.43	0/572	0.73	0/770
2	F	0.39	0/572	0.65	0/770
2	G	0.39	0/572	0.64	0/770
2	H	0.39	0/572	0.63	0/770
2	I	0.59	1/572 (0.2%)	0.78	2/770 (0.3%)
2	J	0.41	0/572	0.62	0/770
2	K	0.35	0/572	0.62	0/770
2	L	0.38	0/572	0.65	0/770
2	M	0.41	0/572	0.63	0/770
2	N	0.38	0/572	0.69	1/770 (0.1%)
All	All	0.45	1/8908 (0.0%)	0.78	13/12153 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	O	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	204	ARG	N-CA	5.13	1.56	1.46

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	P	2	U	P-O3'-C3'	-27.78	86.36	119.70
1	P	3	U	O5'-P-OP2	-17.89	89.23	110.70
1	P	2	U	O3'-P-O5'	8.56	120.27	104.00
1	P	3	U	O5'-P-OP1	-7.25	99.17	105.70
1	P	1	U	P-O3'-C3'	6.98	128.07	119.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	O	6	U	Sidechain

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	O	120	0	61	8	0
1	P	120	0	61	7	0
1	Q	120	0	61	16	0
1	R	120	0	61	8	0
1	S	120	0	61	5	0
1	T	100	0	51	7	0
1	U	120	0	61	5	0
2	A	570	0	599	35	0
2	B	565	0	594	33	0
2	C	565	0	594	39	0
2	D	565	0	594	23	0
2	E	565	0	594	36	0
2	F	565	0	594	32	0
2	G	565	0	594	44	0
2	H	565	0	594	21	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	I	565	0	593	25	0
2	J	565	0	594	23	0
2	K	565	0	594	19	0
2	L	565	0	594	21	0
2	M	565	0	594	24	0
2	N	565	0	594	30	0
3	O	1	0	0	0	0
3	P	1	0	0	0	0
3	Q	1	0	0	0	0
3	R	1	0	0	0	0
3	S	1	0	0	0	0
3	T	1	0	0	0	0
3	U	1	0	0	0	0
4	A	20	0	11	5	0
4	B	20	0	11	10	0
4	C	20	0	11	6	0
4	D	20	0	11	3	0
4	E	20	0	11	3	0
4	F	20	0	11	3	0
4	G	20	0	11	2	0
4	H	20	0	11	3	0
4	I	20	0	11	4	0
4	J	20	0	11	9	0
4	K	20	0	11	9	0
4	L	20	0	11	5	0
4	M	40	0	22	6	0
5	A	9	0	0	2	0
5	B	7	0	0	1	0
5	C	9	0	0	1	0
5	D	16	0	0	1	0
5	E	13	0	0	2	0
5	F	7	0	0	1	0
5	G	11	0	0	1	0
5	H	14	0	0	4	0
5	I	11	0	0	2	0
5	J	11	0	0	2	0
5	K	11	0	0	0	0
5	L	7	0	0	0	0
5	M	17	0	0	1	0
5	N	15	0	0	3	0
5	O	6	0	0	1	0
5	P	7	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	Q	9	0	0	1	0
5	R	8	0	0	1	0
5	S	6	0	0	0	0
5	T	8	0	0	0	0
5	U	3	0	0	0	0
All	All	9227	0	8891	417	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 23.

The worst 5 of 417 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)
2:I:203:GLU:HB2	5:I:421:HOH:O	1.33	1.22
2:C:311:ARG:HG3	2:C:311:ARG:HH11	0.99	1.15
2:H:122:LYS:NZ	2:I:265:ASP:OD1	1.80	1.15
4:J:419:U5P:H3'	4:K:476:U5P:O2P	1.50	1.08
4:J:419:U5P:H3'	4:K:476:U5P:P	1.93	1.07

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
2	A	70/77 (91%)	64 (91%)	4 (6%)	2 (3%)	3 6
2	B	69/77 (90%)	65 (94%)	4 (6%)	0	100 100
2	C	69/77 (90%)	63 (91%)	5 (7%)	1 (1%)	9 19
2	D	69/77 (90%)	63 (91%)	4 (6%)	2 (3%)	3 6
2	E	69/77 (90%)	62 (90%)	6 (9%)	1 (1%)	9 19

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	F	69/77 (90%)	63 (91%)	6 (9%)	0	100 100
2	G	69/77 (90%)	65 (94%)	4 (6%)	0	100 100
2	H	69/77 (90%)	64 (93%)	4 (6%)	1 (1%)	9 19
2	I	69/77 (90%)	63 (91%)	6 (9%)	0	100 100
2	J	69/77 (90%)	63 (91%)	6 (9%)	0	100 100
2	K	69/77 (90%)	65 (94%)	2 (3%)	2 (3%)	3 6
2	L	69/77 (90%)	61 (88%)	8 (12%)	0	100 100
2	M	69/77 (90%)	65 (94%)	4 (6%)	0	100 100
2	N	69/77 (90%)	63 (91%)	6 (9%)	0	100 100
All	All	967/1078 (90%)	889 (92%)	69 (7%)	9 (1%)	14 31

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	103	GLU
2	D	423	LYS
2	H	123	LYS
2	K	423	LYS
2	K	424	GLY

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
2	A	62/65 (95%)	62 (100%)	0	100 100
2	B	62/65 (95%)	61 (98%)	1 (2%)	58 79
2	C	62/65 (95%)	59 (95%)	3 (5%)	21 44
2	D	62/65 (95%)	60 (97%)	2 (3%)	34 60
2	E	62/65 (95%)	57 (92%)	5 (8%)	9 20
2	F	62/65 (95%)	57 (92%)	5 (8%)	9 20
2	G	62/65 (95%)	60 (97%)	2 (3%)	34 60

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	H	62/65 (95%)	61 (98%)	1 (2%)	58 79
2	I	62/65 (95%)	60 (97%)	2 (3%)	34 60
2	J	62/65 (95%)	59 (95%)	3 (5%)	21 44
2	K	62/65 (95%)	59 (95%)	3 (5%)	21 44
2	L	62/65 (95%)	60 (97%)	2 (3%)	34 60
2	M	62/65 (95%)	62 (100%)	0	100 100
2	N	62/65 (95%)	57 (92%)	5 (8%)	9 20
All	All	868/910 (95%)	834 (96%)	34 (4%)	27 53

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	L	526	GLU
2	N	711	ARG
2	N	763	ARG
2	F	605	PRO
2	F	604	ARG

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

Mol	Chain	Res	Type
2	A	149	GLN
2	G	766	ASN
2	H	166	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	O	6/7 (85%)	3 (50%)	2 (33%)
1	P	6/7 (85%)	3 (50%)	1 (16%)
1	Q	6/7 (85%)	3 (50%)	1 (16%)
1	R	6/7 (85%)	4 (66%)	1 (16%)
1	S	5/7 (71%)	5 (100%)	1 (20%)
1	T	4/7 (57%)	4 (100%)	1 (25%)
1	U	5/7 (71%)	3 (60%)	0
All	All	38/49 (77%)	25 (65%)	7 (18%)

5 of 25 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	O	2	U
1	O	5	U
1	O	6	U
1	P	2	U
1	P	3	U

5 of 7 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	Q	1	U
1	R	1	U
1	T	4	U
1	S	4	U
1	P	1	U

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 21 ligands modelled in this entry, 7 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	U5P	A	410	-	18,21,22	0.27	0	25,30,33	0.52	0
4	U5P	J	419	-	18,21,22	0.26	0	25,30,33	0.58	1 (4%)
4	U5P	B	411	-	18,21,22	0.28	0	25,30,33	0.45	0
4	U5P	D	476	-	18,21,22	0.23	0	25,30,33	0.43	0
4	U5P	M	422	-	18,21,22	0.29	0	25,30,33	0.56	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
4	U5P	I	418	-	18,21,22	0.28	0	25,30,33	0.52	0
4	U5P	F	415	-	18,21,22	0.30	0	25,30,33	0.51	0
4	U5P	K	476	-	18,21,22	0.45	0	25,30,33	1.39	3 (12%)
4	U5P	E	414	-	18,21,22	0.33	0	25,30,33	0.45	0
4	U5P	L	421	-	18,21,22	0.32	0	25,30,33	0.50	0
4	U5P	M	423	-	18,21,22	0.33	0	25,30,33	0.37	0
4	U5P	H	417	-	18,21,22	0.29	0	25,30,33	0.56	0
4	U5P	G	416	-	18,21,22	0.28	0	25,30,33	0.54	1 (4%)
4	U5P	C	412	-	18,21,22	0.30	0	25,30,33	0.36	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
4	U5P	A	410	-	-	1/7/25/26	0/2/2/2
4	U5P	J	419	-	-	0/7/25/26	0/2/2/2
4	U5P	B	411	-	-	0/7/25/26	0/2/2/2
4	U5P	D	476	-	-	3/7/25/26	0/2/2/2
4	U5P	M	422	-	-	0/7/25/26	0/2/2/2
4	U5P	I	418	-	-	0/7/25/26	0/2/2/2
4	U5P	F	415	-	-	0/7/25/26	0/2/2/2
4	U5P	K	476	-	-	2/7/25/26	0/2/2/2
4	U5P	E	414	-	-	2/7/25/26	0/2/2/2
4	U5P	L	421	-	-	0/7/25/26	0/2/2/2
4	U5P	M	423	-	-	4/7/25/26	0/2/2/2
4	U5P	H	417	-	-	1/7/25/26	0/2/2/2
4	U5P	G	416	-	-	0/7/25/26	0/2/2/2
4	U5P	C	412	-	-	3/7/25/26	0/2/2/2

