wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 10:39 AM GMT

PDB ID	:	3MEL
Title	:	Crystal Structure of Thiamin pyrophosphokinase family protein from Enterococcus faecalis, Northeast Structural Genomics Consortium Target EfR150
Authors	:	Kuzin, A.; Abasidze, M.; Seetharaman, J.; Mao, M.; Xiao, R.; Ciccosanti, C.; Foote, E.L.; Maglaqui, M.; Zhao, L.; Everett, J.K.; Nair, R.; Acton, T.B.; Rost, B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2010-03-31
Resolution	:	2.79 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

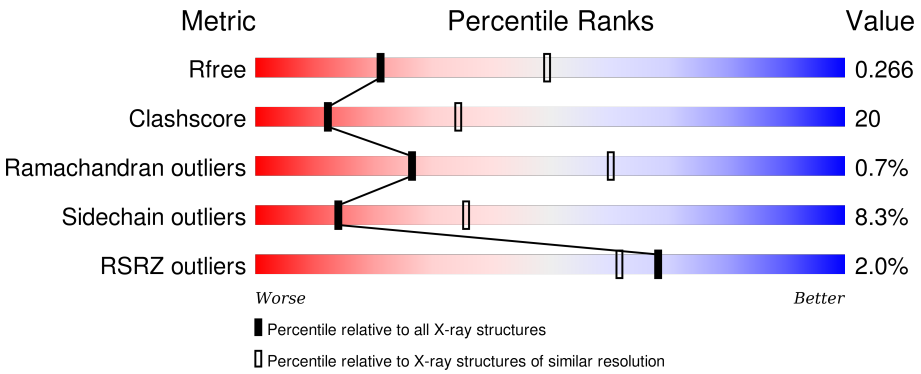
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.79 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	3004 (2.80-2.76)
Clashscore	102246	3480 (2.80-2.76)
Ramachandran outliers	100387	3423 (2.80-2.76)
Sidechain outliers	100360	3425 (2.80-2.76)
RSRZ outliers	91569	3016 (2.80-2.76)

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 <=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	222	64% 27%
1	B	222	61% 32%
1	C	222	% 55% 37%
1	D	222	7% 60% 32%

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	PO4	B	223	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6859 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 Thiamin pyrophosphokinase family protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	213	1693	1079	279	329	4	2	0	0	0
1	B	213	1693	1079	279	329	4	2	0	0	0
1	C	214	1701	1084	280	330	4	3	0	0	0
1	D	213	1693	1079	279	329	4	2	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

