wwPDB X-ray Structure Validation Summary Report (i)

Nov 14, 2023 – 06:57 PM JST

PDB ID	:	5ZZJ
Title	:	Crystal structure of a enzyme from Santalum album
Authors	:	Han, X.; Liu, W.D.; Chen, C.C.; Guo, R.T.
Deposited on	:	2018-06-01
Resolution	:	2.60 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

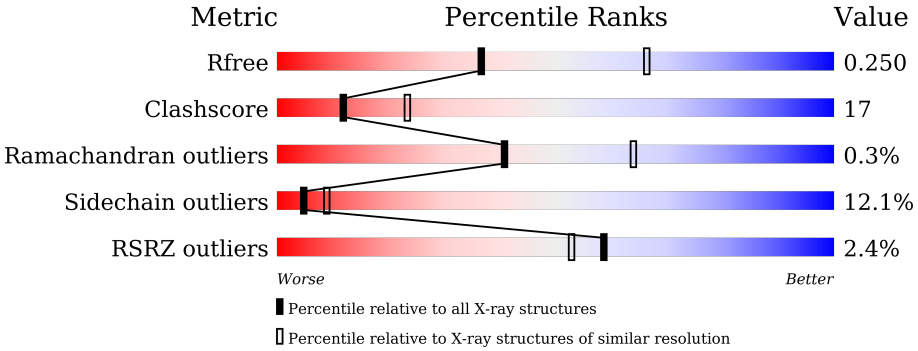
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.60 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}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	569	%, 66%, 22%, •, 8%
1	B	569	2%, 61%, 26%, •, 9%
1	C	569	3%, 62%, 24%, 6%, 9%
1	D	569	3%, 58%, 28%, 6%, 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 17019 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 Santalene synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	522	4231	2736	707	769	19	0	0	0
1	B	520	4215	2728	704	764	19	0	0	0
1	C	519	4211	2726	703	763	19	0	0	0
1	D	521	4227	2734	706	768	19	0	0	0

- Molecule 2 is water.

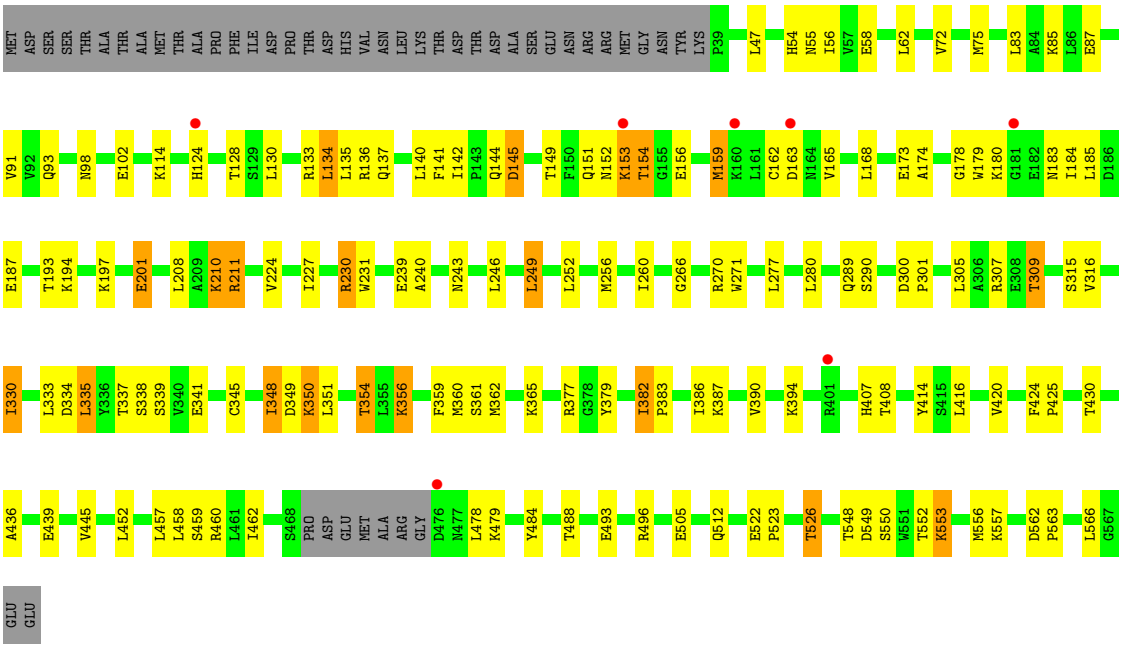
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	45	45	45	0	0
2	B	45	45	45	0	0
2	C	35	35	35	0	0
2	D	10	10	10	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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 (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.

