



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 08:08 AM EDT

PDB ID : 3HP0
Title : Crystal structure of a Putative polyketide biosynthesis enoyl-CoA hydratase (pksH) from *Bacillus subtilis*
Authors : Satyanarayana, L.; Eswaramoorthy, S.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2009-06-03
Resolution : 2.32 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 ⓘ 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
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
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)	:	2.37.1

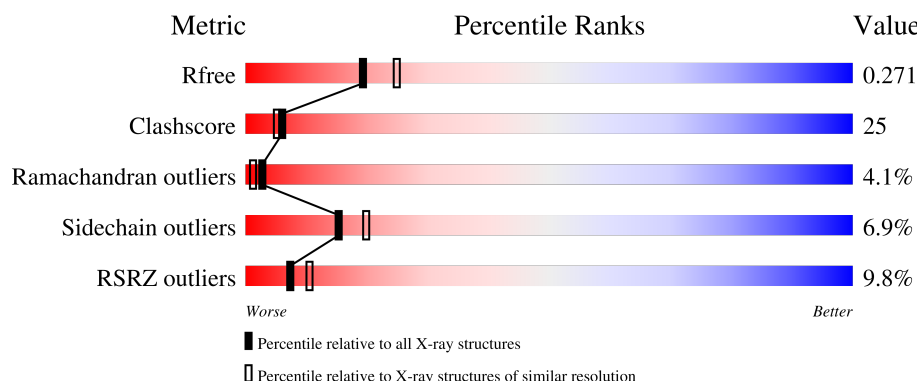
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.32 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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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 $\leq 5\%$. 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	267	<div> <div>14%</div> <div>49%</div> <div>39%</div> <div>5%</div> <div>6%</div> </div>
1	B	267	<div> <div>2%</div> <div>63%</div> <div>27%</div> <div>•</div> <div>6%</div> </div>
1	C	267	<div> <div>•</div> <div>60%</div> <div>27%</div> <div>6%</div> <div>•</div> <div>6%</div> </div>
1	D	267	<div> <div>10%</div> <div>49%</div> <div>38%</div> <div>6%</div> <div>•</div> <div>6%</div> </div>
1	E	267	<div> <div>8%</div> <div>48%</div> <div>41%</div> <div>•</div> <div>•</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	267	<div><div></div><div>18%</div><div>55%</div><div>33%</div><div>6%</div><div>6%</div></div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11858 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 Putative polyketide biosynthesis enoyl-CoA hydratase homolog pksH.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	Se	0	0	0
			1836	1161	321	344	4	6			
1	B	251	Total	C	N	O	S	Se	0	0	0
			1995	1269	345	369	5	7			
1	C	251	Total	C	N	O	S	Se	0	0	0
			1995	1269	345	369	5	7			
1	D	251	Total	C	N	O	S	Se	0	0	0
			1995	1269	345	369	5	7			
1	E	251	Total	C	N	O	S	Se	0	0	0
			1995	1269	345	369	5	7			
1	F	251	Total	C	N	O	S	Se	0	0	0
			1833	1158	322	344	4	5			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	expression tag	UNP P40805
A	2	SER	-	expression tag	UNP P40805
A	260	GLU	-	expression tag	UNP P40805
A	261	GLY	-	expression tag	UNP P40805
A	262	HIS	-	expression tag	UNP P40805
A	263	HIS	-	expression tag	UNP P40805
A	264	HIS	-	expression tag	UNP P40805
A	265	HIS	-	expression tag	UNP P40805
A	266	HIS	-	expression tag	UNP P40805
A	267	HIS	-	expression tag	UNP P40805
B	1	MSE	-	expression tag	UNP P40805
B	2	SER	-	expression tag	UNP P40805
B	260	GLU	-	expression tag	UNP P40805
B	261	GLY	-	expression tag	UNP P40805
B	262	HIS	-	expression tag	UNP P40805
B	263	HIS	-	expression tag	UNP P40805

