Full wwPDB X-ray Structure Validation Report (i)

Aug 6, 2023 – 12:45 AM EDT

PDB ID	:	1KG5
Title	:	Crystal structure of the K142Q mutant of E.coli MutY (core fragment)
Authors	:	Gilboa, R.; Kilshtein, A.; Zharkov, D.O.; Kycia, J.H.; Gerchman, S.E.; Grollman, A.P.; Shoham, G.
Deposited on	:	2001-11-26
Resolution	:	1.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i)) were used in the production of this report:

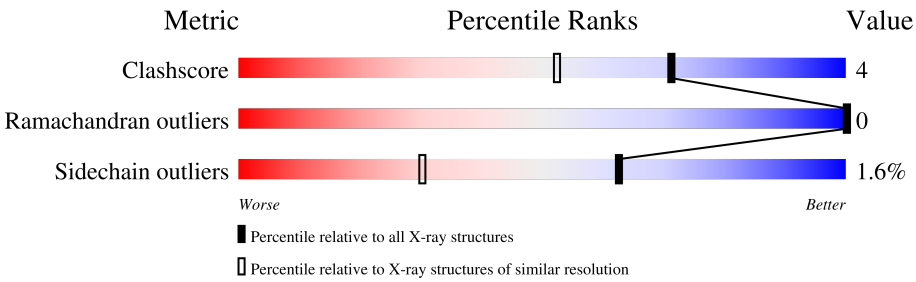
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 1.35 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))
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	225	92% 7% •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2137 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.

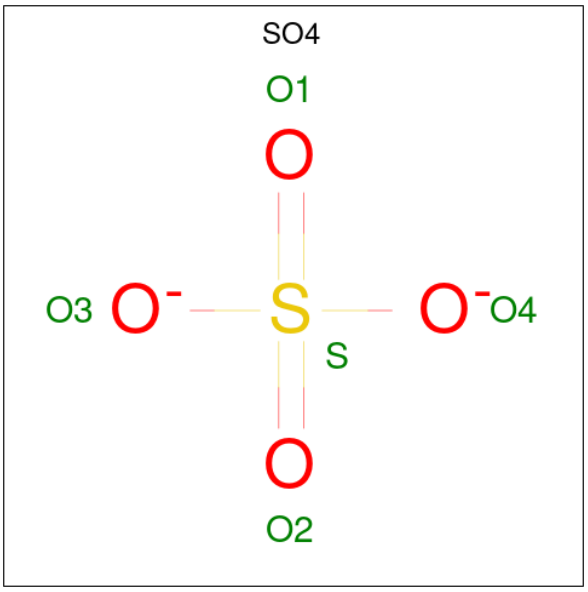
- Molecule 1 is a protein called A/G-specific adenine glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	225	1791	1151	308	321	11	0	9	0

There is a discrepancy between the modelled and reference sequences:

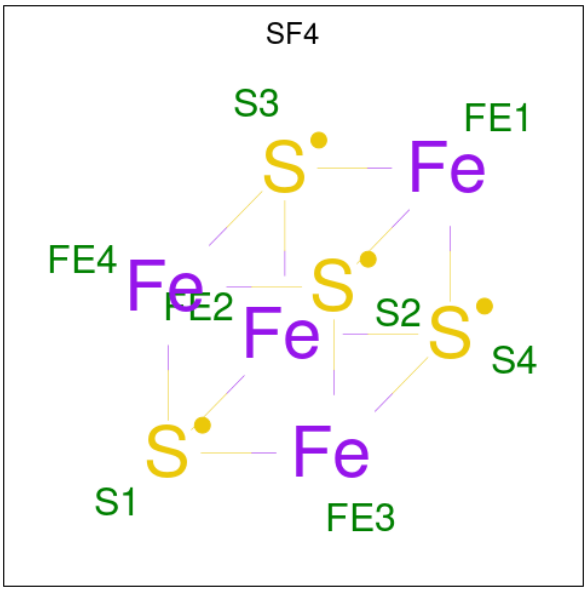
Chain	Residue	Modelled	Actual	Comment	Reference
A	142	GLN	LYS	engineered mutation	UNP P17802

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



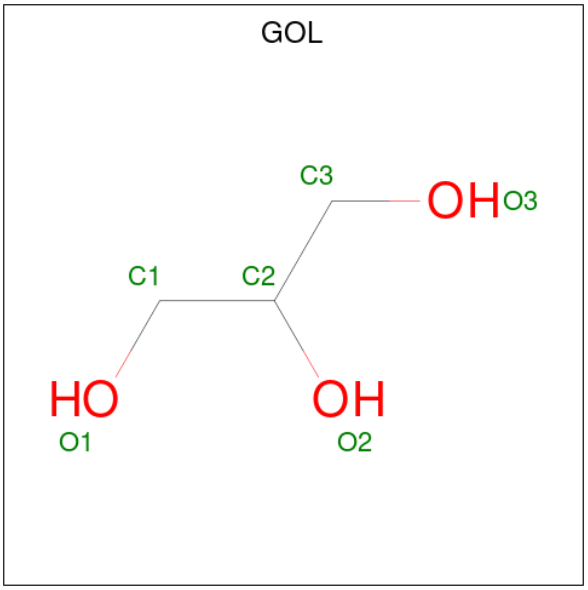
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	0	0

- Molecule 3 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Fe S 8 4 4	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0
4	A	1	Total C O 6 3 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	6	3	3	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	309	309	309	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. 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. 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.

