



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:27 PM GMT

PDB ID : 4F1N
Title : Crystal structure of Kluyveromyces polysporus Argonaute with a guide RNA
Authors : Nakanishi, K.; Weinberg, D.E.; Bartel, D.P.; Patel, D.J.
Deposited on : 2012-05-07
Resolution : 3.19 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
<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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

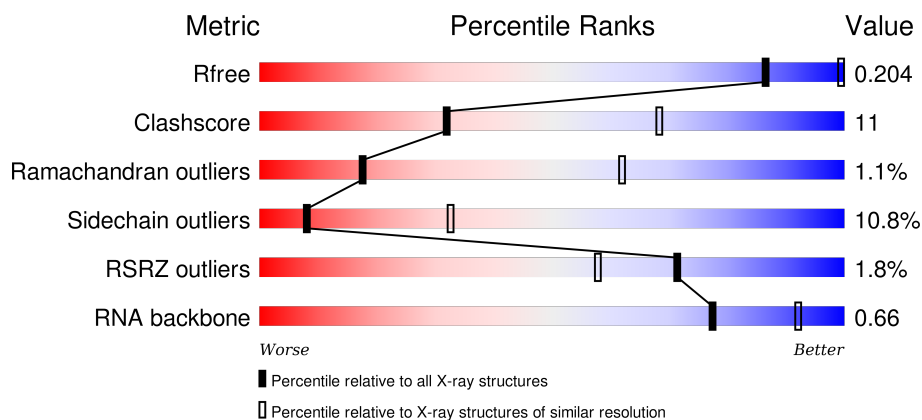
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.19 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}	91344	1115 (3.22-3.14)
Clashscore	102246	1125 (3.20-3.16)
Ramachandran outliers	100387	1105 (3.20-3.16)
Sidechain outliers	100360	1104 (3.20-3.16)
RSRZ outliers	91569	1120 (3.22-3.14)
RNA backbone	2183	1063 (3.66-2.70)

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 ≥ 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	1046	<div> <div>2%</div> <div>59% 23% 14%</div> </div>
1	B	1046	<div> <div>2%</div> <div>56% 22% 18%</div> </div>
2	E	9	<div> <div>33% 44% 22%</div> </div>
2	F	9	<div> <div>22% 44% 33%</div> </div>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 14426 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 KpAGO.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	904	Total	C	N	O	S	0	0	0
			7225	4633	1201	1359	32			
1	B	853	Total	C	N	O	S	0	0	0
			6809	4372	1127	1280	30			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	206	SER	-	EXPRESSION TAG	UNP A7TMA9
B	206	SER	-	EXPRESSION TAG	UNP A7TMA9

- Molecule 2 is a RNA chain called RNA 5'-R(P*UP*AP*AP*AP*AP*AP*AP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	9	Total	C	N	O	P	0	0	0
			179	79	37	54	9			
2	F	9	Total	C	N	O	P	0	0	0
			179	79	37	54	9			

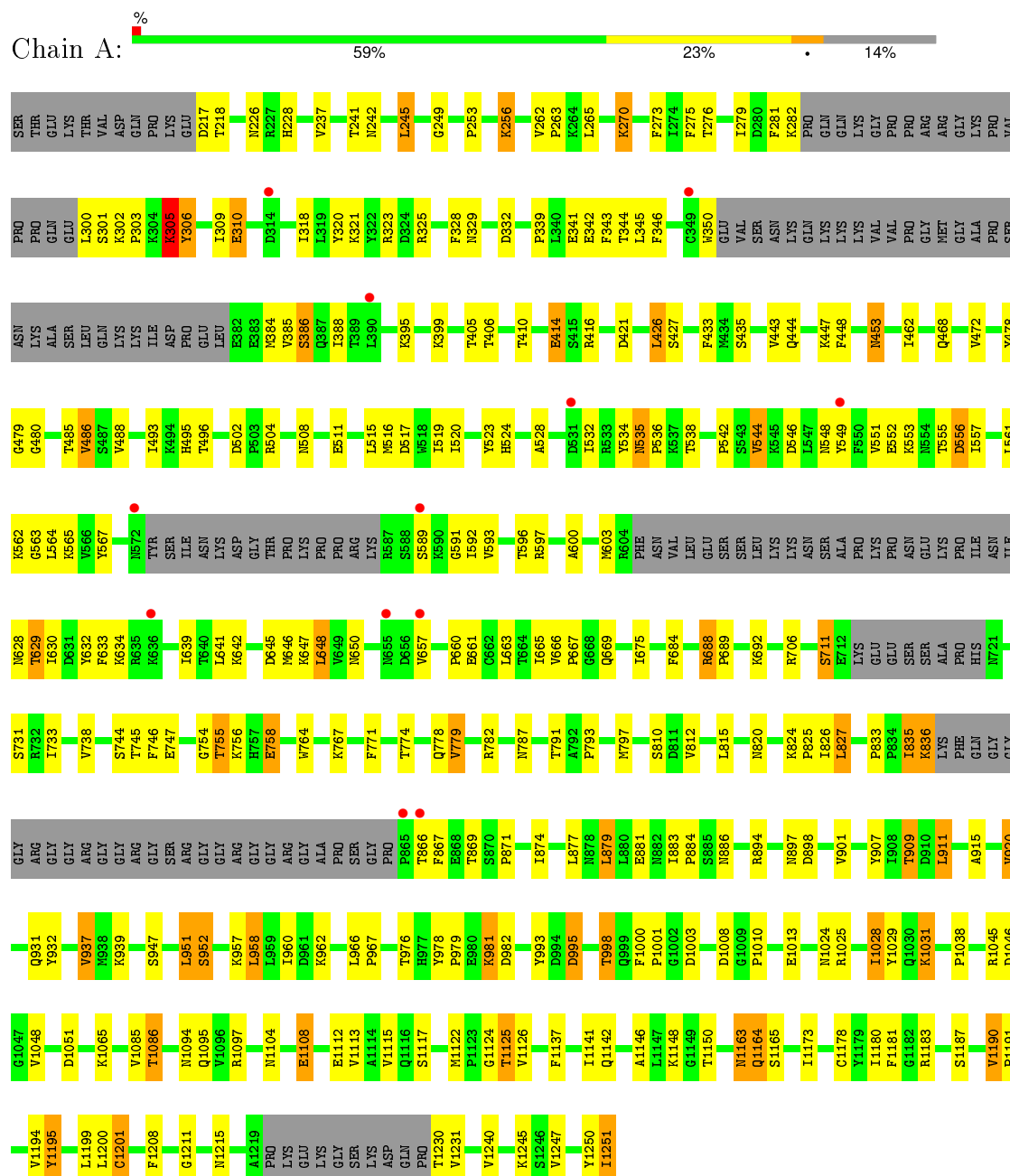
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	20	Total	O	0	0
			20	20		
3	E	1	Total	O	0	0
			1	1		
3	B	12	Total	O	0	0
			12	12		
3	F	1	Total	O	0	0
			1	1		