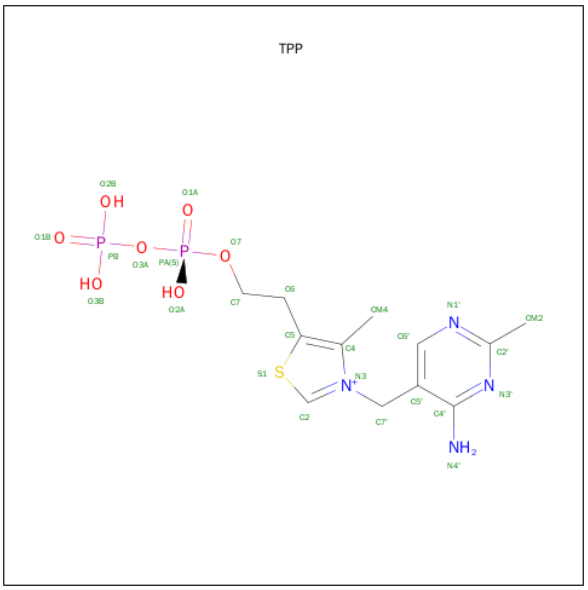
Chain	Residue	Modelled	Actual	Comment	Reference
A	215	LEU	-	expression tag	UNP Q82ZE3
A	216	GLU	-	expression tag	UNP Q82ZE3
A	217	HIS	-	expression tag	UNP Q82ZE3
A	218	HIS	-	expression tag	UNP Q82ZE3
A	219	HIS	-	expression tag	UNP Q82ZE3
A	220	HIS	-	expression tag	UNP Q82ZE3
A	221	HIS	-	expression tag	UNP Q82ZE3
A	222	HIS	-	expression tag	UNP Q82ZE3
B	215	LEU	-	expression tag	UNP Q82ZE3
B	216	GLU	-	expression tag	UNP Q82ZE3
B	217	HIS	-	expression tag	UNP Q82ZE3
B	218	HIS	-	expression tag	UNP Q82ZE3
B	219	HIS	-	expression tag	UNP Q82ZE3
B	220	HIS	-	expression tag	UNP Q82ZE3
B	221	HIS	-	expression tag	UNP Q82ZE3
B	222	HIS	-	expression tag	UNP Q82ZE3
C	215	LEU	-	expression tag	UNP Q82ZE3
C	216	GLU	-	expression tag	UNP Q82ZE3
C	217	HIS	-	expression tag	UNP Q82ZE3
C	218	HIS	-	expression tag	UNP Q82ZE3
C	219	HIS	-	expression tag	UNP Q82ZE3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	220	HIS	-	expression tag	UNP Q82ZE3
C	221	HIS	-	expression tag	UNP Q82ZE3
C	222	HIS	-	expression tag	UNP Q82ZE3
D	215	LEU	-	expression tag	UNP Q82ZE3
D	216	GLU	-	expression tag	UNP Q82ZE3
D	217	HIS	-	expression tag	UNP Q82ZE3
D	218	HIS	-	expression tag	UNP Q82ZE3
D	219	HIS	-	expression tag	UNP Q82ZE3
D	220	HIS	-	expression tag	UNP Q82ZE3
D	221	HIS	-	expression tag	UNP Q82ZE3
D	222	HIS	-	expression tag	UNP Q82ZE3

- Molecule 2 is THIAMINE DIPHOSPHATE (three-letter code: TPP) (formula: $C_{12}H_{19}N_4O_7P_2S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	P	S		
2	A	1	26	12	4	7	2	1	0	0
			Total	C	N	O	P	S		
2	D	1	26	12	4	7	2	1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

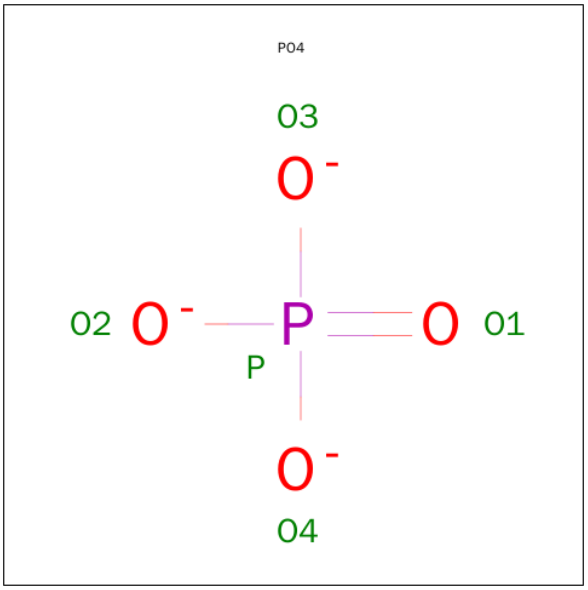
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	B	3	3	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	4	Total Mg 4 4	0	0
3	D	2	Total Mg 2 2	0	0
3	C	2	Total Mg 2 2	0	0

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total O P 5 4 1	0	0
4	C	1	Total O P 5 4 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	2	Total O 2 2	0	0
5	B	1	Total O 1 1	0	0
5	C	2	Total O 2 2	0	0
5	D	1	Total O 1 1	0	0