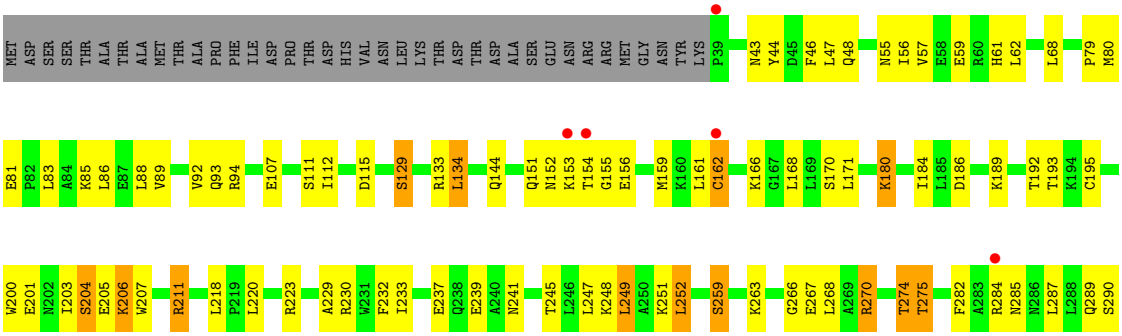
• Molecule 1: Santalene synthase

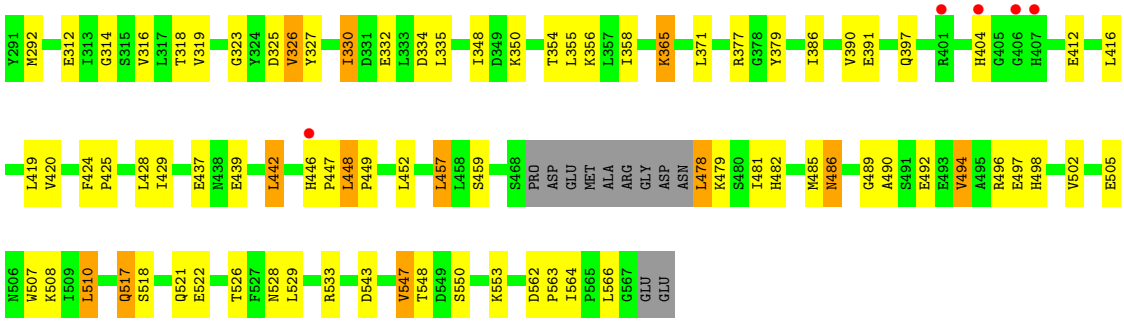
Chain A:



• Molecule 1: Santalene synthase

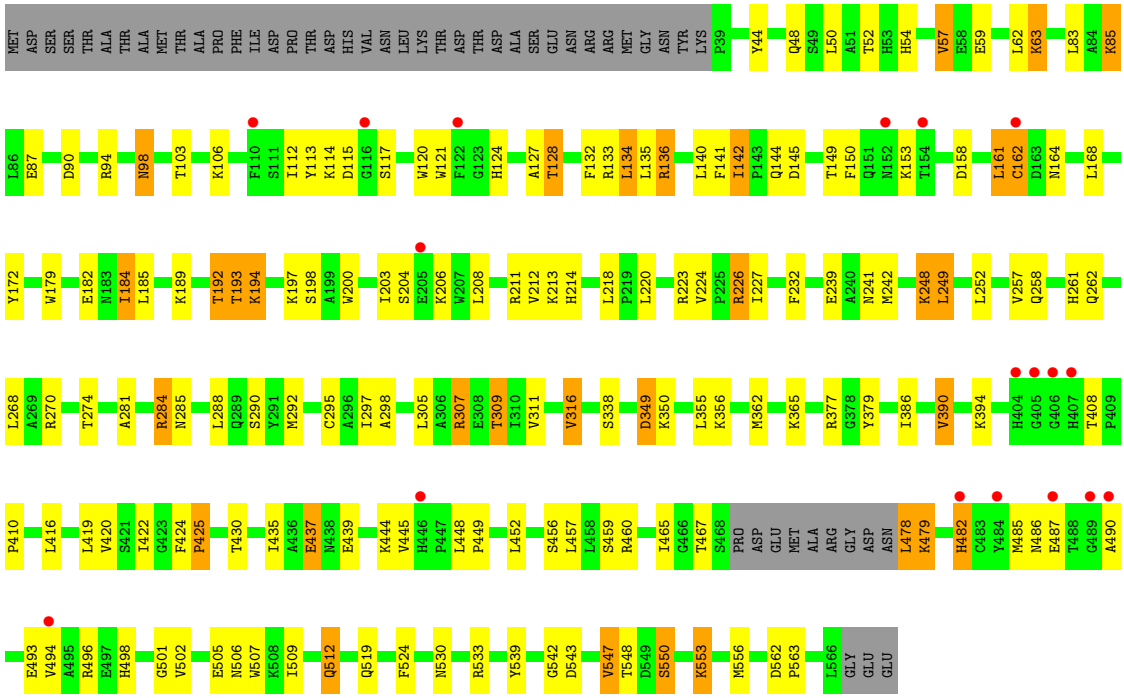
Chain B:





