



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 22, 2023 – 10:28 PM JST

PDB ID : 7XQC
Title : Crystal structure of N-terminal domain of Rv2908c fused with Maltose Binding Protein (MBP)
Authors : Thakur, K.G.; Singh, A.; Kapoor, S.; Deep, A.
Deposited on : 2022-05-07
Resolution : 2.80 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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

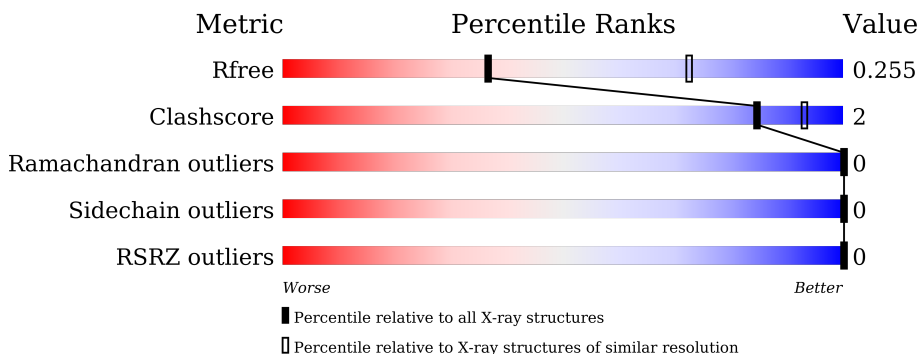
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	450	
1	B	450	
1	C	450	
1	D	450	

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 13042 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 Fusion protein of Maltose-binding periplasmic protein and RNA-binding protein KhpA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	414	3212	2054	534	617	7	0	1	0
1	B	414	3224	2062	538	617	7	0	2	0
1	C	413	3196	2046	531	612	7	0	0	0
1	D	414	3204	2050	532	615	7	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

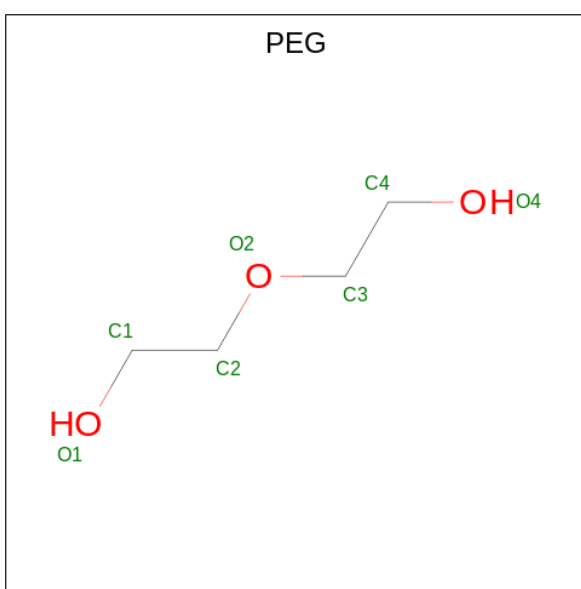
Chain	Residue	Modelled	Actual	Comment	Reference
A	325	ASP	GLN	conflict	UNP P0AEX9
A	362	ALA	LYS	conflict	UNP P0AEX9
A	363	ALA	ASP	conflict	UNP P0AEX9
A	367	ASN	-	linker	UNP P0AEX9
A	368	ALA	-	linker	UNP P0AEX9
A	369	ALA	-	linker	UNP P0AEX9
A	370	HIS	-	linker	UNP P0AEX9
B	325	ASP	GLN	conflict	UNP P0AEX9
B	362	ALA	LYS	conflict	UNP P0AEX9
B	363	ALA	ASP	conflict	UNP P0AEX9
B	367	ASN	-	linker	UNP P0AEX9
B	368	ALA	-	linker	UNP P0AEX9
B	369	ALA	-	linker	UNP P0AEX9
B	370	HIS	-	linker	UNP P0AEX9
C	325	ASP	GLN	conflict	UNP P0AEX9
C	362	ALA	LYS	conflict	UNP P0AEX9
C	363	ALA	ASP	conflict	UNP P0AEX9
C	367	ASN	-	linker	UNP P0AEX9
C	368	ALA	-	linker	UNP P0AEX9
C	369	ALA	-	linker	UNP P0AEX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	370	HIS	-	linker	UNP P0AEX9
D	325	ASP	GLN	conflict	UNP P0AEX9
D	362	ALA	LYS	conflict	UNP P0AEX9
D	363	ALA	ASP	conflict	UNP P0AEX9
D	367	ASN	-	linker	UNP P0AEX9
D	368	ALA	-	linker	UNP P0AEX9
D	369	ALA	-	linker	UNP P0AEX9
D	370	HIS	-	linker	UNP P0AEX9

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C₄H₁₀O₃).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 7 4 3	0	0
2	B	1	Total C O 7 4 3	0	0
2	C	1	Total C O 7 4 3	0	0
2	D	1	Total C O 7 4 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	51	Total O 51 51	0	0

Continued on next page...

