



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:53 PM GMT

PDB ID : 4AST
Title : The apo structure of a bacterial aldo-keto reductase AKR14A1
Authors : Zhu, X.; Ellis, E.M.; Lapthorn, A.
Deposited on : 2012-05-03
Resolution : 2.38 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

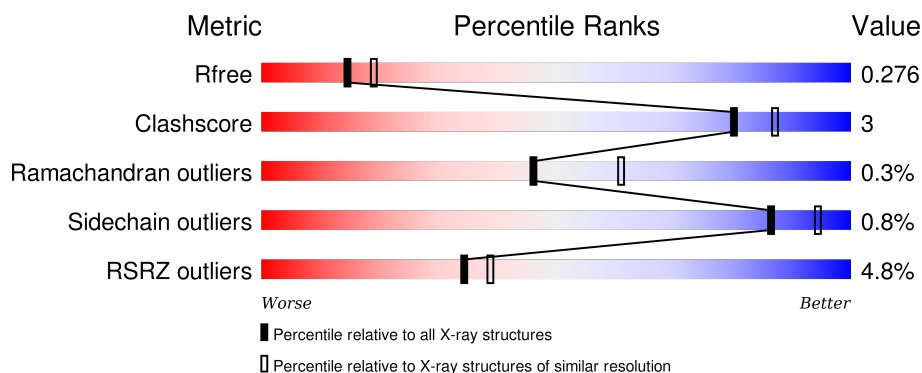
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.38 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	4019 (2.40-2.36)
Clashscore	102246	4595 (2.40-2.36)
Ramachandran outliers	100387	4520 (2.40-2.36)
Sidechain outliers	100360	4522 (2.40-2.36)
RSRZ outliers	91569	4034 (2.40-2.36)

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	346	 83% 5% 11%
1	B	346	 84% 6% 10%
1	C	346	 82% 7% 10%
1	D	346	 81% 8% 11%
1	E	346	 82% 6% 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	346	<p>7% 79% 7% • 12%</p>
1	G	346	<p>10% 80% 7% • 12%</p>
1	H	346	<p>9% 77% 8% • 14%</p>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 20048 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 ALDO-KETO REDUCTASE AKR14A1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	307	Total	C	N	O	S	0	1	0
			2429	1533	428	458	10			
1	B	311	Total	C	N	O	S	0	0	0
			2449	1546	431	462	10			
1	C	310	Total	C	N	O	S	0	1	0
			2453	1548	431	464	10			
1	D	308	Total	C	N	O	S	0	2	0
			2447	1546	431	460	10			
1	E	307	Total	C	N	O	S	0	1	0
			2428	1533	427	458	10			
1	F	303	Total	C	N	O	S	0	1	0
			2403	1518	423	452	10			
1	G	303	Total	C	N	O	S	0	1	0
			2400	1516	422	452	10			
1	H	297	Total	C	N	O	S	0	1	0
			2358	1492	415	441	10			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	113	Total	O	0	0
			113	113		
2	B	114	Total	O	0	0
			114	114		
2	C	93	Total	O	0	0
			93	93		
2	D	96	Total	O	0	0
			96	96		
2	E	86	Total	O	0	0
			86	86		
2	F	63	Total	O	0	0
			63	63		

Continued on next page...

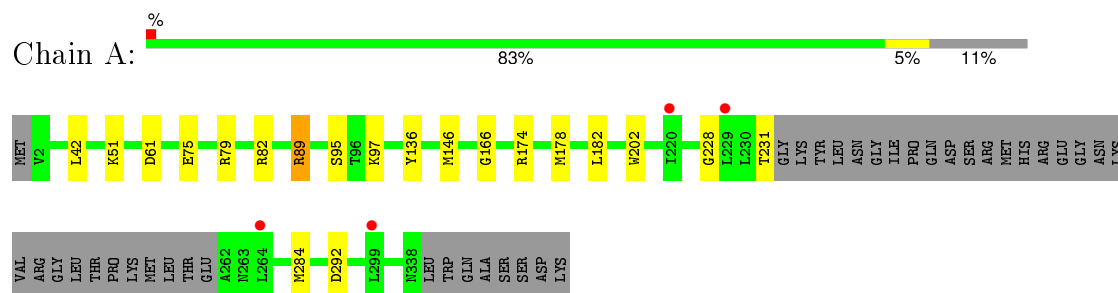
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	54	Total 54	O 54	0	0
2	H	62	Total 62	O 62	0	0

