



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 2, 2017 – 02:52 AM EDT

PDB ID : 3TPC
Title : Crystal structure of a hypothetical protein SMa1452 from Sinorhizobium meliloti 1021
Authors : Agarwal, R.; Chamala, S.; Evans, B.; Foti, R.; Gizzi, A.; Hillerich, B.; Kar, A.; LaFleur, J.; Seidel, R.; Villigas, G.; Zencheck, W.; Almo, S.C.; Swaminathan, S.; New York Structural Genomics Research Consortium (NYSGRG)
Deposited on : unknown
Resolution : 2.34 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

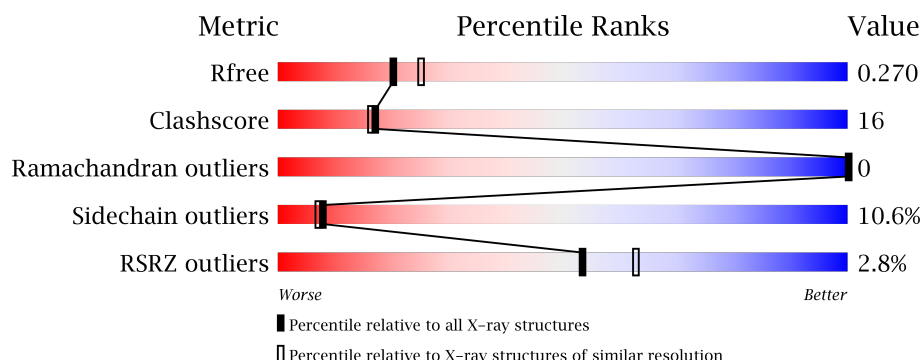
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.34 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}	100719	1570 (2.36-2.32)
Clashscore	112137	1673 (2.36-2.32)
Ramachandran outliers	110173	1654 (2.36-2.32)
Sidechain outliers	110143	1655 (2.36-2.32)
RSRZ outliers	101464	1576 (2.36-2.32)

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	257	
1	B	257	
1	C	257	
1	D	257	
1	E	257	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	257	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>2%64%19%••13%</div></div>
1	G	257	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>5%66%17%•14%</div></div>
1	H	257	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>3%69%15%•14%</div></div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13015 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 Short chain alcohol dehydrogenase-related dehydrogenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	225	Total	C	N	O	S	Se	0	0	0
			1627	1022	295	303	2	5			
1	B	221	Total	C	N	O	S	Se	0	0	0
			1573	989	280	297	2	5			
1	C	224	Total	C	N	O	S	Se	0	0	0
			1604	1007	288	302	2	5			
1	D	226	Total	C	N	O	S	Se	0	0	0
			1597	1007	282	301	2	5			
1	E	225	Total	C	N	O	S	Se	0	0	0
			1624	1021	295	301	2	5			
1	F	223	Total	C	N	O	S	Se	0	0	0
			1593	1000	289	297	2	5			
1	G	222	Total	C	N	O	S	Se	0	0	0
			1542	968	275	292	2	5			
1	H	222	Total	C	N	O	S	Se	0	0	0
			1592	997	290	298	2	5			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
A	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
B	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
B	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
C	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
C	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
D	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
D	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
E	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
E	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
F	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
F	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
G	1	MSE	-	EXPRESSION TAG	UNP Q92YS1

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	2	VAL	-	EXPRESSION TAG	UNP Q92YS1
H	1	MSE	-	EXPRESSION TAG	UNP Q92YS1
H	2	VAL	-	EXPRESSION TAG	UNP Q92YS1

