wwPDB X-ray Structure Validation Summary Report (i)

Mar 23, 2024 – 07:22 PM EDT

PDB ID	:	2AR0
Title	:	Crystal structure of Type I restriction enzyme EcoKI M protein (EC 2.1.1.72) (M.EcoKI)
Authors	:	Rajashankar, K.R.; Kniewel, R.; Lima, C.D.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2005-08-18
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 (i) 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 (1)) were used in the production of this report:

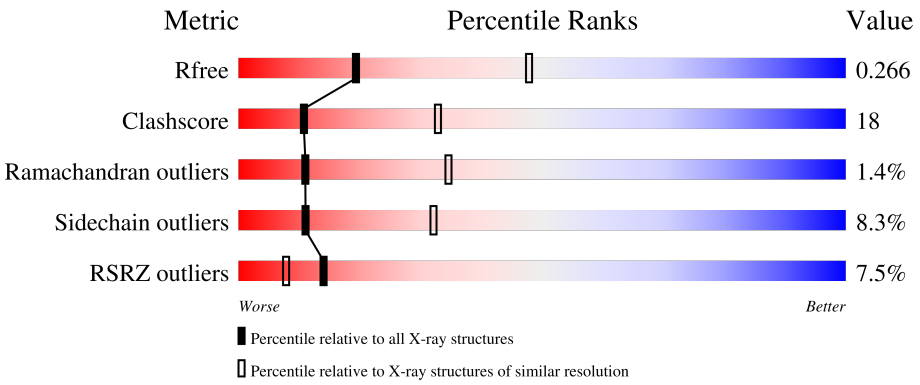
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.1

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 2.80 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 <=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	541	6% 58% 28% • 10%
1	B	541	7% 57% 29% • 11%

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
2	UNX	A	1001	-	-	-	X
2	UNX	A	1003	-	-	-	X
2	UNX	A	1004	-	-	-	X
2	UNX	A	1005	-	-	-	X
2	UNX	A	1006	-	-	-	X
2	UNX	A	1007	-	-	-	X
2	UNX	A	1009	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7828 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 Type I restriction enzyme EcoKI M protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	485	3843	2415	671	742	6	9	0	0	0
1	B	483	3832	2408	669	740	6	9	0	0	0

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	cloning artifact	UNP P08957
A	2	SER	-	cloning artifact	UNP P08957
A	3	LEU	-	cloning artifact	UNP P08957
A	41	MSE	MET	modified residue	UNP P08957
A	76	MSE	MET	modified residue	UNP P08957
A	112	MSE	MET	modified residue	UNP P08957
A	132	MSE	MET	modified residue	UNP P08957
A	229	MSE	MET	modified residue	UNP P08957
A	295	MSE	MET	modified residue	UNP P08957
A	331	MSE	MET	modified residue	UNP P08957
A	386	MSE	MET	modified residue	UNP P08957
A	492	MSE	MET	modified residue	UNP P08957
A	506	MSE	MET	modified residue	UNP P08957
A	530	LYS	-	expression tag	UNP P08957
A	531	GLU	-	expression tag	UNP P08957
A	532	GLU	-	expression tag	UNP P08957
A	533	GLY	-	expression tag	UNP P08957
A	534	GLY	-	expression tag	UNP P08957
A	535	SER	-	expression tag	UNP P08957
A	536	HIS	-	expression tag	UNP P08957
A	537	HIS	-	expression tag	UNP P08957
A	538	HIS	-	expression tag	UNP P08957
A	539	HIS	-	expression tag	UNP P08957
A	540	HIS	-	expression tag	UNP P08957
A	541	HIS	-	expression tag	UNP P08957

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MSE	-	cloning artifact	UNP P08957
B	2	SER	-	cloning artifact	UNP P08957
B	3	LEU	-	cloning artifact	UNP P08957
B	41	MSE	MET	modified residue	UNP P08957
B	76	MSE	MET	modified residue	UNP P08957
B	112	MSE	MET	modified residue	UNP P08957
B	132	MSE	MET	modified residue	UNP P08957
B	229	MSE	MET	modified residue	UNP P08957
B	295	MSE	MET	modified residue	UNP P08957
B	331	MSE	MET	modified residue	UNP P08957
B	386	MSE	MET	modified residue	UNP P08957
B	492	MSE	MET	modified residue	UNP P08957
B	506	MSE	MET	modified residue	UNP P08957
B	530	LYS	-	expression tag	UNP P08957
B	531	GLU	-	expression tag	UNP P08957
B	532	GLU	-	expression tag	UNP P08957
B	533	GLY	-	expression tag	UNP P08957
B	534	GLY	-	expression tag	UNP P08957
B	535	SER	-	expression tag	UNP P08957
B	536	HIS	-	expression tag	UNP P08957
B	537	HIS	-	expression tag	UNP P08957
B	538	HIS	-	expression tag	UNP P08957
B	539	HIS	-	expression tag	UNP P08957
B	540	HIS	-	expression tag	UNP P08957
B	541	HIS	-	expression tag	UNP P08957

- Molecule 2 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	X		
2	A	10	10	10	0	0

- Molecule 3 is water.

