



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 10, 2023 – 06:21 AM EDT

PDB ID : 7SOG
Title : Thermostable actophorin
Authors : Lieberman, R.L.; Quirk, S.
Deposited on : 2021-10-30
Resolution : 1.74 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) : 1.13
EDS : 2.35.1
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) : 2.35.1

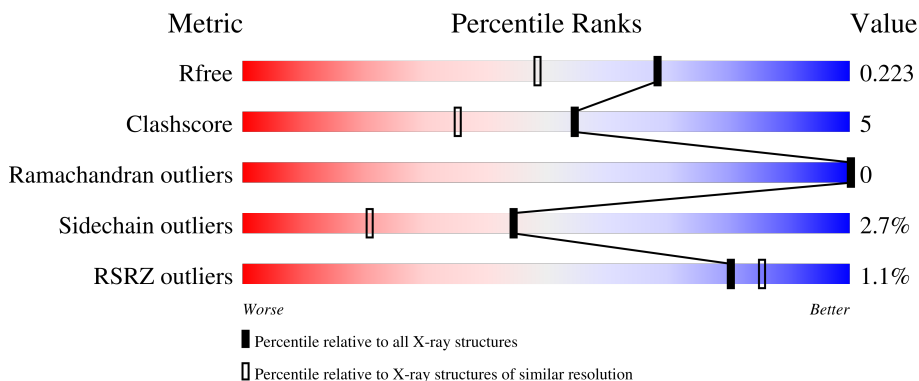
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

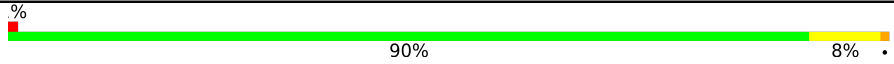
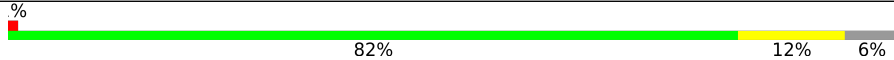
The reported resolution of this entry is 1.74 Å.

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	3764 (1.76-1.72)
Clashscore	141614	3923 (1.76-1.72)
Ramachandran outliers	138981	3878 (1.76-1.72)
Sidechain outliers	138945	3878 (1.76-1.72)
RSRZ outliers	127900	3705 (1.76-1.72)

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	135	
1	B	135	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4458 atoms, of which 2108 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 Actophorin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	135	2234	709	1109	185	226	5	0	12	0
1	B	127	2035	657	999	169	205	5	0	4	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P37167
A	2	PRO	SER	engineered mutation	UNP P37167
A	21	LYS	HIS	engineered mutation	UNP P37167
A	22	LYS	GLN	engineered mutation	UNP P37167
A	27	ILE	THR	engineered mutation	UNP P37167
A	30	LEU	MET	engineered mutation	UNP P37167
A	32	ASP	ALA	engineered mutation	UNP P37167
A	33	ASP	SER	engineered mutation	UNP P37167
A	41	LYS	HIS	engineered mutation	UNP P37167
A	53	LEU	LYS	engineered mutation	UNP P37167
A	54	ALA	SER	engineered mutation	UNP P37167
A	59	ASN	ARG	engineered mutation	UNP P37167
A	71	GLU	GLN	engineered mutation	UNP P37167
A	81	VAL	THR	engineered mutation	UNP P37167
A	99	ALA	THR	engineered mutation	UNP P37167
A	101	SER	THR	engineered mutation	UNP P37167
A	104	ALA	SER	engineered mutation	UNP P37167
A	110	ASP	VAL	engineered mutation	UNP P37167
A	122	ASP	ALA	engineered mutation	UNP P37167
A	130	LYS	SER	engineered mutation	UNP P37167
B	1	GLY	-	expression tag	UNP P37167
B	2	PRO	SER	engineered mutation	UNP P37167
B	21	LYS	HIS	engineered mutation	UNP P37167
B	22	LYS	GLN	engineered mutation	UNP P37167
B	27	ILE	THR	engineered mutation	UNP P37167

