



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 11:09 AM UTC

PDB ID : 2ARH / pdb_00002arh
Title : Crystal Structure of a Protein of Unknown Function AQ1966 from Aquifex
aeolicus VF5
Authors : Qiu, Y.; Kim, Y.; Yang, X.; Collart, F.; Joachimiak, A.; Kossiakoff, A.; Mid-
west Center for Structural Genomics (MCSG)
Deposited on : 2005-08-19
Resolution : 2.46 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

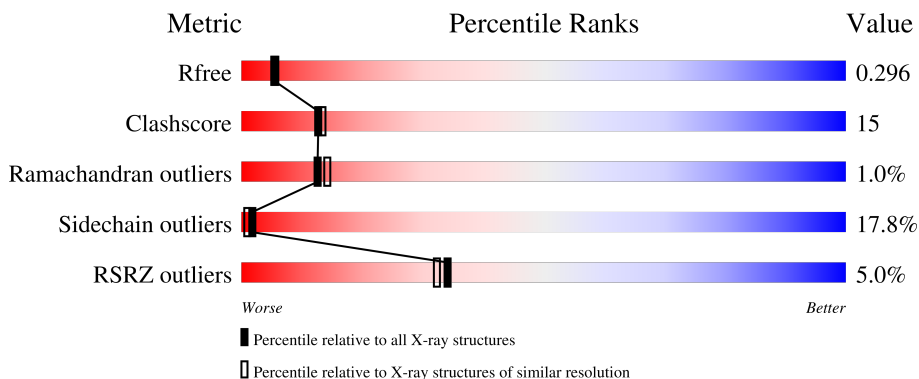
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.46 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	1190 (2.46-2.46)
Clashscore	190562	1229 (2.46-2.46)
Ramachandran outliers	187476	1218 (2.46-2.46)
Sidechain outliers	187428	1218 (2.46-2.46)
RSRZ outliers	180081	1190 (2.46-2.46)

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	203	
1	B	203	
1	C	203	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	SE	C	203	-	-	X	-

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5126 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 hypothetical protein aq_1966.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	198	1662	1082	274	303	1	2	0	0	0
1	B	197	1659	1081	275	300	1	2	0	0	0
1	C	189	1592	1041	264	286	1		0	0	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ASP	-	cloning artifact	UNP O67778
A	0	ALA	-	cloning artifact	UNP O67778
A	1	MSE	MET	modified residue	UNP O67778
A	138	PHE	TYR	engineered mutation	UNP O67778
A	144	MSE	MET	modified residue	UNP O67778
B	-1	ASP	-	cloning artifact	UNP O67778
B	0	ALA	-	cloning artifact	UNP O67778
B	1	MSE	MET	modified residue	UNP O67778
B	138	PHE	TYR	engineered mutation	UNP O67778
B	144	MSE	MET	modified residue	UNP O67778
C	-1	ASP	-	cloning artifact	UNP O67778
C	0	ALA	-	cloning artifact	UNP O67778
C	1	MSE	MET	modified residue	UNP O67778
C	138	PHE	TYR	engineered mutation	UNP O67778
C	144	MSE	MET	modified residue	UNP O67778

- Molecule 2 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



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

- Molecule 3 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0

- Molecule 4 is SELENIUM ATOM (CCD ID: SE) (formula: Se).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	2	Total Se 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	78	Total O 78 78	0	0
5	B	76	Total O 76 76	0	0
5	C	51	Total O 51 51	0	0

LYS	
CYS	
GLU	
ASP	
E182	
G183	
L184	
I185	
K186	
K187	
V188	
N193	
F194	
LEU	
GLU	
HIS	
GLU	
GLY	
SER	
SER	

4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	121.61Å 121.61Å 102.07Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 2.46 20.00 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.7 (20.00-2.46) 99.5 (20.00-2.55)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.70 (at 2.56Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.226 , 0.290 0.234 , 0.296	Depositor DCC
R_{free} test set	1458 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	46.2	Xtrriage
Anisotropy	0.007	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 37.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.027 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	5126	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, SE, 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	1.02	2/1701 (0.1%)	1.20	2/2276 (0.1%)
1	B	1.02	3/1700 (0.2%)	1.20	12/2276 (0.5%)
1	C	0.82	0/1631	1.14	8/2183 (0.4%)
All	All	0.96	5/5032 (0.1%)	1.18	22/6735 (0.3%)