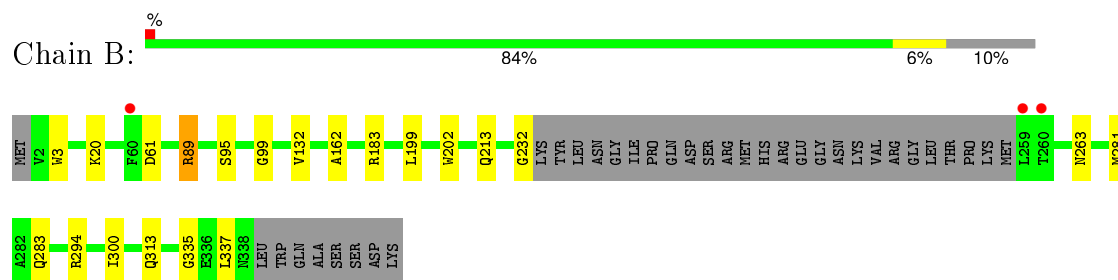
3 Residue-property plots [i](#)

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.

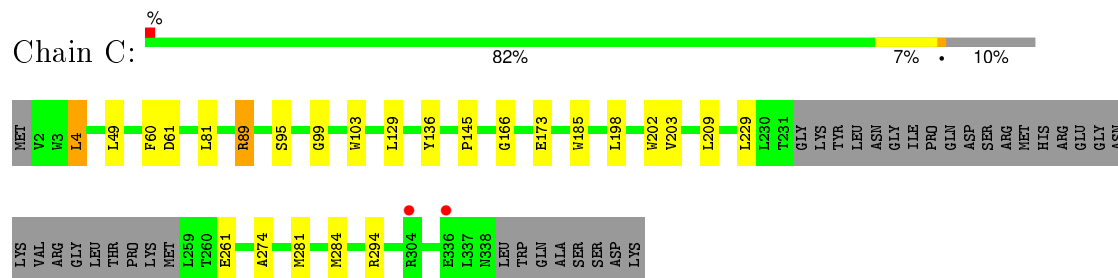
• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



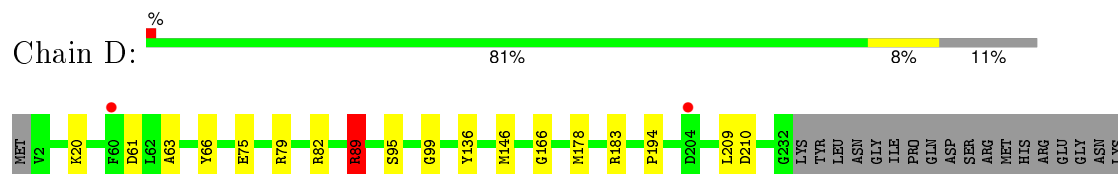
• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



• Molecule 1: ALDO-KETO REDUCTASE AKR14A1

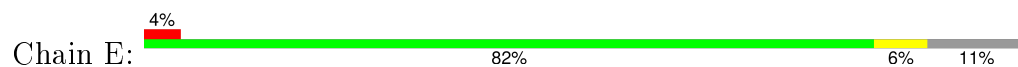


• Molecule 1: ALDO-KETO REDUCTASE AKR14A1

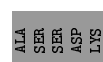
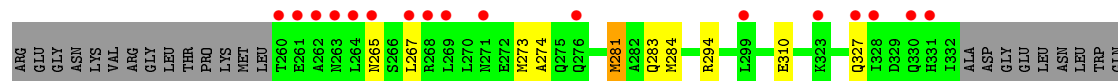
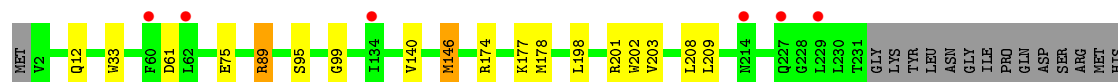
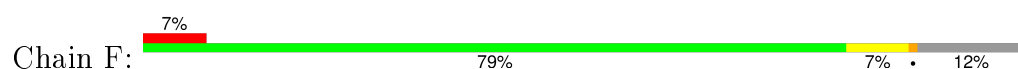




