



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:42 am GMT

PDB ID : 3MJT  
Title : Structure of A-type Ketoreductases from Modular Polyketide Synthase  
Authors : Zheng, J.; Taylor, C.A.; Piasecki, S.K.; Keatinge-Clay, A.T.  
Deposited on : 2010-04-13  
Resolution : 1.60 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

<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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

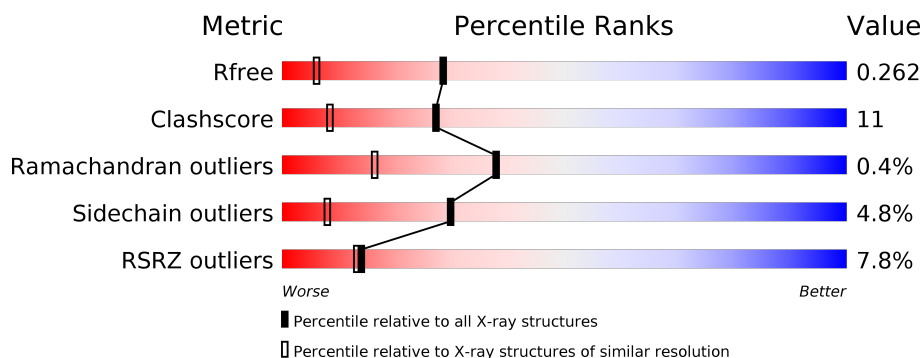
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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))
$R_{free}$	100719	2696 (1.60-1.60)
Clashscore	112137	2967 (1.60-1.60)
Ramachandran outliers	110173	2887 (1.60-1.60)
Sidechain outliers	110143	2886 (1.60-1.60)
RSRZ outliers	101464	2714 (1.60-1.60)

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  $\geq 3$ , 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	496	<div> <div>8%</div> <div> <div></div> <div>76%</div> <div>17%</div> <div>• •</div> </div> </div>
1	B	496	<div> <div>7%</div> <div> <div></div> <div>75%</div> <div>17%</div> <div>• • •</div> </div> </div>

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
2	NDP	A	476	X	-	-	-
2	NDP	B	476	X	-	-	-
3	GOL	A	477	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7581 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 AmphB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	476	Total	C	N	O	S	0	0	0
			3466	2164	636	657	9			
1	B	475	Total	C	N	O	S	0	0	0
			3460	2161	635	655	9			

There are 44 discrepancies between the modelled and reference sequences:

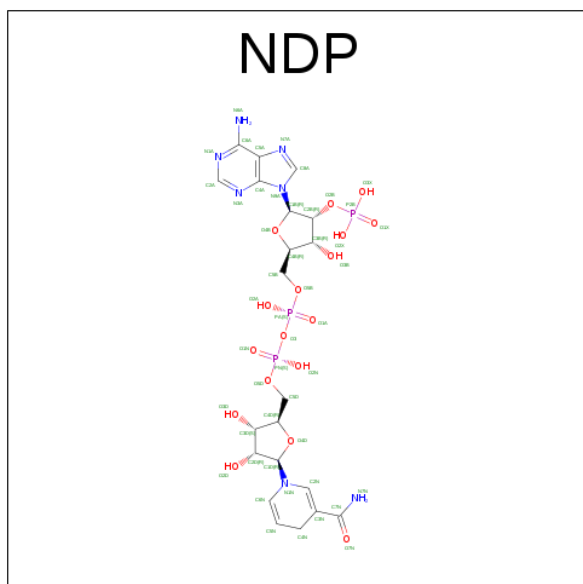
Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	EXPRESSION TAG	UNP Q93NW7
A	-19	GLY	-	EXPRESSION TAG	UNP Q93NW7
A	-18	SER	-	EXPRESSION TAG	UNP Q93NW7
A	-17	SER	-	EXPRESSION TAG	UNP Q93NW7
A	-16	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-15	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-14	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-13	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-12	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-11	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	-10	SER	-	EXPRESSION TAG	UNP Q93NW7
A	-9	SER	-	EXPRESSION TAG	UNP Q93NW7
A	-8	GLY	-	EXPRESSION TAG	UNP Q93NW7
A	-7	LEU	-	EXPRESSION TAG	UNP Q93NW7
A	-6	VAL	-	EXPRESSION TAG	UNP Q93NW7
A	-5	PRO	-	EXPRESSION TAG	UNP Q93NW7
A	-4	ARG	-	EXPRESSION TAG	UNP Q93NW7
A	-3	GLY	-	EXPRESSION TAG	UNP Q93NW7
A	-2	SER	-	EXPRESSION TAG	UNP Q93NW7
A	-1	HIS	-	EXPRESSION TAG	UNP Q93NW7
A	0	MET	-	EXPRESSION TAG	UNP Q93NW7
A	364	HIS	GLN	ENGINEERED	UNP Q93NW7
B	-20	MET	-	EXPRESSION TAG	UNP Q93NW7
B	-19	GLY	-	EXPRESSION TAG	UNP Q93NW7
B	-18	SER	-	EXPRESSION TAG	UNP Q93NW7