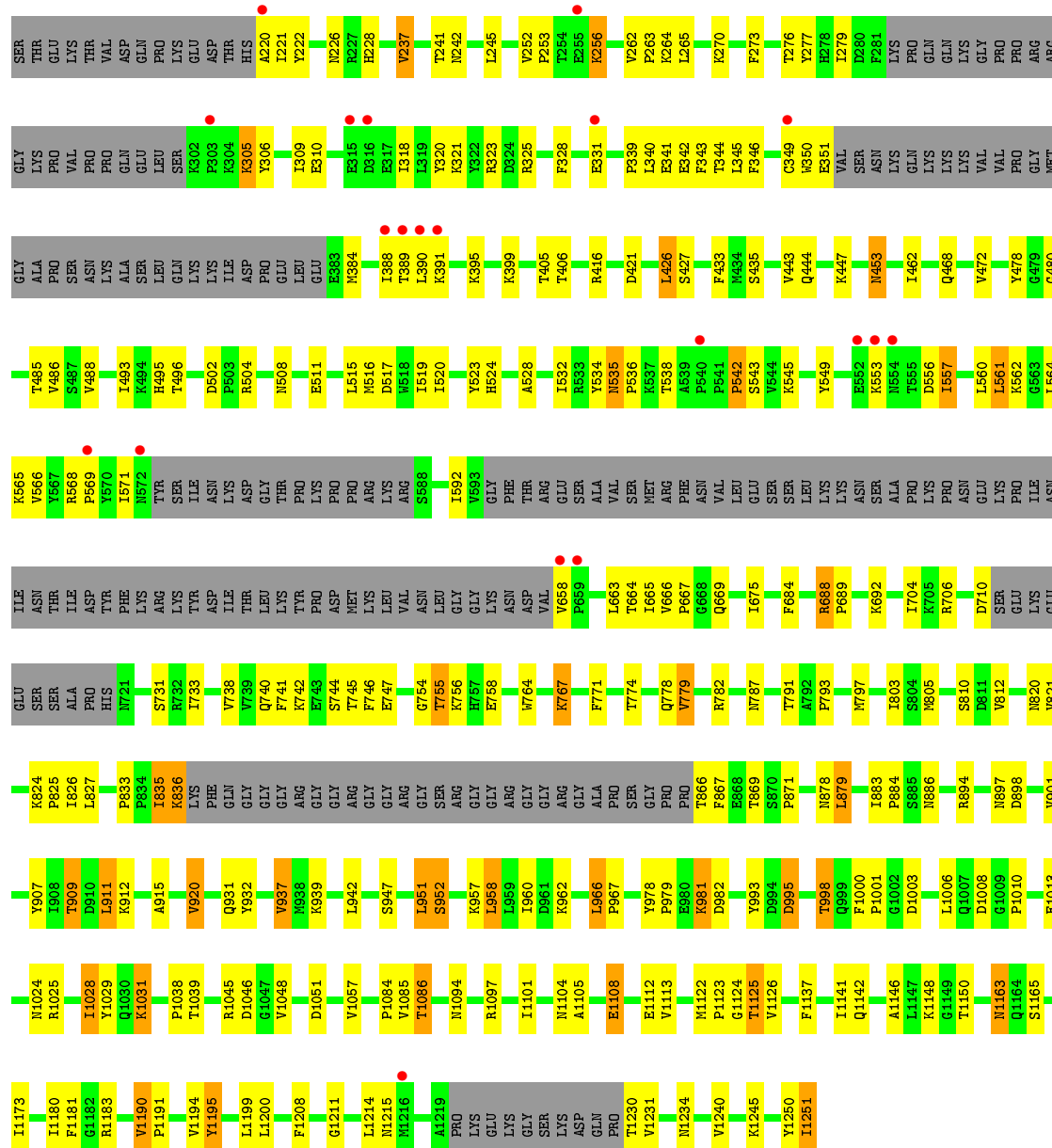
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

