wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:31 PM GMT

PDB ID	:	1TYG
Title	:	Structure of the thiazole synthase/ThiS complex
Authors	:	Settembre, E.C.; Dorrestein, P.C.; Zhai, H.; Chatterjee, A.; McLafferty, F.W.; Begley, T.P.; Ealick, S.E.
Deposited on	:	2004-07-07
Resolution	:	3.15 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:

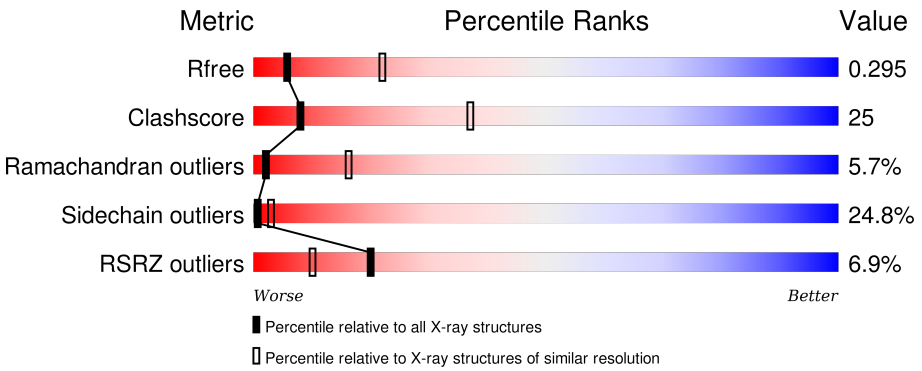
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.15 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	1112 (3.20-3.12)
Clashscore	102246	1249 (3.20-3.12)
Ramachandran outliers	100387	1222 (3.20-3.12)
Sidechain outliers	100360	1221 (3.20-3.12)
RSRZ outliers	91569	1117 (3.20-3.12)

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	B	87	% 33% 25% 14% • 25%
1	G	87	37% 25% 37% 11% 26%
2	A	253	58% 28% 7% • •
2	C	253	4% 40% 37% 14% • 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4555 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 yjbS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	65	527	334	91	100	2	0	0	0
			Total	C	N	O	S			
1	G	64	511	323	87	99	2	0	0	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-20	MET	-	CLONING ARTIFACT	UNP O31617
B	-19	GLY	-	CLONING ARTIFACT	UNP O31617
B	-18	HIS	-	CLONING ARTIFACT	UNP O31617
B	-17	HIS	-	CLONING ARTIFACT	UNP O31617
B	-16	HIS	-	CLONING ARTIFACT	UNP O31617
B	-15	HIS	-	CLONING ARTIFACT	UNP O31617
B	-14	HIS	-	CLONING ARTIFACT	UNP O31617
B	-13	HIS	-	CLONING ARTIFACT	UNP O31617
B	-12	HIS	-	CLONING ARTIFACT	UNP O31617
B	-11	HIS	-	CLONING ARTIFACT	UNP O31617
B	-10	HIS	-	CLONING ARTIFACT	UNP O31617
B	-9	HIS	-	CLONING ARTIFACT	UNP O31617
B	-8	SER	-	CLONING ARTIFACT	UNP O31617
B	-7	SER	-	CLONING ARTIFACT	UNP O31617
B	-6	GLY	-	CLONING ARTIFACT	UNP O31617
B	-5	HIS	-	CLONING ARTIFACT	UNP O31617
B	-4	ILE	-	CLONING ARTIFACT	UNP O31617
B	-3	GLY	-	CLONING ARTIFACT	UNP O31617
B	-2	GLY	-	CLONING ARTIFACT	UNP O31617
B	-1	ARG	-	CLONING ARTIFACT	UNP O31617
B	0	HIS	-	CLONING ARTIFACT	UNP O31617
G	-20	MET	-	CLONING ARTIFACT	UNP O31617
G	-19	GLY	-	CLONING ARTIFACT	UNP O31617
G	-18	HIS	-	CLONING ARTIFACT	UNP O31617
G	-17	HIS	-	CLONING ARTIFACT	UNP O31617

Continued on next page...