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-17	SER	-	EXPRESSION TAG	UNP Q93NW7
B	-16	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-15	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-14	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-13	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-12	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-11	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	-10	SER	-	EXPRESSION TAG	UNP Q93NW7
B	-9	SER	-	EXPRESSION TAG	UNP Q93NW7
B	-8	GLY	-	EXPRESSION TAG	UNP Q93NW7
B	-7	LEU	-	EXPRESSION TAG	UNP Q93NW7
B	-6	VAL	-	EXPRESSION TAG	UNP Q93NW7
B	-5	PRO	-	EXPRESSION TAG	UNP Q93NW7
B	-4	ARG	-	EXPRESSION TAG	UNP Q93NW7
B	-3	GLY	-	EXPRESSION TAG	UNP Q93NW7
B	-2	SER	-	EXPRESSION TAG	UNP Q93NW7
B	-1	HIS	-	EXPRESSION TAG	UNP Q93NW7
B	0	MET	-	EXPRESSION TAG	UNP Q93NW7
B	364	HIS	GLN	ENGINEERED	UNP Q93NW7

- Molecule 2 is NADPH DIHYDRO-NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NDP) (formula:  $C_{21}H_{30}N_7O_{17}P_3$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0
			48	21	7	17	3	

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	N	O	P	0	0
			48	21	7	17	3		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	268	Total	O	0	0
			268	268		
4	B	279	Total	O	0	0
			279	279		

These plots are drawn for all protein, RNA and DNA 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 ( $\text{RSRZ} > 2$ ). 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.

Chain A:

8% 76% 17%

MET GLY SER SER HIS HIS HIS HIS HIS SER SER GLY LEU VAL ARG PRO ARG GLY S-2 H-1 M0 R4 R11 R19 P20 A21 W37 V38 T39 A40 V41 D48 R51 L66 A67 R70 E75 F76 H77 E87 H90 T91 D92 L115 W120 R124 G125 A128 A129 G130 R131 L135 L136 P138 V139 Q140 L143 W144 R148 L164 P165 E166 T167 R171 R175 L180 D187 R192 G193 S194 R200 H203 V208 P209 G210 S211 G212 K213 P216 V217 T224 R233 R236 E240 H245 T256 L262 R263 A264 E265 Q268 T274 D279 E284 A285 L286 L289 L290 P294 E295 D296 A297 P298 L299 H304 D311 R328 L331 L340 T341 A342 D343 L344 F352 A357 V358 H364 P365 G366 Y367 Y373 H380 R381 S389 S390 V391 A392 H392

[illegible]

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.40Å 63.61Å 71.85Å 72.84° 67.20° 89.77°	Depositor
Resolution (Å)	62.79 – 1.60 30.67 – 1.60	Depositor EDS
% Data completeness (in resolution range)	94.7 (62.79-1.60) 86.5 (30.67-1.60)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.32 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.223 , 0.265 0.222 , 0.262	Depositor DCC
$R_{free}$ test set	5935 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.5	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 47.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.109 for h,-k,h-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7581	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.74% 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: GOL, NDP

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	1.26	5/3537 (0.1%)	1.19	21/4832 (0.4%)
1	B	1.27	8/3531 (0.2%)	1.32	32/4824 (0.7%)
All	All	1.27	13/7068 (0.2%)	1.25	53/9656 (0.5%)

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	3

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	120	TRP	CG-CD1	7.90	1.47	1.36
1	A	39	THR	CB-CG2	-7.00	1.29	1.52
1	B	457	ARG	CB-CG	-6.98	1.33	1.52
1	A	357	ALA	CA-CB	6.67	1.66	1.52
1	B	17	THR	N-CA	6.31	1.58	1.46

The worst 5 of 53 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	382	ARG	NE-CZ-NH1	25.26	132.93	120.30
1	B	382	ARG	NE-CZ-NH2	-21.11	109.74	120.30
1	B	121	CYS	CA-CB-SG	-13.34	89.99	114.00
1	A	200	ARG	NE-CZ-NH1	12.27	126.44	120.30
1	A	200	ARG	NE-CZ-NH2	-11.79	114.41	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	16	GLY	Peptide
1	B	17	THR	Peptide
1	B	471	ALA	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	3466	0	3425	66	0
1	B	3460	0	3420	93	0
2	A	48	0	25	1	0
2	B	48	0	25	1	0
3	A	6	0	8	4	0
3	B	6	0	8	0	0
4	A	268	0	0	12	0
4	B	279	0	0	8	0
All	All	7581	0	6911	159	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 11.

