



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 02:17 AM UTC

PDB ID : 7SMC / pdb_00007smc
Title : p107 pocket domain complexed with ARID4A peptide
Authors : Putta, S.; Fernandez, S.M.; Tripathi, S.M.; Muller, G.A.; Rubin, S.M.
Deposited on : 2021-10-25
Resolution : 2.70 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 : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

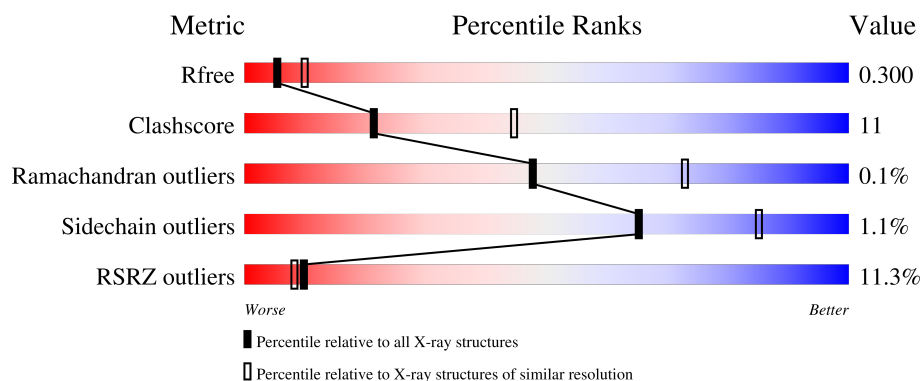
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	180053	3538 (2.70-2.70)
Clashscore	190562	3843 (2.70-2.70)
Ramachandran outliers	187476	3778 (2.70-2.70)
Sidechain outliers	187428	3778 (2.70-2.70)
RSRZ outliers	180081	3538 (2.70-2.70)

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 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	371	7% 67% 23% 8%
1	C	371	12% 67% 26% 6%
2	B	15	20% 53% 20% 27%
2	D	15	40% 13% 27% 7% 53%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5749 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 Retinoblastoma-like protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	340	Total	C	N	O	S	0	0	0
			2761	1776	464	503	18			
1	C	347	Total	C	N	O	S	0	0	0
			2802	1804	471	509	18			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	388	GLY	-	expression tag	UNP P28749
A	389	GLU	-	expression tag	UNP P28749
A	390	PHE	-	expression tag	UNP P28749
C	388	GLY	-	expression tag	UNP P28749
C	389	GLU	-	expression tag	UNP P28749
C	390	PHE	-	expression tag	UNP P28749

- Molecule 2 is a protein called AT-rich interactive domain-containing protein 4A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	S	0	0	0
			87	54	13	19	1			
2	D	7	Total	C	N	O	S	0	0	0
			54	34	9	10	1			

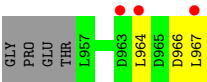
- Molecule 3 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



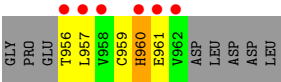
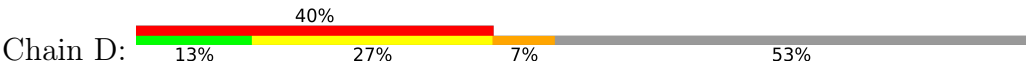
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	10	Total	O	0	0
			10	10		
4	C	20	Total	O	0	0
			20	20		



● Molecule 2: AT-rich interactive domain-containing protein 4A



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	62.86Å 62.92Å 71.27Å 69.89° 76.00° 74.36°	Depositor
Resolution (Å)	47.11 – 2.70 47.11 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.3 (47.11-2.70) 99.3 (47.11-2.70)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.92 (at 2.69Å)	Xtriage
Refinement program	PHENIX (1.17.1_3660: ???)	Depositor
R, R_{free}	0.238 , 0.298 0.239 , 0.300	Depositor DCC
R_{free} test set	1286 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å ²)	38.3	Xtriage
Anisotropy	0.369	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 52.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	5749	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.97% 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

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	0.78	0/2820	1.33	23/3816 (0.6%)
1	C	0.64	0/2863	1.18	19/3878 (0.5%)
2	B	0.77	0/87	1.74	1/118 (0.8%)
2	D	0.75	0/54	1.32	0/73
All	All	0.72	0/5824	1.26	43/7885 (0.5%)

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	A	0	1

There are no bond length outliers.

