



wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 07:50 PM GMT

PDB ID : 2O8F
Title : human MutSalpha (MSH2/MSH6) bound to DNA with a single base T insert
Authors : Warren, J.J.; Pohlhaus, T.J.; Changela, A.; Modrich, P.L.; Beese, L.S.
Deposited on : 2006-12-12
Resolution : 3.25 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

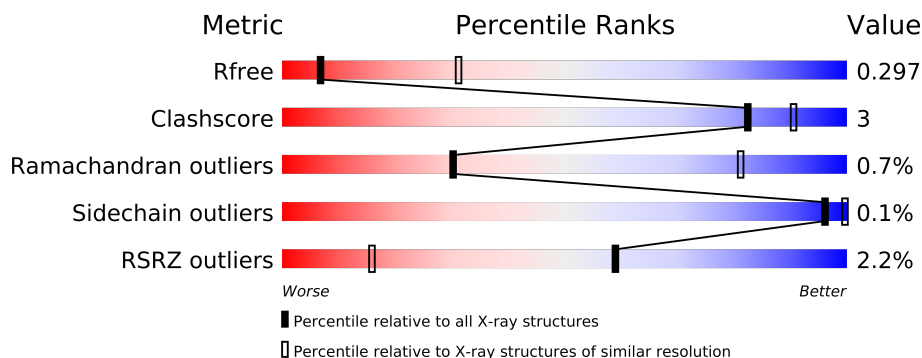
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance





The reported resolution of this entry is 3.25 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}	66092	1085 (3.32-3.20)
Clashscore	79885	1374 (3.32-3.20)
Ramachandran outliers	78287	1348 (3.32-3.20)
Sidechain outliers	78261	1346 (3.32-3.20)
RSRZ outliers	66119	1086 (3.32-3.20)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	E	17	
2	F	18	
3	A	934	
4	B	1022	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 14708 atoms, of which 0 are hydrogen and 0 are deuterium.

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	17	Total	C	N	O	P	0	0	0
			346	163	68	99	16			

- Molecule 2 is a DNA chain called 5'-D(*CP*GP*CP*TP*AP*GP*CP*GP*TP*GP*CP*G
P*GP*CP*CP*GP*TP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	18	Total	C	N	O	P	0	0	0
			365	173	67	108	17			

- Molecule 3 is a protein called DNA mismatch repair protein Msh2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	832	Total	C	N	O	S	0	0	0
			6494	4119	1103	1238	34			

- Molecule 4 is a protein called DNA mismatch repair protein MSH6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	933	Total	C	N	O	S	0	0	0
			7456	4730	1279	1396	51			

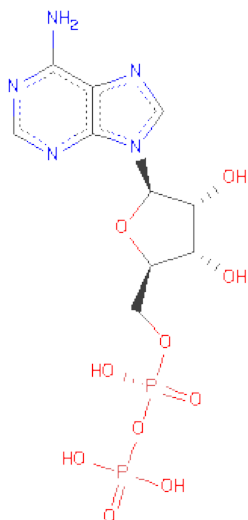
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	339	MET	-	INITIATING METHIONINE	UNP P52701
B	340	GLY	-	CLONING ARTIFACT	UNP P52701

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		

- Molecule 6 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 7 is water.

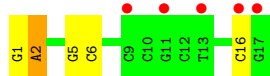
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	O	0	0
			1	1		
7	B	17	Total	O	0	0
			17	17		
7	E	1	Total	O	0	0
			1	1		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

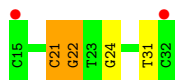
- Molecule 1: 5'-D(*GP*AP*CP*GP*GP*CP*CP*GP*CP*CP*GP*CP*TP*AP*GP*CP*G)-3',

Chain E: 



