



# wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 20, 2023 – 07:13 am BST

PDB ID : 8CJH  
Title : Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin  
Authors : Bonhomme, S.; Dessen, A.; Macheboeuf, P.  
Deposited on : 2023-02-13  
Resolution : 2.98 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.32.2  
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) : 2.32.2

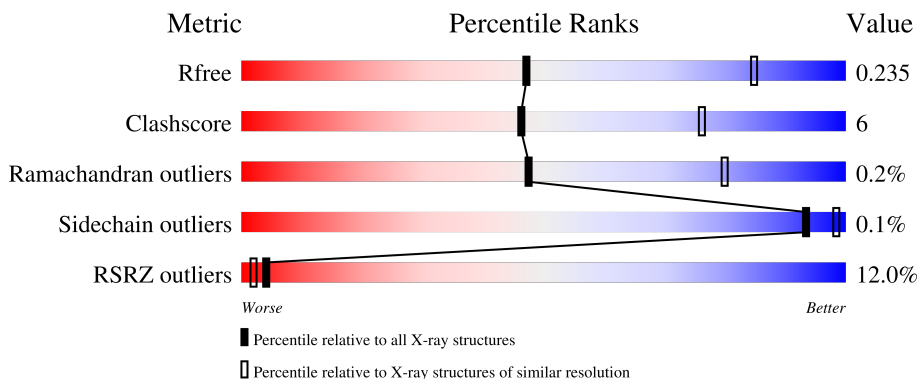
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.98 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.



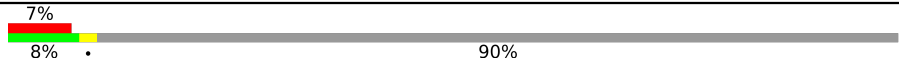
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2754 (3.00-2.96)
Clashscore	141614	3103 (3.00-2.96)
Ramachandran outliers	138981	2993 (3.00-2.96)
Sidechain outliers	138945	2996 (3.00-2.96)
RSRZ outliers	127900	2644 (3.00-2.96)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	787	
1	B	787	
1	C	787	
1	D	787	
1	E	787	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	787	 7% 8% . 90%

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 22022 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.

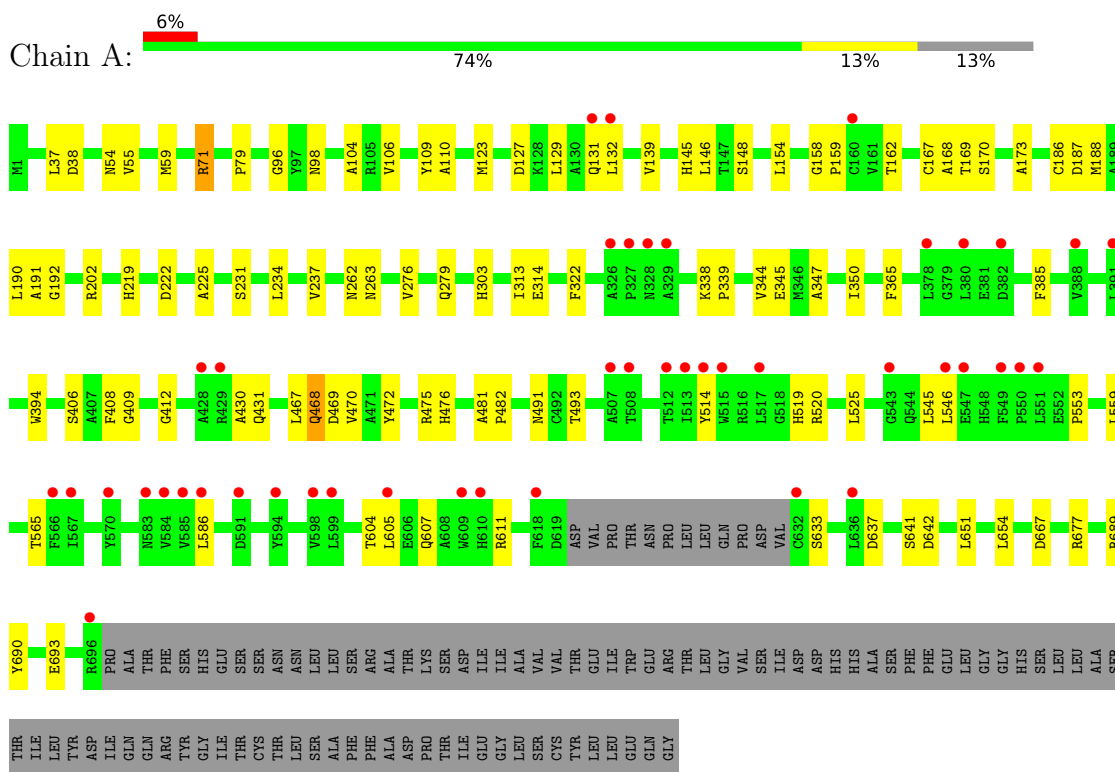
- Molecule 1 is a protein called Colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase ClbK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	684	Total	C	N	O	S	0	0	0
			5202	3264	918	987	33			
1	B	684	Total	C	N	O	S	0	0	0
			5202	3264	918	987	33			
1	C	684	Total	C	N	O	S	0	0	0
			5202	3264	918	987	33			
1	D	684	Total	C	N	O	S	0	0	0
			5202	3264	918	987	33			
1	E	78	Total	C	N	O	S	0	0	0
			607	388	99	118	2			
1	F	78	Total	C	N	O	S	0	0	0
			607	388	99	118	2			

