wwPDB X-ray Structure Validation Summary Report (i)

Sep 21, 2016 – 08:23 PM EDT

PDB ID	:	5DPL
Title	:	The structure of PKMT2 from Rickettsia typhi in complex with AdoHcy
Authors	:	Noinaj, N.; Abeykoon, A.; He, Y.; Yang, D.C.; Buchanan, S.K.
Deposited on	:	2015-09-12
Resolution	:	3.20 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:

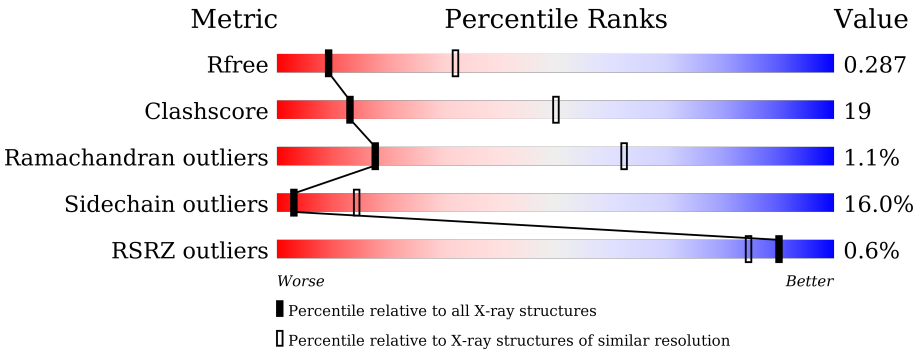
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 3.20 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	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	535	% 53% 37% 6%
1	B	535	53% 35% 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8077 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.

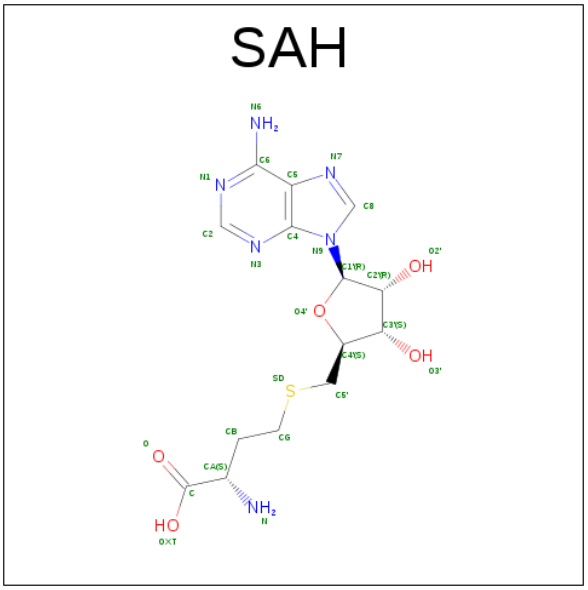
- Molecule 1 is a protein called protein lysine methyltransferase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	520	Total	C	N	O	S	0	1	0
			4055	2615	659	768	13			
1	B	512	Total	C	N	O	S	0	1	0
			3970	2566	643	749	12			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q68XQ5
B	0	GLY	-	expression tag	UNP Q68XQ5

- Molecule 2 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

Continued on next page...

Continued from previous page...

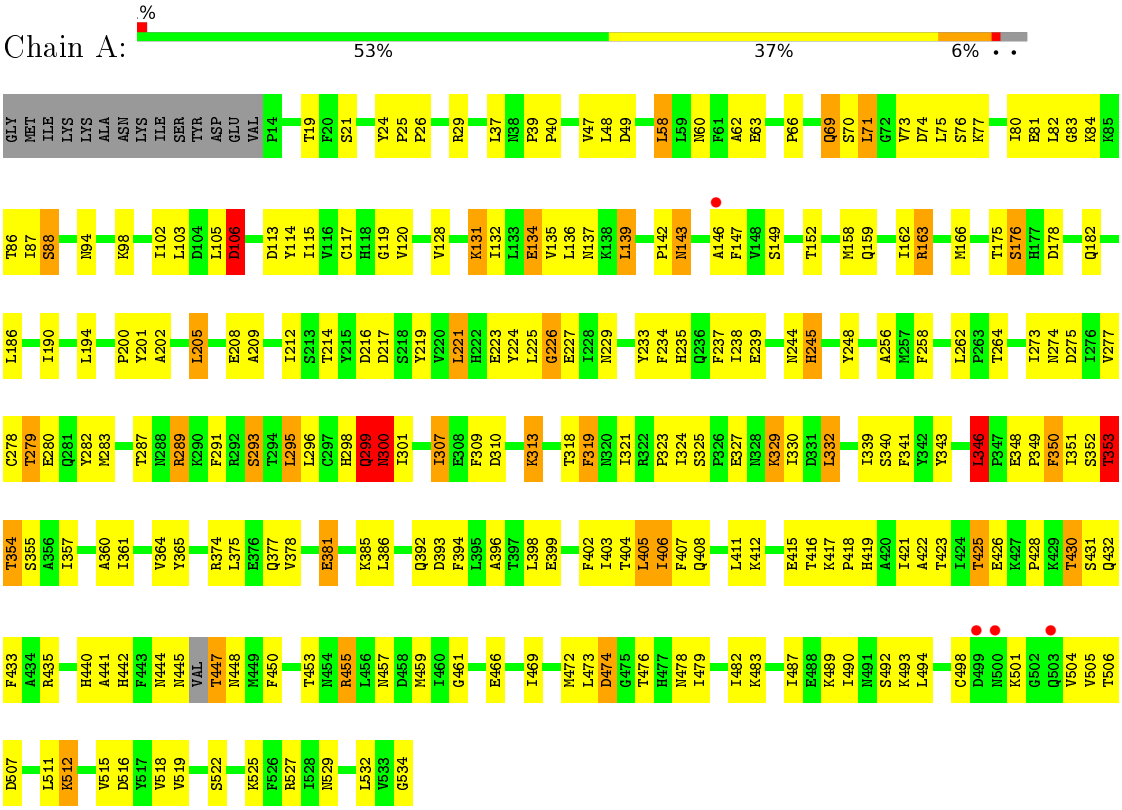
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