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	K	476	U5P	O2'-C2'-C3'	-5.25	94.98	111.82
4	K	476	U5P	O3'-C3'-C2'	3.59	123.31	111.82
4	K	476	U5P	O2'-C2'-C1'	2.22	117.73	110.10
4	G	416	U5P	O3'-C3'-C2'	-2.08	105.16	111.82
4	J	419	U5P	O3'-C3'-C2'	-2.05	105.24	111.82

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	M	423	U5P	O4'-C1'-N1-C2
4	M	423	U5P	C2'-C1'-N1-C6
4	M	423	U5P	C2'-C1'-N1-C2
4	D	476	U5P	C3'-C4'-C5'-O5'
4	K	476	U5P	C3'-C4'-C5'-O5'

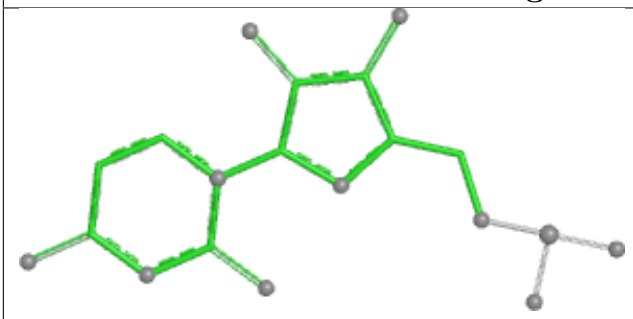
There are no ring outliers.

14 monomers are involved in 43 short contacts:

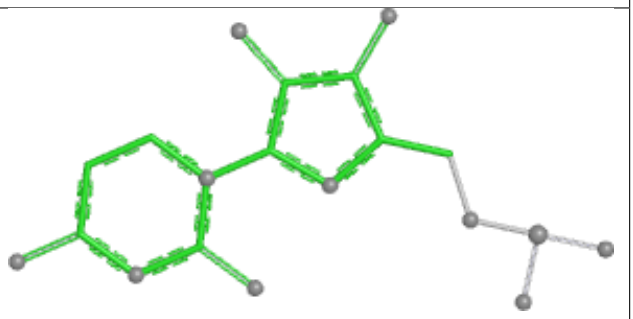
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	410	U5P	5	0
4	J	419	U5P	9	0
4	B	411	U5P	10	0
4	D	476	U5P	3	0
4	M	422	U5P	5	0
4	I	418	U5P	4	0
4	F	415	U5P	3	0
4	K	476	U5P	9	0
4	E	414	U5P	3	0
4	L	421	U5P	5	0
4	M	423	U5P	2	0
4	H	417	U5P	3	0
4	G	416	U5P	2	0
4	C	412	U5P	6	0

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.