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	264	HIS	-	expression tag	UNP P40805
B	265	HIS	-	expression tag	UNP P40805
B	266	HIS	-	expression tag	UNP P40805
B	267	HIS	-	expression tag	UNP P40805
C	1	MSE	-	expression tag	UNP P40805
C	2	SER	-	expression tag	UNP P40805
C	260	GLU	-	expression tag	UNP P40805
C	261	GLY	-	expression tag	UNP P40805
C	262	HIS	-	expression tag	UNP P40805
C	263	HIS	-	expression tag	UNP P40805
C	264	HIS	-	expression tag	UNP P40805
C	265	HIS	-	expression tag	UNP P40805
C	266	HIS	-	expression tag	UNP P40805
C	267	HIS	-	expression tag	UNP P40805
D	1	MSE	-	expression tag	UNP P40805
D	2	SER	-	expression tag	UNP P40805
D	260	GLU	-	expression tag	UNP P40805
D	261	GLY	-	expression tag	UNP P40805
D	262	HIS	-	expression tag	UNP P40805
D	263	HIS	-	expression tag	UNP P40805
D	264	HIS	-	expression tag	UNP P40805
D	265	HIS	-	expression tag	UNP P40805
D	266	HIS	-	expression tag	UNP P40805
D	267	HIS	-	expression tag	UNP P40805
E	1	MSE	-	expression tag	UNP P40805
E	2	SER	-	expression tag	UNP P40805
E	260	GLU	-	expression tag	UNP P40805
E	261	GLY	-	expression tag	UNP P40805
E	262	HIS	-	expression tag	UNP P40805
E	263	HIS	-	expression tag	UNP P40805
E	264	HIS	-	expression tag	UNP P40805
E	265	HIS	-	expression tag	UNP P40805
E	266	HIS	-	expression tag	UNP P40805
E	267	HIS	-	expression tag	UNP P40805
F	1	MSE	-	expression tag	UNP P40805
F	2	SER	-	expression tag	UNP P40805
F	260	GLU	-	expression tag	UNP P40805
F	261	GLY	-	expression tag	UNP P40805
F	262	HIS	-	expression tag	UNP P40805
F	263	HIS	-	expression tag	UNP P40805
F	264	HIS	-	expression tag	UNP P40805
F	265	HIS	-	expression tag	UNP P40805

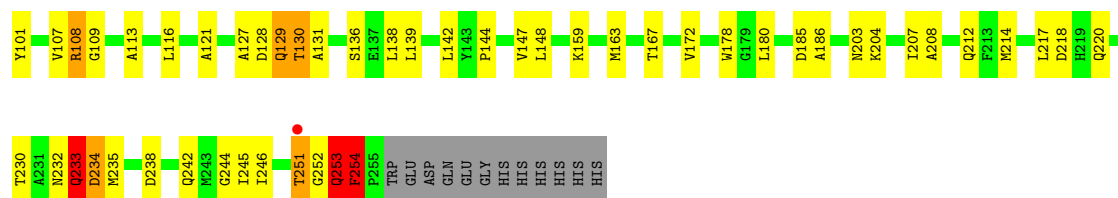
Continued on next page...

Continued from previous page...

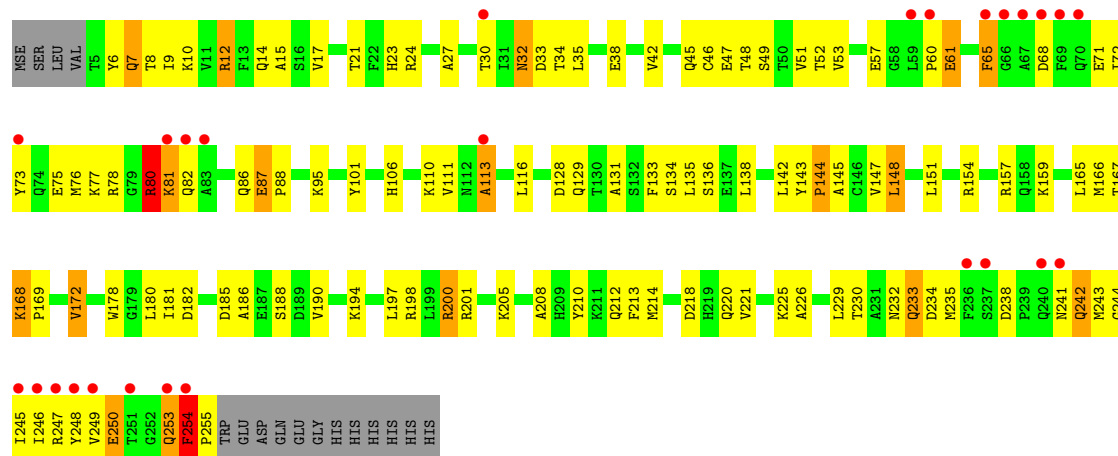
Chain	Residue	Modelled	Actual	Comment	Reference
F	266	HIS	-	expression tag	UNP P40805
F	267	HIS	-	expression tag	UNP P40805

- Molecule 2 is water.

