



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 2, 2021 – 09:08 PM EDT

PDB ID : 3K0J  
Title : Crystal structure of the E. coli ThiM riboswitch in complex with thiamine pyrophosphate and the U1A crystallization module  
Authors : Kulshina, N.; Edwards, T.E.; Ferre-D'Amare, A.R.  
Deposited on : 2009-09-24  
Resolution : 3.10 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)  
Xtrriage (Phenix) : 1.13  
EDS : 2.23.2  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.2

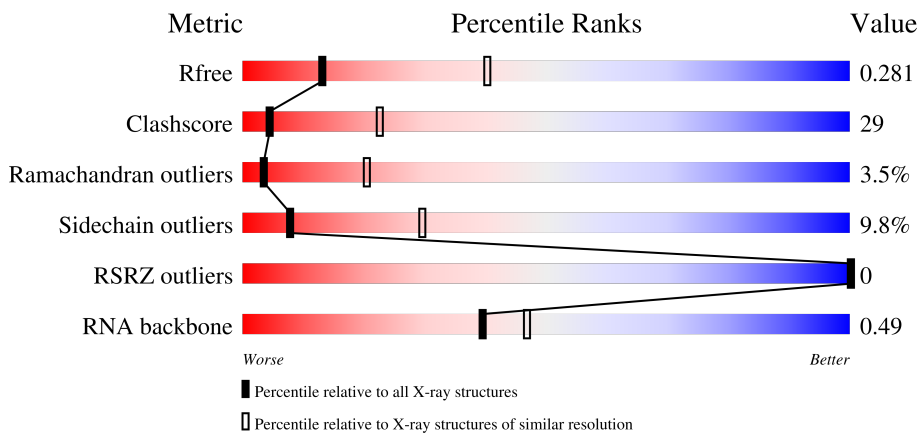
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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.



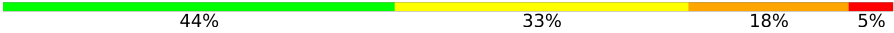
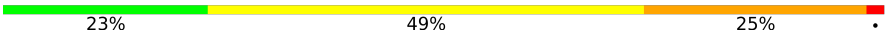
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)
RNA backbone	3102	1116 (3.40-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	96	 42% 43% 11% .
1	B	96	 41% 50% . . .
1	C	96	 51% 31% 6% . 10%
1	D	96	 35% 43% 7% 15%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	E	87	 44% 33% 18% 5%
2	F	87	 23% 49% 25% •

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6477 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 U1 small nuclear ribonucleoprotein A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	92	740	474	129	133	4	0	0	0
1	B	92	725	465	124	132	4	0	0	0
1	C	86	672	435	116	119	2	0	0	0
1	D	82	638	410	108	117	3	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	231	HIS	TYR	engineered mutation	UNP P09012
A	236	ARG	GLN	engineered mutation	UNP P09012
B	331	HIS	TYR	engineered mutation	UNP P09012
B	336	ARG	GLN	engineered mutation	UNP P09012
C	431	HIS	TYR	engineered mutation	UNP P09012
C	436	ARG	GLN	engineered mutation	UNP P09012
D	531	HIS	TYR	engineered mutation	UNP P09012
D	536	ARG	GLN	engineered mutation	UNP P09012

- Molecule 2 is a RNA chain called RNA (87-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	E	87	1829	816	326	601	86	0	0	0
2	F	87	1813	808	322	597	86	0	0	0

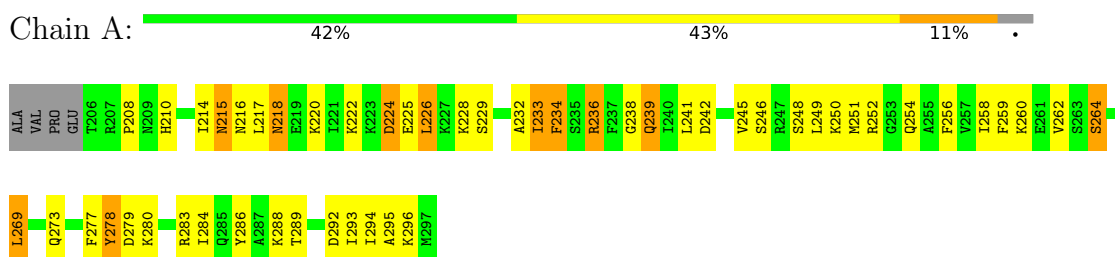
- Molecule 3 is THIAMINE DIPHOSPHATE (three-letter code: TPP) (formula: C<sub>12</sub>H<sub>19</sub>N<sub>4</sub>O<sub>7</sub>P<sub>2</sub>S).



