



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 17, 2024 – 06:24 AM EDT

PDB ID : 3GKU
Title : Crystal structure of a probable RNA-binding protein from *Clostridium symbiosum* ATCC 14940
Authors : Tan, K.; Keigher, L.; Jedrzejczak, R.; Babnigg, G.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on : 2009-03-11
Resolution : 2.95 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
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 2.37.1
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) : 2.37.1

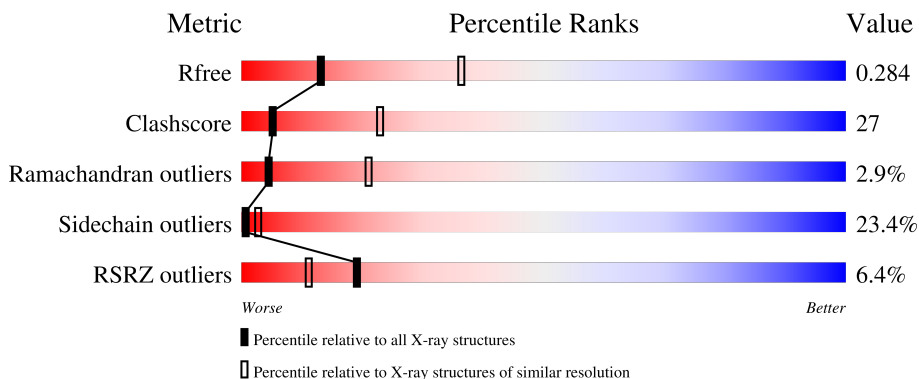
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	225	
1	B	225	
1	C	225	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4131 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 Probable RNA-binding protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	206	Total 1636	C 1020	N 287	O 323	Se 6	0	0	0
1	B	193	Total 1500	C 934	N 263	O 297	Se 6	0	0	0
1	C	134	Total 995	C 619	N 163	O 208	Se 5	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

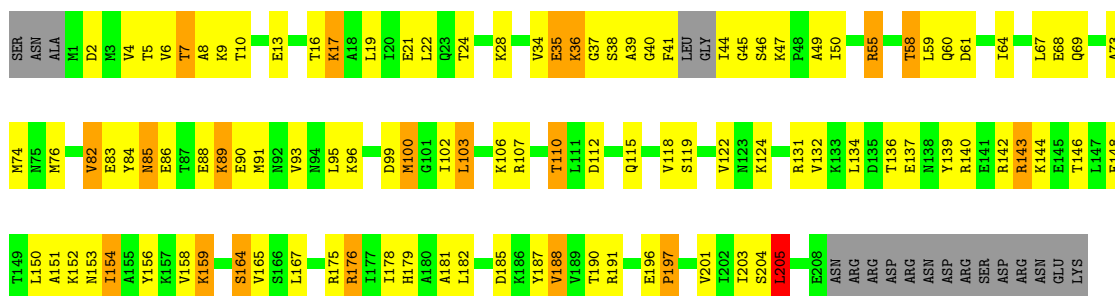
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	EXPRESSION TAG	PDB 3GKU
A	-1	ASN	-	EXPRESSION TAG	PDB 3GKU
A	0	ALA	-	EXPRESSION TAG	PDB 3GKU
B	-2	SER	-	EXPRESSION TAG	PDB 3GKU
B	-1	ASN	-	EXPRESSION TAG	PDB 3GKU
B	0	ALA	-	EXPRESSION TAG	PDB 3GKU
C	-2	SER	-	EXPRESSION TAG	PDB 3GKU
C	-1	ASN	-	EXPRESSION TAG	PDB 3GKU
C	0	ALA	-	EXPRESSION TAG	PDB 3GKU

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.

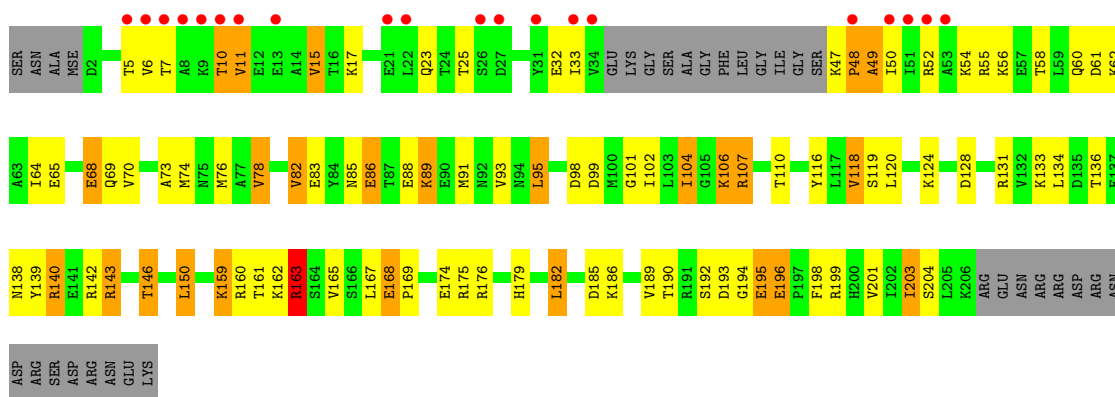
- Molecule 1: Probable RNA-binding protein

Chain A: 