3 Residue-property plots (i)

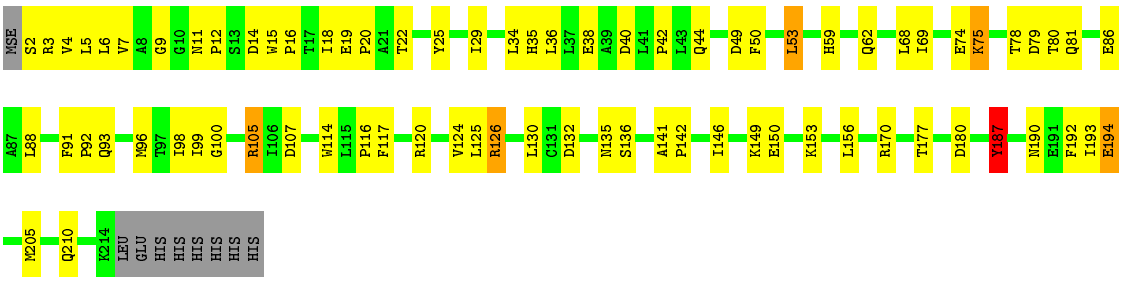
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Thiamin pyrophosphokinase family protein

Chain A:

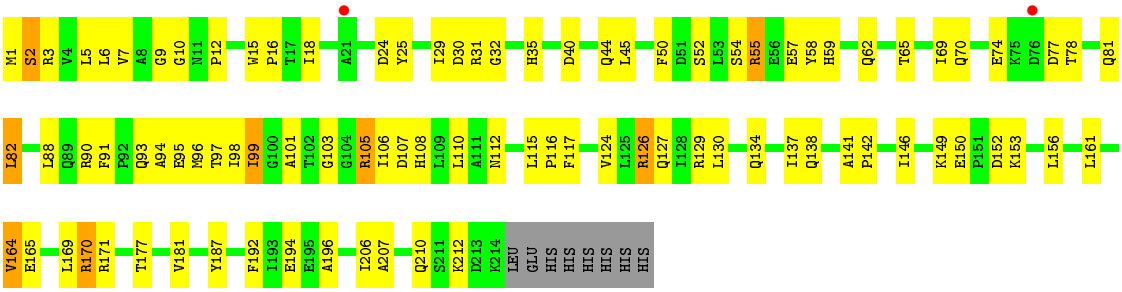
- Molecule 1: Thiamin pyrophosphokinase family protein

Chain B:



- Molecule 1: Thiamin pyrophosphokinase family protein

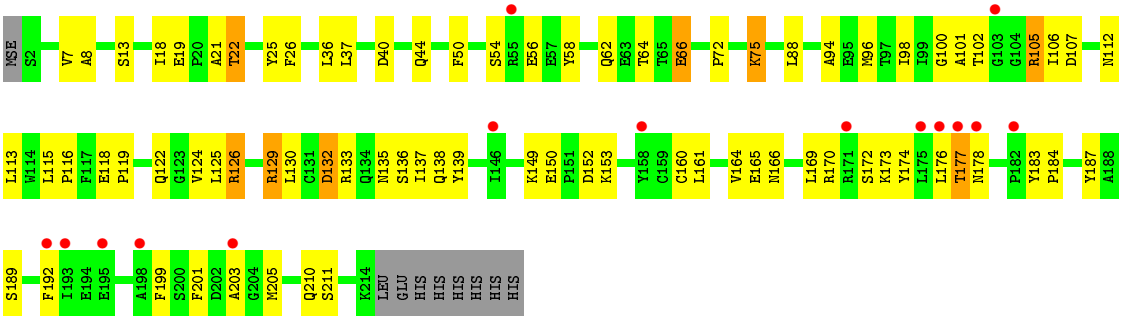
Chain C:



- Molecule 1: Thiamin pyrophosphokinase family protein

Chain D:

7%
60%
32%


NSE S2 V7 A8 S13 I18 E19 P20 A21 T22 Y25 F26 L36 L37 D40 Q44 F50 S54 R55 E56 E57 Y58 Q62 E63 T64 T65 E66 P72 K75 L88 A94 E95 M96 T97 I98 I99 G100 A101 T102 G103 G104 R105 I106 D107 N112
L113 W114 L115 P116 F117 E118 P119 Q122 G123 V124 L125 R126 R129 L130 C131 D132 R133 Q134 N135 S136 I137 Q138 Y139 I146 K149 E150 P151 D152 K153 Y158 C159 C160 L161 V164 E165 N166 L169 R170 R171 S172 K173 Y174 L175 L176 T177 N178 P182 Y183 P184 Y187 A188
S189 F192 I193 E194 E195 A198 F199 S200 F201 D202 A203 G204 M205 Q210 S211 K214 LEU GLU HIS HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.29Å 102.11Å 94.28Å 90.00° 98.13° 90.00°	Depositor
Resolution (Å)	48.53 – 2.79 48.53 – 2.79	Depositor EDS
% Data completeness (in resolution range)	99.1 (48.53-2.79) 99.2 (48.53-2.79)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	7.83 (at 2.77Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2)	Depositor
R, R_{free}	0.207 , 0.275 0.200 , 0.266	Depositor DCC
R_{free} test set	1476 reflections (5.09%)	DCC
Wilson B-factor (Å^2)	42.0	Xtriage
Anisotropy	0.634	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 35.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 29025 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6859	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.20% 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, 0.375 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: PO4, 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	0.48	0/1731	0.64	1/2356 (0.0%)
1	B	0.46	0/1731	0.64	0/2356
1	C	0.42	0/1739	0.60	0/2366
1	D	0.42	0/1731	0.59	0/2356
All	All	0.44	0/6932	0.62	1/9434 (0.0%)

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	B	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	169	LEU	CA-CB-CG	6.99	131.37	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	187	TYR	Peptide

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	1693	0	1644	55	0
1	B	1693	0	1644	67	0
1	C	1701	0	1656	90	0
1	D	1693	0	1644	65	0
2	A	26	0	16	1	0
2	D	26	0	16	3	0
3	A	4	0	0	0	0
3	B	3	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
4	B	5	0	0	0	0
4	C	5	0	0	1	0
5	A	2	0	0	0	0
5	B	1	0	0	0	0
5	C	2	0	0	0	0
5	D	1	0	0	0	0
All	All	6859	0	6620	266	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 20.

The worst 5 of 266 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)
1:C:1:MSE:H2	1:C:2:SER:HA	1.07	1.18
1:B:193:ILE:HG22	1:B:194:GLU:HG3	1.30	1.11
1:C:1:MSE:H2	1:C:2:SER:CA	1.68	1.05
1:D:105:ARG:HG2	1:D:105:ARG:HH11	1.19	1.04
1:A:105:ARG:HG2	1:A:105:ARG:HH11	1.21	1.02

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	211/222 (95%)	203 (96%)	8 (4%)	0	100	100
1	B	211/222 (95%)	201 (95%)	9 (4%)	1 (0%)	34	68
1	C	212/222 (96%)	194 (92%)	16 (8%)	2 (1%)	21	53
1	D	211/222 (95%)	198 (94%)	10 (5%)	3 (1%)	14	39
All	All	845/888 (95%)	796 (94%)	43 (5%)	6 (1%)	26	60

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	21	ALA
1	C	103	GLY
1	D	13	SER
1	C	10	GLY
1	B	187	TYR

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	183/189 (97%)	167 (91%)	16 (9%)	13	33
1	B	183/189 (97%)	168 (92%)	15 (8%)	14	36
1	C	184/189 (97%)	167 (91%)	17 (9%)	11	30
1	D	183/189 (97%)	170 (93%)	13 (7%)	18	44
All	All	733/756 (97%)	672 (92%)	61 (8%)	14	36