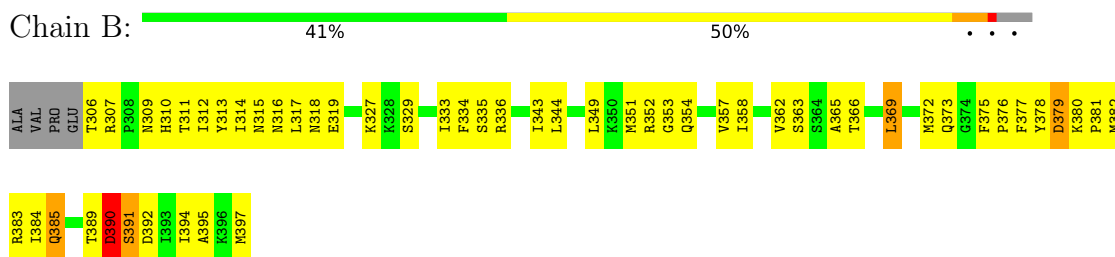
### 3 Residue-property plots [i](#)

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.

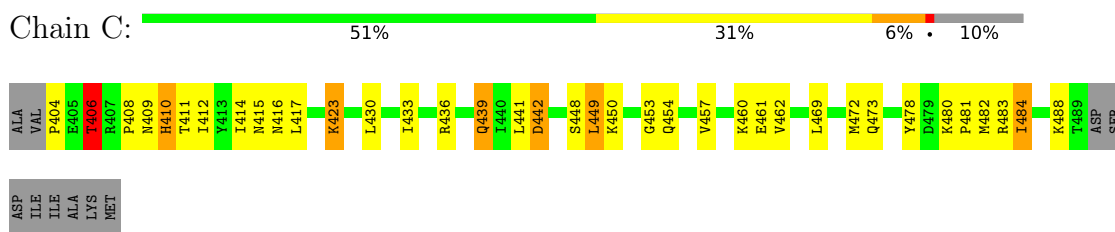
- Molecule 1: U1 small nuclear ribonucleoprotein A



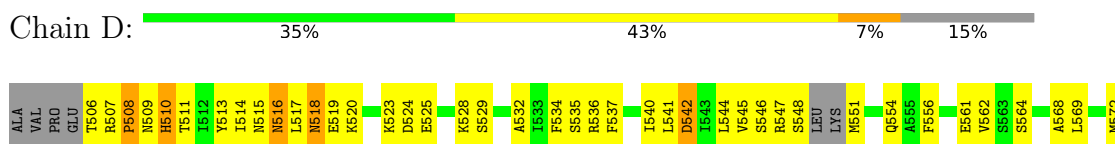
- Molecule 1: U1 small nuclear ribonucleoprotein A



- Molecule 1: U1 small nuclear ribonucleoprotein A

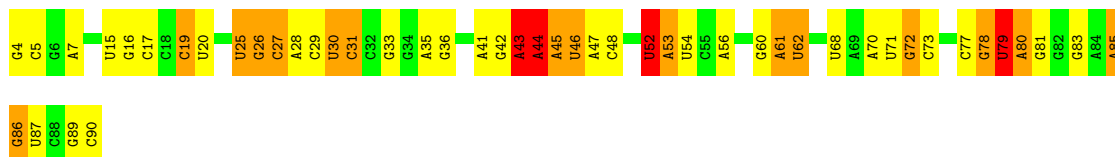


- Molecule 1: U1 small nuclear ribonucleoprotein A

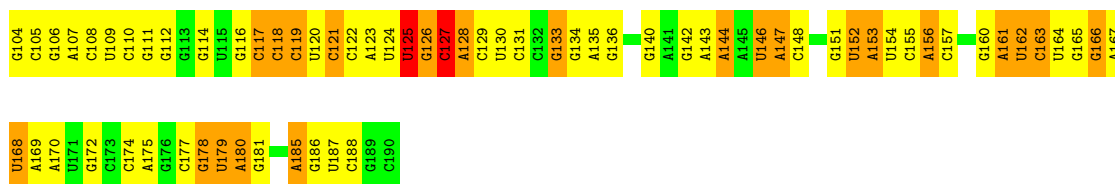




- Molecule 2: RNA (87-MER)



- Molecule 2: RNA (87-MER)



