Full wwPDB X-ray Structure Validation Report (i)

Nov 21, 2023 – 01:27 AM JST

PDB ID	:	7EFA
Title	:	Crystal structure of the complex between the C-terminal domain of mouse MUTYH and human PCNA
Authors	:	Nakamura, T.; Nakabeppu, Y.; Yamagata, Y.
Deposited on	:	2021-03-21
Resolution	:	2.70 Å(reported)

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

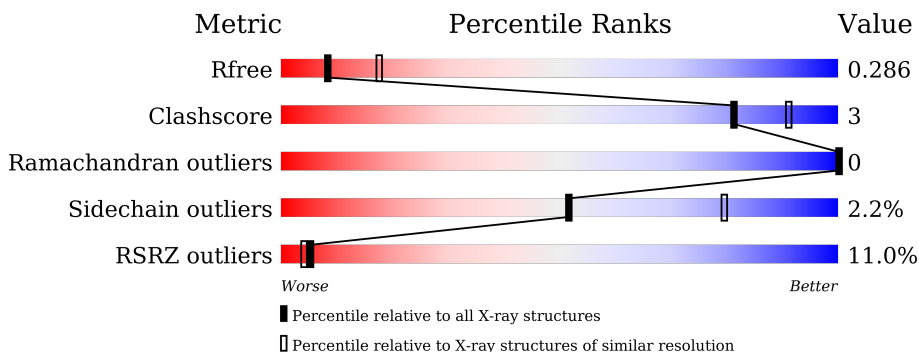|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.70 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
Rfree		0.286
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		2.2%
RSRZ outliers		11.0%

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))
R_{free}	130704	2808 (2.70-2.70)
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 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	261	3% 87% 11% •
2	B	189	20% 74% • 23%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 2852 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 Proliferating cell nuclear antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	256	Total	C	N	O	S	0	0	0
			1968	1236	323	393	16			

- Molecule 2 is a protein called Adenine DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	145	Total	C	N	O	S	0	0	0
			883	553	162	166	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	327	GLY	-	expression tag	UNP Q99P21
B	328	PRO	-	expression tag	UNP Q99P21
B	329	LEU	-	expression tag	UNP Q99P21
B	330	GLY	-	expression tag	UNP Q99P21

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	O	0	0
			1	1		

3 Residue-property plots ⓘ

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: Proliferating cell nuclear antigen

Chain A:
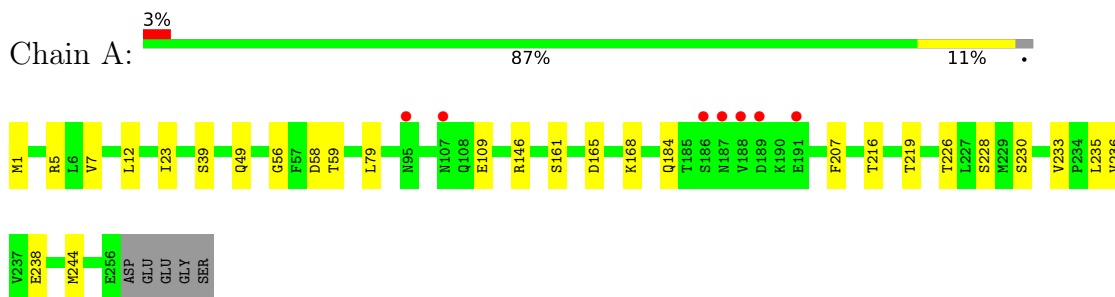
• Molecule 2: Adenine DNA glycosylase

Chain B:
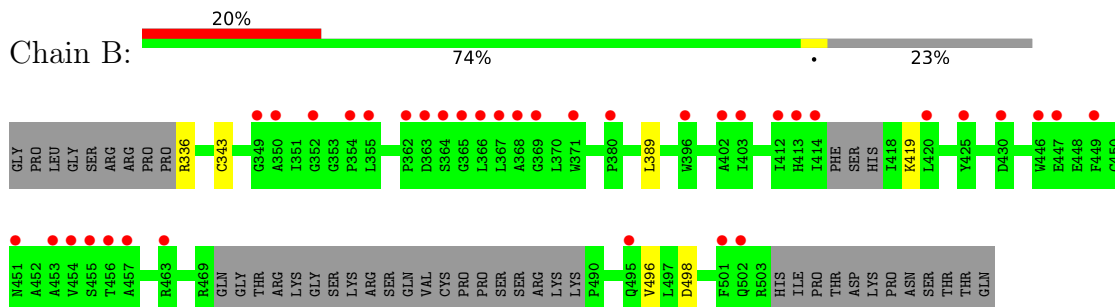
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	87.78Å 87.78Å 124.38Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.40 – 2.70 48.13 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (36.40-2.70) 99.9 (48.13-2.70)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.94 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, R_{free}	0.235 , 0.286 0.235 , 0.286	Depositor DCC
R_{free} test set	749 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	92.8	Xtriage
Anisotropy	0.179	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 85.0	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.063 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2852	wwPDB-VP
Average B, all atoms (Å^2)	122.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.98% 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 ⓘ