5 of 61 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	149	LYS
1	C	40	ASP
1	D	129	ARG
1	B	180	ASP
1	B	194	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	93	GLN
1	C	210	GLN
1	D	179	GLN
1	C	108	HIS
1	C	178	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 15 ligands modelled in this entry, 11 are monoatomic - leaving 4 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	$\#\|Z\| > 2$
2	TPP	A	902	3	20,27,27	5.41	9 (45%)	31,40,40	4.62	20 (64%)
4	PO4	B	223	-	4,4,4	0.29	0	6,6,6	0.29	0
4	PO4	C	223	-	4,4,4	0.24	0	6,6,6	0.27	0
2	TPP	D	902	3	20,27,27	5.24	9 (45%)	31,40,40	4.71	20 (64%)

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
2	TPP	A	902	3	-	0/16/17/17	0/2/2/2
4	PO4	B	223	-	-	0/0/0/0	0/0/0/0
4	PO4	C	223	-	-	0/0/0/0	0/0/0/0
2	TPP	D	902	3	-	0/16/17/17	0/2/2/2

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	902	TPP	C7'-C5'	2.23	1.56	1.51
2	D	902	TPP	C7'-C5'	2.68	1.56	1.51
2	D	902	TPP	C4'-N4'	5.08	1.46	1.34
2	A	902	TPP	C4'-N4'	5.27	1.47	1.34
2	A	902	TPP	C6'-N1'	5.70	1.46	1.34

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	902	TPP	C6-C5-C4	-12.33	116.52	127.56
2	D	902	TPP	C6-C5-C4	-11.73	117.05	127.56
2	D	902	TPP	C5'-C4'-N3'	-8.48	107.08	121.23
2	D	902	TPP	C5'-C4'-N4'	-8.12	110.33	122.25
2	D	902	TPP	CM4-C4-N3	-7.71	112.31	122.59

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	902	TPP	1	0
4	C	223	PO4	1	0
2	D	902	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	211/222 (95%)	-0.03	0	100	100	25, 36, 49, 59	0
1	B	211/222 (95%)	-0.15	0	100	100	24, 40, 56, 81	0
1	C	211/222 (95%)	0.09	2 (0%)	85	81	35, 48, 66, 86	0
1	D	211/222 (95%)	0.41	15 (7%)	19	12	31, 50, 74, 80	0
All	All	844/888 (95%)	0.08	17 (2%)	68	62	24, 43, 68, 86	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	193	ILE	4.0
1	D	103	GLY	3.9
1	D	55	ARG	3.2
1	D	182	PRO	2.8
1	D	178	ASN	2.7

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 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(Å^2)	Q<0.9
4	PO4	B	223	5/5	0.89	0.28	2.85	51,52,63,69	0
2	TPP	A	902	26/26	0.92	0.20	1.15	32,48,129,185	0
2	TPP	D	902	26/26	0.90	0.22	1.00	38,52,129,185	0
4	PO4	C	223	5/5	0.93	0.17	0.31	53,53,63,68	0
3	MG	C	224	1/1	0.96	0.37	-	43,43,43,43	0
3	MG	C	225	1/1	0.95	0.22	-	25,25,25,25	0
3	MG	A	225	1/1	0.83	0.25	-	63,63,63,63	0
3	MG	D	223	1/1	0.87	0.33	-	61,61,61,61	0
3	MG	B	225	1/1	0.91	0.17	-	47,47,47,47	0
3	MG	D	224	1/1	0.93	0.32	-	62,62,62,62	0
3	MG	B	224	1/1	0.74	0.30	-	50,50,50,50	0
3	MG	A	224	1/1	0.98	0.18	-	33,33,33,33	0
3	MG	A	223	1/1	0.69	0.18	-	61,61,61,61	0
3	MG	A	226	1/1	0.92	0.35	-	40,40,40,40	0
3	MG	B	226	1/1	0.95	0.30	-	48,48,48,48	0

6.5 Other polymers (i)

There are no such residues in this entry.