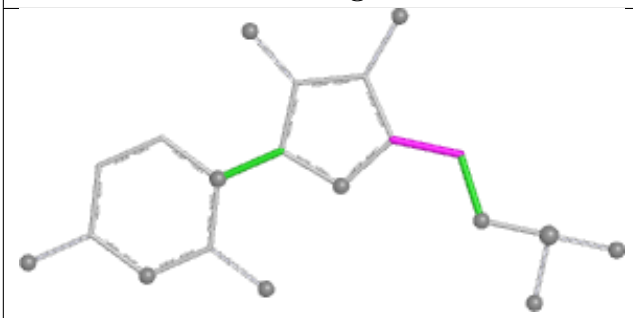
Ligand U5P A 410



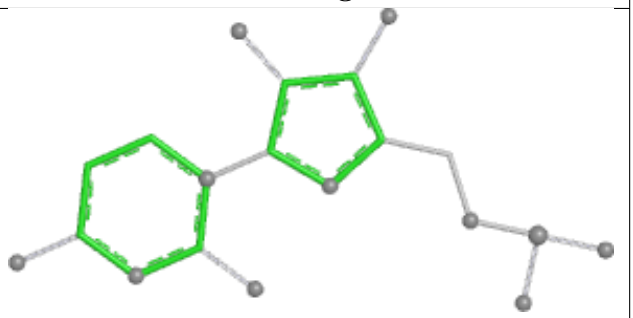
Bond lengths



Bond angles

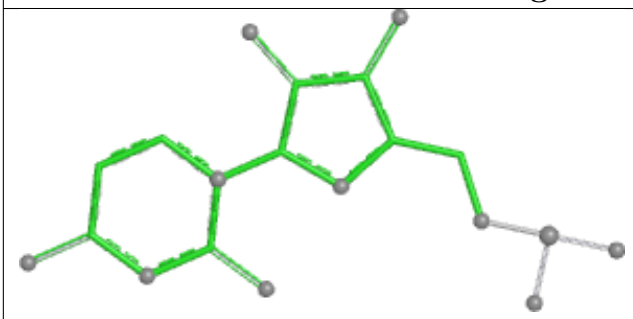


Torsions

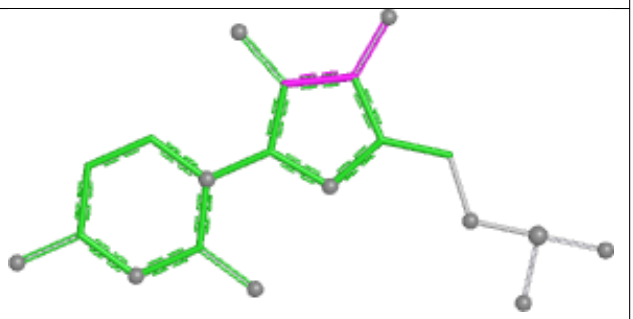


Rings

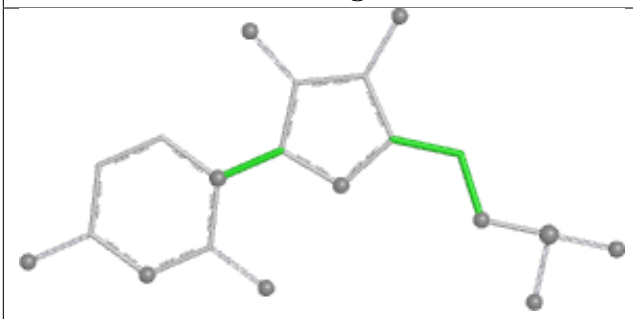
Ligand U5P J 419



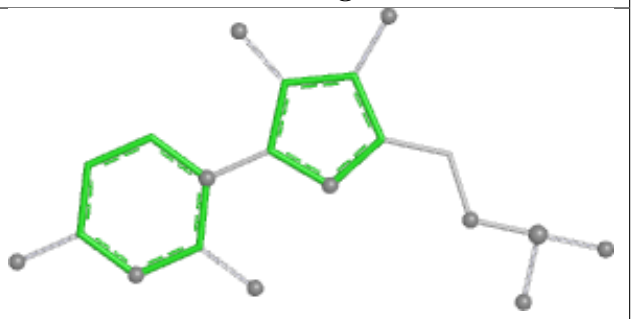
Bond lengths



Bond angles

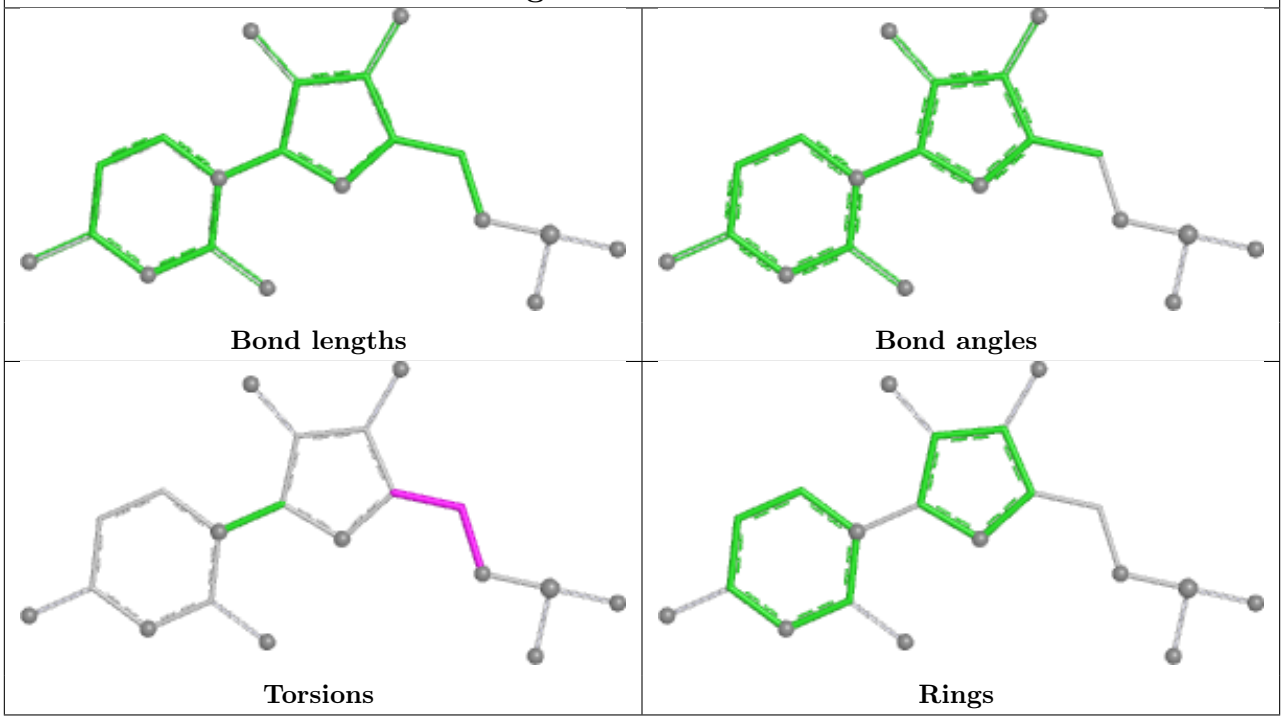


Torsions

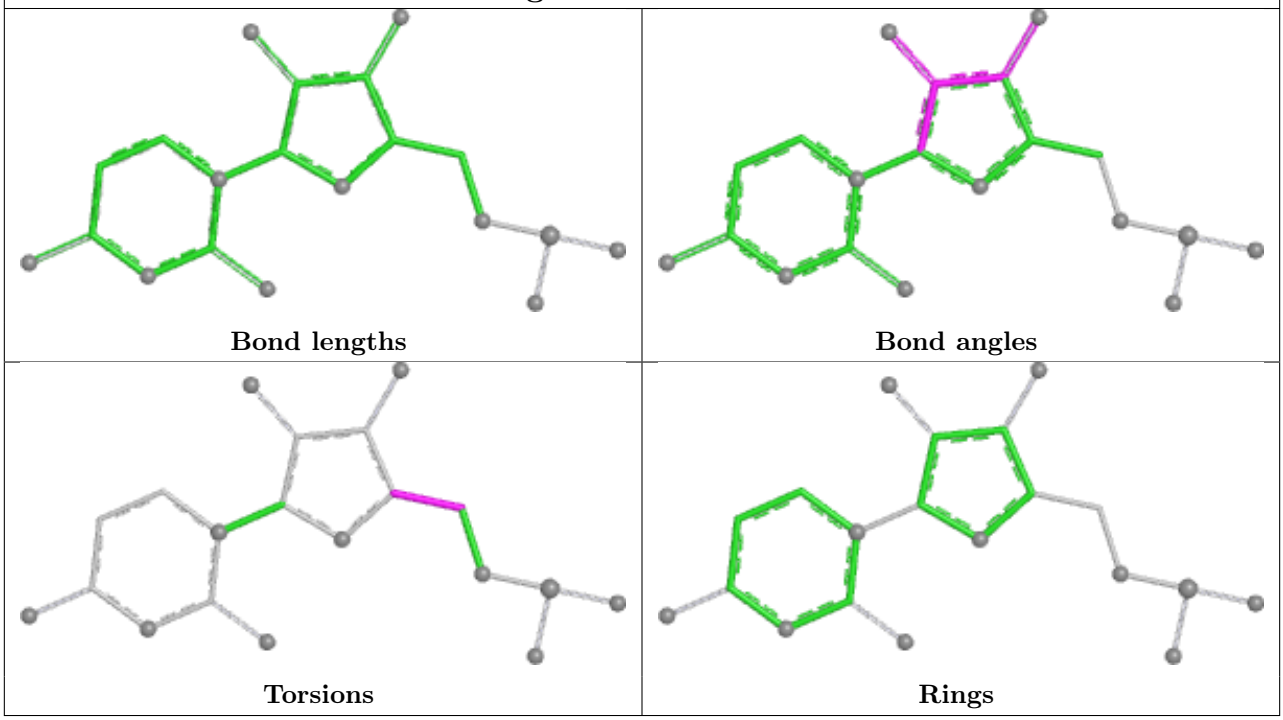


Rings

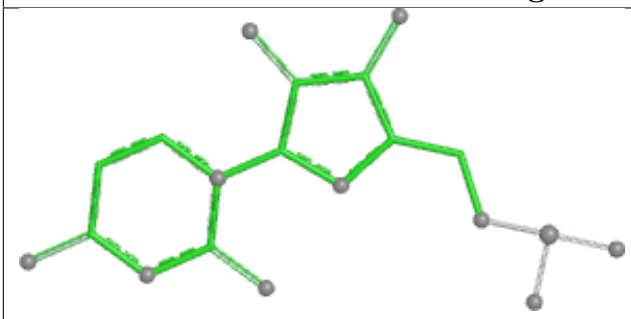
Ligand U5P D 476



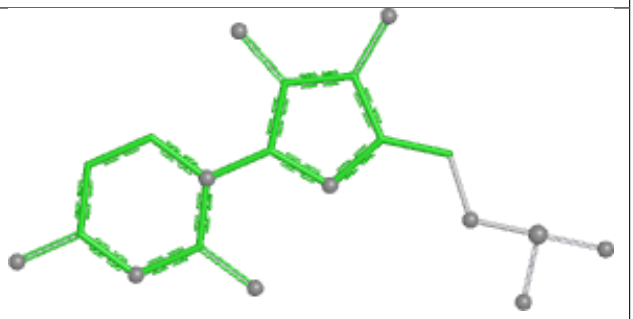
Ligand U5P K 476



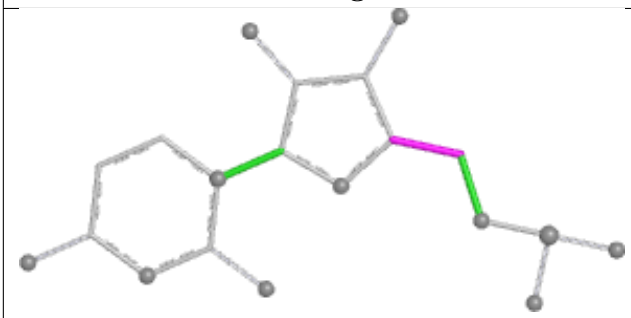
Ligand U5P E 414



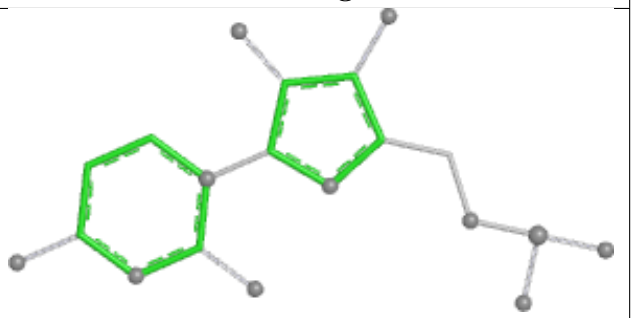
Bond lengths



Bond angles

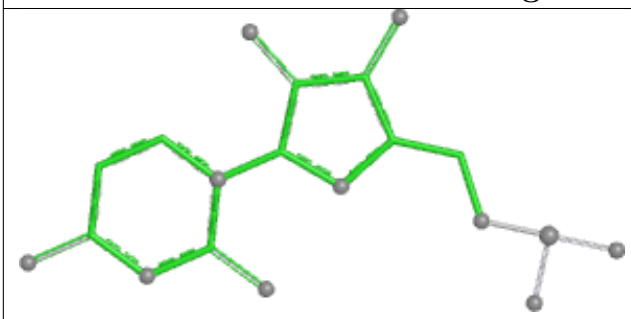


Torsions

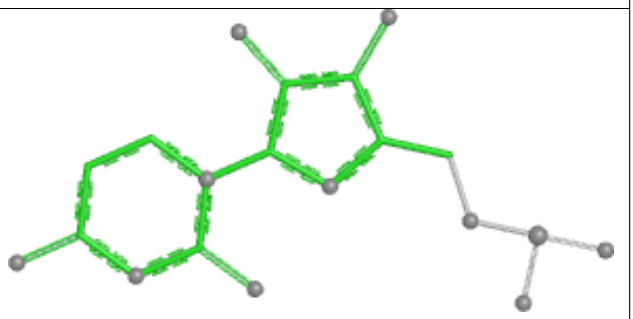


Rings

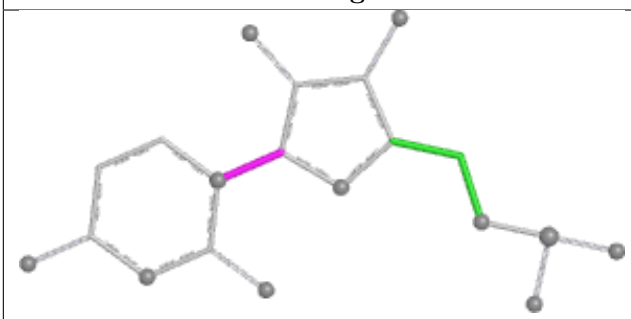
Ligand U5P M 423



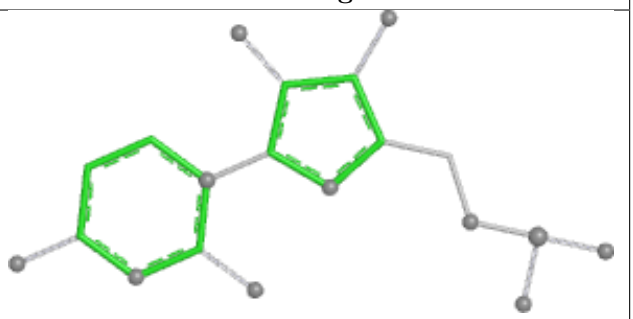
Bond lengths



Bond angles

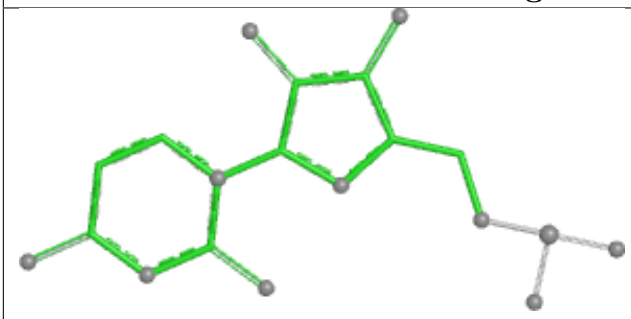


Torsions

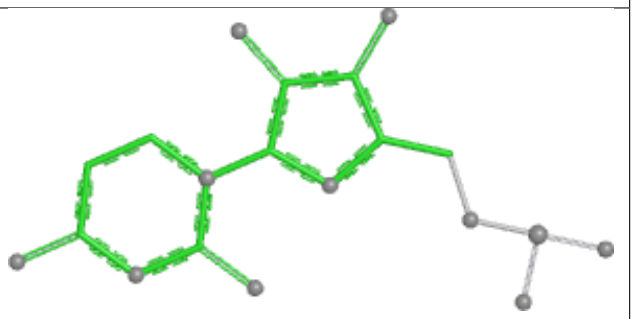


Rings

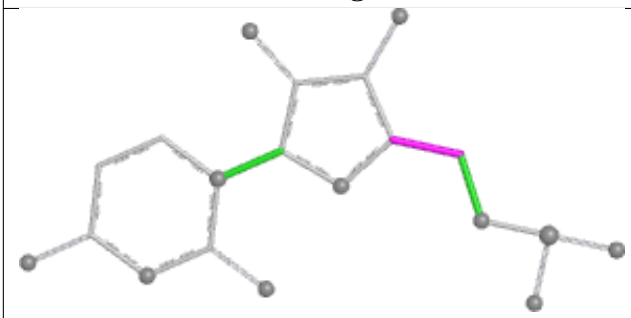
Ligand U5P H 417



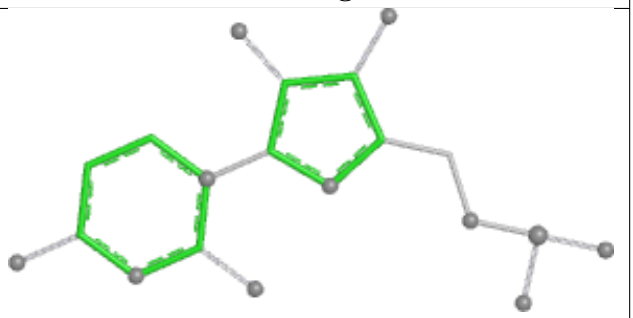
Bond lengths



Bond angles

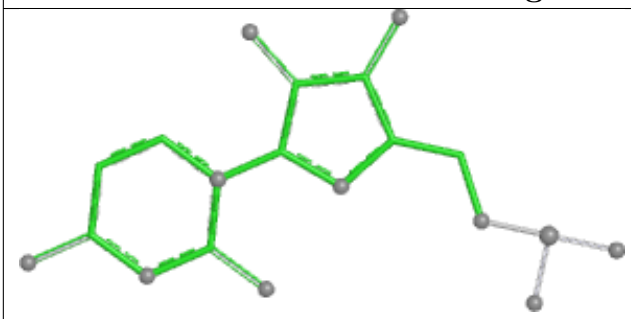


Torsions

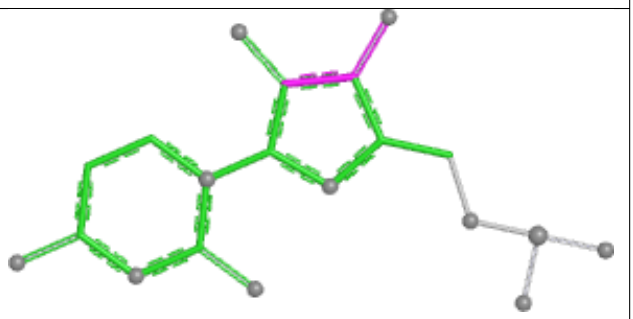


Rings

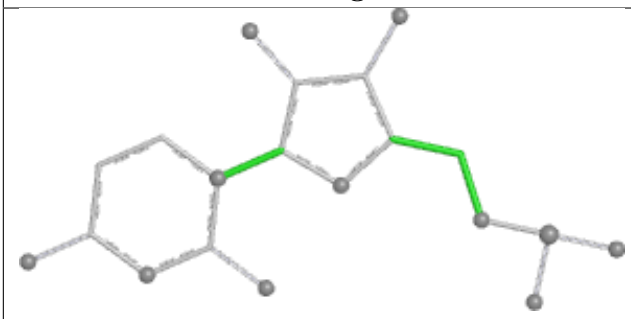
Ligand U5P G 416



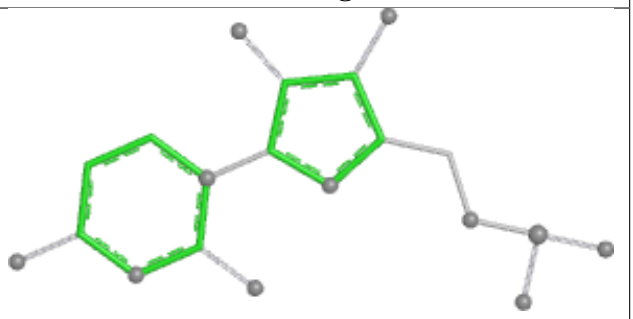