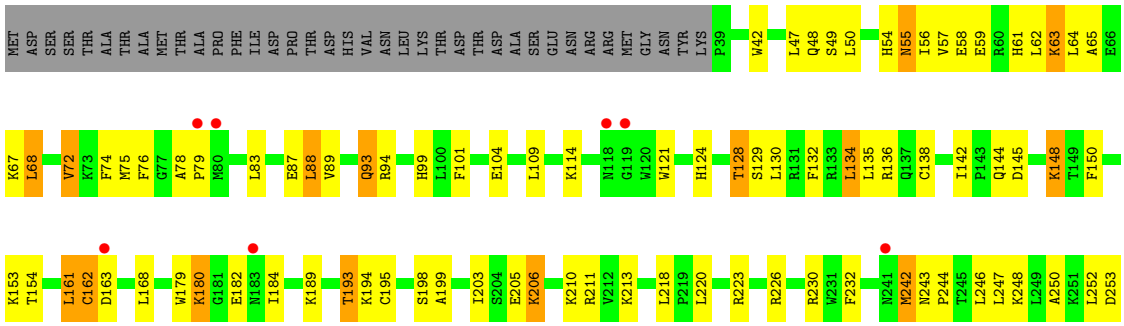
- Molecule 1: Santalene synthase

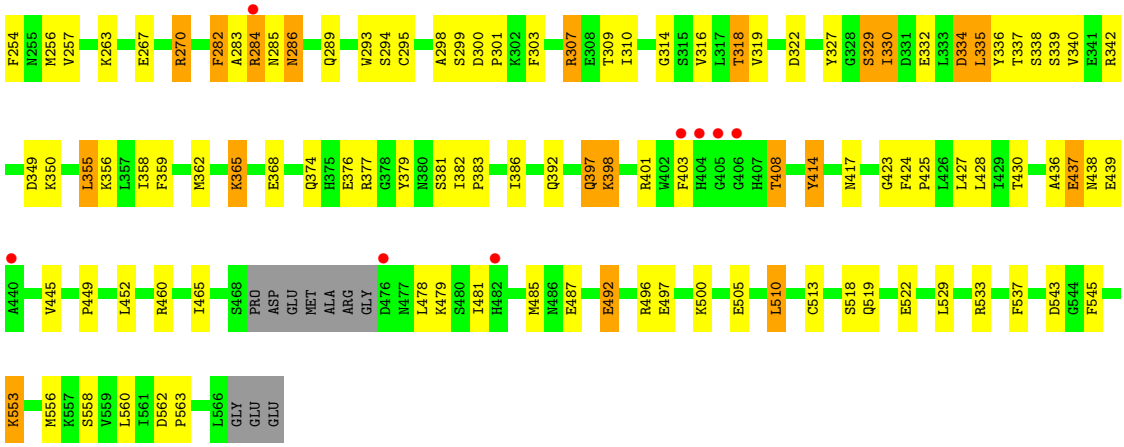
Chain C:



- Molecule 1: Santalene synthase

Chain D:




F254 N255 M256 V257 K263 E267 R270 F282 A283 R284 N285 N286 Q289 W293 S294 C295 A298 S299 D300 P301 K302 F303 R307 E308 T309 I310 G314 S315 V316 L317 T318 V319 D322 Y327 G328 S329 I330 D331 E332 L333 D334 L335 Y336 T337 S338 S339 V340 E341 R342
D349 K350 L355 K356 L357 I358 F359 M362 K365 E368 Q374 H375 E376 R377 G378 Y379 N380 S381 I382 P383 I386 Q392 Q397 K398 R401 W402 F403 H404 G405 G406 H407 T408 Y414 N417 G423 F424 P425 L426 L427 L428 I429 T430 A436 E437 N438 E439
A440 V445 P449 L452 R460 I465 S468 PRO ASP GLU MET ALA ARG GLY D476 M477 L478 K479 S480 I481 H482 M485 M486 E487 E492 R496 E497 K500 E505 L510 C513 S518 Q519 E522 L529 R533 F537 D543 G544 F545
K553 M556 K557 S558 V559 L560 I561 D562 P563 L566 GLY GLU GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	122.51Å 155.11Å 164.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 2.60 24.99 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.9 (25.00-2.60) 99.1 (24.99-2.60)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.43 (at 2.60Å)	Xtriage
Refinement program	REFMAC 5.8.0222	Depositor
R, R_{free}	0.189 , 0.234 0.217 , 0.250	Depositor DCC
R_{free} test set	4938 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	59.5	Xtriage
Anisotropy	0.017	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 39.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	17019	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.63% 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 (i)

5.1 Standard geometry (i)

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	0.82	0/4341	0.69	0/5879
1	B	0.79	0/4325	0.69	0/5857
1	C	0.76	0/4321	0.68	0/5852
1	D	0.72	0/4337	0.71	0/5874
All	All	0.77	0/17324	0.69	0/23462

There are no bond length outliers.

There are no bond angle outliers.

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	4231	0	4176	107	0
1	B	4215	0	4166	144	0
1	C	4211	0	4161	149	0
1	D	4227	0	4173	187	0
2	A	45	0	0	3	0
2	B	45	0	0	2	0
2	C	35	0	0	2	0
2	D	10	0	0	1	0
All	All	17019	0	16676	580	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 17.

The worst 5 of 580 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:162:CYS:HB3	1:B:195:CYS:SG	1.47	1.50
1:C:141:PHE:O	1:C:142:ILE:HD13	1.30	1.26
1:D:76:PHE:CD2	1:D:104:GLU:HB3	1.71	1.25
1:B:162:CYS:CB	1:B:195:CYS:SG	2.31	1.17
1:C:208:LEU:O	1:C:212:VAL:HG23	1.45	1.16

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	518/569 (91%)	503 (97%)	14 (3%)	1 (0%)	47	71
1	B	516/569 (91%)	503 (98%)	10 (2%)	3 (1%)	25	47
1	C	515/569 (90%)	497 (96%)	17 (3%)	1 (0%)	47	71
1	D	517/569 (91%)	491 (95%)	24 (5%)	2 (0%)	34	57
All	All	2066/2276 (91%)	1994 (96%)	65 (3%)	7 (0%)	41	64

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	240	ALA
1	B	115	ASP
1	C	164	ASN
1	D	284	ARG
1	D	55	ASN

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	459/500 (92%)	409 (89%)	50 (11%)	6 11
1	B	457/500 (91%)	409 (90%)	48 (10%)	7 13
1	C	457/500 (91%)	395 (86%)	62 (14%)	3 6
1	D	459/500 (92%)	397 (86%)	62 (14%)	4 6
All	All	1832/2000 (92%)	1610 (88%)	222 (12%)	5 9

5 of 222 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	162	CYS
1	D	553	LYS
1	C	425	PRO
1	D	510	LEU
1	D	330	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	486	ASN
1	B	517	GLN
1	C	506	ASN
1	D	506	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	522/569 (91%)	-0.29	7 (1%)	77	73	35, 55, 91, 134	0
1	B	520/569 (91%)	-0.27	10 (1%)	66	62	30, 57, 95, 134	0
1	C	519/569 (91%)	-0.13	18 (3%)	44	36	34, 66, 108, 125	0
1	D	521/569 (91%)	-0.05	15 (2%)	51	45	30, 71, 107, 138	0
All	All	2082/2276 (91%)	-0.19	50 (2%)	59	53	30, 63, 101, 138	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	446	HIS	5.9
1	B	284	ARG	4.8
1	C	405	GLY	4.5
1	C	489	GLY	4.4
1	D	404	HIS	4.3

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.