



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 14, 2018 – 01:17 am GMT

PDB ID : 1VH1  
Title : Crystal structure of CMP-KDO synthetase  
Authors : Structural GenomiX  
Deposited on : 2003-12-01  
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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : trunk31020  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk31020

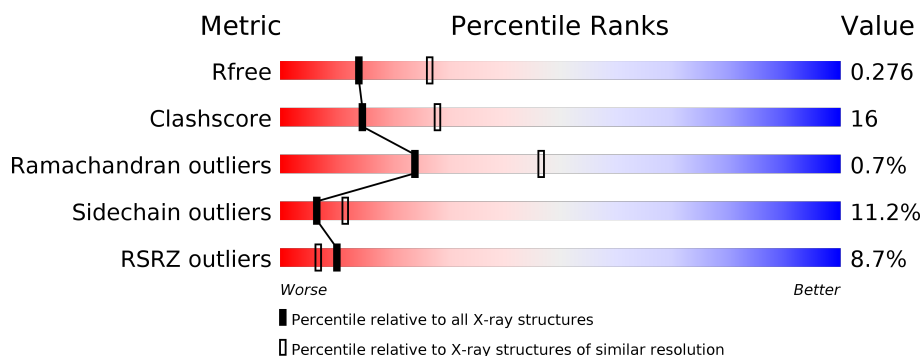
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
$R_{free}$	111664	2767 (2.60-2.60)
Clashscore	122126	3110 (2.60-2.60)
Ramachandran outliers	120053	3062 (2.60-2.60)
Sidechain outliers	120020	3062 (2.60-2.60)
RSRZ outliers	108989	2706 (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 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	256	<div> <div>8%</div> <div>64%</div> <div>24%</div> <div>5%</div> <div>7%</div> </div>
1	B	256	<div> <div>11%</div> <div>51%</div> <div>27%</div> <div>8%</div> <div>13%</div> </div>
1	C	256	<div> <div>5%</div> <div>64%</div> <div>22%</div> <div>5%</div> <div>9%</div> </div>
1	D	256	<div> <div>6%</div> <div>60%</div> <div>29%</div> <div>5%</div> <div>6%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7321 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 3-deoxy-manno-octulosonate cytidyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	239	Total	C	N	O	S	Se	0	0	0
			1826	1157	324	337	2	6			
1	B	224	Total	C	N	O	S	Se	0	0	0
			1729	1096	310	315	2	6			
1	C	233	Total	C	N	O	S	Se	0	0	0
			1782	1133	313	328	2	6			
1	D	241	Total	C	N	O	S	Se	0	0	0
			1828	1161	321	338	2	6			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	cloning artifact	UNP P04951
A	29	MSE	MET	modified residue	UNP P04951
A	67	MSE	MET	modified residue	UNP P04951
A	103	MSE	MET	modified residue	UNP P04951
A	123	MSE	MET	modified residue	UNP P04951
A	208	MSE	MET	modified residue	UNP P04951
A	247	MSE	MET	modified residue	UNP P04951
A	249	GLY	-	cloning artifact	UNP P04951
A	250	SER	-	cloning artifact	UNP P04951
A	251	HIS	-	cloning artifact	UNP P04951
A	252	HIS	-	cloning artifact	UNP P04951
A	253	HIS	-	cloning artifact	UNP P04951
A	254	HIS	-	cloning artifact	UNP P04951
A	255	HIS	-	cloning artifact	UNP P04951
A	256	HIS	-	cloning artifact	UNP P04951
B	1	MSE	-	cloning artifact	UNP P04951
B	29	MSE	MET	modified residue	UNP P04951
B	67	MSE	MET	modified residue	UNP P04951
B	103	MSE	MET	modified residue	UNP P04951
B	123	MSE	MET	modified residue	UNP P04951
B	208	MSE	MET	modified residue	UNP P04951