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	C	0	5

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	139	ILE	CA-CB	7.91	1.64	1.54
1	A	66	VAL	CA-CB	6.01	1.61	1.54
1	B	188	VAL	CA-CB	5.55	1.61	1.54
1	B	76	VAL	CA-CB	5.26	1.61	1.54
1	A	1	MSE	SE-CE	5.05	2.10	1.95

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	57	LYS	CA-C-N	-11.73	106.76	119.19
1	C	57	LYS	C-N-CA	-11.73	106.76	119.19
1	B	141	GLU	N-CA-C	-9.85	96.61	110.50
1	A	115	VAL	N-CA-CB	7.64	117.86	109.99
1	C	4	TYR	N-CA-C	-7.51	100.42	110.55

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	12	GLU	Peptide
1	C	193	ASN	Peptide
1	C	3	LYS	Peptide
1	C	47	LEU	Peptide
1	C	56	ARG	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	1662	0	1653	53	1
1	B	1659	0	1651	52	0
1	C	1592	0	1586	52	0
2	A	5	0	0	0	0
3	A	1	0	0	0	0
4	C	2	0	0	3	0
5	A	78	0	0	4	8
5	B	76	0	0	9	7
5	C	51	0	0	2	6
All	All	5126	0	4890	149	12

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 15.

The worst 5 of 149 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:47:LEU:HD13	1:A:48:LEU:HD13	1.40	1.04
1:A:156:LYS:H	1:A:156:LYS:HZ2	1.06	1.01
1:B:197:HIS:NE2	5:B:245:HOH:O	1.97	0.95
1:A:158:GLU:OE2	1:A:162:LYS:HE3	1.66	0.94
1:B:56:ARG:HD3	5:B:223:HOH:O	1.70	0.91

The worst 5 of 12 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)
5:A:205:HOH:O	5:B:232:HOH:O[4_556]	1.26	0.94
5:A:204:HOH:O	5:B:226:HOH:O[4_556]	1.59	0.61
5:A:242:HOH:O	5:B:235:HOH:O[6_656]	1.65	0.55
5:A:239:HOH:O	5:B:221:HOH:O[6_656]	1.74	0.46
5:B:205:HOH:O	5:B:230:HOH:O[4_556]	1.75	0.45

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	196/203 (97%)	188 (96%)	7 (4%)	1 (0%)	24	32
1	B	195/203 (96%)	183 (94%)	11 (6%)	1 (0%)	24	32
1	C	185/203 (91%)	169 (91%)	12 (6%)	4 (2%)	5	3
All	All	576/609 (95%)	540 (94%)	30 (5%)	6 (1%)	12	14

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	195	LEU
1	B	57	LYS
1	C	5	GLU
1	C	16	ASN
1	C	17	SER

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	177/179 (99%)	145 (82%)	32 (18%)	2	1
1	B	177/179 (99%)	149 (84%)	28 (16%)	2	1
1	C	169/179 (94%)	136 (80%)	33 (20%)	1	0
All	All	523/537 (97%)	430 (82%)	93 (18%)	2	1

5 of 93 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	190	GLU
1	C	50	LEU
1	B	197	HIS
1	C	16	ASN
1	C	74	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	75	ASN
1	C	167	ASN
1	B	72	ASN
1	B	75	ASN
1	B	112	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](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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	$\# Z > 2$
2	SO4	A	202	-	4,4,4	0.24	0	6,6,6	0.45	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 [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 [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	196/203 (96%)	0.65	7 (3%) 46 46	28, 39, 53, 61	0
1	B	195/203 (96%)	0.13	5 (2%) 57 58	29, 42, 58, 63	0
1	C	188/203 (92%)	0.85	17 (9%) 15 13	35, 55, 78, 82	0
All	All	579/609 (95%)	0.54	29 (5%) 34 32	28, 44, 71, 82	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	2	VAL	5.6
1	C	184	LEU	3.7
1	C	17	SER	3.6
1	C	11	LEU	3.3
1	A	196	GLU	3.3

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(\AA^2)	Q<0.9
4	SE	C	202	1/1	0.86	0.42	157,157,157,157	1
4	SE	C	203	1/1	0.94	0.09	78,78,78,78	1
3	CA	A	203	1/1	0.97	0.15	55,55,55,55	0
2	SO4	A	202	5/5	0.98	0.07	37,38,39,41	5

6.5 Other polymers [i](#)

There are no such residues in this entry.