WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 12, 2024 – 02:42 PM EDT

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
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:

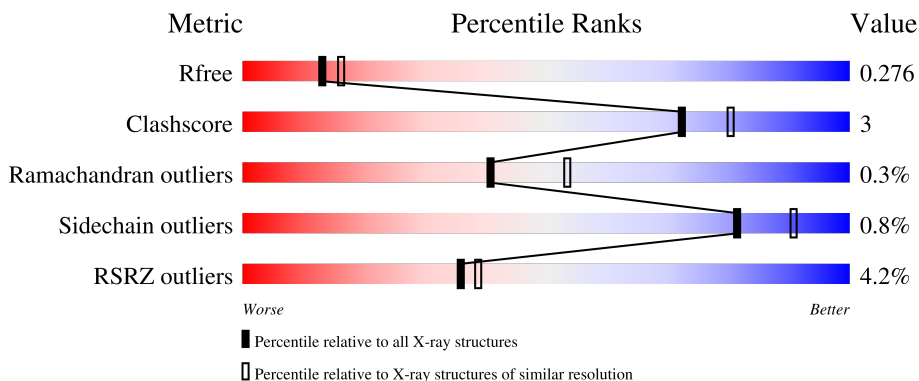|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.38 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.
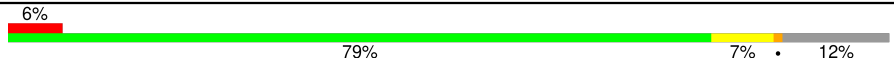
Metric	Percentile Ranks	Value
Rfree		0.276
Clashscore		3
Ramachandran outliers		0.3%
Sidechain outliers		0.8%
RSRZ outliers		4.2%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (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 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 <=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	3% 82% 6% 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	346	6% 79% 7% · 12%
1	G	346	9% 80% 7% · 12%
1	H	346	8% 77% 8% · 14%

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

- Molecule 2 is water.

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

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

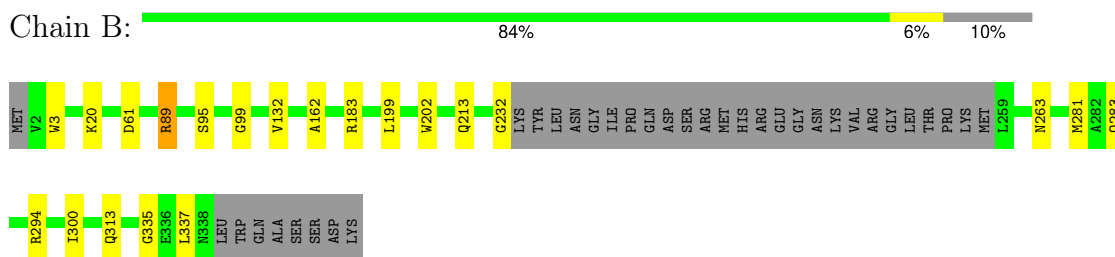| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	54	Total 54	O 54	0	0
2	H	62	Total 62	O 62	0	0

3 Residue-property plots ⓘ

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.
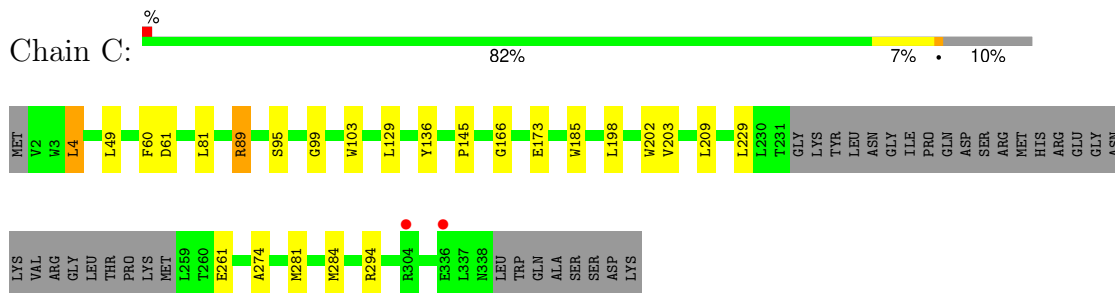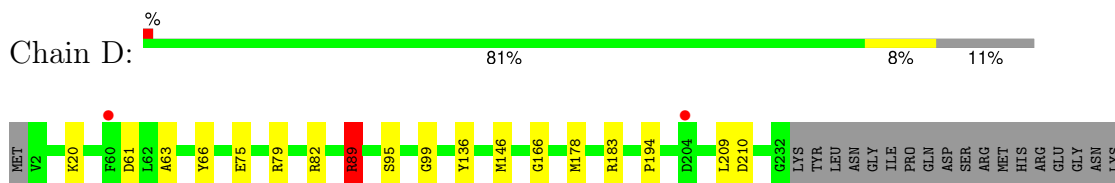
- Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain A:

- Molecule 1: ALDO-KETO REDUCTASE AKR14A1
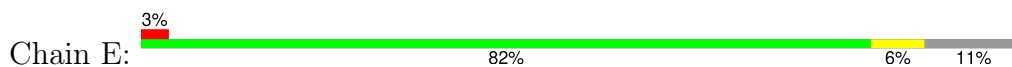
Chain B:

- Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain C:

- Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain D:

VAL ARG GLY LEU THR PRO LYS MET LEU THR GLU A262 N263 L264 L267 R268 Q279 Q283 Q313 R294 Q313 N317 L318 L325 G335 N338 LEU TRP GLN ALA SER SER ASP LYS

● Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain E:

3%

82%　　　　　6%　　11%

MET V2 D61 L62 A63 Y66 L81 R89 S95 G99 L129 Y136 M146 G166 K177 M178 W202 V203 D204 L208 L209 I220 T231 GLY LYS TYR LEU ASN GLY ILE PRO GLN ASP SER ARG MET HIS ARG GLU GLY ASN LYS VAL ARG GLY

LEU THR PRO LYS MET L259 T260 E261 L264 N265 S266 L267 R268 L269 Q283 R294 L299 Q307 Q313 Q327 A333 D334 G335 GLU LEU ASN LEU TRP GLN ALA SER SER ASP LYS

● Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain F:

6%

79%　　　　　7%　·　12%

MET V2 Q12 W33 F60 D61 L62 E75 R89 S95 G99 I134 V140 M146 R174 K177 M178 L198 R201 W202 V203 L208 L209 N214 Q227 G228 L229 L230 T231 GLY LYS TYR LEU ASN GLY ILE PRO GLN ASP SER ARG MET HIS

ARG GLU GLY ASN LYS VAL ARG GLY LEU THR PRO LYS MET LEU T260 E261 A262 N263 L264 N265 S266 L267 R268 L269 M273 A274 Q275 Q276 M281 A282 Q283 M284 R294 L299 E310 Q327 I328 D329 Q330 H331 I332 ALA ASP GLY GLU LEU ASN LEU TRP GLN ALA SER SER ASP
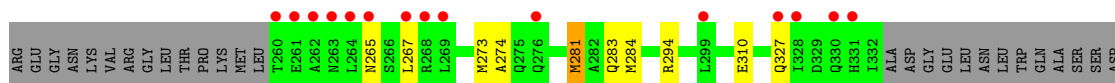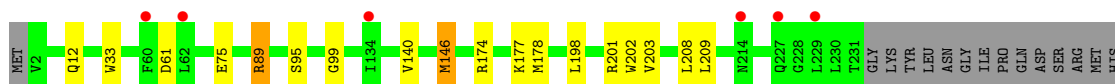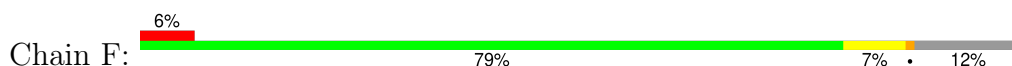
LYS

● Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain G:

9%

80%　　　　　7%　·　12%

MET V2 L28 W33 F60 D61 R82 R89 L92 S95 T96 K97 A98 G99 I134 F135 Y136 M146 G166 E173 R174 K177 M178 R183 L189 L190 I191 L199 W202 V203 D204 K205 Q213 I220 L229 L230 T231 GLY LYS TYR LEU

ASN GLY ILE PRO GLN ASP SER ARG MET HIS ARG GLU GLY ASN LYS VAL ARG GLY LEU THR PRO LYS MET LEU THR GLU ALA ASN LEU N265 S266 L267 R268 L269 L270 Q276 Q283 V298 L299 S303 R304 V312 Q313 T319 F320 S321 T322 K323 E324 L325 Q330 H331 I332