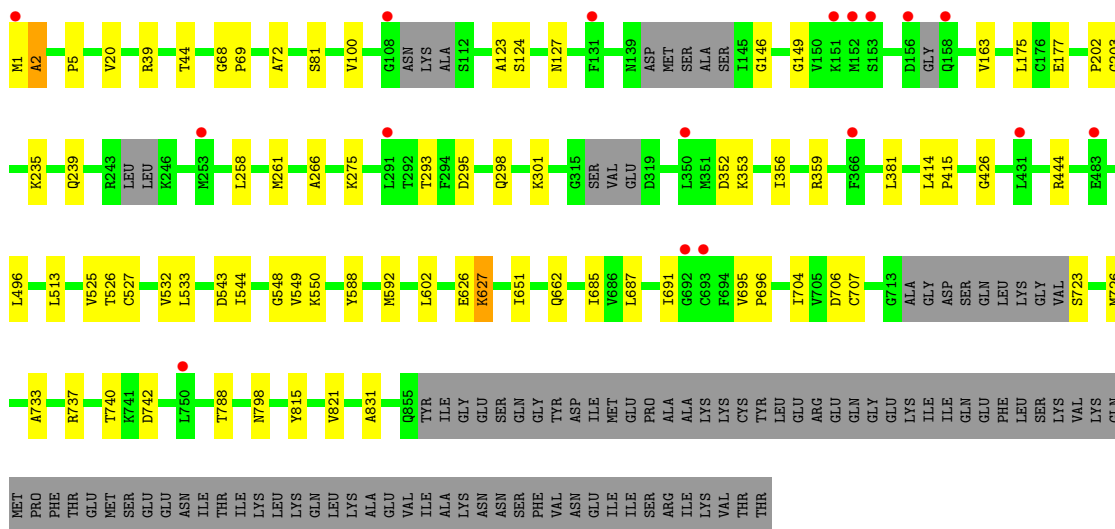
- Molecule 2: 5'-D(*CP*GP*CP*TP*AP*GP*CP*GP*TP*GP*CP*GP*GP*CP*CP*GP*TP*C)-3'

Chain F: 



- Molecule 3: DNA mismatch repair protein Msh2

Chain A: 



- Molecule 4: DNA mismatch repair protein MSH6

Chain B: 

GLU	CYS	THR	LYS	T757	P531	MET
CYS	GLU	PHE	THR	L758	GLY	GLY
ASP	GLU	PHE	PHE	I795	K543	SER
PRO	GLN	GLY	D1106	E796	S550	ALA
SER	GLN	SER	L798	D797	GLY	PRO
GLU	THR	I1109	I1109	L798	H552	ASN
THR	ILE	P1110	P1110	I805	T553	SER
ILE	T1284	I1113	I1113	K833	R554	GLU
F1285	F1285	N1124	N1124	V558	GLN	ALA
L1308	L1308	GLY	GLY	H837	VAL	HIS
P1309	P1309	K1126	K1126	P838	Q572	VAL
L1332	L1332	A1127	A1127	F882	D576	SER
E1335	E1335	P1135	P1135	L887	L585	GLY
VAL	VAL	R1145	R1145	I891	V586	GLY
CYS	CYS	L1149	L1149	K930	P590	ASP
LEU	LEU	M1153	M1153	F933	P591	SER
ALA	ALA	V1160	V1160	ASP	E604	ARG
SER	SER	P1161	P1161	S935	L613	P362
GLU	GLU	L1177	L1177	D936	L647	D380
ARG	ARG	G1178	G1178	ALA	I651	E381
VAL	VAL	ALA	ALA	L955	GLY	R384
ASP	ASP	SER	SER	E956	V653	T395
ALA	ALA	ASP	ASP	K957	E665	L396
HIS	HIS	ASP	ASP	Y969	S866	Y397
LYS	LYS	ILE	ILE	ARG	A680	P408
LEU	LEU	MET	MET	ARG	E699	K412
LEU	LEU	SER	SER	THR	L981	T416
THR	THR	GLY	GLY	P982	P882	I416
ILE	ILE	GLU	GLU	P991	T716	H437
LYS	LYS	S1188	S1188	GLU	V717	T716
GLU	GLU	E993	E993	THR	S718	A440
LEU	LEU	E1193	E1193	THR	THR	L447
		L1194	L1194	ARG	ARG	L447
		L1209	L1209	SER	THR	N456
		L1235	L1235	GLY	GLY	W456
		I1239	I1239	ALA	ALA	A457
		R1263	R1263	ILE	ILE	D471
		G1269	G1269	PHE	PHE	S472
		N1270	N1270	THR	THR	L473
		VAL	VAL	K728	K728	V474
		ASN	ASN	N742	N742	Q475
		GLU	GLU	L1091	L1091	K476
		ASN	ASN	S1094	S1094	Y505
				L747	L747	Y524
				T1100	T1100	