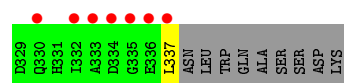
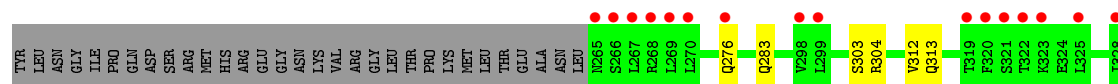
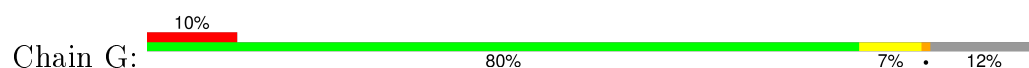
• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



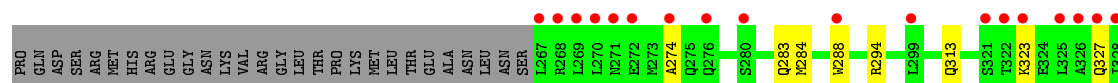
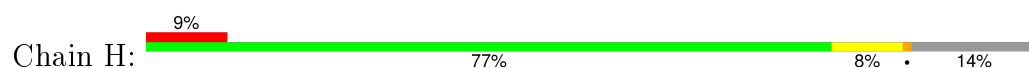
• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



• Molecule 1: ALDO-KETO REDUCTASE AKR14A1



D329	
Q330	
H331	
I332	
A333	
D334	
GLY	
GLU	
LEU	
ASN	
LEU	
TRP	
GLN	
ALA	
SER	
SER	
ASP	
LYS	

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	98.81Å 144.32Å 113.90Å 90.00° 96.51° 90.00°	Depositor
Resolution (Å)	113.16 – 2.38 33.95 – 2.38	Depositor EDS
% Data completeness (in resolution range)	93.8 (113.16-2.38) 93.9 (33.95-2.38)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.76 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.239 , 0.279 0.237 , 0.276	Depositor DCC
R_{free} test set	6339 reflections (5.62%)	DCC
Wilson B-factor (Å ²)	36.8	Xtriage
Anisotropy	0.031	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 35.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 118669 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	20048	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.04% 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	0.92	2/2479 (0.1%)	0.92	5/3357 (0.1%)
1	B	0.94	3/2499 (0.1%)	0.93	6/3384 (0.2%)
1	C	0.89	3/2503 (0.1%)	0.92	6/3390 (0.2%)
1	D	0.88	0/2499	0.89	3/3385 (0.1%)
1	E	0.87	1/2478 (0.0%)	0.92	7/3356 (0.2%)
1	F	0.89	2/2453 (0.1%)	0.97	11/3322 (0.3%)
1	G	0.83	2/2450 (0.1%)	0.86	4/3317 (0.1%)
1	H	0.84	2/2408 (0.1%)	0.91	8/3260 (0.2%)
All	All	0.88	15/19769 (0.1%)	0.91	50/26771 (0.2%)

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	D	0	1

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	202	TRP	CD2-CE2	7.02	1.49	1.41
1	F	33	TRP	CD2-CE2	6.75	1.49	1.41
1	G	33	TRP	CD2-CE2	6.67	1.49	1.41
1	G	202	TRP	CD2-CE2	6.65	1.49	1.41
1	C	202	TRP	CD2-CE2	6.23	1.48	1.41

The worst 5 of 50 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	294	ARG	NE-CZ-NH1	12.64	126.62	120.30
1	E	294	ARG	NE-CZ-NH1	11.78	126.19	120.30
1	F	146	MET	CB-CA-C	8.06	126.53	110.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	294	ARG	NE-CZ-NH1	7.92	124.26	120.30
1	H	146	MET	CB-CA-C	7.87	126.14	110.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	89	ARG	Sidechain

5.2 Too-close contacts ⓘ

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	2429	0	2390	11	0
1	B	2449	0	2414	6	0
1	C	2453	0	2414	13	1
1	D	2447	0	2402	20	1
1	E	2428	0	2391	15	0
1	F	2403	0	2368	30	0
1	G	2400	0	2362	16	1
1	H	2358	0	2324	27	0
2	A	113	0	0	1	0
2	B	114	0	0	0	1
2	C	93	0	0	0	0
2	D	96	0	0	2	0
2	E	86	0	0	0	0
2	F	63	0	0	3	0
2	G	54	0	0	0	0
2	H	62	0	0	1	0
All	All	20048	0	19065	132	2

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 3.

