Full wwPDB X-ray Structure Validation Report ⓘ

Jun 2, 2025 – 10:05 am BST

PDB ID	:	8RAX / pdb_00008rax
Title	:	The crystal structure of DNA-bound human MutSbeta (MSH2_G674D/MSH3) in the canonical mismatch bound conformation, nucleotide free
Authors	:	Thomsen, M.; Costanzi, E.
Deposited on	:	2023-12-01
Resolution	:	3.16 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 ⓘ 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:

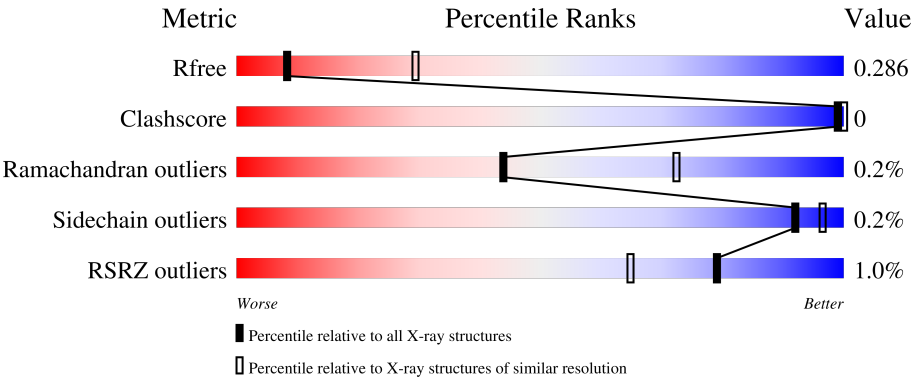
MolProbity	:	4-5-2 with Phenix2.0rc1
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 3.16 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}	164625	2168 (3.20-3.12)
Clashscore	180529	2333 (3.20-3.12)
Ramachandran outliers	177936	2266 (3.20-3.12)
Sidechain outliers	177891	2265 (3.20-3.12)
RSRZ outliers	164620	2169 (3.20-3.12)

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	934	% 86% · 13%
2	B	918	% 89% · 9%
3	C	24	75% 25%
4	D	24	92% 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13983 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 DNA mismatch repair protein Msh2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	813	Total	C	N	O	S	477	0	0
			6454	4111	1096	1215	32			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	674	ASP	GLY	engineered mutation	UNP P43246

- Molecule 2 is a protein called DNA mismatch repair protein Msh3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	831	Total	C	N	O	S	320	3	0
			6678	4262	1139	1245	32			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	217	GLY	-	expression tag	UNP P20585
B	218	PRO	-	expression tag	UNP P20585
B	949	ARG	GLN	conflict	UNP P20585

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*TP*CP*TP*GP*AP*AP*GP*CP*CP *GP*AP*TP*CP*GP*AP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	18	Total	C	N	O	P	10	0	0
			371	176	70	107	18			

- Molecule 4 is a DNA chain called DNA (5'-D(*TP*CP*AP*TP*CP*GP*AP*TP*CP*GP* CP*AP*GP*CP*TP*TP*CP*AP*GP*AP*TP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	22	Total	C	N	O	P	0	0	0
			445	214	80	130	21			

- Molecule 5 is water.