4 Data and refinement statistics

Property	Value	Source
Space group	P 43 3 2	Depositor
Cell constants a, b, c, α , β , γ	259.55Å 259.55Å 259.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.25 82.08 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.6 (20.00-3.25) 99.5 (82.08-3.25)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	0.13	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.47 (at 3.26Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.243 , 0.293 0.253 , 0.297	Depositor DCC
R_{free} test set	2382 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	85.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 101.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.24$	Xtriage
Outliers	0 of 47211 reflections	Xtriage
F_o, F_c correlation	0.83	EDS
Total number of atoms	14708	wwPDB-VP
Average B, all atoms (Å ²)	111.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.32% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

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

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	E	0.61	0/388	1.32	5/597 (0.8%)
2	F	0.61	0/408	1.46	6/628 (1.0%)
3	A	0.30	0/6595	0.56	0/8898
4	B	0.32	0/7600	0.58	0/10244
All	All	0.33	0/14991	0.65	11/20367 (0.1%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	22	DG	O3'-P-O5'	-11.73	81.71	104.00
2	F	22	DG	OP1-P-O3'	-7.96	87.68	105.20
2	F	22	DG	OP2-P-O3'	-7.58	88.53	105.20
1	E	2	DA	P-O3'-C3'	7.20	128.34	119.70
1	E	1	DG	P-O3'-C3'	7.11	128.23	119.70

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	346	0	190	1	0
2	F	365	0	203	2	0
3	A	6494	0	6491	43	0
4	B	7456	0	7431	48	0
5	A	1	0	0	0	0
6	A	27	0	12	0	0
7	A	1	0	0	0	0
7	B	17	0	0	0	0
7	E	1	0	0	0	0
All	All	14708	0	14327	90	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

The worst 5 of 90 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:A:39:ARG:HE	3:A:44:THR:HG21	1.52	0.74
4:B:936:ASP:HA	4:B:939:GLN:HB3	1.70	0.73
4:B:747:LEU:HA	4:B:757:THR:HG21	1.70	0.73
4:B:554:ARG:HH22	4:B:604:GLU:HG3	1.55	0.72
4:B:699:GLU:HG2	4:B:833:LYS:HD2	1.75	0.67

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	
3	A	818/934 (88%)	723 (88%)	86 (10%)	9 (1%)	21	74
4	B	913/1022 (89%)	830 (91%)	80 (9%)	3 (0%)	50	92
All	All	1731/1956 (88%)	1553 (90%)	166 (10%)	12 (1%)	30	81

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	533	LEU
3	A	526	THR
3	A	627	LYS
3	A	2	ALA
4	B	930	LYS

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	
3	A	704/808 (87%)	704 (100%)	0	100	100
4	B	821/899 (91%)	819 (100%)	2 (0%)	96	99
All	All	1525/1707 (89%)	1523 (100%)	2 (0%)	96	99

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	576	ASP
4	B	647	LEU

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

Mol	Chain	Res	Type
3	A	798	ASN

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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	$\# Z > 2$
6	ADP	A	936	5	29,29,29	1.06	2 (6%)	45,45,45	1.81	8 (17%)

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
6	ADP	A	936	5	-	0/16/32/32	0/1/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	936	ADP	C5-C4	3.29	1.47	1.40
6	A	936	ADP	C4-N9	-2.12	1.34	1.37

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	936	ADP	N3-C2-N1	-6.18	123.54	128.71
6	A	936	ADP	N3-C4-N9	5.55	135.45	125.43
6	A	936	ADP	C5-C4-N3	-3.37	118.36	125.70
6	A	936	ADP	PA-O3A-PB	-3.29	122.03	131.68
6	A	936	ADP	C4-C5-N7	-3.20	106.78	109.52

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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, 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	E	17/17 (100%)	1.49	5 (29%)	1 1	105, 113, 127, 132	0
2	F	18/18 (100%)	0.72	2 (11%)	6 2	102, 112, 118, 118	0
3	A	832/934 (89%)	0.42	17 (2%)	62 18	38, 111, 114, 120	0
4	B	933/1022 (91%)	0.46	16 (1%)	67 21	38, 110, 121, 126	0
All	All	1800/1991 (90%)	0.45	40 (2%)	59 16	38, 111, 118, 132	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	17	DG	4.6
4	B	651	ILE	4.0
3	A	152	MET	3.4
3	A	291	LEU	3.1
3	A	1	MET	3.0

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 ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
5	MG	A	935	1/1	0.32	1.12	92,92,92,92	0
6	ADP	A	936	27/27	0.23	-1.02	110,110,111,111	0

6.5 Other polymers ⓘ

There are no such residues in this entry.