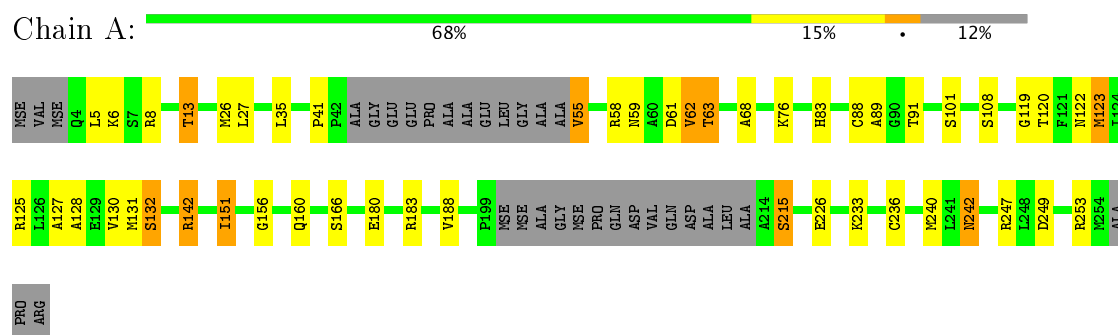
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	32	Total 32	O 32	0	0
2	B	36	Total 36	O 36	0	0
2	C	42	Total 42	O 42	0	0
2	D	39	Total 39	O 39	0	0
2	E	41	Total 41	O 41	0	0
2	F	30	Total 30	O 30	0	0
2	G	24	Total 24	O 24	0	0
2	H	19	Total 19	O 19	0	0

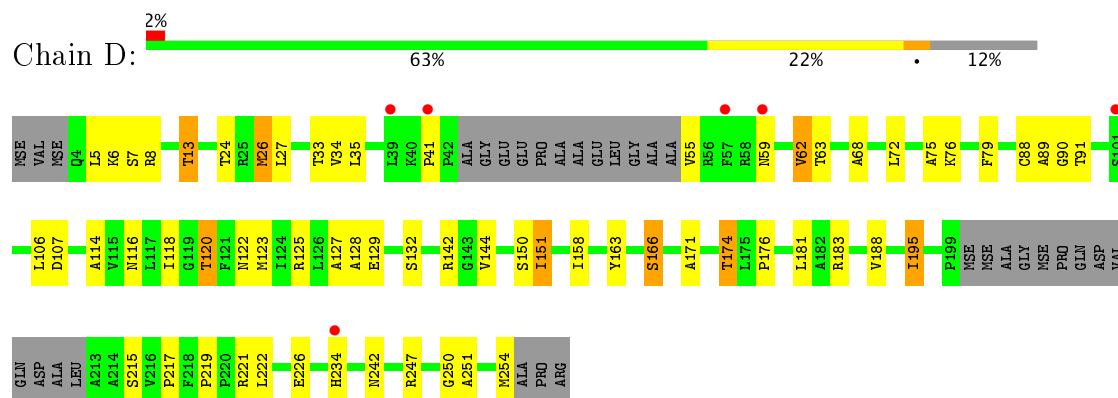
3 Residue-property plots [i](#)

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

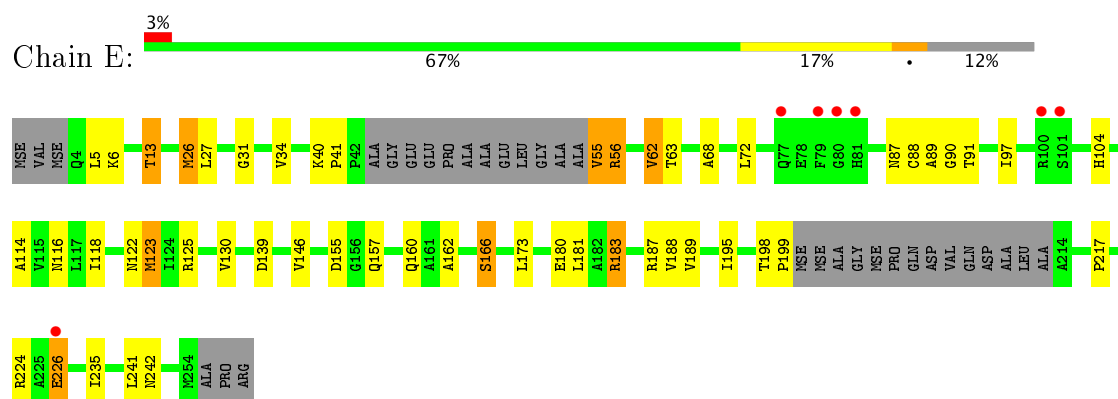
- Molecule 1: Short chain alcohol dehydrogenase-related dehydrogenase