The worst 5 of 132 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:146:MET:HE3	1:D:178:MET:CE	1.74	1.16
1:G:146:MET:CE	1:G:178:MET:HE2	1.76	1.15
1:F:146:MET:CE	1:F:177:LYS:HB3	1.78	1.14
1:E:146:MET:HE3	1:E:178:MET:CE	1.76	1.14
1:F:146:MET:HG3	1:F:178:MET:CE	1.79	1.12

All (2) 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:210:ASP:OD2	2:B:2108:HOH:O[1_556]	2.15	0.05
1:C:173:GLU:OE2	1:G:276:GLN:OE1[2_545]	2.19	0.01

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	304/346 (88%)	298 (98%)	6 (2%)	0	100	100
1	B	307/346 (89%)	301 (98%)	5 (2%)	1 (0%)	46	61
1	C	307/346 (89%)	301 (98%)	5 (2%)	1 (0%)	46	61
1	D	306/346 (88%)	301 (98%)	3 (1%)	2 (1%)	26	36
1	E	304/346 (88%)	297 (98%)	6 (2%)	1 (0%)	46	61
1	F	300/346 (87%)	293 (98%)	6 (2%)	1 (0%)	46	61
1	G	300/346 (87%)	293 (98%)	6 (2%)	1 (0%)	46	61
1	H	294/346 (85%)	290 (99%)	3 (1%)	1 (0%)	46	61
All	All	2422/2768 (88%)	2374 (98%)	40 (2%)	8 (0%)	46	61

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	194	PRO
1	E	99	GLY
1	G	99	GLY
1	B	99	GLY
1	F	99	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	257/290 (89%)	256 (100%)	1 (0%)	93	98
1	B	259/290 (89%)	258 (100%)	1 (0%)	93	98
1	C	260/290 (90%)	256 (98%)	4 (2%)	72	86
1	D	258/290 (89%)	254 (98%)	4 (2%)	70	85
1	E	257/290 (89%)	254 (99%)	3 (1%)	78	90
1	F	255/290 (88%)	254 (100%)	1 (0%)	93	98
1	G	254/290 (88%)	253 (100%)	1 (0%)	93	98
1	H	249/290 (86%)	247 (99%)	2 (1%)	86	94
All	All	2049/2320 (88%)	2032 (99%)	17 (1%)	86	94

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	264	LEU
1	D	318	LEU
1	F	89	ARG
1	D	89	ARG
1	G	89	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	338	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	H	327	GLN
1	D	331	HIS
1	A	338	ASN
1	G	76	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 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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	307/346 (88%)	-0.06	4 (1%) 79 81	16, 25, 48, 71	1 (0%)
1	B	311/346 (89%)	-0.11	3 (0%) 84 86	16, 26, 48, 69	1 (0%)
1	C	310/346 (89%)	-0.03	2 (0%) 90 91	18, 30, 52, 79	1 (0%)
1	D	308/346 (89%)	-0.00	5 (1%) 74 77	18, 29, 56, 78	1 (0%)
1	E	307/346 (88%)	0.18	15 (4%) 33 38	20, 33, 64, 94	1 (0%)
1	F	303/346 (87%)	0.34	23 (7%) 17 19	20, 34, 71, 92	1 (0%)
1	G	303/346 (87%)	0.60	35 (11%) 6 7	24, 39, 80, 99	1 (0%)
1	H	297/346 (85%)	0.51	30 (10%) 9 10	20, 40, 75, 95	1 (0%)
All	All	2446/2768 (88%)	0.18	117 (4%) 34 39	16, 32, 67, 99	8 (0%)

The worst 5 of 117 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	267	LEU	5.4
1	G	337	LEU	5.0
1	G	333	ALA	4.8
1	H	268	ARG	4.5
1	G	335	GLY	4.4

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.