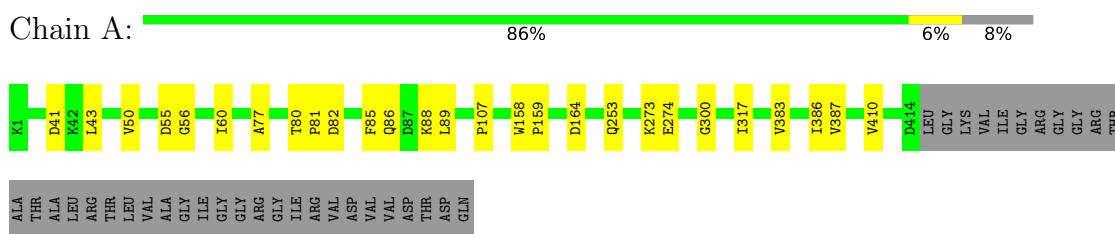
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	36	Total 36	O 36	0	0
3	C	52	Total 52	O 52	0	0
3	D	39	Total 39	O 39	0	0

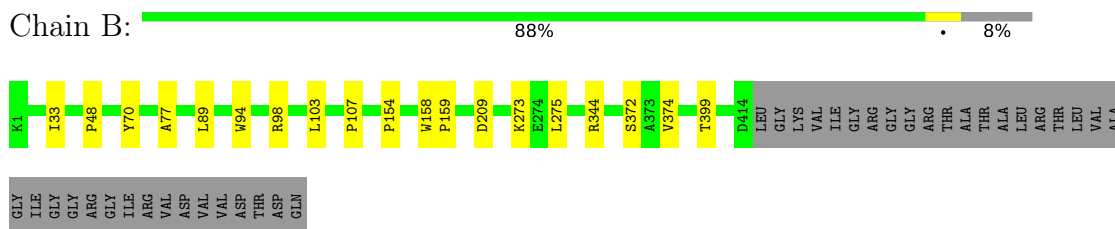
3 Residue-property plots [i](#)

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.

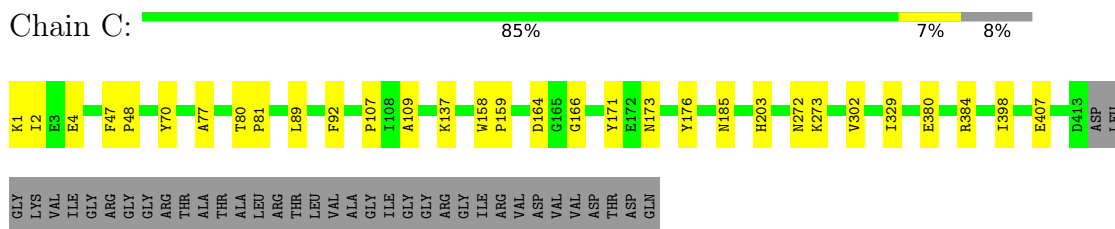
- Molecule 1: Fusion protein of Maltose-binding periplasmic protein and RNA-binding protein KhpA



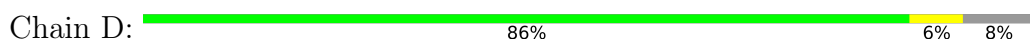
- Molecule 1: Fusion protein of Maltose-binding periplasmic protein and RNA-binding protein KhpA



- Molecule 1: Fusion protein of Maltose-binding periplasmic protein and RNA-binding protein KhpA



- Molecule 1: Fusion protein of Maltose-binding periplasmic protein and RNA-binding protein KhpA