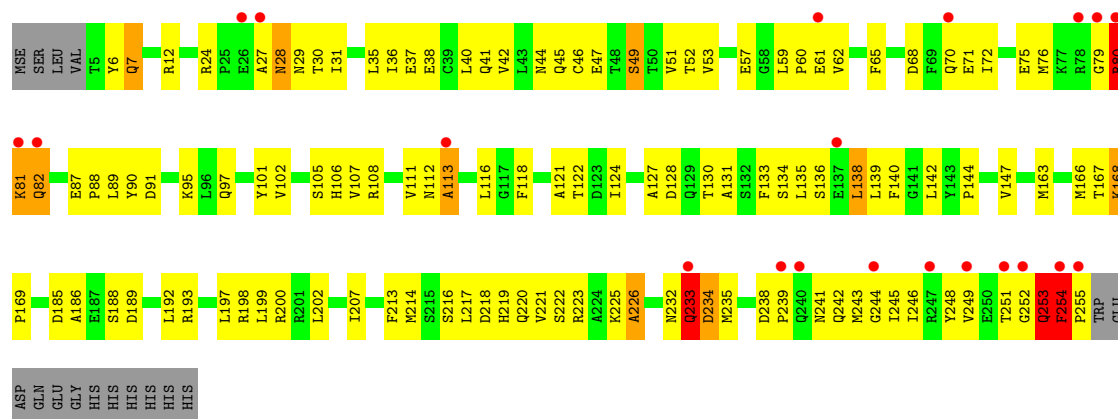
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	33	Total O 33 33	0	0
2	B	44	Total O 44 44	0	0
2	C	49	Total O 49 49	0	0
2	D	23	Total O 23 23	0	0
2	E	31	Total O 31 31	0	0
2	F	29	Total O 29 29	0	0