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.

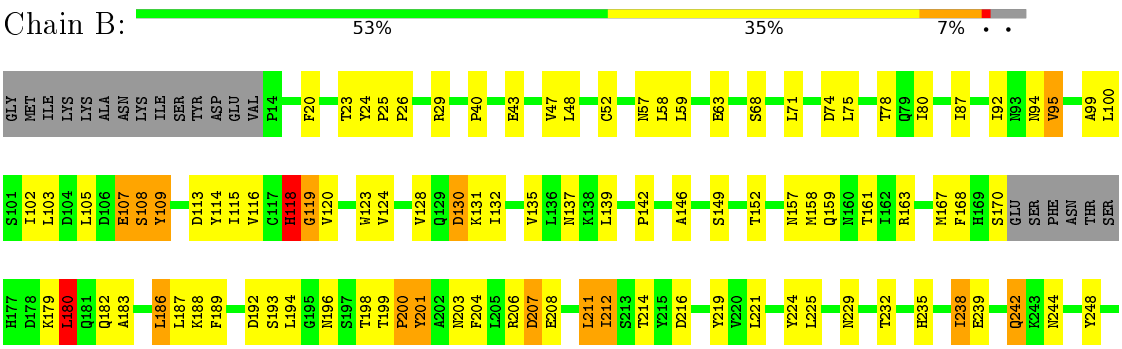
- Molecule 1: protein lysine methyltransferase 2

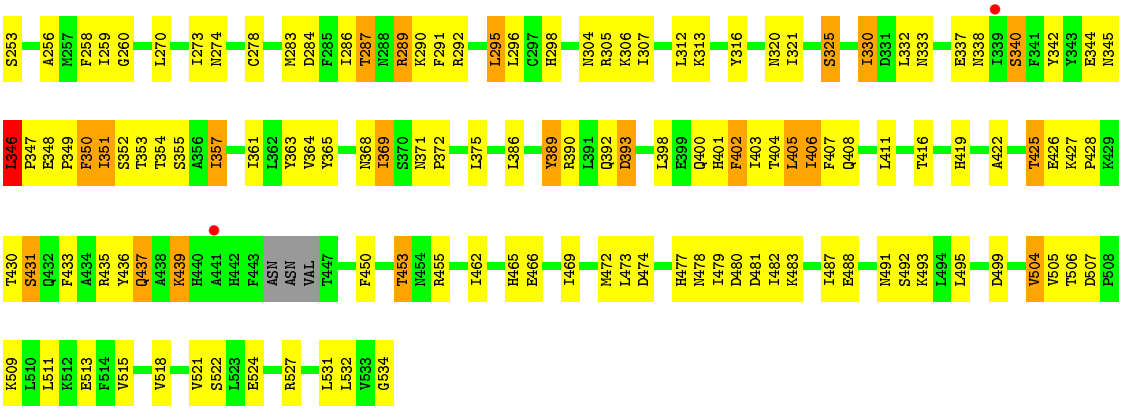
Chain A:



- Molecule 1: protein lysine methyltransferase 2

Chain B:




S253 A256 M257 F258 I259 G260 L270 I273 N274 C278 M283 D284 F285 I286 T287 N288 R289 K290 F291 R292 L295 L296 C297 H298 N304 R305 K306 I307 L312 K313 Y316 N320 I321 S325 I330 D331 L332 N333 E337 N338 I339 S340 F341 Y342 Y343 E344 N345
L346 P347 E348 P349 F350 I351 S352 T353 T354 S355 A356 I357 I361 L362 Y363 V364 Y365 N368 I369 S370 N371 P372 L375 L386 Y389 R390 L391 Q392 D393 L398 E399 Q400 H401 F402 I403 T404 I405 I406 F407 Q408 L411 T416 H419 A422 T425 E426 K427 P428 K429
T430 S431 Q432 F433 A434 R435 Y436 Q437 A438 K439 H440 A441 H442 F443 ASN ASN VAL T447 F450 T453 N454 R455 I462 H465 E466 I469 M472 L473 D474 H477 N478 I479 D480 D481 I482 K483 I487 E488 N491 S492 K493 L494 L495 D499 V504 V505 T506 D507 P508
K509 L510 L511 K512 E513 F514 V515 V518 V521 S522 L523 E524 R527 L531 L532 V533 G534