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	247	MSE	MET	modified residue	UNP P04951
B	249	GLY	-	cloning artifact	UNP P04951
B	250	SER	-	cloning artifact	UNP P04951
B	251	HIS	-	cloning artifact	UNP P04951
B	252	HIS	-	cloning artifact	UNP P04951
B	253	HIS	-	cloning artifact	UNP P04951
B	254	HIS	-	cloning artifact	UNP P04951
B	255	HIS	-	cloning artifact	UNP P04951
B	256	HIS	-	cloning artifact	UNP P04951
C	1	MSE	-	cloning artifact	UNP P04951
C	29	MSE	MET	modified residue	UNP P04951
C	67	MSE	MET	modified residue	UNP P04951
C	103	MSE	MET	modified residue	UNP P04951
C	123	MSE	MET	modified residue	UNP P04951
C	208	MSE	MET	modified residue	UNP P04951
C	247	MSE	MET	modified residue	UNP P04951
C	249	GLY	-	cloning artifact	UNP P04951
C	250	SER	-	cloning artifact	UNP P04951
C	251	HIS	-	cloning artifact	UNP P04951
C	252	HIS	-	cloning artifact	UNP P04951
C	253	HIS	-	cloning artifact	UNP P04951
C	254	HIS	-	cloning artifact	UNP P04951
C	255	HIS	-	cloning artifact	UNP P04951
C	256	HIS	-	cloning artifact	UNP P04951
D	1	MSE	-	cloning artifact	UNP P04951
D	29	MSE	MET	modified residue	UNP P04951
D	67	MSE	MET	modified residue	UNP P04951
D	103	MSE	MET	modified residue	UNP P04951
D	123	MSE	MET	modified residue	UNP P04951
D	208	MSE	MET	modified residue	UNP P04951
D	247	MSE	MET	modified residue	UNP P04951
D	249	GLY	-	cloning artifact	UNP P04951
D	250	SER	-	cloning artifact	UNP P04951
D	251	HIS	-	cloning artifact	UNP P04951
D	252	HIS	-	cloning artifact	UNP P04951
D	253	HIS	-	cloning artifact	UNP P04951
D	254	HIS	-	cloning artifact	UNP P04951
D	255	HIS	-	cloning artifact	UNP P04951
D	256	HIS	-	cloning artifact	UNP P04951

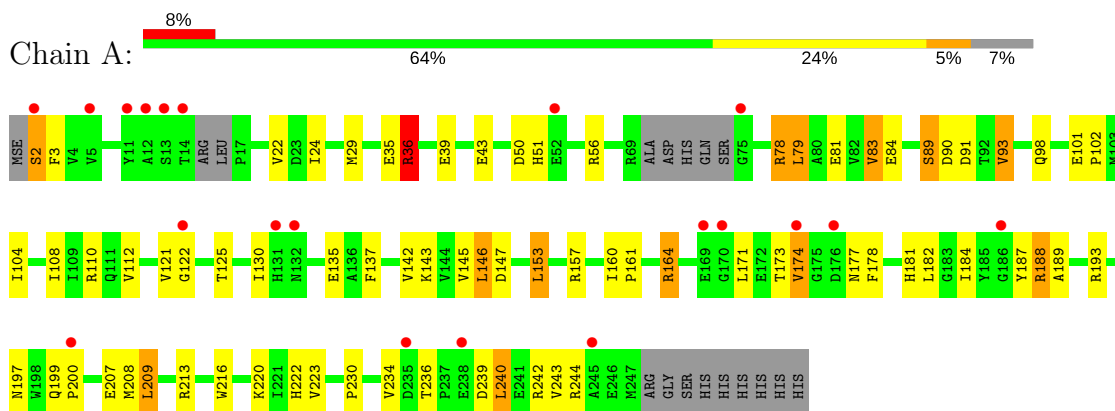
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	45	Total 45	O 45	0	0
2	B	34	Total 34	O 34	0	0
2	C	42	Total 42	O 42	0	0
2	D	35	Total 35	O 35	0	0