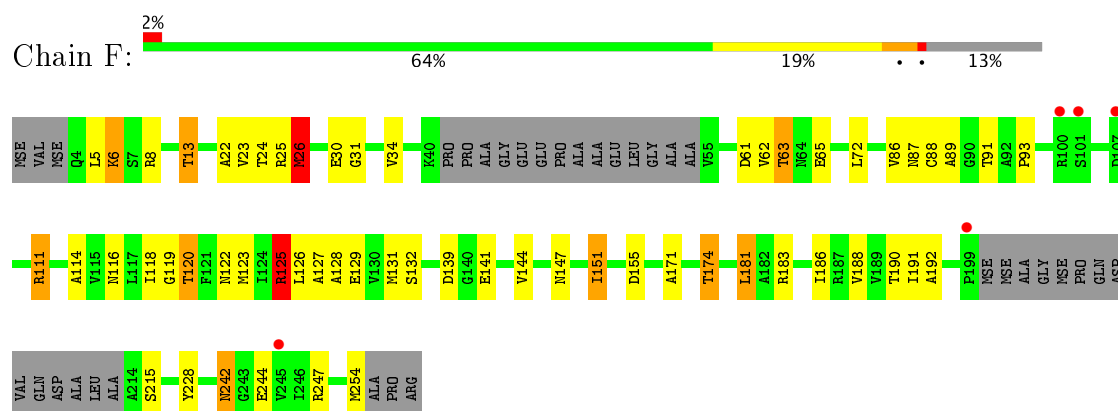
- Molecule 1: Short chain alcohol dehydrogenase-related dehydrogenase



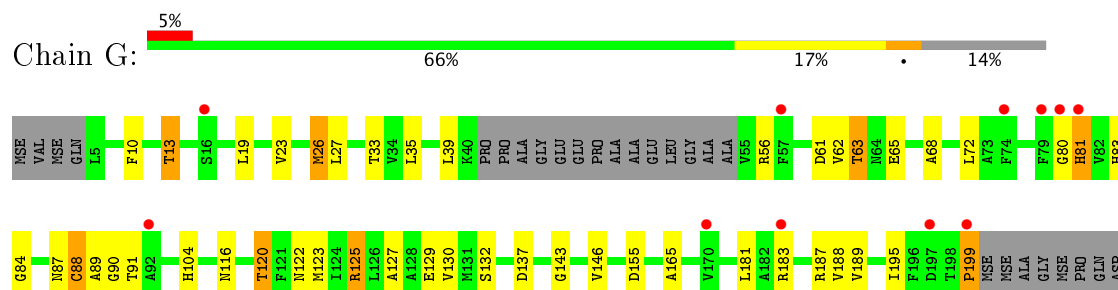
- Molecule 1: Short chain alcohol dehydrogenase-related dehydrogenase



- Molecule 1: Short chain alcohol dehydrogenase-related dehydrogenase



- Molecule 1: Short chain alcohol dehydrogenase-related dehydrogenase





4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	69.15Å 69.38Å 121.85Å 98.36° 92.72° 113.01°	Depositor
Resolution (Å)	46.20 – 2.34 46.20 – 2.34	Depositor EDS
% Data completeness (in resolution range)	97.4 (46.20-2.34) 91.4 (46.20-2.34)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.212 , 0.270 0.212 , 0.270	Depositor DCC
R_{free} test set	4193 reflections (5.27%)	DCC
Wilson B-factor (Å ²)	35.5	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 41.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	13015	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.20% 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

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	1.02	2/1646 (0.1%)	0.94	2/2223 (0.1%)
1	B	0.95	1/1589 (0.1%)	0.95	2/2149 (0.1%)
1	C	1.04	1/1622 (0.1%)	0.94	2/2193 (0.1%)
1	D	0.93	0/1616	0.92	2/2189 (0.1%)
1	E	0.95	0/1643	0.89	0/2219
1	F	0.94	1/1610 (0.1%)	0.92	3/2174 (0.1%)
1	G	0.86	0/1558	0.88	1/2112 (0.0%)
1	H	0.92	0/1609	0.87	0/2173
All	All	0.95	5/12893 (0.0%)	0.91	12/17432 (0.1%)

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	1
1	C	0	1
All	All	0	2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	107	ASP	CB-CG	5.79	1.64	1.51
1	A	226	GLU	CG-CD	5.72	1.60	1.51
1	F	244	GLU	CG-CD	5.67	1.60	1.51
1	B	9	VAL	N-CA	5.20	1.56	1.46
1	A	130	VAL	CB-CG1	-5.02	1.42	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	155	ASP	CB-CG-OD1	5.93	123.64	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	80	GLY	N-CA-C	5.87	127.77	113.10
1	C	107	ASP	CB-CG-OD2	5.70	123.43	118.30
1	F	125	ARG	NE-CZ-NH2	-5.57	117.52	120.30
1	A	132	SER	N-CA-CB	-5.28	102.58	110.50