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	83.35Å 91.03Å 105.83Å 90.00° 112.30° 90.00°	Depositor
Resolution (Å)	48.95 – 3.20 48.96 – 3.19	Depositor EDS
% Data completeness (in resolution range)	97.4 (48.95-3.20) 93.5 (48.96-3.19)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.14	Depositor
$< I/\sigma(I) >$ [1]	1.93 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.10_2142: ???)	Depositor
R, R_{free}	0.244 , 0.280 0.249 , 0.287	Depositor DCC
R_{free} test set	1912 reflections (8.35%)	DCC
Wilson B-factor (Å^2)	65.2	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , -13.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.107 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	8077	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.50% 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: SAH

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	0.32	0/4145	0.62	3/5643 (0.1%)
1	B	0.32	0/4059	0.63	2/5530 (0.0%)
All	All	0.32	0/8204	0.63	5/11173 (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	A	0	6
1	B	0	5
All	All	0	11

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	346	LEU	CA-CB-CG	7.28	132.05	115.30
1	B	118	HIS	C-N-CA	-6.63	108.38	122.30
1	B	180	LEU	CA-CB-CG	6.23	129.63	115.30
1	A	205	LEU	CA-CB-CG	5.06	126.95	115.30
1	A	353	THR	C-N-CA	5.02	134.25	121.70

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	106	ASP	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	176	SER	Peptide
1	A	226	GLY	Peptide
1	A	299	GLN	Peptide
1	A	501	LYS	Peptide

5.2 Too-close contacts

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	4055	0	3881	151	0
1	B	3970	0	3781	151	0
2	A	26	0	19	2	0
2	B	26	0	19	2	0
All	All	8077	0	7700	300	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 19.

The worst 5 of 300 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:B:320:ASN:HA	1:B:369:ILE:HD11	1.36	1.03
1:A:435:ARG:NH1	1:A:473:LEU:O	2.01	0.92
1:B:186:LEU:O	1:B:189:PHE:HB3	1.68	0.92
1:B:393:ASP:OD1	1:B:393:ASP:N	2.05	0.89
1:A:421:ILE:HG22	1:A:423:THR:H	1.42	0.84

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	517/535 (97%)	445 (86%)	67 (13%)	5 (1%)	19	65
1	B	507/535 (95%)	443 (87%)	58 (11%)	6 (1%)	16	60
All	All	1024/1070 (96%)	888 (87%)	125 (12%)	11 (1%)	17	62

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	300	ASN
1	A	504	VAL
1	B	505	VAL
1	A	299	GLN
1	B	200	PRO

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	426/484 (88%)	358 (84%)	68 (16%)	3	14
1	B	414/484 (86%)	348 (84%)	66 (16%)	3	14
All	All	840/968 (87%)	706 (84%)	134 (16%)	3	14

5 of 134 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	440	HIS
1	B	100	LEU
1	B	425	THR
1	A	453	THR
1	B	24	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	69	GLN
1	A	137	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

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

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	$\#\|Z\|>2$
2	SAH	A	601	-	22,28,28	1.07	2 (9%)	18,40,40	2.90	3 (16%)
2	SAH	B	601	-	22,28,28	1.07	2 (9%)	18,40,40	2.91	3 (16%)

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	SAH	A	601	-	-	0/7/31/31	0/3/3/3
2	SAH	B	601	-	-	0/7/31/31	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	601	SAH	C2-N1	2.33	1.38	1.33
2	A	601	SAH	C2-N1	2.37	1.38	1.33
2	A	601	SAH	C2-N3	3.62	1.38	1.32
2	B	601	SAH	C2-N3	3.63	1.38	1.32

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	SAH	N3-C2-N1	-10.93	120.29	128.87
2	A	601	SAH	N3-C2-N1	-10.79	120.39	128.87
2	B	601	SAH	C5'-SD-CG	-4.32	89.29	102.42
2	A	601	SAH	C5'-SD-CG	-3.69	91.22	102.42
2	B	601	SAH	O4'-C1'-N9	2.44	112.72	108.11

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	SAH	2	0
2	B	601	SAH	2	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

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	A	520/535 (97%)	-0.45	4 (0%)	87	80	33, 68, 101, 126	0
1	B	512/535 (95%)	-0.41	2 (0%)	93	90	44, 77, 105, 121	0
All	All	1032/1070 (96%)	-0.43	6 (0%)	90	84	33, 72, 104, 126	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	499	ASP	4.6
1	A	503	GLN	2.4
1	A	500	ASN	2.2
1	B	441	ALA	2.2
1	A	146	ALA	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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, 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	LLDF	B-factors(Å^2)	Q<0.9
2	SAH	A	601	26/26	0.92	0.18	-0.23	66,76,87,113	0
2	SAH	B	601	26/26	0.95	0.13	-1.01	53,67,78,85	0

6.5 Other polymers (i)

There are no such residues in this entry.