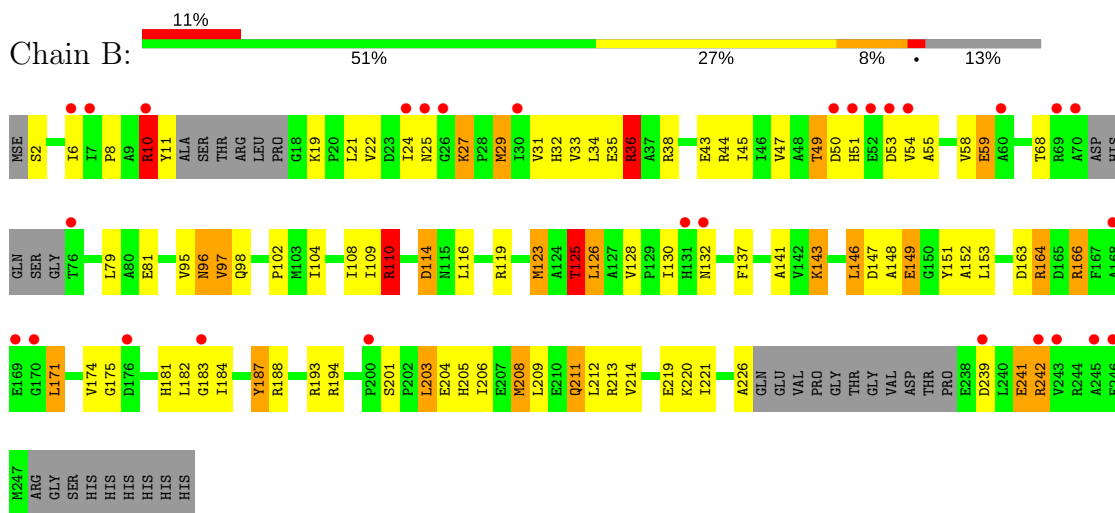
### 3 Residue-property plots

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 ( $RSRZ > 2$ ). 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.

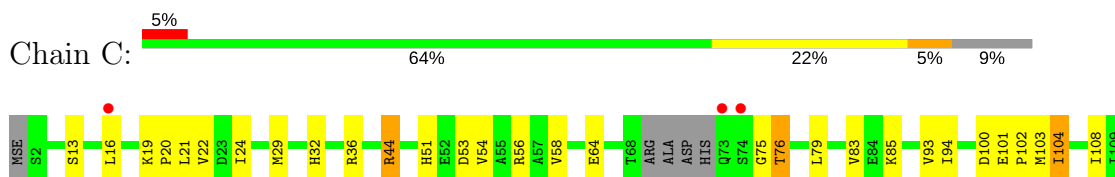
#### • Molecule 1: 3-deoxy-manno-octulosonate cytidyltransferase

