wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 05:06 am BST

PDB ID	:	1F9N
Title	:	CRYSTAL STRUCTURE OF AHRC, THE ARGININE REPRES-SOR/ACTIVATOR PROTEIN FROM BACILLUS SUBTILIS
Authors	:	Dennis, C.A.; Glykos, N.M.; Parsons, M.R.; Phillips, S.E.V.
Deposited on	:	2000-07-11
Resolution	:	2.70 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

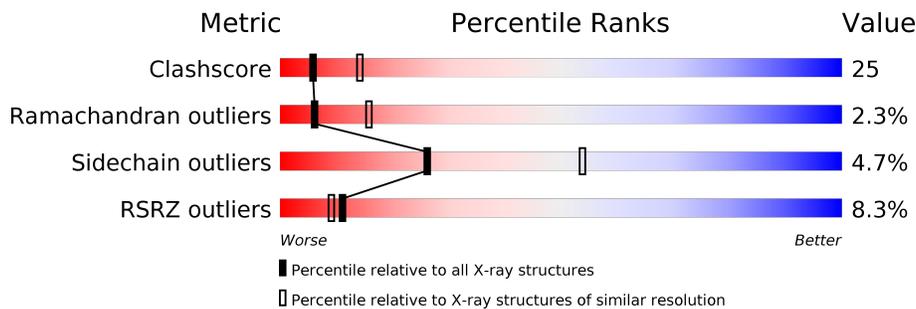|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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	Percentile Ranks	Value
Clashscore		25
Ramachandran outliers		2.3%
Sidechain outliers		4.7%
RSRZ outliers		8.3%

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))
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (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 on 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	149	8% / 49% 49% ••
1	B	149	3% / 54% 42% 5%
1	C	149	20% / 51% 39% 8% ••
1	D	149	% / 58% 36% • •
1	E	149	% / 63% 35% ••
1	F	149	16% / 54% 42% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7108 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 ARGININE REPRESSOR/ACTIVATOR PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	147	Total 1143	C 715	N 195	O 224	S 9	0	0	0
1	B	149	Total 1157	C 725	N 196	O 226	S 10	0	0	0
1	C	148	Total 1132	C 708	N 193	O 222	S 9	0	0	0
1	D	146	Total 1124	C 703	N 189	O 223	S 9	0	0	0
1	E	148	Total 1159	C 725	N 199	O 226	S 9	0	0	0
1	F	149	Total 1153	C 724	N 195	O 224	S 10	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total 21	O 21	0	0
2	B	56	Total 56	O 56	0	0
2	C	34	Total 34	O 34	0	0
2	D	36	Total 36	O 36	0	0
2	E	60	Total 60	O 60	0	0
2	F	33	Total 33	O 33	0	0

3 Residue-property plots ⓘ

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: ARGININE REPRESSOR/ACTIVATOR PROTEIN

- Molecule 1: ARGININE REPRESSOR/ACTIVATOR PROTEIN

- Molecule 1: ARGININE REPRESSOR/ACTIVATOR PROTEIN

- Molecule 1: ARGININE REPRESSOR/ACTIVATOR PROTEIN

MET ASN LYS G4 H7 R11 E12 I13 I14 T15 I19 E20 T21 Q22 D23 E24 L25 V26 D27 Q31 K35 V36 T37 T40 V41 S42 R43 D44 I45 K46 E47 L48 K52 V53 P54 T55 N56 N57 Y60 K61 Y62 S63 L64 D67 N71 P72 L73 S74 K75 L76

K77 M81 D82 S91 H92 M93 I94 V95 L96 K97 T98 Q104 A105 I106 L109 L113 D116 T121 I122 C123 G124 I128 L129 I130 I131 C132 R133 T134 P135 L149

• Molecule 1: ARGININE REPRESSOR/ACTIVATOR PROTEIN

Chain E:

MET N2 R6 K9 I10 E20 T21 Q22 L25 V26 L29 G33 V36 T37 T40 V41 I45 L48 H49 L50 P54 N57 Y60 P65 F70 N71 P72 L73 S74 K75 A79 F84 V85 K86 I87 D88 S89 A90 S91 H92 M93 I94 V95

L96 K97 N102 I106 G107 M110 L113 D116 M119 I122 I128 I129 I130 I131 P135 E136 D137 V141 K142 N143 E147 L148 L149

• Molecule 1: ARGININE REPRESSOR/ACTIVATOR PROTEIN

Chain F:

M1 N2 Q5 R11 E12 I13 I14 T15 S16 I19 E20 T21 Q22 D23 E24 L25 V26 D27 M28 V36 T40 V41 I45 L48 H49 L50 V51 K52 V53 P54 T55 N56 N57 G58 S59 Y60 K61 Y62 S63 L64 P65 A66 D67 Q68 R69 F70 N71 P72 L73 L76 K77