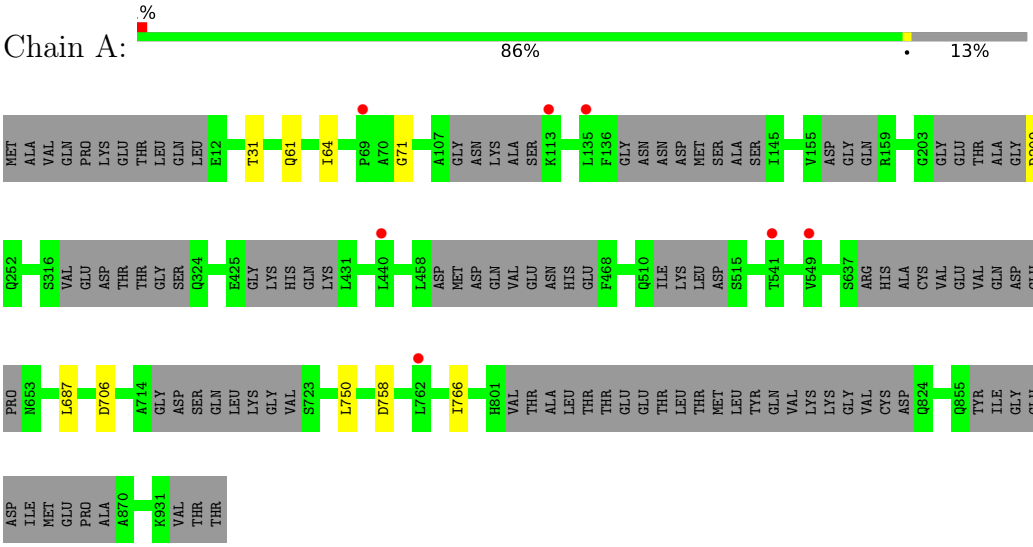
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	17	Total	O	0	0
			17	17		
5	B	13	Total	O	0	0
			13	13		
5	C	3	Total	O	0	0
			3	3		
5	D	2	Total	O	0	0
			2	2		

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.

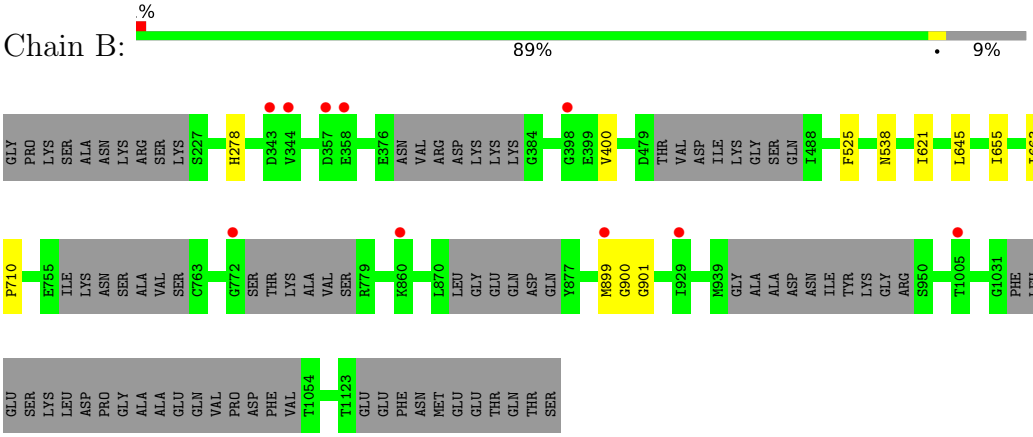
• Molecule 1: DNA mismatch repair protein Msh2

Chain A:



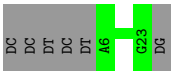
• Molecule 2: DNA mismatch repair protein Msh3

Chain B:



• Molecule 3: DNA (5'-D(P*AP*TP*CP*TP*GP*AP*AP*GP*CP*CP*GP*AP*TP*CP*GP*AP*TP*G)-3')

Chain C:



• Molecule 4: DNA (5'-D(*TP*CP*AP*TP*CP*GP*AP*TP*CP*GP*CP*AP*GP*CP*TP*TP *CP*AP*GP*AP*TP*A)-3')

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	66.90Å 91.00Å 96.16Å 67.68° 86.73° 73.86°	Depositor
Resolution (Å)	75.93 – 3.16 75.93 – 3.16	Depositor EDS
% Data completeness (in resolution range)	90.7 (75.93-3.16) 90.7 (75.93-3.16)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.57 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.257 , 0.292 0.255 , 0.286	Depositor DCC
R_{free} test set	1696 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	89.3	Xtriage
Anisotropy	0.496	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 76.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	13983	wwPDB-VP
Average B, all atoms (Å^2)	118.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.51% 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	1.05	0/6548	1.59	3/8805 (0.0%)
2	B	1.06	0/6798	1.57	6/9174 (0.1%)
3	C	0.30	0/416	0.73	0/640
4	D	0.30	0/498	0.68	0/766
All	All	1.02	0/14260	1.53	9/19385 (0.0%)