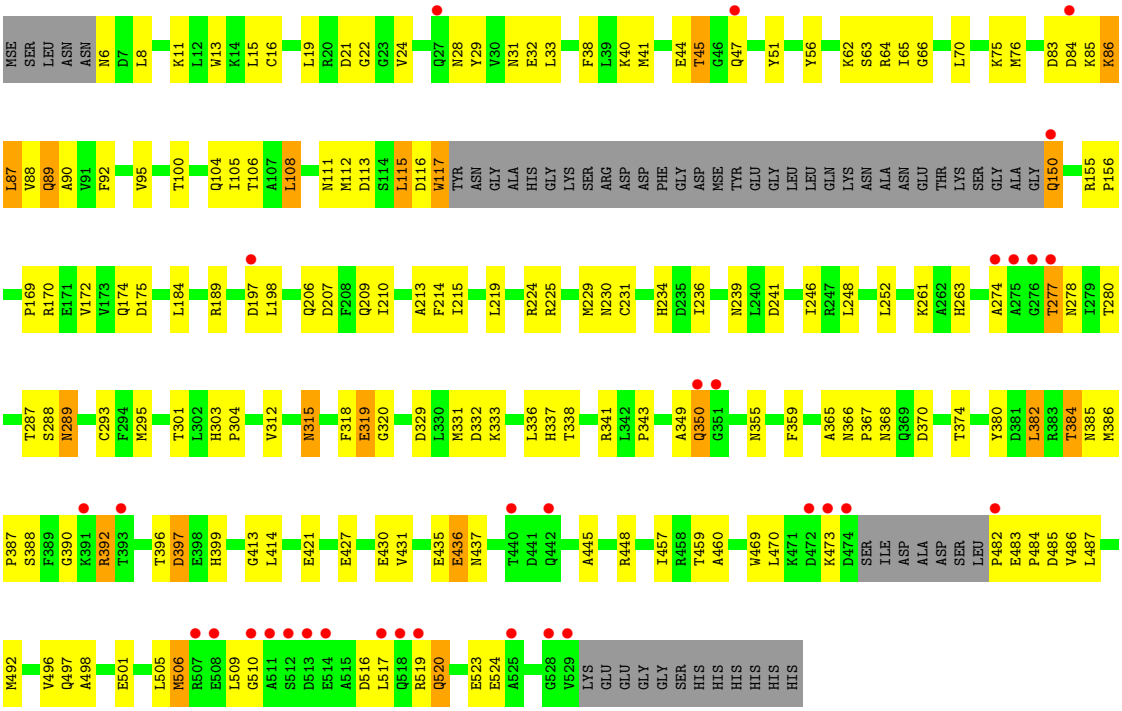
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	86	86	86	0	0
			Total	O		
3	B	57	57	57	0	0

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.

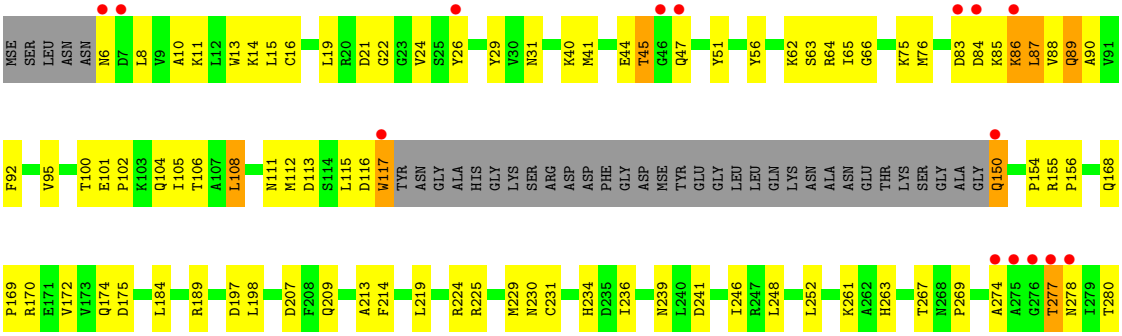
• Molecule 1: Type I restriction enzyme EcoKI M protein