M81 K86 I87 A90 S91 H92 M93 I94 K97 N102 A103 Q104 M110 L113 D114 E117 M118 M119 G120 T121 I122 D126 T127 I128 L129 I130 P135 V141 L145 L146 E147 L148 L149

4   Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	230.31Å 73.86Å 138.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.70 34.73 – 2.70	Depositor EDS
% Data completeness (in resolution range)	88.7 (30.00-2.70) 85.0 (34.73-2.70)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.36 (at 2.68Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.220 , 0.270 0.215 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	56.4	Xtriage
Anisotropy	0.044	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 87.4	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7108	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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 (i)

5.1 Standard geometry (i)

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	0.26	0/1157	0.52	0/1563
1	B	0.24	0/1171	0.50	0/1580
1	C	0.29	0/1145	0.58	0/1548
1	D	0.24	0/1138	0.49	0/1541
1	E	0.26	0/1173	0.52	0/1582
1	F	0.37	0/1167	0.61	1/1574 (0.1%)
All	All	0.28	0/6951	0.54	1/9388 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	56	ASN	CA-CB-CG	-5.83	100.57	113.40

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	1143	0	1153	56	1
1	B	1157	0	1176	59	0
1	C	1132	0	1135	88	0
1	D	1124	0	1118	42	0
1	E	1159	0	1181	40	0
1	F	1153	0	1169	76	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	21	0	0	0	0
2	B	56	0	0	1	0
2	C	34	0	0	0	0
2	D	36	0	0	0	0
2	E	60	0	0	1	0
2	F	33	0	0	2	0
All	All	7108	0	6932	347	1

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 25.

The worst 5 of 347 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:C:48:LEU:O	1:C:65:PRO:HG2	1.64	0.97
1:F:56:ASN:ND2	2:F:171:HOH:O	1.98	0.95
1:F:51:VAL:HG12	1:F:52:LYS:N	1.82	0.93
1:C:62:TYR:HD1	1:C:62:TYR:H	1.04	0.93
1:C:62:TYR:N	1:C:62:TYR:CD1	2.34	0.90

All (1) 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:A:144:ARG:NH2	1:A:144:ARG:NH2[3_555]	1.76	0.44

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	145/149 (97%)	137 (94%)	7 (5%)	1 (1%)	22 46

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	147/149 (99%)	136 (92%)	7 (5%)	4 (3%)	5	12
1	C	146/149 (98%)	125 (86%)	11 (8%)	10 (7%)	1	1
1	D	144/149 (97%)	135 (94%)	8 (6%)	1 (1%)	22	46
1	E	146/149 (98%)	141 (97%)	3 (2%)	2 (1%)	11	28
1	F	147/149 (99%)	132 (90%)	13 (9%)	2 (1%)	11	28
All	All	875/894 (98%)	806 (92%)	49 (6%)	20 (2%)	6	16

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	91	SER
1	B	2	ASN
1	F	72	PRO
1	B	91	SER
1	C	20	GLU

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	128/133 (96%)	123 (96%)	5 (4%)	32	61
1	B	130/133 (98%)	124 (95%)	6 (5%)	27	54
1	C	125/133 (94%)	115 (92%)	10 (8%)	12	27
1	D	125/133 (94%)	121 (97%)	4 (3%)	39	68
1	E	131/133 (98%)	128 (98%)	3 (2%)	50	78
1	F	128/133 (96%)	120 (94%)	8 (6%)	18	40
All	All	767/798 (96%)	731 (95%)	36 (5%)	26	54

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	109	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	142	LYS
1	F	114	ASP
1	C	126	ASP
1	D	35	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	143	ASN
1	C	68	GLN
1	F	38	GLN
1	B	92	HIS
1	B	102	ASN

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 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 (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(Å²)	Q<0.9
1	A	147/149 (98%)	0.38	12 (8%)	11	9	35, 67, 133, 146	0
1	B	149/149 (100%)	0.21	4 (2%)	54	55	21, 47, 91, 124	0
1	C	148/149 (99%)	1.06	30 (20%)	1	0	33, 66, 150, 164	0
1	D	146/149 (97%)	0.12	2 (1%)	75	77	31, 59, 109, 127	0
1	E	148/149 (99%)	0.08	2 (1%)	75	77	24, 43, 74, 83	0
1	F	149/149 (100%)	0.88	24 (16%)	1	1	39, 76, 151, 165	0
All	All	887/894 (99%)	0.46	74 (8%)	11	9	21, 59, 130, 165	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	59	SER	7.8
1	A	39	ALA	7.7
1	C	53	VAL	7.5
1	C	54	PRO	7.4
1	C	70	PHE	7.2

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 carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.