Bond lengths



Bond angles

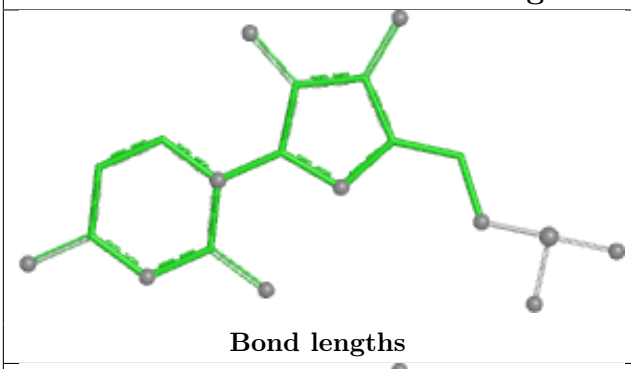


Torsions

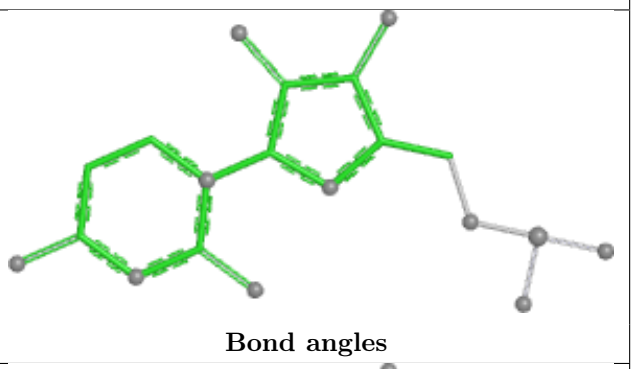


Rings

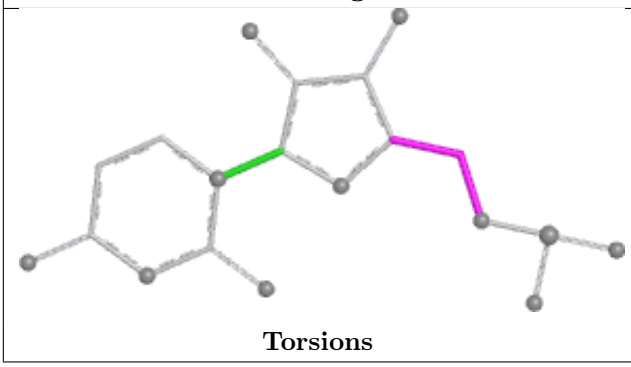
Ligand U5P C 412



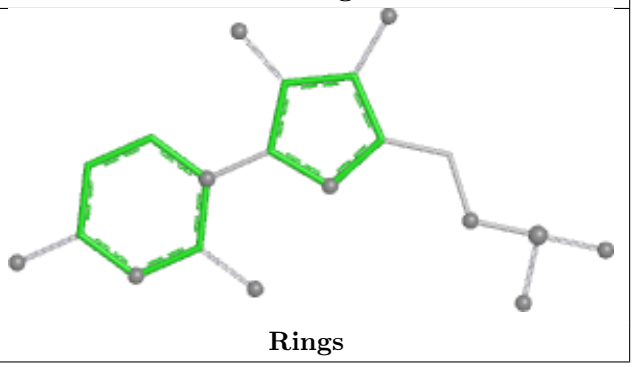
Bond lengths



Bond angles



Torsions



Rings

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	O	6/7 (85%)	1.52	2 (33%) 1 1	45, 56, 66, 87	2 (33%)
1	P	6/7 (85%)	1.83	2 (33%) 1 1	46, 50, 75, 76	4 (66%)
1	Q	6/7 (85%)	1.45	2 (33%) 1 1	42, 54, 69, 72	3 (50%)
1	R	6/7 (85%)	1.41	0 100 100	44, 50, 65, 73	2 (33%)
1	S	6/7 (85%)	2.06	2 (33%) 1 1	52, 57, 71, 72	3 (50%)
1	T	5/7 (71%)	1.67	2 (40%) 1 1	52, 59, 64, 80	4 (80%)
1	U	6/7 (85%)	1.81	2 (33%) 1 1	44, 53, 69, 72	3 (50%)
2	A	72/77 (93%)	-0.49	1 (1%) 73 68	19, 32, 52, 57	3 (4%)
2	B	71/77 (92%)	-0.48	0 100 100	16, 30, 52, 56	10 (14%)
2	C	71/77 (92%)	-0.59	0 100 100	12, 28, 49, 54	2 (2%)
2	D	71/77 (92%)	-0.57	0 100 100	13, 28, 47, 61	7 (9%)
2	E	71/77 (92%)	-0.47	0 100 100	16, 26, 47, 50	9 (12%)
2	F	71/77 (92%)	-0.43	1 (1%) 73 68	19, 30, 55, 73	4 (5%)
2	G	71/77 (92%)	-0.44	0 100 100	17, 31, 51, 62	6 (8%)
2	H	71/77 (92%)	-0.56	0 100 100	16, 29, 48, 55	2 (2%)
2	I	71/77 (92%)	-0.44	1 (1%) 73 68	16, 31, 52, 59	7 (9%)
2	J	71/77 (92%)	-0.44	0 100 100	17, 30, 55, 64	6 (8%)
2	K	71/77 (92%)	-0.58	0 100 100	18, 27, 51, 59	5 (7%)
2	L	71/77 (92%)	-0.50	0 100 100	16, 28, 51, 63	7 (9%)