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	211	ALA	Peptide
1	C	198	THR	Peptide

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	1627	0	1636	43	0
1	B	1573	0	1566	82	0
1	C	1604	0	1598	41	0
1	D	1597	0	1578	65	0
1	E	1624	0	1634	45	0
1	F	1593	0	1586	47	0
1	G	1542	0	1498	47	0
1	H	1592	0	1581	47	0
2	A	32	0	0	4	0
2	B	36	0	0	3	0
2	C	42	0	0	2	0
2	D	39	0	0	1	0
2	E	41	0	0	2	0
2	F	30	0	0	2	0
2	G	24	0	0	0	0
2	H	19	0	0	0	0
All	All	13015	0	12677	397	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 16.

The worst 5 of 397 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:234:HIS:ND1	1:D:234:HIS:HB3	1.37	1.37
1:B:234:HIS:CE1	1:D:234:HIS:HB3	1.74	1.20
1:G:68:ALA:CB	1:G:123:MSE:HE1	1.74	1.18
1:A:123:MSE:CE	1:A:123:MSE:HA	1.74	1.17
1:G:68:ALA:HB1	1:G:123:MSE:HE1	1.28	1.14

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	219/257 (85%)	213 (97%)	6 (3%)	0	100	100
1	B	215/257 (84%)	206 (96%)	9 (4%)	0	100	100
1	C	218/257 (85%)	208 (95%)	10 (5%)	0	100	100
1	D	220/257 (86%)	211 (96%)	9 (4%)	0	100	100
1	E	219/257 (85%)	206 (94%)	13 (6%)	0	100	100
1	F	217/257 (84%)	204 (94%)	13 (6%)	0	100	100
1	G	216/257 (84%)	202 (94%)	14 (6%)	0	100	100
1	H	216/257 (84%)	204 (94%)	12 (6%)	0	100	100
All	All	1740/2056 (85%)	1654 (95%)	86 (5%)	0	100	100

There are no Ramachandran outliers to report.

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	160/173 (92%)	144 (90%)	16 (10%)	9	8
1	B	151/173 (87%)	135 (89%)	16 (11%)	8	7
1	C	156/173 (90%)	139 (89%)	17 (11%)	7	7
1	D	153/173 (88%)	136 (89%)	17 (11%)	7	6
1	E	159/173 (92%)	143 (90%)	16 (10%)	9	8
1	F	153/173 (88%)	137 (90%)	16 (10%)	8	7
1	G	143/173 (83%)	127 (89%)	16 (11%)	7	6
1	H	153/173 (88%)	137 (90%)	16 (10%)	8	7
All	All	1228/1384 (89%)	1098 (89%)	130 (11%)	8	7

5 of 130 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	150	SER
1	E	91	THR
1	H	91	THR
1	D	166	SER
1	E	13	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	242	ASN
1	F	87	ASN
1	H	147	ASN
1	E	87	ASN
1	E	147	ASN

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 [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	220/257 (85%)	-0.22	0 100 100	17, 29, 45, 58	0
1	B	216/257 (84%)	0.13	8 (3%) 42 53	18, 32, 50, 58	0
1	C	219/257 (85%)	0.09	4 (1%) 69 77	17, 30, 49, 60	0
1	D	221/257 (85%)	0.03	6 (2%) 55 64	19, 32, 52, 61	0
1	E	220/257 (85%)	0.12	7 (3%) 48 59	22, 35, 59, 69	0
1	F	218/257 (84%)	0.01	5 (2%) 61 70	21, 33, 48, 64	0
1	G	217/257 (84%)	0.28	12 (5%) 26 36	21, 38, 55, 61	0
1	H	217/257 (84%)	0.21	7 (3%) 48 59	23, 39, 63, 68	0
All	All	1748/2056 (85%)	0.08	49 (2%) 53 63	17, 33, 55, 69	0

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	80	GLY	4.8
1	C	100	ARG	3.9
1	B	79	PHE	3.9
1	G	199	PRO	3.8
1	F	100	ARG	3.7

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.