The worst 5 of 159 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:121:CYS:HB3	4:B:756:HOH:O	1.11	1.27
1:B:121:CYS:SG	1:B:121:CYS:O	2.04	1.14
1:B:461:LEU:O	1:B:461:LEU:HD13	1.54	1.08
1:B:470:SER:O	1:B:471:ALA:HB3	1.55	1.03
1:B:121:CYS:SG	1:B:161:LEU:HD23	2.02	0.98

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	474/496 (96%)	459 (97%)	13 (3%)	2 (0%)	38	16
1	B	473/496 (95%)	459 (97%)	12 (2%)	2 (0%)	38	16
All	All	947/992 (96%)	918 (97%)	25 (3%)	4 (0%)	38	16

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	17	THR
1	B	471	ALA
1	A	394	GLY
1	A	405	PRO

### 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	334/351 (95%)	320 (96%)	14 (4%)	34	11
1	B	333/351 (95%)	315 (95%)	18 (5%)	26	6
All	All	667/702 (95%)	635 (95%)	32 (5%)	30	8

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	-1	HIS
1	B	82	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	421	PRO
1	B	73	GLU
1	B	124	ARG

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	A	336	HIS
1	A	456	ASN
1	B	380	HIS
1	A	380	HIS
1	A	414	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](#)

4 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	NDP	A	476	-	43,52,52	1.36	7 (16%)	49,80,80	2.97	14 (28%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	GOL	A	477	-	5,5,5	0.45	0	5,5,5	0.46	0
2	NDP	B	476	-	43,52,52	1.50	7 (16%)	49,80,80	2.69	15 (30%)
3	GOL	B	477	-	5,5,5	0.54	0	5,5,5	0.73	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
2	NDP	A	476	-	1/1/14/17	0/30/77/77	0/5/5/5
3	GOL	A	477	-	-	0/4/4/4	0/0/0/0
2	NDP	B	476	-	1/1/14/17	0/30/77/77	0/5/5/5
3	GOL	B	477	-	-	0/4/4/4	0/0/0/0

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	476	NDP	C4N-C5N	-4.41	1.39	1.49
2	A	476	NDP	C4N-C5N	-3.84	1.40	1.49
2	A	476	NDP	C5A-C4A	-2.73	1.34	1.40
2	B	476	NDP	C5A-N7A	-2.24	1.31	1.39
2	A	476	NDP	C2D-C3D	-2.14	1.47	1.53

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	476	NDP	N3A-C2A-N1A	-12.49	117.98	128.86
2	B	476	NDP	N3A-C2A-N1A	-10.81	119.44	128.86
2	A	476	NDP	C4B-O4B-C1B	-10.66	98.43	109.77
2	B	476	NDP	C4B-O4B-C1B	-8.28	100.95	109.77
2	A	476	NDP	C2B-C3B-C4B	-5.12	90.31	101.95

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	476	NDP	C4B
2	A	476	NDP	C4B

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	476	NDP	1	0
3	A	477	GOL	4	0
2	B	476	NDP	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, 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	476/496 (95%)	0.65	39 (8%) 12 12	9, 17, 34, 60	0
1	B	475/496 (95%)	0.55	35 (7%) 15 15	8, 16, 34, 61	0
All	All	951/992 (95%)	0.60	74 (7%) 14 13	8, 17, 34, 61	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	411	LEU	8.5
1	B	402	ALA	7.6
1	B	405	PRO	6.9
1	B	471	ALA	6.7
1	A	412	VAL	6.3

### 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, 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	A	477	6/6	0.82	0.14	1.44	32,33,38,40	0
3	GOL	B	477	6/6	0.88	0.12	0.33	23,24,28,29	0
2	NDP	B	476	48/48	0.94	0.10	-0.35	11,15,19,20	0
2	NDP	A	476	48/48	0.93	0.11	-0.39	12,18,22,22	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.