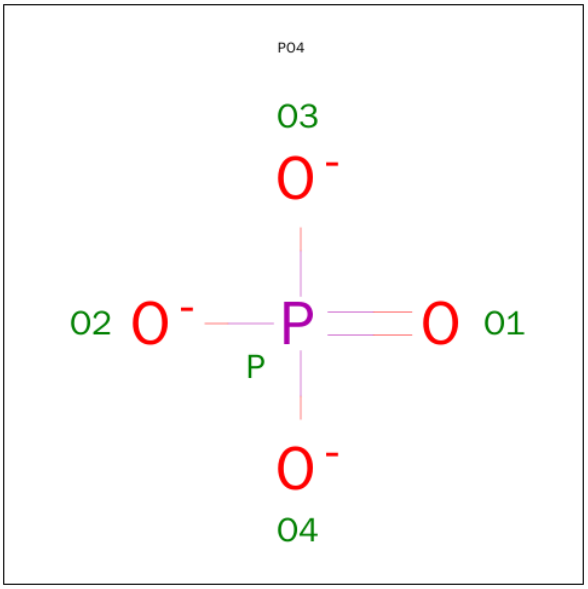
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	-16	HIS	-	CLONING ARTIFACT	UNP O31617
G	-15	HIS	-	CLONING ARTIFACT	UNP O31617
G	-14	HIS	-	CLONING ARTIFACT	UNP O31617
G	-13	HIS	-	CLONING ARTIFACT	UNP O31617
G	-12	HIS	-	CLONING ARTIFACT	UNP O31617
G	-11	HIS	-	CLONING ARTIFACT	UNP O31617
G	-10	HIS	-	CLONING ARTIFACT	UNP O31617
G	-9	HIS	-	CLONING ARTIFACT	UNP O31617
G	-8	SER	-	CLONING ARTIFACT	UNP O31617
G	-7	SER	-	CLONING ARTIFACT	UNP O31617
G	-6	GLY	-	CLONING ARTIFACT	UNP O31617
G	-5	HIS	-	CLONING ARTIFACT	UNP O31617
G	-4	ILE	-	CLONING ARTIFACT	UNP O31617
G	-3	GLY	-	CLONING ARTIFACT	UNP O31617
G	-2	GLY	-	CLONING ARTIFACT	UNP O31617
G	-1	ARG	-	CLONING ARTIFACT	UNP O31617
G	0	HIS	-	CLONING ARTIFACT	UNP O31617

- Molecule 2 is a protein called Thiazole biosynthesis protein thiG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	242	1784	1135	297	343	9	0	0	0
			Total	C	N	O	S			
2	C	237	1718	1090	283	338	7	0	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

- Molecule 4 is water.

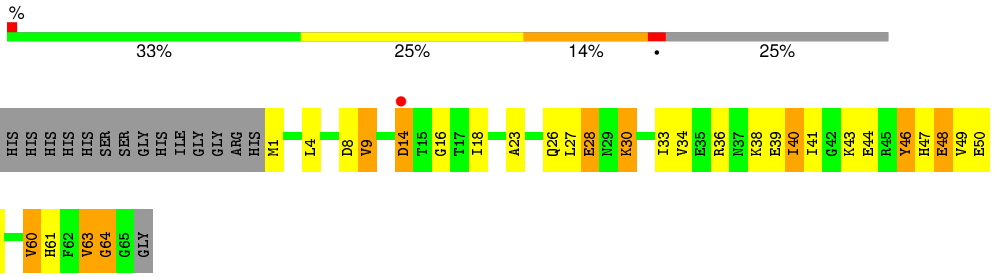
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	4	4	4	0	0
			Total	O		
4	C	1	1	1	0	0

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.

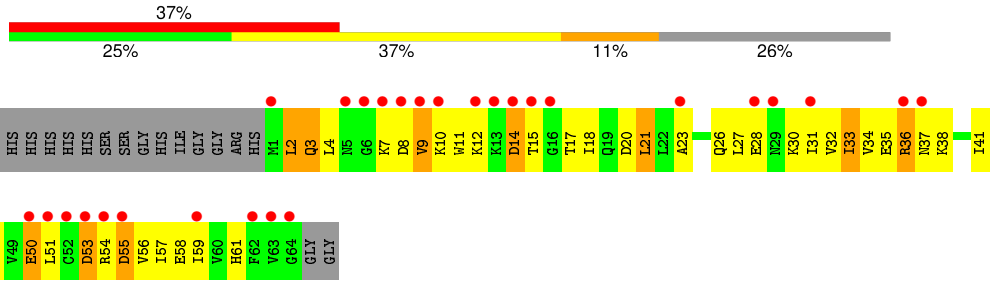
- Molecule 1: yjbS

Chain B:



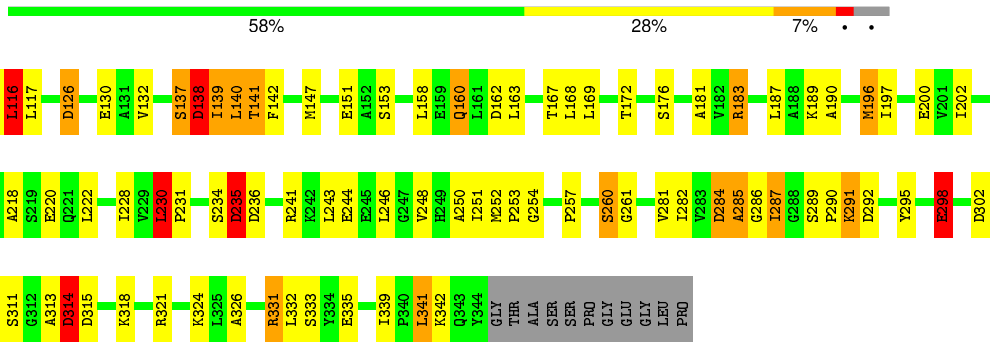
- Molecule 1: yjbS

Chain G:



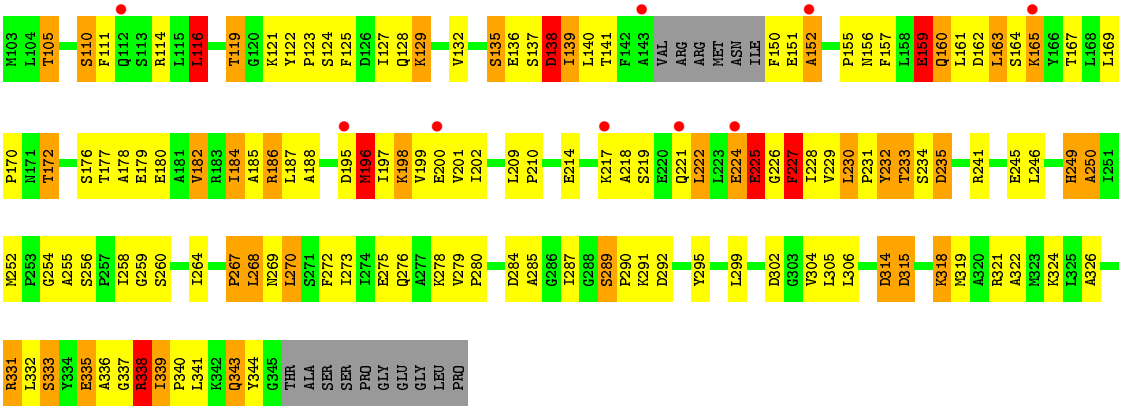
- Molecule 2: Thiazole biosynthesis protein thiG

Chain A:



- Molecule 2: Thiazole biosynthesis protein thiG

Chain C:


M103 L104 T105 S110 F111 Q112 S113 R114 L115 L116 T119 G120 K121 Y122 P123 S124 F125 D126 I127 Q128 K129 V132 S135 E136 S137 D138 I139 L140 T141 F142 A143 VAL ARG ARG MET ASN ILE F150 E151 A152 P155 N156 F157 L158 E159 Q160 L161 D162 L163 S164 K165 Y166 T167 L168 L169
P170 N171 T172 S176 T177 A178 E179 E180 A181 V182 R183 I184 A185 R186 L187 A188 D195 M196 I197 K198 V199 E200 V201 I202 L209 P210 E214 K217 A218 S219 E220 Q221 L222 L223 E224 E225 G226 F227 I228 V229 L230 P231 Y232 T233 S234 D235 R241 E245 L246 H249 A250 I251
M252 P253 G254 A255 S256 P257 I258 G259 S260 I264 P267 L268 N269 L270 S271 F272 I273 I274 E275 Q276 A277 K278 V279 P280 D284 A285 G286 I287 G288 S289 P290 K291 D292 Y295 L299 D302 G303 V304 L305 L306 D314 D315 K318 M319 A320 R321 A322 M323 K324 L325 A326
R331 L332 S333 Y334 E335 A336 G337 R338 I339 P340 L341 K342 Q343 Y344 G345 THR ALA SER SER PRO GLY GLU GLY LEU PRO

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	91.65Å 91.65Å 401.30Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 3.15 14.99 – 3.15	Depositor EDS
% Data completeness (in resolution range)	95.2 (15.00-3.15) 95.2 (14.99-3.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.50 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.242 , 0.305 0.229 , 0.295	Depositor DCC
R_{free} test set	1261 reflections (7.89%)	DCC
Wilson B-factor (Å^2)	89.4	Xtriage
Anisotropy	0.023	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 66.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 17236 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	4555	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 2.95% 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