A333 D334 G335 E336 L337 ASN LEU TRP GLN ALA SER SER ASP LYS

● Molecule 1: ALDO-KETO REDUCTASE AKR14A1

Chain H:
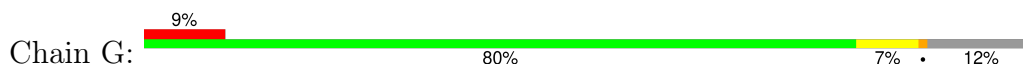
8%

77%　　　　　8%　·　14%

MET V2 W3 L28 S29 K51 D61 R79 R89 S95 G99 Y136 P145 M146 L163 G166 Y170 R174 M178 Y196 N197 L198 L199 W202 V203 D204 L209 Q213 I220 L229 L230 THR GLY LYS TYR LEU ASN GLY ILE PRO

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
$< I/\sigma(I) >$ [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	5959 reflections (5.02%)	wwPDB-VP
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
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	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.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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(^{o})	Ideal(^{o})
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:E:146:MET:HE3	1:E:178:MET:CE	1.76	1.14
1:F:146:MET:CE	1:F:177:LYS:HB3	1.78	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%)	41	53
1	C	307/346 (89%)	301 (98%)	5 (2%)	1 (0%)	41	53
1	D	306/346 (88%)	301 (98%)	3 (1%)	2 (1%)	22	30
1	E	304/346 (88%)	297 (98%)	6 (2%)	1 (0%)	41	53
1	F	300/346 (87%)	293 (98%)	6 (2%)	1 (0%)	41	53
1	G	300/346 (87%)	293 (98%)	6 (2%)	1 (0%)	41	53
1	H	294/346 (85%)	290 (99%)	3 (1%)	1 (0%)	41	53
All	All	2422/2768 (88%)	2374 (98%)	40 (2%)	8 (0%)	41	53

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%)	91	96
1	B	259/290 (89%)	258 (100%)	1 (0%)	91	96
1	C	260/290 (90%)	256 (98%)	4 (2%)	65	79
1	D	258/290 (89%)	254 (98%)	4 (2%)	62	78
1	E	257/290 (89%)	254 (99%)	3 (1%)	71	84
1	F	255/290 (88%)	254 (100%)	1 (0%)	91	96
1	G	254/290 (88%)	253 (100%)	1 (0%)	91	96
1	H	249/290 (86%)	247 (99%)	2 (1%)	81	91
All	All	2049/2320 (88%)	2032 (99%)	17 (1%)	81	91

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

Mol	Chain	Res	Type
1	G	89	ARG
1	H	229	LEU
1	D	264	LEU
1	D	318	LEU
1	D	325	LEU

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

Mol	Chain	Res	Type
1	D	331	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	76	ASN
1	H	327	GLN
1	A	338	ASN
1	A	193	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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(Å²)	Q<0.9
1	A	307/346 (88%)	-0.11	3 (0%)	82	83	16, 25, 48, 71	1 (0%)
1	B	311/346 (89%)	-0.16	0	100	100	16, 26, 48, 69	1 (0%)
1	C	310/346 (89%)	-0.08	2 (0%)	89	89	18, 30, 52, 79	1 (0%)
1	D	308/346 (89%)	-0.05	5 (1%)	72	73	18, 29, 56, 78	1 (0%)
1	E	307/346 (88%)	0.12	12 (3%)	39	42	20, 33, 64, 94	1 (0%)
1	F	303/346 (87%)	0.28	21 (6%)	16	18	20, 34, 71, 92	1 (0%)
1	G	303/346 (87%)	0.53	32 (10%)	6	7	24, 39, 80, 99	1 (0%)
1	H	297/346 (85%)	0.45	28 (9%)	8	9	20, 40, 75, 95	1 (0%)
All	All	2446/2768 (88%)	0.12	103 (4%)	36	39	16, 32, 67, 99	8 (0%)

The worst 5 of 103 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	267	LEU	5.2
1	G	337	LEU	4.9
1	G	333	ALA	4.6
1	H	268	ARG	4.4
1	G	335	GLY	4.3

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 monosaccharides 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.