2	M	71/77 (92%)	-0.59	0 100 100	15, 28, 42, 51	8 (11%)
2	N	71/77 (92%)	-0.40	2 (2%) 55 49	17, 28, 56, 69	8 (11%)
All	All	1036/1127 (91%)	-0.41	17 (1%) 70 65	12, 30, 55, 87	105 (10%)

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	U	3	U	4.0
1	Q	3	U	4.0
1	S	3	U	3.9
1	P	6	U	3.7
1	U	6	U	3.7

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
4	U5P	L	421	20/21	0.75	0.15	50,79,110,110	3
4	U5P	F	415	20/21	0.76	0.18	46,82,116,117	4
4	U5P	M	422	20/21	0.76	0.15	51,79,101,101	3
4	U5P	I	418	20/21	0.78	0.15	36,73,105,106	5
4	U5P	C	412	20/21	0.79	0.13	33,61,106,106	4
4	U5P	K	476	20/21	0.79	0.14	28,69,109,109	2
4	U5P	G	416	20/21	0.80	0.14	49,79,104,105	2
4	U5P	H	417	20/21	0.81	0.14	50,75,114,116	4
4	U5P	E	414	20/21	0.82	0.14	31,60,99,102	5
4	U5P	B	411	20/21	0.82	0.13	44,72,107,107	5
4	U5P	D	476	20/21	0.83	0.14	37,70,112,112	5
4	U5P	M	423	20/21	0.83	0.15	61,91,124,124	3
4	U5P	A	410	20/21	0.84	0.14	27,72,103,104	4
3	CA	R	390	1/1	0.88	0.08	51,51,51,51	0
4	U5P	J	419	20/21	0.88	0.13	52,77,111,112	6
3	CA	P	330	1/1	0.89	0.07	53,53,53,53	0
3	CA	O	310	1/1	0.92	0.07	45,45,45,45	0
3	CA	Q	380	1/1	0.94	0.06	46,46,46,46	0
3	CA	S	340	1/1	0.94	0.07	51,51,51,51	0

Continued on next page...

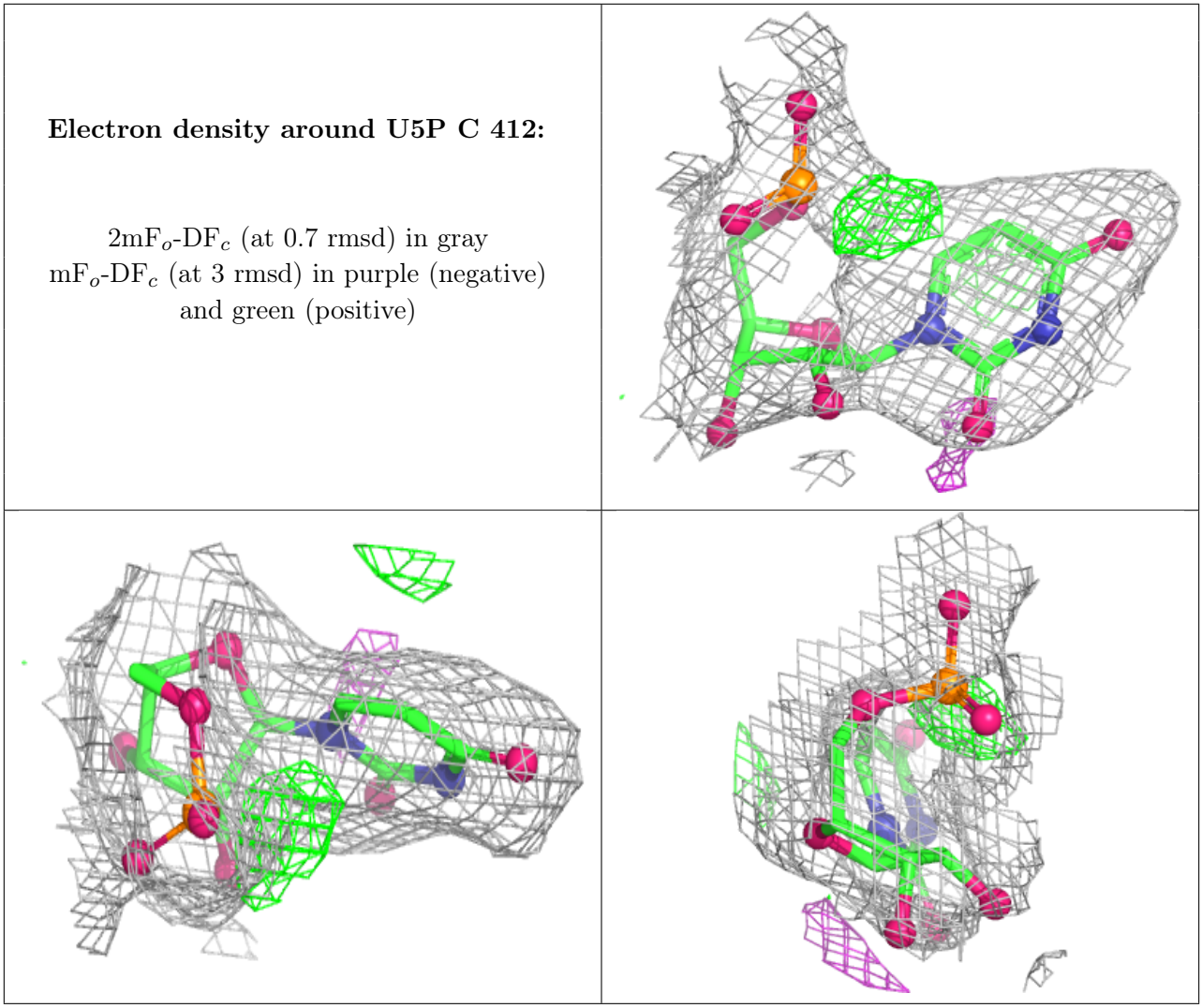
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CA	T	320	1/1	0.95	0.07	48,48,48,48	0
3	CA	U	370	1/1	0.96	0.06	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.