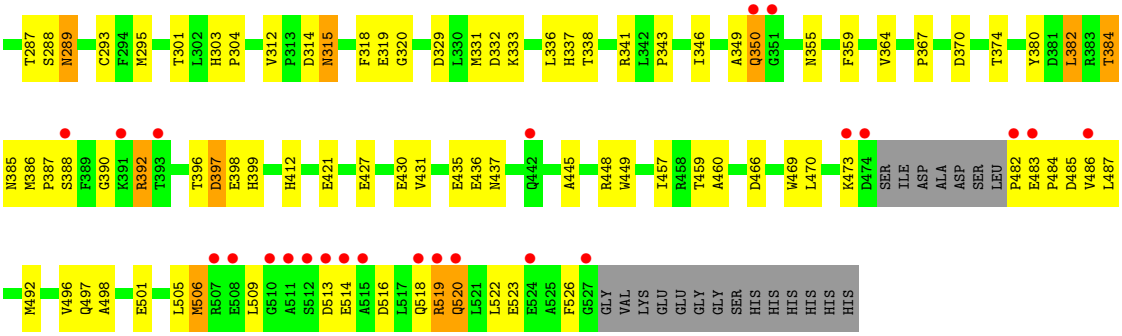
Chain A: 6% 58% 28% • 10%



• Molecule 1: Type I restriction enzyme EcoKI M protein

Chain B: 7% 57% 29% • 11%




T287 S288 N289 C293 F294 M295 T301 L302 H303 P304 V312 P313 D314 N315 F318 E319 G320 D329 L330 M331 D332 K333 L336 H337 T338 R341 L342 P343 I346 A349 Q350 G351 N355 F359 V364 P367 D370 T374 Y380 D381 L382 R383 T384
N385 M386 P387 S388 F389 G390 K391 R392 T393 T396 D397 E398 H399 H412 E421 E427 E430 V431 E435 E436 N437 Q442 A445 R448 W449 I457 R458 T459 A460 D466 W469 L470 K473 D474 SER ILE ASP ALA ASP SER LEU P482 E483 P484 D485 V486 L487
M492 V496 Q497 A498 E501 L505 M506 R507 E508 L509 G510 A511 S512 D513 E514 A515 D516 L517 Q518 R519 Q520 L521 L522 E523 E524 A525 F526 G527 GLY VAL LYS GLU GLU GLY GLY SER HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	105.43Å 105.43Å 138.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.81 – 2.80 30.03 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.3 (19.81-2.80) 98.9 (30.03-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.224 , 0.261 0.229 , 0.266	Depositor DCC
R_{free} test set	3534 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	38.9	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 55.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.044 for h,-k,-l	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	7828	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.73% 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)

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

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	0.44	0/3915	0.59	0/5292
1	B	0.44	0/3904	0.59	0/5277
All	All	0.44	0/7819	0.59	0/10569

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	3843	0	3743	145	0
1	B	3832	0	3732	140	0
2	A	10	0	0	0	0
3	A	86	0	0	4	0
3	B	57	0	0	2	0
All	All	7828	0	7475	278	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 18.

The worst 5 of 278 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:B:287:THR:HG22	1:B:289:ASN:H	1.23	1.03
1:A:287:THR:HG22	1:A:289:ASN:H	1.23	1.01
1:B:117:TRP:HB2	3:B:563:HOH:O	1.65	0.95
1:B:219:LEU:HD22	1:B:277:THR:HG23	1.53	0.90
1:A:219:LEU:HD22	1:A:277:THR:HG23	1.57	0.86

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	479/541 (88%)	445 (93%)	27 (6%)	7 (2%)	10	33
1	B	477/541 (88%)	439 (92%)	32 (7%)	6 (1%)	12	36
All	All	956/1082 (88%)	884 (92%)	59 (6%)	13 (1%)	11	34

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	86	LYS
1	A	349	ALA
1	B	86	LYS
1	B	349	ALA
1	A	350	GLN

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	418/451 (93%)	385 (92%)	33 (8%)	12	34
1	B	417/451 (92%)	381 (91%)	36 (9%)	10	30
All	All	835/902 (93%)	766 (92%)	69 (8%)	11	32

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	397	ASP
1	B	427	GLU
1	B	519	ARG
1	A	397	ASP
1	A	392	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	89	GLN
1	B	263	HIS
1	B	94	ASN
1	B	239	ASN
1	B	296	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)

Of 10 ligands modelled in this entry, 10 are unknown - leaving 0 for Mogul analysis.

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	476/541 (87%)	0.23	32 (6%)	17	10	12, 40, 78, 87	0
1	B	474/541 (87%)	0.22	39 (8%)	11	6	13, 40, 79, 88	0
All	All	950/1082 (87%)	0.23	71 (7%)	14	8	12, 40, 79, 88	0

The worst 5 of 71 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	276	GLY	7.6
1	B	511	ALA	7.1
1	A	513	ASP	6.5
1	A	511	ALA	6.1
1	A	474	ASP	5.6

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	UNX	A	1009	1/1	-0.10	0.63	49,49,49,49	0
2	UNX	A	1007	1/1	0.62	0.51	31,31,31,31	0
2	UNX	A	1006	1/1	0.64	0.89	47,47,47,47	0
2	UNX	A	1001	1/1	0.68	0.59	42,42,42,42	0
2	UNX	A	1005	1/1	0.72	0.43	27,27,27,27	0
2	UNX	A	1003	1/1	0.74	1.03	52,52,52,52	0
2	UNX	A	1004	1/1	0.77	0.71	34,34,34,34	0
2	UNX	A	1002	1/1	0.80	0.34	25,25,25,25	0
2	UNX	A	1010	1/1	0.86	0.88	43,43,43,43	0
2	UNX	A	1008	1/1	0.90	0.84	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.