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.25Å 71.60Å 128.35Å 90.00° 94.62° 90.00°	Depositor
Resolution (Å)	29.50 – 3.10 29.50 – 3.08	Depositor EDS
% Data completeness (in resolution range)	93.1 (29.50-3.10) 92.5 (29.50-3.08)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.98 (at 3.05Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.197 , 0.279 0.198 , 0.281	Depositor DCC
$R_{free}$ test set	1595 reflections (9.14%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.6	Xtrriage
Anisotropy	0.664	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 57.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	6477	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% 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: MG, TPP

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	0.51	0/753	0.82	2/1008 (0.2%)
1	B	0.44	0/738	0.65	0/993
1	C	0.52	1/686 (0.1%)	0.71	1/924 (0.1%)
1	D	0.39	0/650	0.62	0/877
2	E	0.51	0/2044	0.84	4/3185 (0.1%)
2	F	0.44	0/2026	0.79	2/3157 (0.1%)
All	All	0.47	1/6897 (0.0%)	0.78	9/10144 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	404	PRO	C-N	-5.46	1.21	1.34

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	127	C	C2'-C3'-O3'	7.66	126.35	109.50
1	A	278	TYR	CB-CG-CD2	-5.62	117.63	121.00
2	E	44	A	O4'-C1'-N9	5.46	112.56	108.20
2	E	52	U	N1-C1'-C2'	5.36	120.97	114.00
2	E	43	A	N9-C1'-C2'	5.24	120.81	114.00

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	740	0	758	57	0
1	B	725	0	725	62	0
1	C	672	0	673	31	0
1	D	638	0	616	51	0
2	E	1829	0	928	63	0
2	F	1813	0	920	72	0
3	E	26	0	16	7	0
3	F	26	0	16	3	0
4	E	6	0	0	0	0
4	F	2	0	0	0	0
All	All	6477	0	4652	316	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 29.

The worst 5 of 316 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:E:78:G:H4'	2:E:79:U:H5'	1.24	1.15
2:F:146:U:H3'	2:F:147:A:H5'	1.35	1.06
1:D:514:ILE:HD11	1:D:534:PHE:CZ	1.91	1.05
1:C:406:THR:O	1:C:406:THR:OG1	1.77	0.98
1:D:514:ILE:HG12	1:D:584:ILE:HG23	1.46	0.98

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	90/96 (94%)	81 (90%)	8 (9%)	1 (1%)	14 46

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	90/96 (94%)	75 (83%)	12 (13%)	3 (3%)	4	21
1	C	84/96 (88%)	73 (87%)	9 (11%)	2 (2%)	6	27
1	D	78/96 (81%)	65 (83%)	7 (9%)	6 (8%)	1	5
All	All	342/384 (89%)	294 (86%)	36 (10%)	12 (4%)	3	20

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	547	ARG
1	A	236	ARG
1	B	318	ASN
1	B	352	ARG
1	C	449	LEU

### 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	81/86 (94%)	73 (90%)	8 (10%)	8	29
1	B	78/86 (91%)	71 (91%)	7 (9%)	9	34
1	C	71/86 (83%)	63 (89%)	8 (11%)	6	23
1	D	67/86 (78%)	61 (91%)	6 (9%)	9	34
All	All	297/344 (86%)	268 (90%)	29 (10%)	8	29

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	397	MET
1	D	561	GLU
1	C	423	LYS
1	D	518	ASN
1	C	410	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	373	GLN
1	D	518	ASN
1	B	385	GLN
1	D	573	GLN
1	C	439	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	E	86/87 (98%)	20 (23%)	9 (10%)
2	F	86/87 (98%)	25 (29%)	8 (9%)
All	All	172/174 (98%)	45 (26%)	17 (9%)

5 of 45 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	E	19	C
2	E	26	G
2	E	27	C
2	E	30	U
2	E	31	C

5 of 17 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	F	162	U
2	F	179	U
2	E	79	U
2	E	85	A
2	F	118	C

## 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

Of 10 ligands modelled in this entry, 8 are monoatomic - leaving 2 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
3	TPP	E	601	4	22,27,27	3.55	7 (31%)	29,40,40	2.66	13 (44%)
3	TPP	F	602	4	22,27,27	2.10	3 (13%)	29,40,40	2.52	12 (41%)

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
3	TPP	E	601	4	-	3/16/17/17	0/2/2/2
3	TPP	F	602	4	-	1/16/17/17	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	601	TPP	C4-N3	-11.37	1.29	1.39
3	E	601	TPP	C6-C5	-10.11	1.46	1.50
3	F	602	TPP	C4-N3	-7.14	1.33	1.39
3	F	602	TPP	C6-C5	-4.41	1.49	1.50
3	E	601	TPP	CM4-C4	2.45	1.54	1.49

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	602	TPP	PA-O3A-PB	-7.33	107.67	132.83
3	E	601	TPP	PA-O3A-PB	-7.04	108.68	132.83
3	E	601	TPP	C5-C4-N3	5.69	118.96	107.57
3	F	602	TPP	C5-C4-N3	4.98	117.53	107.57
3	E	601	TPP	C6-C5-C4	-4.74	123.63	127.43

There are no chirality outliers.