K1	ARG	K170	LEU
G5	THR	Y171	GLY
F27	LEU	E172	LYS
T31	VAL	I178	VAL
F47	ALA	N272	ILE
P48	GLY	K273	GLY
D55	ILE	Q335	ARG
A77	ARG	R354	GLY
L89	VAL	E380	THR
Y90	ASP	V383	ALA
P91	VAL	R384	THR
P107	VAL	V387	ALA
W158	ASP	D414	LEU
P159	THR		GLY
K170	GLN		VAL
Y171			ILE
E172			GLY
I178			ARG
N272			GLY
K273			THR
Q335			ALA
R354			THR
E380			ALA
V383			LEU
R384			
V387			
D414			

4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	68.86Å 189.58Å 72.13Å 90.00° 114.73° 90.00°	Depositor
Resolution (Å)	65.60 – 2.80 65.51 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.9 (65.60-2.80) 99.9 (65.51-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.19	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.27 (at 2.81Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.196 , 0.254 0.200 , 0.255	Depositor DCC
R_{free} test set	1986 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å ²)	27.4	Xtrriage
Anisotropy	0.030	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 25.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	0.041 for l,-k,h	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	13042	wwPDB-VP
Average B, all atoms (Å ²)	29.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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: PEG

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	0.68	0/3287	0.74	0/4469
1	B	0.68	0/3299	0.75	0/4483
1	C	0.67	0/3271	0.76	0/4447
1	D	0.68	0/3279	0.75	0/4458
All	All	0.68	0/13136	0.75	0/17857

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	3212	0	3170	20	0
1	B	3224	0	3189	11	0
1	C	3196	0	3161	17	0
1	D	3204	0	3165	14	0
2	A	7	0	10	0	0
2	B	7	0	10	0	0
2	C	7	0	10	0	0
2	D	7	0	10	0	0
3	A	51	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	36	0	0	1	0
3	C	52	0	0	0	0
3	D	39	0	0	0	0
All	All	13042	0	12725	60	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 2.

The worst 5 of 60 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:D:383:VAL:O	1:D:387:VAL:HG23	1.94	0.68
1:C:171:TYR:HB2	1:C:176:TYR:CE1	2.34	0.63
1:C:171:TYR:CZ	1:C:173:ASN:O	2.52	0.63
1:A:383:VAL:O	1:A:387:VAL:HG23	1.99	0.62
1:B:33:ILE:HD13	1:B:275:LEU:HD22	1.82	0.60

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	413/450 (92%)	405 (98%)	8 (2%)	0	100	100
1	B	414/450 (92%)	405 (98%)	9 (2%)	0	100	100
1	C	411/450 (91%)	401 (98%)	10 (2%)	0	100	100
1	D	412/450 (92%)	402 (98%)	10 (2%)	0	100	100
All	All	1650/1800 (92%)	1613 (98%)	37 (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	334/358 (93%)	334 (100%)	0	100	100
1	B	335/358 (94%)	335 (100%)	0	100	100
1	C	332/358 (93%)	332 (100%)	0	100	100
1	D	333/358 (93%)	333 (100%)	0	100	100
All	All	1334/1432 (93%)	1334 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	409	HIS

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](#)

4 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
2	PEG	D	501	-	6,6,6	0.18	0	5,5,5	0.16	0
2	PEG	C	501	-	6,6,6	0.11	0	5,5,5	0.13	0
2	PEG	A	501	-	6,6,6	0.12	0	5,5,5	0.17	0
2	PEG	B	501	-	6,6,6	0.28	0	5,5,5	0.12	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PEG	D	501	-	-	1/4/4/4	-
2	PEG	C	501	-	-	0/4/4/4	-
2	PEG	A	501	-	-	1/4/4/4	-
2	PEG	B	501	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	501	PEG	O2-C3-C4-O4
2	A	501	PEG	O1-C1-C2-O2
2	B	501	PEG	O2-C3-C4-O4

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

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	414/450 (92%)	-0.55	0 100 100	14, 23, 38, 61	0
1	B	414/450 (92%)	-0.38	0 100 100	16, 29, 50, 68	0
1	C	413/450 (91%)	-0.45	0 100 100	14, 25, 39, 51	0
1	D	414/450 (92%)	-0.35	0 100 100	17, 32, 59, 73	0
All	All	1655/1800 (91%)	-0.43	0 100 100	14, 27, 48, 73	0

There are no RSRZ outliers to report.

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	PEG	B	501	7/7	0.85	0.16	42,45,53,56	0
2	PEG	C	501	7/7	0.88	0.23	20,20,20,20	0
2	PEG	A	501	7/7	0.89	0.20	34,35,36,37	0
2	PEG	D	501	7/7	0.89	0.19	33,36,37,38	0

6.5 Other polymers [i](#)

There are no such residues in this entry.