



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:52 AM GMT

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
A user guide is available at
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

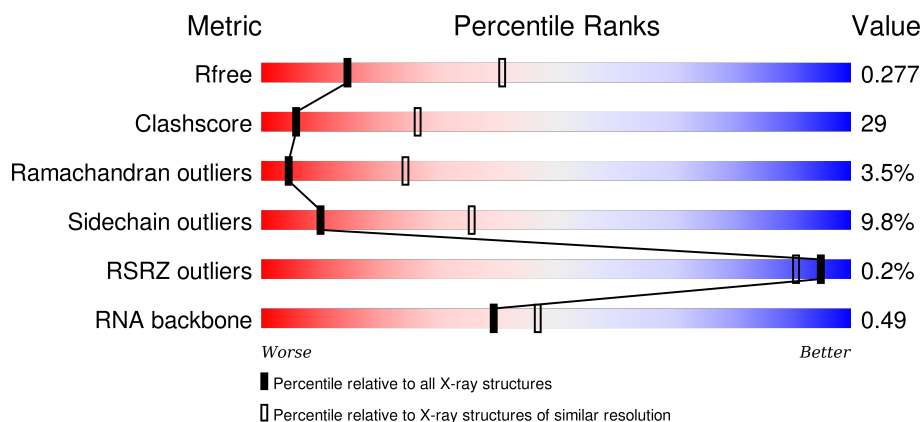
1 Overall quality at a glance

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}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)
RNA backbone	2183	1010 (3.52-2.68)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	
1	B	96	
1	C	96	
1	D	96	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	87	
2	F	87	

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
3	TPP	E	601	-	-	-	X
3	TPP	F	602	-	-	-	X
4	MG	E	707	-	-	-	X

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

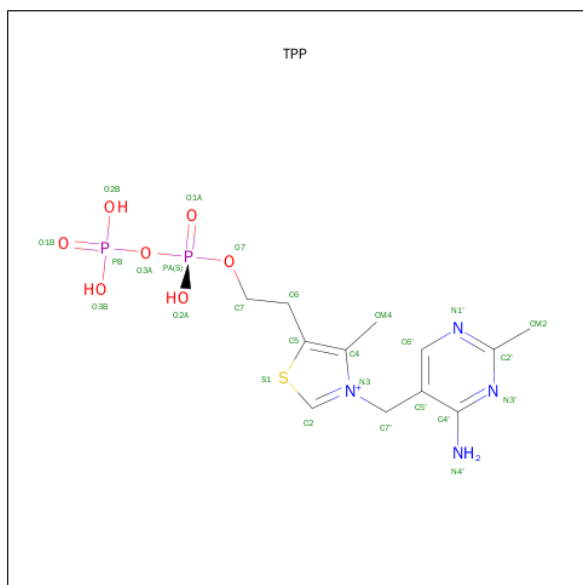
There are 8 discrepancies between the modelled and reference sequences:

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

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

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

- Molecule 3 is THIAMINE DIPHOSPHATE (three-letter code: TPP) (formula: C₁₂H₁₉N₄O₇P₂S).

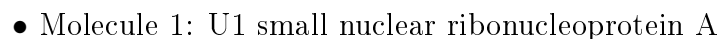
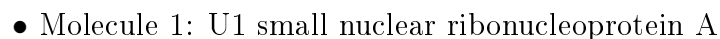
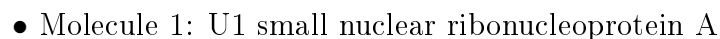


Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	E	1	Total 26	C 12	N 4	O 7	P 2	S 1	0	0
3	F	1	Total 26	C 12	N 4	O 7	P 2	S 1	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

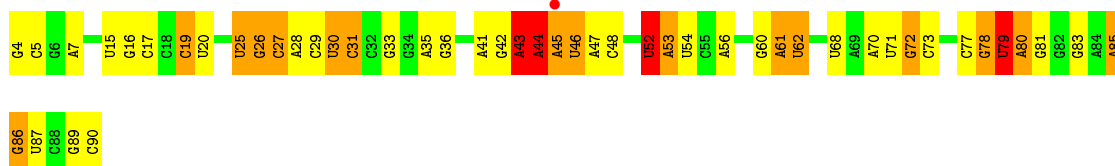
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	F	2	Total Mg 2 2	0	0
4	E	6	Total Mg 6 6	0	0