All (4) torsion outliers are listed below:

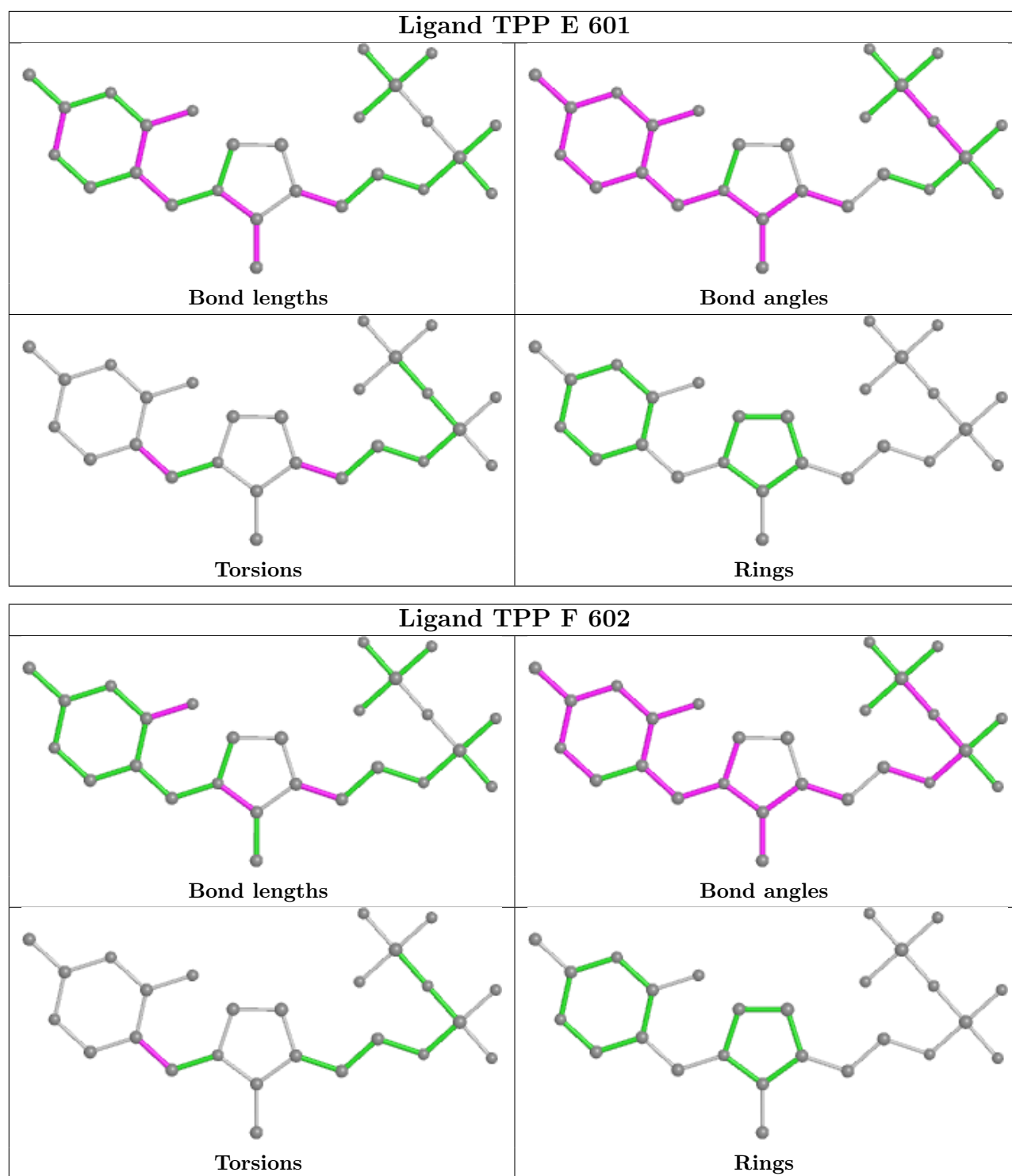
Mol	Chain	Res	Type	Atoms
3	E	601	TPP	C6'-C5'-C7'-N3
3	F	602	TPP	C6'-C5'-C7'-N3
3	E	601	TPP	C4'-C5'-C7'-N3
3	E	601	TPP	C4-C5-C6-C7

There are no ring outliers.