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	B	0.92	0/533	1.14	3/716 (0.4%)
1	G	0.92	0/517	1.03	3/699 (0.4%)
2	A	1.00	3/1810 (0.2%)	1.15	8/2455 (0.3%)
2	C	1.35	24/1741 (1.4%)	1.12	11/2364 (0.5%)
All	All	1.13	27/4601 (0.6%)	1.12	25/6234 (0.4%)

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
2	A	0	3
All	All	0	4

The worst 5 of 27 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	186	ARG	CZ-NH1	24.68	1.65	1.33
2	C	225	GLU	CG-CD	15.40	1.75	1.51
2	C	164	SER	CB-OG	14.96	1.61	1.42
2	C	225	GLU	CB-CG	13.71	1.78	1.52
2	C	159	GLU	CD-OE1	8.38	1.34	1.25

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	186	ARG	NE-CZ-NH2	-22.77	108.92	120.30
2	C	186	ARG	NE-CZ-NH1	11.74	126.17	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	236	ASP	CB-CG-OD2	8.21	125.69	118.30
2	C	138	ASP	CB-CG-OD2	7.76	125.28	118.30
1	G	55	ASP	CB-CG-OD2	7.07	124.67	118.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	235	ASP	Peptide
2	A	260	SER	Peptide
2	A	314	ASP	Peptide
1	B	53	ASP	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	B	527	0	535	35	0
1	G	511	0	499	24	0
2	A	1784	0	1821	83	0
2	C	1718	0	1736	103	0
3	A	5	0	0	1	0
3	C	5	0	0	0	0
4	A	4	0	0	0	0
4	C	1	0	0	0	0
All	All	4555	0	4591	226	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 25.

The worst 5 of 226 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:C:225:GLU:CB	2:C:225:GLU:CG	1.78	1.61
2:C:225:GLU:CG	2:C:225:GLU:CD	1.75	1.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:36:ARG:NH2	1:B:52:CYS:HB2	1.73	1.03
1:B:18:ILE:HG13	1:B:49:VAL:HG13	1.47	0.94
1:B:36:ARG:HH22	1:B:52:CYS:HB2	1.25	0.94

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	B	63/87 (72%)	47 (75%)	9 (14%)	7 (11%)	0 3
1	G	62/87 (71%)	51 (82%)	7 (11%)	4 (6%)	1 12
2	A	240/253 (95%)	209 (87%)	22 (9%)	9 (4%)	4 26
2	C	233/253 (92%)	188 (81%)	31 (13%)	14 (6%)	2 14
All	All	598/680 (88%)	495 (83%)	69 (12%)	34 (6%)	2 16

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	47	HIS
1	B	51	LEU
1	B	53	ASP
2	A	138	ASP
2	A	314	ASP

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	B	58/76 (76%)	40 (69%)	18 (31%)	0 1
1	G	55/76 (72%)	34 (62%)	21 (38%)	0 0
2	A	188/201 (94%)	158 (84%)	30 (16%)	3 14
2	C	179/201 (89%)	129 (72%)	50 (28%)	0 1
All	All	480/554 (87%)	361 (75%)	119 (25%)	1 3

5 of 119 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	121	LYS
2	C	182	VAL
1	G	32	VAL
2	C	135	SER
2	C	161	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	A	249	HIS
1	G	61	HIS
1	G	5	ASN
1	B	47	HIS
1	G	26	GLN

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)

2 ligands are modelled in this entry.

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	PO4	A	1400	-	4,4,4	0.41	0	6,6,6	0.36	0
3	PO4	C	2400	-	4,4,4	0.77	0	6,6,6	0.31	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
3	PO4	A	1400	-	-	0/0/0/0	0/0/0/0
3	PO4	C	2400	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1400	PO4	1	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	B	65/87 (74%)	-0.47	1 (1%) 76 62	24, 34, 38, 40	0
1	G	64/87 (73%)	2.16	32 (50%) 0 0	137, 140, 146, 149	0
2	A	242/253 (95%)	-0.66	0 100 100	27, 33, 38, 44	0
2	C	237/253 (93%)	-0.14	9 (3%) 44 27	28, 33, 37, 54	0
All	All	608/680 (89%)	-0.14	42 (6%) 20 10	24, 34, 141, 149	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	8	ASP	8.6
1	G	63	VAL	7.1
1	G	14	ASP	6.7
1	G	15	THR	5.6
1	G	47	HIS	5.4

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

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

6.3 Carbohydrates (i)

There are no 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
3	PO4	A	1400	5/5	0.98	0.11	-1.28	25,26,29,30	0
3	PO4	C	2400	5/5	0.96	0.11	-1.73	24,24,25,25	0

6.5 Other polymers (i)

There are no such residues in this entry.