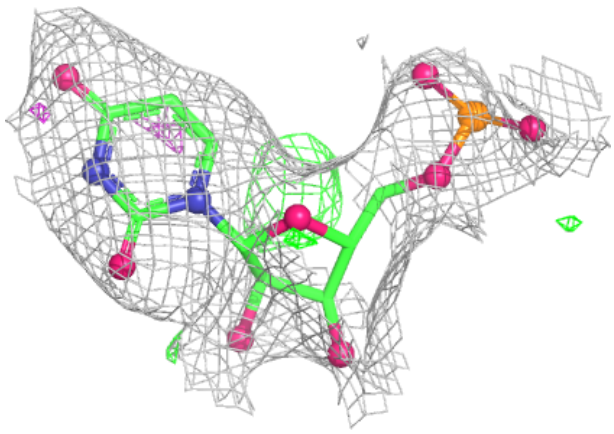
Electron density around U5P C 412:

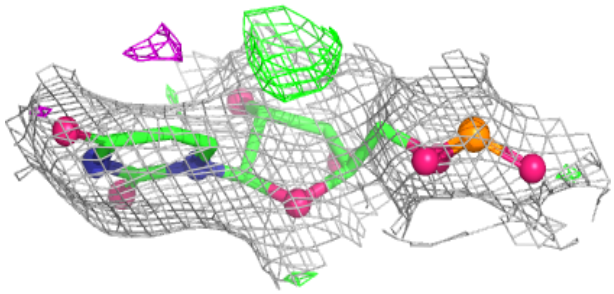
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

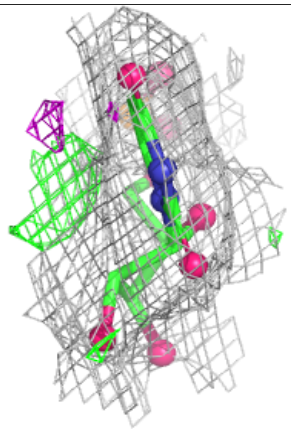


Electron density around U5P K 476:

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

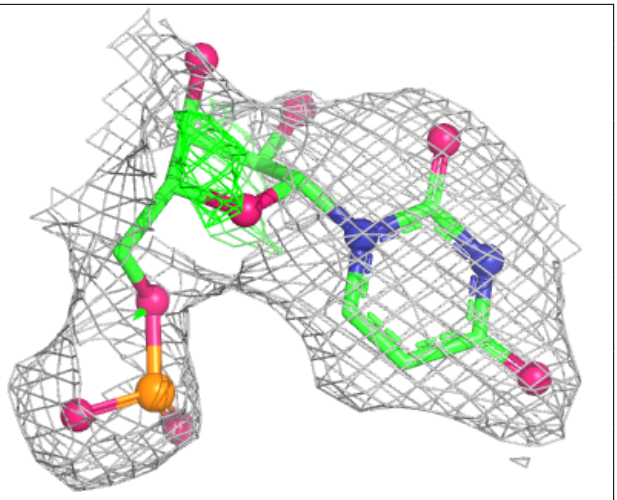


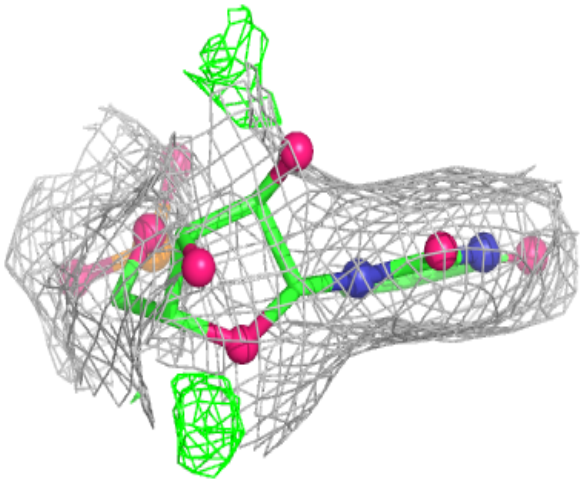


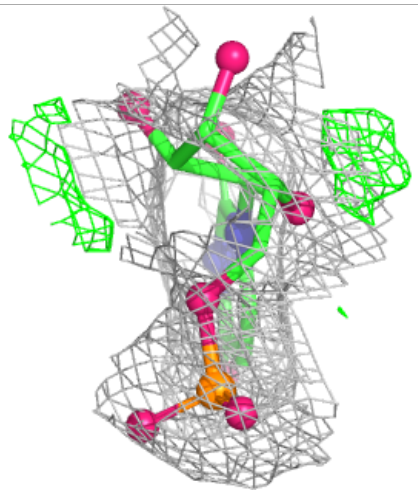


Electron density around U5P G 416:

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

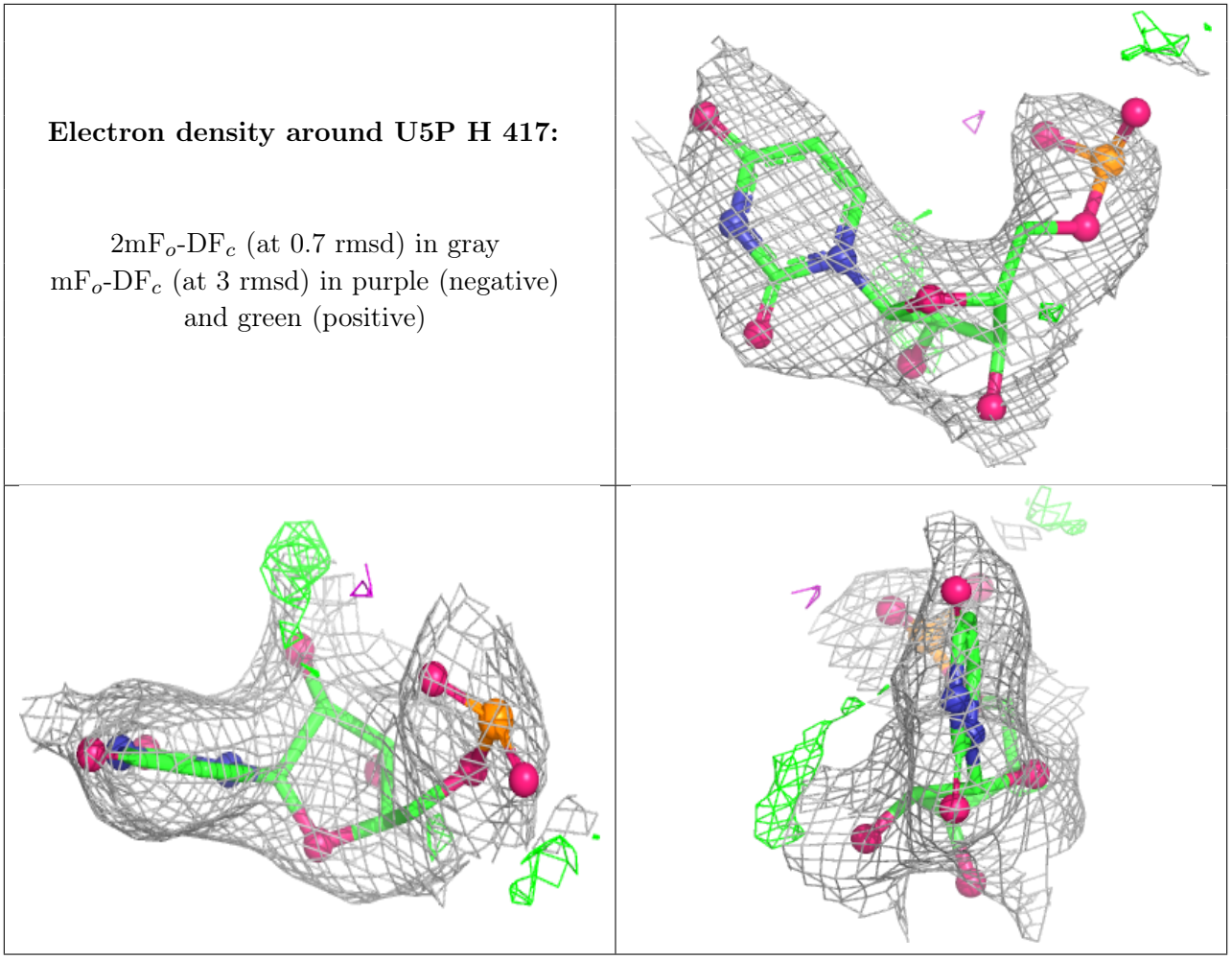






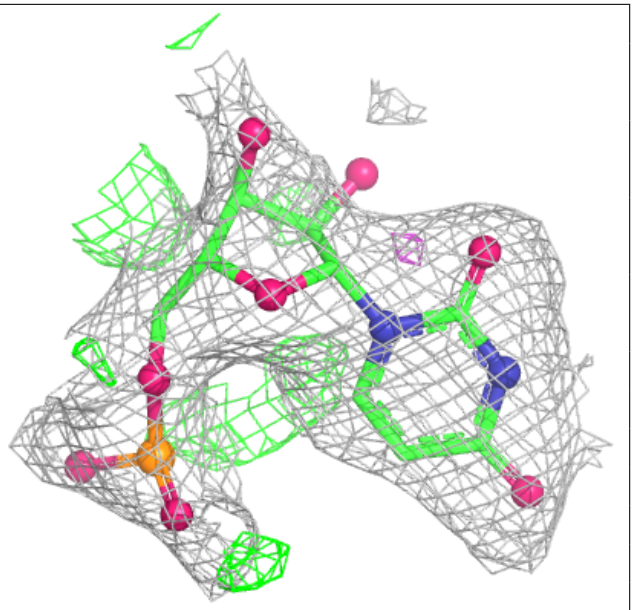
Electron density around U5P H 417:

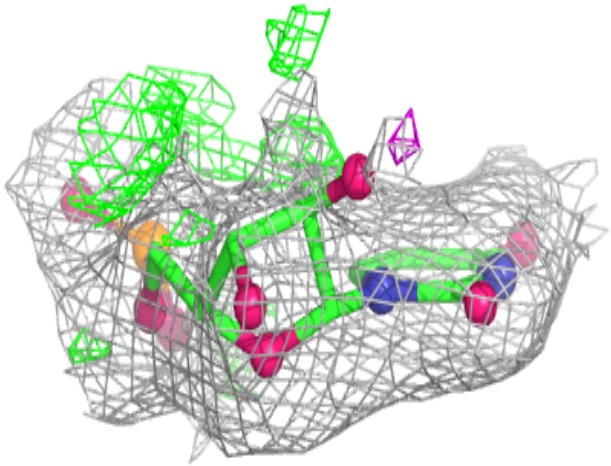
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

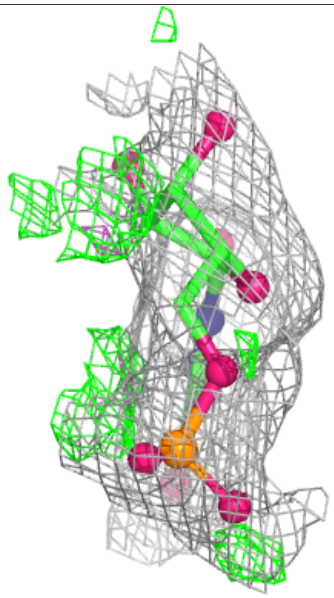


Electron density around U5P E 414:

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

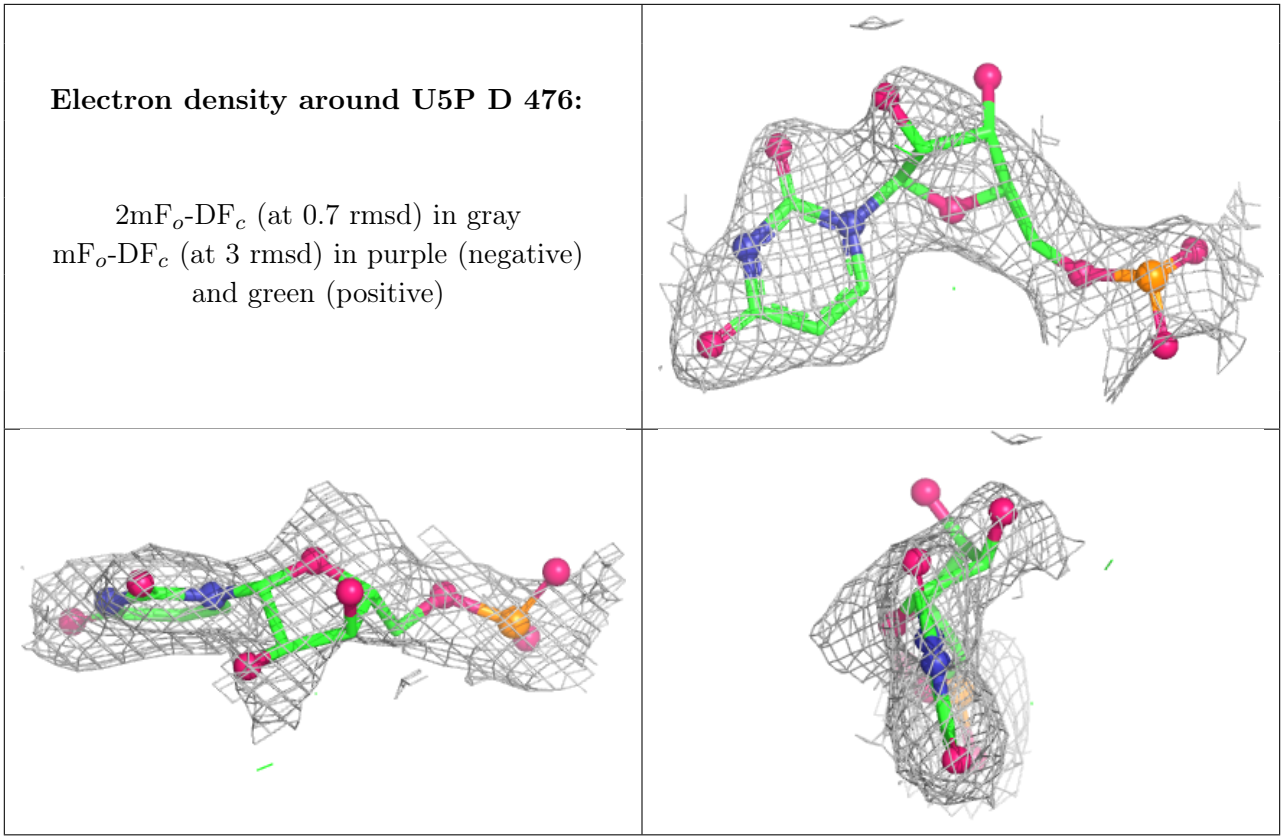






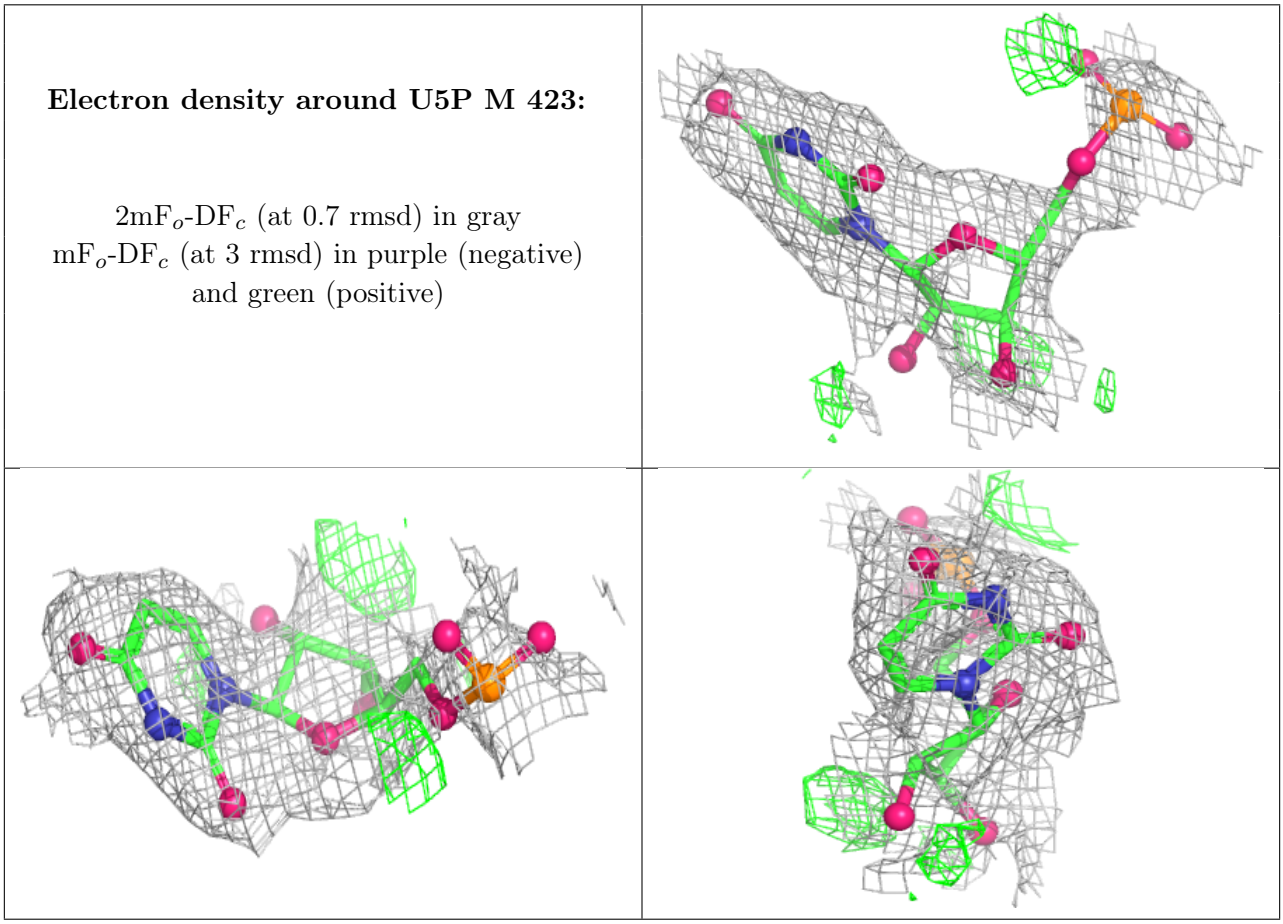
Electron density around U5P D 476:

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



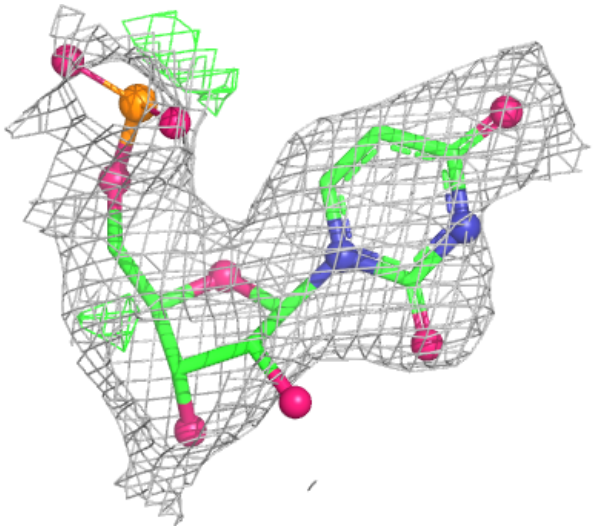
Electron density around U5P M 423:

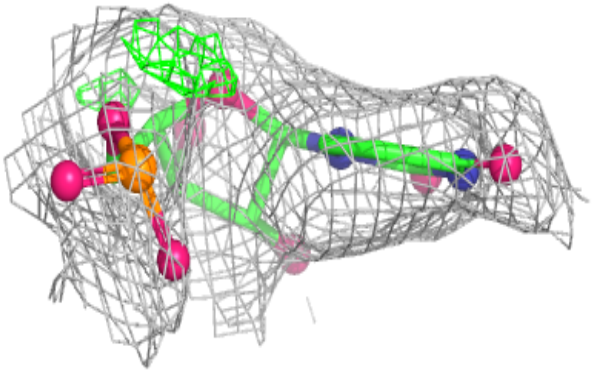
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

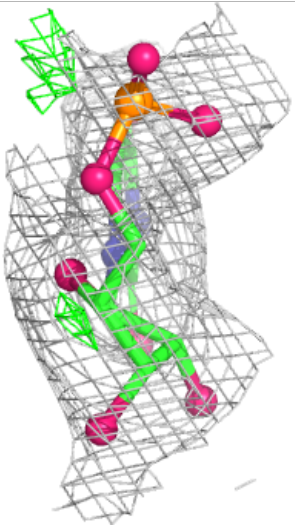


Electron density around U5P A 410:

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

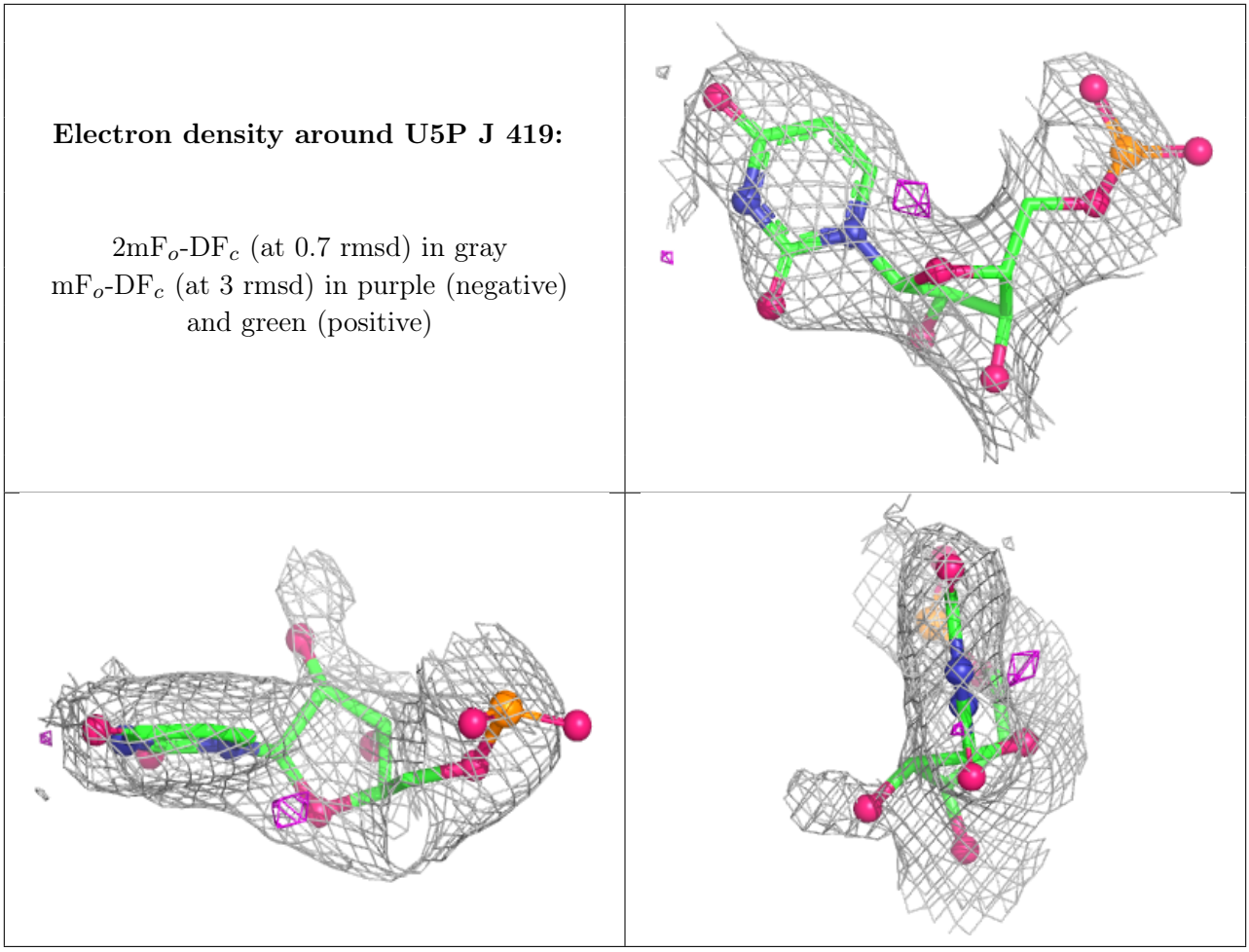






Electron density around U5P J 419:

$2mF_o$-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.