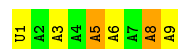
• Molecule 1: KpAGO



Chain B:



Chain E:



Chain F:



U1	A2	A3	A4	A5	A6	A7	A8	A9
----	----	----	----	----	----	----	----	----

4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α , β , γ	171.55Å 171.55Å 83.18Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.12 – 3.19 49.52 – 3.19	Depositor EDS
% Data completeness (in resolution range)	99.5 (38.12-3.19) 95.8 (49.52-3.19)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.26 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.168 , 0.216 0.157 , 0.204	Depositor DCC
R_{free} test set	1910 reflections (4.37%)	DCC
Wilson B-factor (Å ²)	43.7	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 55.5	EDS
Estimated twinning fraction	0.019 for -h,-k,l 0.027 for h,-h-k,-l 0.018 for -k,-h,-l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 45518 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	14426	wwPDB-VP
Average B, all atoms (Å ²)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	0.50	0/7382	0.67	2/9981 (0.0%)
1	B	0.51	0/6958	0.68	2/9413 (0.0%)
2	E	1.01	1/201 (0.5%)	1.16	1/310 (0.3%)
2	F	1.08	1/201 (0.5%)	1.22	4/310 (1.3%)
All	All	0.52	2/14742 (0.0%)	0.70	9/20014 (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	B	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	1	U	OP3-P	-10.60	1.48	1.61
2	E	1	U	OP3-P	-10.32	1.48	1.61

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	399	LYS	CD-CE-NZ	8.05	130.21	111.70
2	F	5	A	N1-C6-N6	6.13	122.28	118.60
1	B	305	LYS	CA-CB-CG	6.01	126.63	113.40
2	E	5	A	N1-C6-N6	5.93	122.16	118.60
1	A	329	ASN	C-N-CA	5.32	133.47	122.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	220	ALA	Peptide
1	B	542	PRO	Peptide

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	7225	0	7219	158	0
1	B	6809	0	6799	144	0
2	E	179	0	87	5	0
2	F	179	0	87	5	0
3	A	20	0	0	0	0
3	B	12	0	0	1	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
All	All	14426	0	14192	303	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 303 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:1104:ASN:HB3	1:B:1112:GLU:HG3	1.60	0.84
1:A:447:LYS:HD3	1:A:488:VAL:HG11	1.61	0.82
1:A:1104:ASN:HB3	1:A:1112:GLU:HG3	1.64	0.80
1:A:1126:VAL:HG22	1:A:1141:ILE:HG12	1.64	0.79
1:B:447:LYS:HD3	1:B:488:VAL:HG11	1.64	0.77

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	888/1046 (85%)	804 (90%)	74 (8%)	10 (1%)	17	62
1	B	837/1046 (80%)	758 (91%)	70 (8%)	9 (1%)	17	62
All	All	1725/2092 (82%)	1562 (91%)	144 (8%)	19 (1%)	17	62

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	544	VAL
1	A	711	SER
1	B	221	ILE
1	B	981	LYS
1	A	480	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	812/929 (87%)	721 (89%)	91 (11%)	7	31
1	B	765/929 (82%)	686 (90%)	79 (10%)	9	35
All	All	1577/1858 (85%)	1407 (89%)	170 (11%)	8	33

5 of 170 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1113	VAL
1	B	245	LEU
1	B	1113	VAL
1	A	1163	ASN
1	A	1231	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	548	ASN
1	A	780	ASN
1	A	820	ASN
1	B	878	ASN
1	B	882	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	E	7/9 (77%)	1 (14%)	1 (14%)
2	F	7/9 (77%)	1 (14%)	1 (14%)
All	All	14/18 (77%)	2 (14%)	2 (14%)

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	E	8	A
2	F	8	A

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	E	8	A
2	F	8	A

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, 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	904/1046 (86%)	-0.32	12 (1%) 79 67	11, 44, 111, 148	0
1	B	853/1046 (81%)	-0.32	20 (2%) 64 49	8, 44, 108, 150	0
2	E	9/9 (100%)	-0.49	0 100 100	30, 35, 67, 175	0
2	F	9/9 (100%)	-0.52	0 100 100	28, 35, 66, 178	0
All	All	1775/2110 (84%)	-0.33	32 (1%) 71 58	8, 44, 110, 178	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	331	GLU	4.1
1	B	553	LYS	3.8
1	B	552	GLU	3.4
1	B	554	ASN	3.1
1	B	569	PRO	3.0

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.