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	655	ILE	CA-C-N	6.09	124.08	120.24
2	B	655	ILE	C-N-CA	6.09	124.08	120.24
1	A	209	ASP	CA-CB-CG	5.62	118.22	112.60
2	B	538	ASN	CA-C-N	5.30	125.86	119.98
2	B	538	ASN	C-N-CA	5.30	125.86	119.98
1	A	758	ASP	CA-C-N	5.26	125.81	119.98
1	A	758	ASP	C-N-CA	5.26	125.81	119.98
2	B	663	ILE	CA-C-N	5.14	127.17	120.28
2	B	663	ILE	C-N-CA	5.14	127.17	120.28

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	6454	0	6533	2	0
2	B	6678	0	6775	6	0
3	C	371	0	203	0	0
4	D	445	0	250	0	0
5	A	17	0	0	0	0
5	B	13	0	0	0	0
5	C	3	0	0	0	0
5	D	2	0	0	0	0
All	All	13983	0	13761	8	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 0.

All (8) 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:750:LEU:HD21	1:A:766:ILE:HD13	1.96	0.48
2:B:400:VAL:HG11	2:B:525:PHE:CE1	2.51	0.46
2:B:899[A]:MET:HG2	2:B:900[A]:GLY:N	2.30	0.46
2:B:400:VAL:HG11	2:B:525:PHE:CZ	2.51	0.45
1:A:61:GLN:HG3	1:A:64:ILE:HD12	2.01	0.43
2:B:680:SER:N	2:B:681:PRO:CD	2.83	0.41
2:B:899[A]:MET:HG2	2:B:901[A]:GLY:H	1.85	0.41
2:B:621:ILE:HG21	2:B:645:LEU:HD21	2.04	0.40

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	785/934 (84%)	741 (94%)	43 (6%)	1 (0%)	48 77
2	B	818/918 (89%)	771 (94%)	45 (6%)	2 (0%)	44 72

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1603/1852 (87%)	1512 (94%)	88 (6%)	3 (0%)	44 72

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	278	HIS
2	B	710	PRO
1	A	71	GLY

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	709/809 (88%)	706 (100%)	3 (0%)	89 94
2	B	745/818 (91%)	745 (100%)	0	100 100
All	All	1454/1627 (89%)	1451 (100%)	3 (0%)	92 96

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

Mol	Chain	Res	Type
1	A	31	THR
1	A	687	LEU
1	A	706	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (17) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	74	ASN
1	A	183	GLN
1	A	263	ASN
1	A	377	GLN
1	A	397	GLN
1	A	400	ASN
1	A	653	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	885	GLN
2	B	394	GLN
2	B	467	HIS
2	B	524	ASN
2	B	550	ASN
2	B	591	ASN
2	B	880	ASN
2	B	908	GLN
2	B	984	HIS
2	B	1027	ASN

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 oligosaccharides 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

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, 95^{th} 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	813/934 (87%)	-0.14	7 (0%) 81 67	45, 107, 160, 221	147 (18%)
2	B	831/918 (90%)	-0.22	10 (1%) 76 60	44, 104, 155, 217	107 (12%)
3	C	18/24 (75%)	-0.14	0 100 100	102, 136, 216, 233	1 (5%)
4	D	22/24 (91%)	-0.28	0 100 100	93, 140, 221, 250	0
All	All	1684/1900 (88%)	-0.18	17 (1%) 79 65	44, 106, 160, 250	255 (15%)

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	899[A]	MET	3.5
1	A	762	LEU	2.8
2	B	344	VAL	2.6
2	B	398	GLY	2.5
1	A	541	THR	2.5
2	B	358	GLU	2.4
1	A	135	LEU	2.4
1	A	69	PRO	2.4
1	A	549	VAL	2.3
2	B	343	ASP	2.2
2	B	860	LYS	2.2
2	B	1005	THR	2.2
1	A	113	LYS	2.2
2	B	929	ILE	2.1
2	B	772	GLY	2.1
1	A	440	LEU	2.0
2	B	357	ASP	2.0

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 (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.