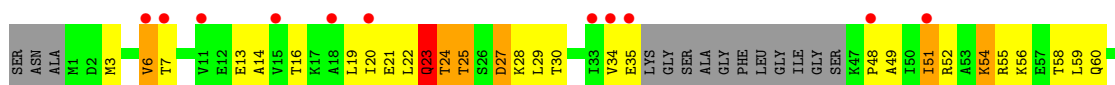
- Molecule 1: Probable RNA-binding protein

Chain B: 



- Molecule 1: Probable RNA-binding protein

Chain C: 



Q69	K144	SER
M76	E145	LEU
A77	THR	LYS
V78	LEU	ARG
D79	GLU	GLU
I80	THR	ASN
S81	LEU	ARG
V82	ALA	ARG
E83	LYS	ASP
Y84	ASN	ARG
N85	ILE	ASN
E86	ALA	ASP
T87	TYR	ARG
E88	LYS	SER
K89	VAL	ASP
E90	LYS	ARG
M91	THR	ASN
M92	LYS	GLU
V93	ARG	LYS
K96	SER	SER
G97	VAL	VAL
D98	SER	SER
D99	LEU	LEU
M100	GLU	GLU
G101	PRO	PRO
I102	MSE	ASN
L103	ASN	ASN
I104	PRO	PRO
G105	TYR	TYR
K106	GLU	GLU
R107	ARG	ARG
G108	ARG	ARG
Q109	ILE	ILE
T110	ILE	ILE
S113	HIS	HIS
L114	ALA	ALA
Q115	ALA	ALA
Y116	LEU	LEU
L117	GLN	GLN
V118	ASN	ASN
S119	ASP	ASP
L120	ASP	ASP
V121	LYS	LYS
Y129	TYR	TYR
I130	VAL	VAL
R131	VAL	VAL
V132	THR	THR
K133	ARG	ARG
L134	SER	SER
D135	ASP	ASP
T136	GLY	GLY
E137	GLU	GLU
E141	PRO	PRO
R142	PHE	PHE
R143	E137	ARG
	HIS	HIS
	VAL	VAL
	ILE	ILE
	ILE	ILE

SER
LEU
LYS
ARG
GLU
ASN
ARG
ARG
ASP
ARG
ASN
ASP
SER
SER
ASP
ARG
ASN
GLU
LYS

4 Data and refinement statistics i

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	126.99Å 126.99Å 106.19Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.72 – 2.95 40.73 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (40.72-2.95) 99.8 (40.73-2.95)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.43 (at 2.95Å)	Xtrriage
Refinement program	REFMAC 5.5.0054	Depositor
R, R_{free}	0.220 , 0.287 0.219 , 0.284	Depositor DCC
R_{free} test set	1084 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å ²)	59.8	Xtrriage
Anisotropy	0.104	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 56.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.035 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	4131	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% 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](#)

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	1.03	0/1645	0.98	2/2200 (0.1%)
1	B	0.82	0/1508	0.84	0/2027
1	C	0.67	0/994	0.80	1/1337 (0.1%)
All	All	0.88	0/4147	0.89	3/5564 (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	C	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	97	GLY	N-CA-C	6.51	129.37	113.10
1	A	205	LEU	CA-CB-CG	5.38	127.69	115.30
1	A	67	LEU	CB-CG-CD1	-5.36	101.89	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	24	THR	Peptide

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	1636	0	1673	87	0
1	B	1500	0	1483	65	0
1	C	995	0	949	73	0
All	All	4131	0	4105	221	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 27.

The worst 5 of 221 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:103:LEU:HD23	1:A:103:LEU:H	1.02	1.13
1:C:100:MSE:HE2	1:C:136:THR:HG22	1.24	1.12
1:A:110:THR:HG22	1:B:116:TYR:HE2	1.08	1.10
1:C:97:GLY:HA2	1:C:100:MSE:HG2	1.29	1.10
1:C:100:MSE:HE2	1:C:136:THR:CG2	1.86	1.05

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	202/225 (90%)	179 (89%)	22 (11%)	1 (0%)	29 64
1	B	189/225 (84%)	161 (85%)	17 (9%)	11 (6%)	1 7
1	C	130/225 (58%)	110 (85%)	17 (13%)	3 (2%)	6 27
All	All	521/675 (77%)	450 (86%)	56 (11%)	15 (3%)	4 21

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	45	GLY
1	B	107	ARG
1	C	23	GLN
1	C	48	PRO
1	B	17	LYS

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	181/195 (93%)	147 (81%)	34 (19%)	1	7
1	B	161/195 (83%)	117 (73%)	44 (27%)	0	1
1	C	103/195 (53%)	77 (75%)	26 (25%)	0	2
All	All	445/585 (76%)	341 (77%)	104 (23%)	1	3

5 of 104 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	128	ASP
1	B	176	ARG
1	C	115	GLN
1	B	133	LYS
1	B	159	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	200	HIS
1	B	75	ASN
1	B	200	HIS
1	B	153	ASN
1	A	183	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](#)

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(Å ²)	Q<0.9
1	A	199/225 (88%)	-0.18	0 100 100	17, 32, 56, 70	0
1	B	187/225 (83%)	0.34	20 (10%) 6 3	27, 46, 82, 92	0
1	C	128/225 (56%)	0.39	13 (10%) 6 4	38, 56, 71, 76	0
All	All	514/675 (76%)	0.15	33 (6%) 19 11	17, 44, 72, 92	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	6	VAL	5.5
1	B	8	ALA	4.8
1	B	9	LYS	4.7
1	B	7	THR	4.4
1	B	51	ILE	4.3

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 monosaccharides 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.