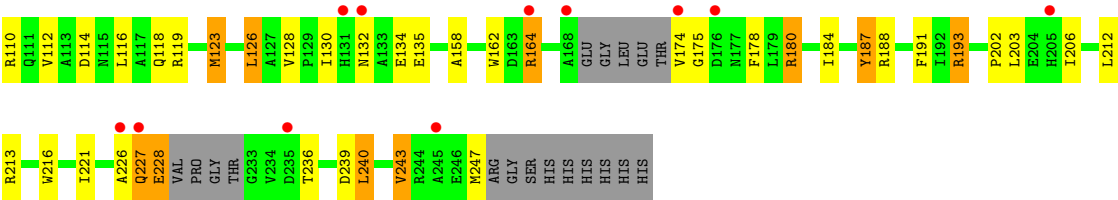


#### • Molecule 1: 3-deoxy-manno-octulosonate cytidyltransferase

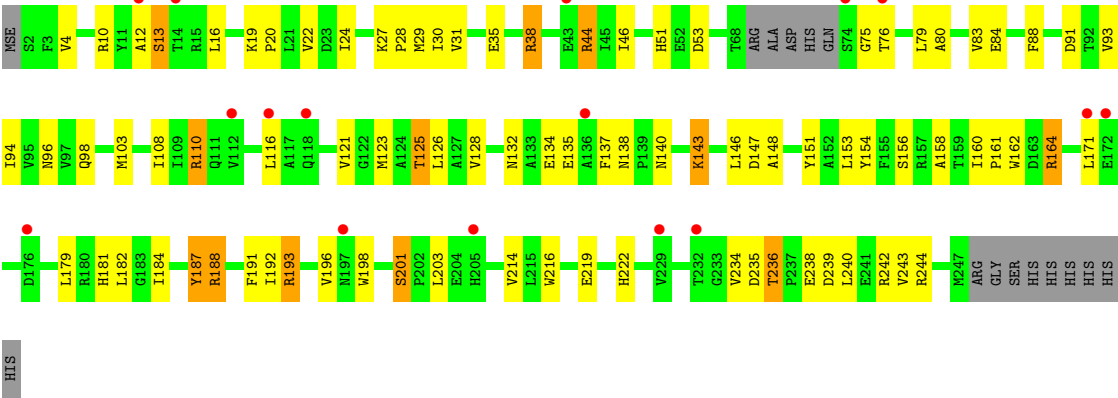


#### • Molecule 1: 3-deoxy-manno-octulosonate cytidyltransferase





• Molecule 1: 3-deoxy-manno-octulosonate cytidyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.59Å 130.81Å 159.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.92 – 2.60 19.92 – 2.60	Depositor EDS
% Data completeness (in resolution range)	(Not available) (19.92-2.60) 100.0 (19.92-2.60)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.88 (at 2.59Å)	Xtriage
Refinement program	REFMAC 4.0	Depositor
R, $R_{free}$	0.240 , 0.302 0.220 , 0.276	Depositor DCC
$R_{free}$ test set	1718 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.7	Xtriage
Anisotropy	0.304	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 64.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7321	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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](#)

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	0.61	0/1856	1.11	13/2517 (0.5%)
1	B	0.60	0/1755	1.15	14/2374 (0.6%)
1	C	0.62	0/1810	1.12	9/2453 (0.4%)
1	D	0.61	0/1859	1.15	12/2525 (0.5%)
All	All	0.61	0/7280	1.13	48/9869 (0.5%)

There are no bond length outliers.

The worst 5 of 48 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	38	ARG	CD-NE-CZ	12.28	140.79	123.60
1	C	193	ARG	CD-NE-CZ	10.70	138.58	123.60
1	D	38	ARG	NE-CZ-NH1	10.33	125.46	120.30
1	A	78	ARG	CD-NE-CZ	10.05	137.67	123.60
1	D	188	ARG	NE-CZ-NH1	-9.77	115.42	120.30

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	1826	0	1785	45	0
1	B	1729	0	1693	80	0
1	C	1782	0	1741	54	0
1	D	1828	0	1786	61	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	45	0	0	2	0
2	B	34	0	0	4	0
2	C	42	0	0	3	0
2	D	35	0	0	5	0
All	All	7321	0	7005	232	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 16.

The worst 5 of 232 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:103:MSE:HG3	1:C:247:MSE:HE2	1.46	0.97
1:B:49:THR:HG21	1:B:54:VAL:HB	1.54	0.89
1:D:236:THR:HG22	1:D:239:ASP:H	1.37	0.88
1:B:125:THR:HG21	1:B:181:HIS:HE1	1.40	0.87
1:A:29:MSE:HG3	1:A:102:PRO:HG3	1.57	0.86

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	233/256 (91%)	225 (97%)	8 (3%)	0	100	100
1	B	216/256 (84%)	202 (94%)	11 (5%)	3 (1%)	12	24
1	C	225/256 (88%)	215 (96%)	9 (4%)	1 (0%)	36	60
1	D	237/256 (93%)	230 (97%)	5 (2%)	2 (1%)	21	42
All	All	911/1024 (89%)	872 (96%)	33 (4%)	6 (1%)	24	46

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	227	GLN
1	D	13	SER
1	B	96	ASN
1	B	97	VAL
1	D	75	GLY

### 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	185/203 (91%)	166 (90%)	19 (10%)	8	14
1	B	174/203 (86%)	143 (82%)	31 (18%)	2	3
1	C	180/203 (89%)	165 (92%)	15 (8%)	12	24
1	D	185/203 (91%)	169 (91%)	16 (9%)	11	22
All	All	724/812 (89%)	643 (89%)	81 (11%)	6	11

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	143	LYS
1	B	203	LEU
1	D	146	LEU
1	B	146	LEU
1	B	166	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	120	GLN
1	B	181	HIS
1	C	98	GLN
1	B	115	ASN
1	C	120	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](#)

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, 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	233/256 (91%)	0.46	20 (8%) 10 7	51, 64, 94, 120	0
1	B	218/256 (85%)	0.65	29 (13%) 3 2	49, 68, 101, 124	0
1	C	227/256 (88%)	0.38	14 (6%) 20 15	49, 63, 95, 104	0
1	D	235/256 (91%)	0.51	16 (6%) 17 12	51, 68, 99, 111	0
All	All	913/1024 (89%)	0.50	79 (8%) 10 7	49, 66, 98, 124	0

The worst 5 of 79 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	74	SER	5.8
1	B	24	ILE	5.3
1	D	14	THR	5.1
1	C	73	GLN	4.9
1	D	176	ASP	4.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.