2 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	E	601	TPP	7	0
3	F	602	TPP	3	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

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	92/96 (95%)	-0.60	0 100 100	8, 26, 47, 65	0
1	B	92/96 (95%)	-0.53	0 100 100	14, 33, 62, 77	0
1	C	86/96 (89%)	-0.52	0 100 100	13, 32, 60, 93	0
1	D	82/96 (85%)	-0.42	0 100 100	18, 44, 66, 83	0
2	E	87/87 (100%)	-0.25	0 100 100	17, 34, 78, 92	0
2	F	87/87 (100%)	0.10	0 100 100	27, 56, 87, 99	0
All	All	526/558 (94%)	-0.37	0 100 100	8, 37, 74, 99	0

There are no RSRZ outliers to report.

### 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.

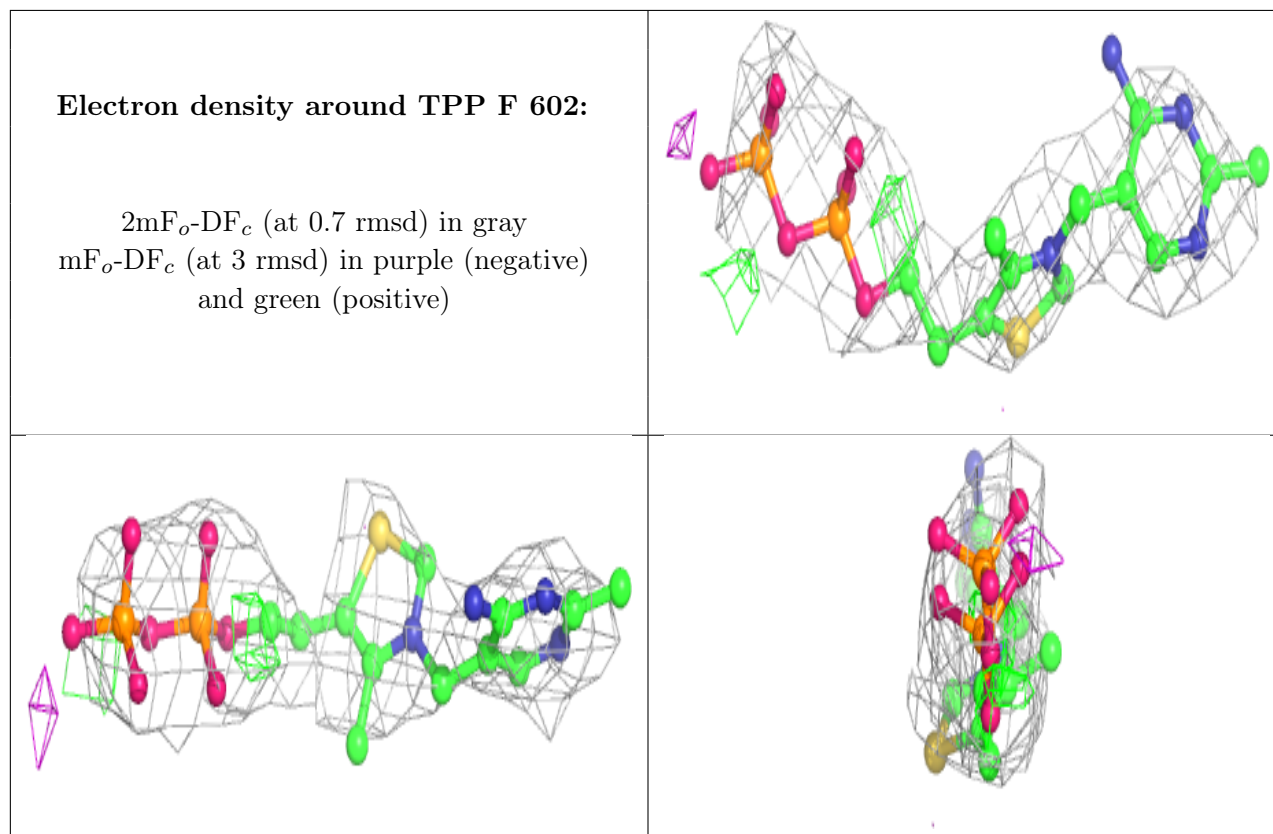
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
4	MG	E	707	1/1	0.73	0.24	43,43,43,43	0