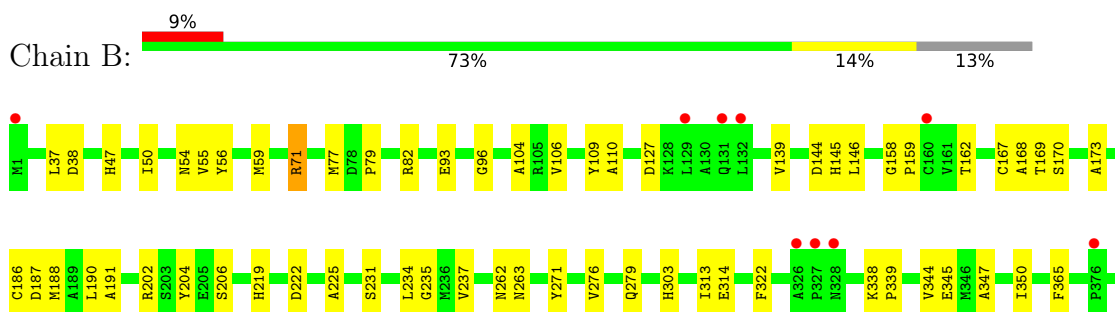
### 3 Residue-property plots [i](#)

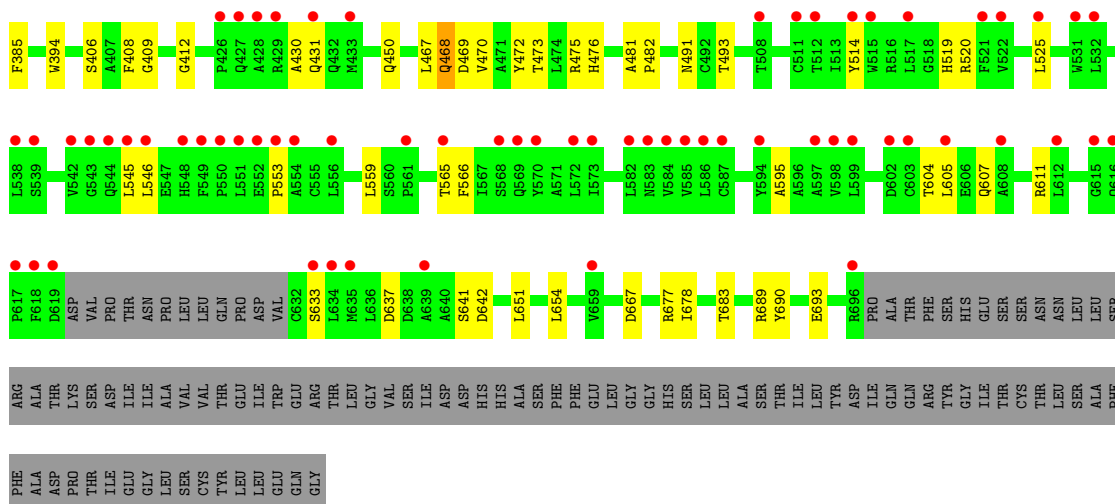
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase C1bK