- Molecule 1: U1 small nuclear ribonucleoprotein A



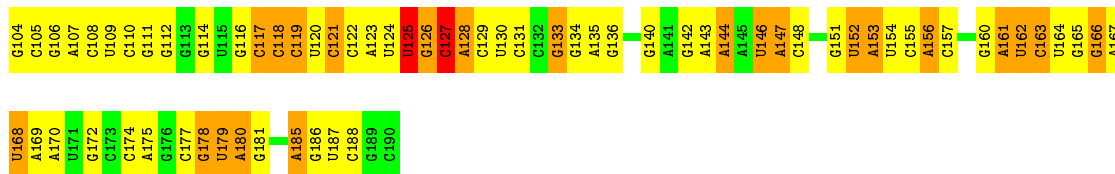
- Molecule 2: RNA (87-MER)

Chain E: 



- Molecule 2: RNA (87-MER)

Chain F: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	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
$\langle I/\sigma(I) \rangle$ ¹	0.98 (at 3.05Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.197 , 0.279 0.199 , 0.277	Depositor DCC
R_{free} test set	1595 reflections (10.96%)	DCC
Wilson B-factor (Å ²)	42.6	Xtriage
Anisotropy	0.664	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 57.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 17443 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	6477	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage'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.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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:OG1	1:C:406:THR:O	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%)	17 55

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%)	5	26
1	C	84/96 (88%)	73 (87%)	9 (11%)	2 (2%)	7	33
1	D	78/96 (81%)	65 (83%)	7 (9%)	6 (8%)	1	6
All	All	342/384 (89%)	294 (86%)	36 (10%)	12 (4%)	4	24

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 ⓘ

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%)	10	34
1	B	78/86 (91%)	71 (91%)	7 (9%)	12	41
1	C	71/86 (83%)	63 (89%)	8 (11%)	7	28
1	D	67/86 (78%)	61 (91%)	6 (9%)	12	41
All	All	297/344 (86%)	268 (90%)	29 (10%)	10	36

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

Mol	Chain	Res	Type
1	B	391	SER
1	C	410	HIS
1	D	544	LEU
1	B	397	MET
1	C	423	LYS

Some 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	367	ASN
1	B	373	GLN
1	D	509	ASN
1	B	354	GLN
1	C	439	GLN

5.3.3 RNA ⓘ

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	E	79	U
2	E	85	A
2	F	161	A
2	E	78	G
2	F	162	U

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry

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	20,27,27	2.91	4 (20%)	31,40,40	3.10	12 (38%)
3	TPP	F	602	4	20,27,27	1.92	2 (10%)	31,40,40	2.54	12 (38%)

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	-	0/16/17/17	0/2/2/2
3	TPP	F	602	4	-	0/16/17/17	0/2/2/2

The worst 5 of 6 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	F	602	TPP	C4-N3	-7.14	1.33	1.39
3	E	601	TPP	C7'-C5'	-2.35	1.46	1.51
3	E	601	TPP	C4'-N4'	2.15	1.39	1.34
3	F	602	TPP	C4'-N4'	2.27	1.39	1.34

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	602	TPP	PA-O3A-PB	-7.46	107.67	132.67
3	E	601	TPP	PA-O3A-PB	-7.15	108.68	132.67
3	E	601	TPP	CM4-C4-C5	-4.42	118.96	128.90
3	E	601	TPP	C6-C5-C4	-4.39	123.63	127.56
3	F	602	TPP	CM4-C4-C5	-3.68	120.63	128.90

There are no chirality outliers.

There are no torsion outliers.

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

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, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2		OWAB(Å ²)	Q<0.9
1	A	92/96 (95%)	-0.59	0	100 100	8, 26, 47, 65	0
1	B	92/96 (95%)	-0.51	0	100 100	14, 33, 62, 77	0
1	C	86/96 (89%)	-0.51	0	100 100	13, 32, 60, 93	0
1	D	82/96 (85%)	-0.40	0	100 100	18, 44, 66, 83	0
2	E	87/87 (100%)	-0.23	1 (1%)	82 66	17, 34, 78, 92	0
2	F	87/87 (100%)	0.14	0	100 100	27, 57, 87, 99	0
All	All	526/558 (94%)	-0.35	1 (0%)	95 91	8, 37, 74, 99	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	45	A	2.1

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

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

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
3	TPP	E	601	26/26	0.88	0.30	4.42	60,74,95,104	0
4	MG	E	707	1/1	0.73	0.24	3.56	43,43,43,43	0
3	TPP	F	602	26/26	0.88	0.27	2.77	45,68,88,91	0
4	MG	E	702	1/1	0.92	0.17	-0.48	32,32,32,32	0
4	MG	F	703	1/1	0.85	0.13	-	21,21,21,21	0
4	MG	F	708	1/1	0.94	0.11	-	11,11,11,11	0
4	MG	E	701	1/1	0.93	0.11	-	22,22,22,22	0
4	MG	E	704	1/1	0.86	0.13	-	28,28,28,28	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

6.5 Other polymers [i](#)

There are no such residues in this entry.