5.1   Standard geometry ⓘ

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	0/1994	0.46	0/2694
2	B	0.24	0/903	0.41	0/1250
All	All	0.25	0/2897	0.44	0/3944

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 ⓘ

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	1968	0	1975	11	0
2	B	883	0	620	3	0
3	A	1	0	0	0	0
All	All	2852	0	2595	14	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 3.

All (14) 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:161:SER:HB2	1:A:168:LYS:HB3	1.77	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:496:VAL:HG12	2:B:498:ASP:H	1.64	0.61
1:A:56:GLY:HA3	1:A:244:MET:HG2	1.83	0.61
2:B:343:CYS:HB2	2:B:389:LEU:HD21	1.85	0.58
1:A:228:SER:HB2	1:A:236:VAL:HB	1.93	0.51
1:A:7:VAL:HG23	1:A:58:ASP:HB2	1.97	0.46
1:A:226:THR:OG1	1:A:238:GLU:HB3	2.16	0.46
2:B:336:ARG:O	2:B:419:LYS:N	2.42	0.44
1:A:207:PHE:CZ	1:A:235:LEU:HB2	2.53	0.43
1:A:5:ARG:HB3	1:A:59:THR:HB	2.01	0.42
1:A:23:ILE:HD13	1:A:39:SER:HB3	2.01	0.41
1:A:216:THR:O	1:A:219:THR:OG1	2.27	0.41
1:A:230:SER:HB2	1:A:233:VAL:HB	2.03	0.41
1:A:12:LEU:HG	1:A:79:LEU:HD11	2.04	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	254/261 (97%)	237 (93%)	17 (7%)	0	100	100
2	B	139/189 (74%)	124 (89%)	15 (11%)	0	100	100
All	All	393/450 (87%)	361 (92%)	32 (8%)	0	100	100

There are no Ramachandran outliers to report.

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	223/228 (98%)	217 (97%)	6 (3%)	44	74
2	B	47/163 (29%)	47 (100%)	0	100	100
All	All	270/391 (69%)	264 (98%)	6 (2%)	52	79

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1	MET
1	A	49	GLN
1	A	109	GLU
1	A	146	ARG
1	A	165	ASP
1	A	184	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	256/261 (98%)	0.27	7 (2%)	54	55	63, 95, 168, 266	0
2	B	145/189 (76%)	1.17	37 (25%)	0	0	106, 165, 213, 249	0
All	All	401/450 (89%)	0.59	44 (10%)	5	4	63, 117, 204, 266	0

All (44) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	364	SER	10.3
1	A	188	VAL	8.3
1	A	187	ASN	7.4
2	B	454	VAL	6.8
2	B	412	ILE	6.6
2	B	350	ALA	6.4
2	B	414	ILE	6.3
1	A	186	SER	6.2
2	B	369	GLY	5.8
2	B	366	LEU	5.2
2	B	371	TRP	5.1
2	B	352	GLY	5.0
2	B	362	PRO	4.8
2	B	354	PRO	4.2
2	B	501	PHE	4.0
2	B	457	ALA	4.0
1	A	189	ASP	3.6
2	B	447	GLU	3.5
2	B	446	TRP	3.4
2	B	502	GLN	3.4
2	B	367	LEU	3.4
2	B	380	PRO	3.3
2	B	449	PHE	3.3
2	B	365	GLY	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	403	ILE	3.1
2	B	402	ALA	3.1
2	B	413	HIS	3.1
2	B	420	LEU	3.0
2	B	463	ARG	3.0
2	B	456	THR	3.0
2	B	368	ALA	2.9
1	A	107	ASN	2.8
2	B	349	GLY	2.8
2	B	396	TRP	2.7
2	B	363	ASP	2.7
1	A	191	GLU	2.6
2	B	453	ALA	2.5
2	B	455	SER	2.5
1	A	95	ASN	2.3
2	B	425	TYR	2.3
2	B	495	GLN	2.2
2	B	430	ASP	2.1
2	B	451	ASN	2.1
2	B	355	LEU	2.1

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.