● Molecule 1: Putative polyketide biosynthesis enoyl-CoA hydratase homolog pksH

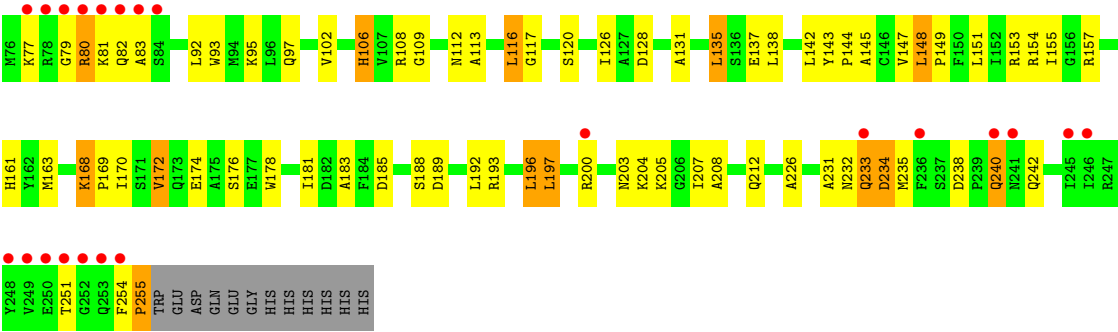


● Molecule 1: Putative polyketide biosynthesis enoyl-CoA hydratase homolog pksH



● Molecule 1: Putative polyketide biosynthesis enoyl-CoA hydratase homolog pksH





4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α , β , γ	80.91Å 80.91Å 218.59Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.09 – 2.32 43.09 – 2.32	Depositor EDS
% Data completeness (in resolution range)	96.4 (43.09-2.32) 96.4 (43.09-2.32)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.90 (at 2.32Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.232 , 0.270 0.232 , 0.271	Depositor DCC
R_{free} test set	5004 reflections (3.75%)	wwPDB-VP
Wilson B-factor (Å ²)	33.5	Xtriage
Anisotropy	0.146	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 49.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.010 for -h,-k,l 0.037 for h,-h-k,-l 0.058 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	11858	wwPDB-VP
Average B, all atoms (Å ²)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.40	0/1859	0.73	4/2514 (0.2%)
1	B	0.42	0/2028	0.65	0/2732
1	C	0.41	0/2028	0.71	3/2732 (0.1%)
1	D	0.38	0/2028	0.62	0/2732
1	E	0.38	0/2028	0.65	1/2732 (0.0%)
1	F	0.37	0/1856	0.69	5/2511 (0.2%)
All	All	0.39	0/11827	0.68	13/15953 (0.1%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	233	GLN	N-CA-C	-6.97	92.19	111.00
1	A	200	ARG	N-CA-C	-6.47	93.52	111.00
1	F	233	GLN	N-CA-C	-5.83	95.25	111.00
1	F	60	PRO	N-CA-CB	5.82	110.28	103.30
1	F	255	PRO	N-CA-CB	5.78	110.23	103.30

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	1836	0	1734	94	0
1	B	1995	0	1997	88	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1995	0	1997	92	0
1	D	1995	0	1997	127	0
1	E	1995	0	1997	113	0
1	F	1833	0	1720	83	0
2	A	33	0	0	2	0
2	B	44	0	0	3	0
2	C	49	0	0	6	0
2	D	23	0	0	8	0
2	E	31	0	0	7	0
2	F	29	0	0	7	0
All	All	11858	0	11442	573	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 25.

The worst 5 of 573 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:D:76:MSE:HE3	1:D:242:GLN:HG3	1.39	1.02
1:B:82:GLN:HE21	1:E:219:HIS:HB3	1.25	1.01
1:F:205:LYS:H	1:F:205:LYS:HD2	1.19	1.01
1:D:254:PHE:HB2	1:D:255:PRO:CD	1.92	1.00
1:D:113:ALA:HB2	1:D:135:LEU:HA	1.39	1.00

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	248/267 (93%)	219 (88%)	16 (6%)	13 (5%)	2 1
1	B	249/267 (93%)	231 (93%)	14 (6%)	4 (2%)	9 8

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	249/267 (93%)	223 (90%)	13 (5%)	13 (5%)	2	1
1	D	249/267 (93%)	215 (86%)	24 (10%)	10 (4%)	3	1
1	E	249/267 (93%)	221 (89%)	18 (7%)	10 (4%)	3	1
1	F	249/267 (93%)	223 (90%)	15 (6%)	11 (4%)	2	1
All	All	1493/1602 (93%)	1332 (89%)	100 (7%)	61 (4%)	3	1

5 of 61 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	234	ASP
1	A	240	GLN
1	B	82	GLN
1	B	167	THR
1	C	27	ALA

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	178/225 (79%)	171 (96%)	7 (4%)	32	45
1	B	218/225 (97%)	204 (94%)	14 (6%)	17	23
1	C	218/225 (97%)	199 (91%)	19 (9%)	10	12
1	D	218/225 (97%)	198 (91%)	20 (9%)	9	10
1	E	218/225 (97%)	206 (94%)	12 (6%)	21	29
1	F	176/225 (78%)	164 (93%)	12 (7%)	16	20
All	All	1226/1350 (91%)	1142 (93%)	84 (7%)	15	20

5 of 84 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	242	GLN
1	E	254	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	254	PHE
1	E	81	LYS
1	F	106	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	41	GLN
1	F	161	HIS
1	E	44	ASN
1	E	129	GLN
1	F	233	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 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	243/267 (91%)	1.01	38 (15%) 2 3	16, 40, 79, 85	0
1	B	244/267 (91%)	0.12	5 (2%) 65 72	14, 30, 58, 73	0
1	C	244/267 (91%)	0.15	4 (1%) 72 78	13, 32, 58, 74	0
1	D	244/267 (91%)	0.63	26 (10%) 6 8	19, 45, 72, 77	0
1	E	244/267 (91%)	0.56	21 (8%) 10 14	17, 46, 68, 77	0
1	F	244/267 (91%)	1.14	49 (20%) 1 1	16, 42, 82, 87	0
All	All	1463/1602 (91%)	0.60	143 (9%) 7 10	13, 39, 74, 87	0

The worst 5 of 143 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	77	LYS	13.6
1	A	58	GLY	11.5
1	F	79	GLY	10.8
1	F	70	GLN	9.6
1	F	73	TYR	8.8

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.