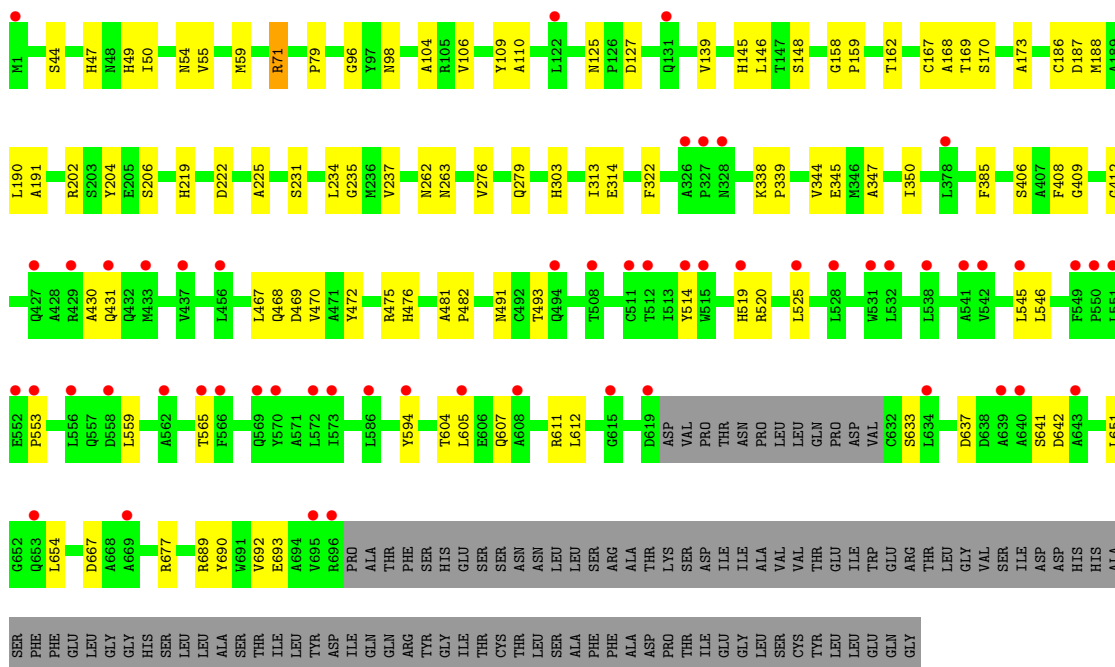
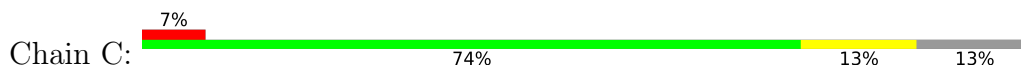


- Molecule 1: Colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase C1bK

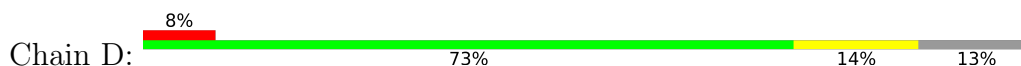




- Molecule 1: Colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase ClbK



- Molecule 1: Colibactin hybrid non-ribosomal peptide synthetase/type I polyketide synthase ClbK









## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	187.64Å 233.37Å 256.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.40 – 2.98 49.35 – 2.98	Depositor EDS
% Data completeness (in resolution range)	68.5 (49.40-2.98) 68.5 (49.35-2.98)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.17 (at 2.96Å)	Xtrriage
Refinement program	REFMAC 5.8.0403	Depositor
R, $R_{free}$	0.217 , 0.238 0.216 , 0.235	Depositor DCC
$R_{free}$ test set	1940 reflections (2.48%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	81.7	Xtrriage
Anisotropy	0.031	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 49.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	22022	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	96.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.33% 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	0.37	0/5311	0.62	1/7233 (0.0%)
1	B	0.38	0/5311	0.62	1/7233 (0.0%)
1	C	0.39	0/5311	0.63	0/7233
1	D	0.38	0/5311	0.61	0/7233
1	E	0.37	0/619	0.55	0/842
1	F	0.36	0/619	0.55	0/842
All	All	0.38	0/22482	0.62	2/30616 (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	1
1	B	0	1
1	C	0	1
1	D	0	1
All	All	0	4

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	468	GLN	CB-CA-C	-5.24	99.92	110.40
1	B	468	GLN	CB-CA-C	-5.03	100.35	110.40

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	71	ARG	Sidechain

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	71	ARG	Sidechain
1	C	71	ARG	Sidechain
1	D	71	ARG	Sidechain

## 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	5202	0	5102	61	0
1	B	5202	0	5102	66	0
1	C	5202	0	5102	63	0
1	D	5202	0	5102	67	1
1	E	607	0	593	11	0
1	F	607	0	593	13	0
All	All	22022	0	21594	275	1

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 6.

The worst 5 of 275 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:F:755:TYR:O	1:F:759:GLN:HG2	1.75	0.85
1:E:755:TYR:O	1:E:759:GLN:HG2	1.78	0.84
1:D:525:LEU:HD21	1:D:553:PRO:HB3	1.70	0.73
1:B:525:LEU:HD21	1:B:553:PRO:HB3	1.71	0.70
1:D:219:HIS:HB3	1:D:222:ASP:HB3	1.75	0.69

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:424:SER:O	1:D:557:GLN:NE2[3_454]	1.91	0.29

## 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	680/787 (86%)	632 (93%)	46 (7%)	2 (0%)	41	74
1	B	680/787 (86%)	630 (93%)	48 (7%)	2 (0%)	41	74
1	C	680/787 (86%)	632 (93%)	46 (7%)	2 (0%)	41	74
1	D	680/787 (86%)	634 (93%)	45 (7%)	1 (0%)	51	83
1	E	76/787 (10%)	72 (95%)	4 (5%)	0	100	100
1	F	76/787 (10%)	72 (95%)	4 (5%)	0	100	100
All	All	2872/4722 (61%)	2672 (93%)	193 (7%)	7 (0%)	47	80

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	127	ASP
1	B	127	ASP
1	C	127	ASP
1	D	127	ASP
1	A	430	ALA

### 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	550/640 (86%)	549 (100%)	1 (0%)	93	98
1	B	550/640 (86%)	549 (100%)	1 (0%)	93	98
1	C	550/640 (86%)	550 (100%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	550/640 (86%)	550 (100%)	0	100	100
1	E	67/640 (10%)	66 (98%)	1 (2%)	65	86
1	F	67/640 (10%)	67 (100%)	0	100	100
All	All	2334/3840 (61%)	2331 (100%)	3 (0%)	93	98

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	586	LEU
1	B	144	ASP
1	E	715	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	131	GLN
1	D	343	HIS
1	D	303	HIS
1	D	446	ASN
1	B	303	HIS

### 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, 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	684/787 (86%)	0.32	45 (6%) 18 9	48, 83, 154, 180	0
1	B	684/787 (86%)	0.50	74 (10%) 5 3	43, 75, 168, 192	0
1	C	684/787 (86%)	0.45	56 (8%) 11 5	50, 81, 161, 193	0
1	D	684/787 (86%)	0.55	63 (9%) 9 5	47, 84, 160, 196	0
1	E	78/787 (9%)	2.91	54 (69%) 0 0	126, 149, 178, 189	0
1	F	78/787 (9%)	3.66	55 (70%) 0 0	147, 178, 203, 213	0
All	All	2892/4722 (61%)	0.61	347 (11%) 4 2	43, 83, 168, 213	0

The worst 5 of 347 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	763	ILE	8.9
1	F	709	LEU	8.1
1	F	781	CYS	7.8
1	E	781	CYS	7.3
1	C	639	ALA	7.1

### 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.