The worst 5 of 43 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	583	ASN	CA-C-N	-8.36	111.65	122.77
1	A	583	ASN	C-N-CA	-8.36	111.65	122.77
1	C	859	ARG	CA-C-N	-7.60	109.44	123.05
1	C	859	ARG	C-N-CA	-7.60	109.44	123.05
1	A	443	THR	CA-C-N	-7.52	110.88	122.62

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	519	PHE	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	2761	0	2734	56	0
1	C	2802	0	2758	71	0
2	B	87	0	80	6	0
2	D	54	0	53	13	0
3	A	15	0	0	0	0
4	A	10	0	0	0	0
4	C	20	0	0	0	0
All	All	5749	0	5625	124	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 124 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:A:866:LYS:HG2	2:B:967:LEU:HD22	1.22	1.11
1:C:594:PRO:HG3	1:C:793:ARG:HH12	1.29	0.95
1:A:397:LEU:O	1:A:401:VAL:HG22	1.67	0.94
1:A:866:LYS:NZ	1:A:866:LYS:HB3	1.82	0.94
1:A:866:LYS:HB3	1:A:866:LYS:HZ1	1.35	0.88

There are no symmetry-related clashes.

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	332/371 (90%)	319 (96%)	12 (4%)	1 (0%)	36	60
1	C	339/371 (91%)	326 (96%)	13 (4%)	0	100	100
2	B	9/15 (60%)	9 (100%)	0	0	100	100
2	D	5/15 (33%)	4 (80%)	1 (20%)	0	100	100
All	All	685/772 (89%)	658 (96%)	26 (4%)	1 (0%)	48	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	520	PRO

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	307/342 (90%)	303 (99%)	4 (1%)	61	83
1	C	309/342 (90%)	307 (99%)	2 (1%)	78	91
2	B	11/14 (79%)	11 (100%)	0	100	100
2	D	7/14 (50%)	6 (86%)	1 (14%)	3	8
All	All	634/712 (89%)	627 (99%)	7 (1%)	65	85

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

Mol	Chain	Res	Type
1	A	943	LYS
1	C	389	GLU
2	D	960	HIS
1	C	939	VAL
1	A	859	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	595	ASN
1	C	814	ASN
2	D	960	HIS
1	C	830	HIS
1	C	411	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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	SO4	A	1001	-	4,4,4	0.61	0	6,6,6	0.16	0
3	SO4	A	1003	-	4,4,4	0.42	0	6,6,6	0.08	0
3	SO4	A	1002	-	4,4,4	0.55	0	6,6,6	0.08	0

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	340/371 (91%)	0.84	26 (7%) 20 17	21, 40, 74, 99	0
1	C	347/371 (93%)	0.98	45 (12%) 7 6	19, 43, 80, 101	0
2	B	11/15 (73%)	1.82	3 (27%) 1 1	53, 61, 76, 84	0
2	D	7/15 (46%)	2.77	6 (85%) 0 0	79, 85, 93, 104	0
All	All	705/772 (91%)	0.94	80 (11%) 10 8	19, 42, 79, 104	0

The worst 5 of 80 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	886	SER	6.0
1	C	970	LYS	5.6
2	D	962	VAL	5.1
1	C	962	PRO	4.8
2	B	967	LEU	4.7

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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
3	SO4	A	1003	5/5	0.77	0.20	93,101,102,102	0
3	SO4	A	1002	5/5	0.98	0.05	34,35,40,42	0
3	SO4	A	1001	5/5	0.98	0.07	37,41,45,46	0

6.5 Other polymers [i](#)

There are no such residues in this entry.