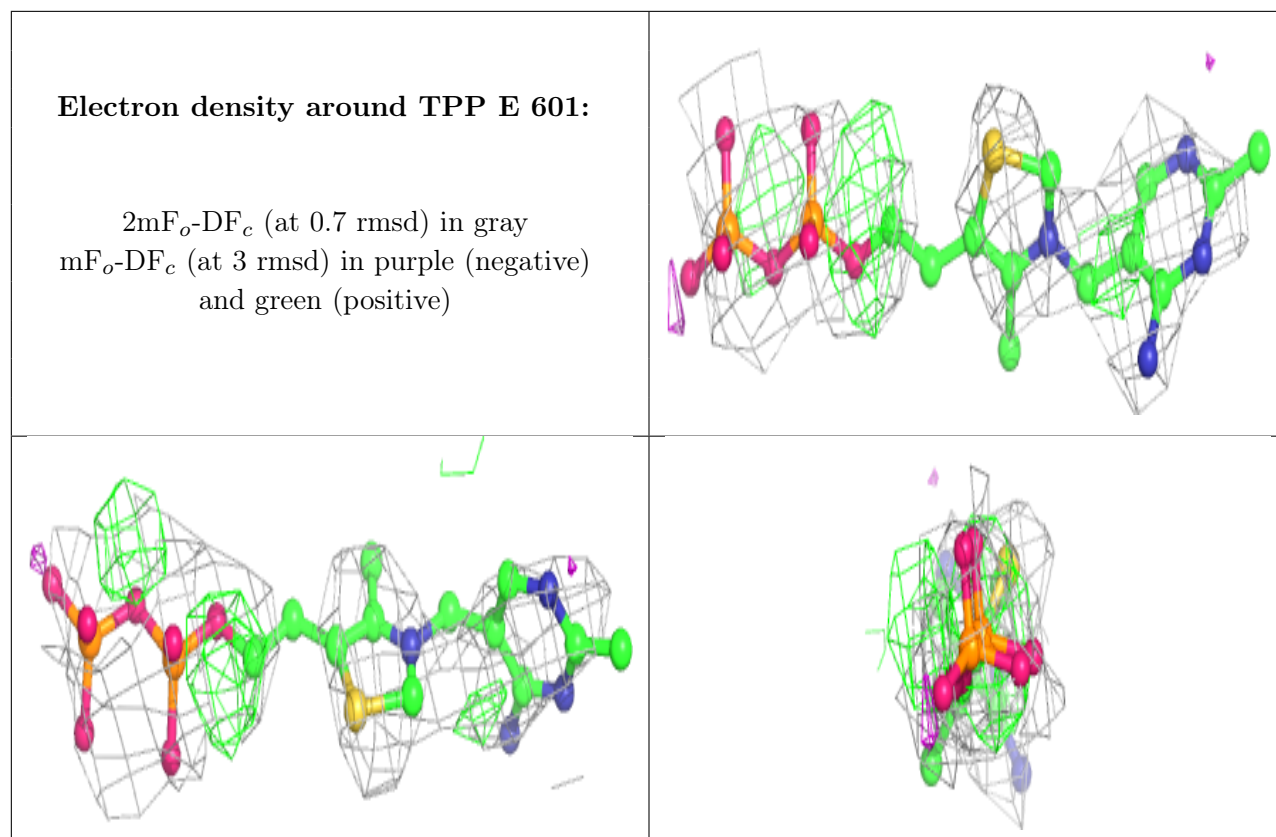
*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	MG	F	703	1/1	0.85	0.13	21,21,21,21	0
4	MG	E	704	1/1	0.86	0.13	28,28,28,28	0
3	TPP	F	602	26/26	0.88	0.27	45,68,88,91	0
3	TPP	E	601	26/26	0.88	0.31	60,74,95,104	0
4	MG	E	706	1/1	0.90	0.22	39,39,39,39	0
4	MG	E	705	1/1	0.91	0.13	26,26,26,26	0
4	MG	E	702	1/1	0.93	0.17	32,32,32,32	0
4	MG	E	701	1/1	0.93	0.11	22,22,22,22	0
4	MG	F	708	1/1	0.94	0.11	11,11,11,11	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.





## 6.5 Other polymers [i](#)

There are no such residues in this entry.