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	LEU	MET	engineered mutation	UNP P37167
B	32	ASP	ALA	engineered mutation	UNP P37167
B	33	ASP	SER	engineered mutation	UNP P37167
B	41	LYS	HIS	engineered mutation	UNP P37167
B	53	LEU	LYS	engineered mutation	UNP P37167
B	54	ALA	SER	engineered mutation	UNP P37167
B	59	ASN	ARG	engineered mutation	UNP P37167
B	71	GLU	GLN	engineered mutation	UNP P37167
B	81	VAL	THR	engineered mutation	UNP P37167
B	99	ALA	THR	engineered mutation	UNP P37167
B	101	SER	THR	engineered mutation	UNP P37167
B	104	ALA	SER	engineered mutation	UNP P37167
B	110	ASP	VAL	engineered mutation	UNP P37167
B	122	ASP	ALA	engineered mutation	UNP P37167
B	130	LYS	SER	engineered mutation	UNP P37167

- Molecule 2 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Ni 2 2	0	0
2	B	2	Total Ni 2 2	0	0

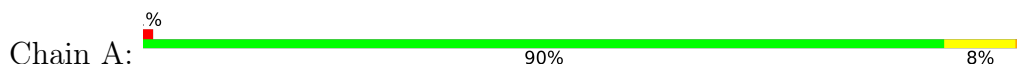
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	108	Total O 108 108	0	0
3	B	77	Total O 77 77	0	0

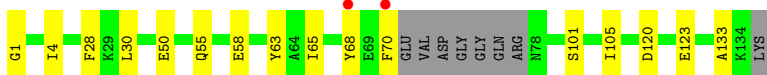
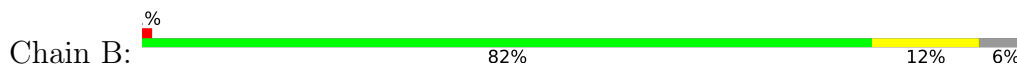
3 Residue-property plots [i](#)

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: Actophorin



- Molecule 1: Actophorin



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	53.72Å 41.29Å 55.50Å 90.00° 95.84° 90.00°	Depositor
Resolution (Å)	36.58 – 1.74 55.21 – 1.74	Depositor EDS
% Data completeness (in resolution range)	94.9 (36.58-1.74) 94.4 (55.21-1.74)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.51 (at 1.74Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.191 , 0.225 0.189 , 0.223	Depositor DCC
R_{free} test set	2012 reflections (8.27%)	wwPDB-VP
Wilson B-factor (Å ²)	22.8	Xtrriage
Anisotropy	0.316	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 46.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.026 for l,-k,h	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4458	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 29.83 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.4497e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NI

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	0.41	0/1178	0.55	0/1585
1	B	0.42	0/1053	0.53	0/1417
All	All	0.42	0/2231	0.54	0/3002

There are no bond length outliers.

There are no bond angle outliers.

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	1125	1109	1077	7	0
1	B	1036	999	1020	15	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	108	0	0	2	0
3	B	77	0	0	5	0
All	All	2350	2108	2097	22	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 5.

The worst 5 of 22 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:55[A]:GLN:OE1	3:B:301:HOH:O	1.95	0.83
1:A:58[B]:GLU:OE1	3:A:301:HOH:O	2.01	0.79
1:B:50:GLU:OE1	3:B:302:HOH:O	2.02	0.77
1:B:68:TYR:CE2	1:B:70:PHE:HB3	2.20	0.77
1:B:58:GLU:OE2	3:B:303:HOH:O	2.14	0.65

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	145/135 (107%)	143 (99%)	2 (1%)	0	100	100
1	B	127/135 (94%)	124 (98%)	3 (2%)	0	100	100
All	All	272/270 (101%)	267 (98%)	5 (2%)	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	126/115 (110%)	119 (94%)	7 (6%)	21	4
1	B	113/115 (98%)	113 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	239/230 (104%)	232 (97%)	7 (3%)	44 18

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	102	LYS
1	A	127	ASP
1	A	134	LYS
1	A	132	ARG
1	A	58[B]	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	34	ASN
1	A	117	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](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 [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	135/135 (100%)	-0.02	1 (0%) 87 91	16, 27, 48, 64	0
1	B	127/135 (94%)	0.04	2 (1%) 72 78	16, 28, 57, 74	0
All	All	262/270 (97%)	0.01	3 (1%) 80 85	16, 28, 53, 74	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	70	PHE	2.6
1	B	68	TYR	2.6
1	A	135	LYS	2.4

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	NI	A	202	1/1	0.96	0.07	35,35,35,35	0
2	NI	A	201	1/1	0.97	0.09	34,34,34,34	1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	NI	B	202	1/1	0.97	0.03	34,34,34,34	1
2	NI	B	201	1/1	0.98	0.09	34,34,34,34	0

6.5 Other polymers [i](#)

There are no such residues in this entry.