Note EDS was not executed.

- Molecule 1: A/G-specific adenine glycosylase

Chain A:

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	83.49Å 50.12Å 70.38Å 90.00° 122.68° 90.00°	Depositor
Resolution (Å)	40.00 – 1.35	Depositor
% Data completeness (in resolution range)	91.7 (40.00-1.35)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
Refinement program	SHELXL-97, CNS	Depositor
R, R_{free}	0.128 , 0.169	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2137	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

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: GOL, SF4, SO4

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	0.79	0/1874	1.14	16/2545 (0.6%)

There are no bond length outliers.

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	194	ARG	NE-CZ-NH1	17.35	128.98	120.30
1	A	194	ARG	NE-CZ-NH2	-8.83	115.89	120.30
1	A	179	ARG	CD-NE-CZ	8.24	135.13	123.60
1	A	194	ARG	CD-NE-CZ	8.21	135.09	123.60
1	A	104	PHE	CB-CG-CD1	-7.31	115.69	120.80
1	A	53	GLU	OE1-CD-OE2	-7.19	114.67	123.30
1	A	194	ARG	CB-CG-CD	6.90	129.53	111.60
1	A	219	TYR	CB-CG-CD2	6.29	124.77	121.00
1	A	143	ARG	CD-NE-CZ	6.02	132.03	123.60
1	A	104	PHE	CB-CG-CD2	5.96	124.98	120.80
1	A	54	ARG	NE-CZ-NH1	5.90	123.25	120.30
1	A	179	ARG	NE-CZ-NH1	5.84	123.22	120.30
1	A	216	TRP	CH2-CZ2-CE2	5.67	123.07	117.40
1	A	178	GLU	OE1-CD-OE2	5.66	130.09	123.30
1	A	143	ARG	NE-CZ-NH1	5.44	123.02	120.30
1	A	108	PHE	CB-CG-CD1	-5.04	117.27	120.80

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	1791	0	1790	13	0
2	A	5	0	0	0	0
3	A	8	0	0	0	0
4	A	24	0	32	0	0
5	A	309	0	0	2	0
All	All	2137	0	1822	13	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 4.

All (13) 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:194:ARG:HH11	1:A:194:ARG:HG2	1.49	0.77
1:A:158:LYS:HZ2	1:A:159:GLU:H	1.32	0.74
1:A:72:GLU:HG3	5:A:696:HOH:O	1.90	0.71
1:A:109[B]:GLU:HG2	5:A:625:HOH:O	1.99	0.62
1:A:149:TYR:CD2	1:A:167[B]:LEU:HD21	2.38	0.59
1:A:158:LYS:HG2	1:A:159:GLU:N	2.19	0.57
1:A:16[A]:LYS:HD2	1:A:17:TYR:CE2	2.40	0.57
1:A:194:ARG:CD	1:A:194:ARG:H	2.18	0.56
1:A:194:ARG:HG2	1:A:194:ARG:NH1	2.14	0.51
1:A:158:LYS:HE3	1:A:158:LYS:H	1.77	0.50
1:A:194:ARG:H	1:A:194:ARG:HD3	1.83	0.43
1:A:158:LYS:H	1:A:158:LYS:CD	2.32	0.43
1:A:158:LYS:HE3	1:A:158:LYS:N	2.34	0.43

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	232/225 (103%)	231 (100%)	1 (0%)	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	194/185 (105%)	191 (98%)	3 (2%)	65	33

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	104	PHE
1	A	158	LYS
1	A	194	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	5	GLN
1	A	42	GLN
1	A	88	ASN
1	A	95	GLN
1	A	205	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)

6 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	$\#\|Z\| > 2$
4	GOL	A	303	-	5,5,5	0.22	0	5,5,5	0.64	0
4	GOL	A	301	-	5,5,5	0.26	0	5,5,5	1.01	0
3	SF4	A	300	1	0,12,12	-	-	-		
4	GOL	A	302	-	5,5,5	0.49	0	5,5,5	0.67	0
2	SO4	A	310	-	4,4,4	0.10	0	6,6,6	0.31	0
4	GOL	A	304	-	5,5,5	0.60	0	5,5,5	0.30	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
4	GOL	A	303	-	-	0/4/4/4	-
4	GOL	A	301	-	-	1/4/4/4	-
4	GOL	A	302	-	-	0/4/4/4	-
3	SF4	A	300	1	-	-	0/6/5/5
4	GOL	A	304	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	301	GOL	C1-C2-C3-O3

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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.