



wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 05:55 AM GMT

PDB ID : 2O8D
Title : human MutSalpha (MSH2/MSH6) bound to ADP and a G dU mispair
Authors : Warren, J.J.; Pohlhaus, T.J.; Changela, A.; Modrich, P.L.; Beese, L.S.
Deposited on : 2006-12-12
Resolution : 3.00 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>

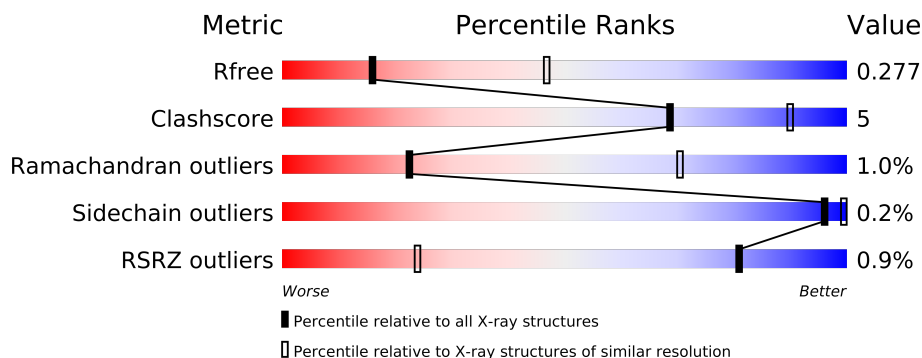
The following versions of software and data (see [references](#)) were used in the production of this report:

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 3.00 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	1216 (3.00-3.00)
Clashscore	79885	1594 (3.00-3.00)
Ramachandran outliers	78287	1537 (3.00-3.00)
Sidechain outliers	78261	1540 (3.00-3.00)
RSRZ outliers	66119	1217 (3.00-3.00)

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	15	
2	F	15	
3	A	934	
4	B	1022	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
5	MG	A	935	-	X

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 14593 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*AP*CP*CP*GP*CP*GP*CP*GP*CP*T
P*AP*GP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	15	Total	C	N	O	P	0	0	0
			307	145	62	86	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	15	Total	C	N	O	P	0	0	0
			302	144	53	91	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	830	Total	C	N	O	S	0	0	0
			6439	4085	1092	1228	34			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	932	Total	C	N	O	S	0	0	0
			7443	4721	1277	1394	51			

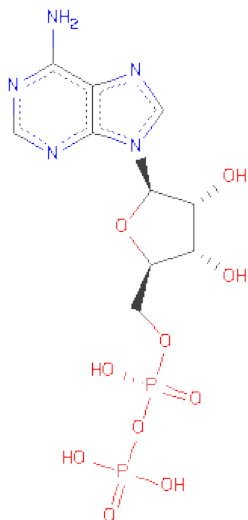
There are 2 discrepancies between the modelled and reference sequences:

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	B	1	Total Mg 1 1	0	0
5	A	1	Total Mg 1 1	0	0

- 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 27 10 5 10 2	0	0
6	B	1	Total C N O P 27 10 5 10 2	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	2	Total O 2 2	0	0
7	B	40	Total O 40 40	0	0
7	E	1	Total O 1 1	0	0
7	F	3	Total O 3 3	0	0

I1313	R1176	E993	T767	R554	MET
E1335	L1177	K997	K775	D575	GLY
VAL	ALA	R1005	L782	D576	SER
CYS	SER	Y1006	A787	R581	ALA
LEU	ASP	W1007	I788	L585	PRO
ALA	ARG	R1024	I795	V586	GLN
ALA	ILE	Y1038	L798	Q593	ASN
SER	MET	K1042	V801	E604	SER
ARG	GLY	L1061	I805	S611	GLU
THR	SER	A1064	V878	L613	GLN
VAL	GLU	I1079	Q889	E641	ALA
S1188	S1188	L1080	S892	I651	HIS
E1193	E1193	L1081	N897	V653	VAL
D1213	D1213	P1082	E899	V658	VAL
R1217	R1217	F1088	G900	M662	SER
E1234	E1234	G1098	R901	T663	GLY
L1235	L1235	ILE	K920	S664	GLY
T1238	T1238	THR	L926	E685	ASP
T1243	T1243	PHE	I927	S666	ASP
L1244	L1244	G1105	A931	D667	ASP
F1245	F1245	I1109	G932	L395	SER
S1246	S1246	P1110	F933	Y397	SER
M1259	M1259	C1117	ASP	N404	ARG
V1260	V1260	E1118	S935	D380	P362
A1261	A1261	E1119	D936	E381	
V1263	V1263	M1124	Y937	R384	
M1270	M1270	GLY	D938	R385	
VAL	VAL	K1126	A940	P386	
GLU	GLU	A1127	R945	T395	
ASN	ASN	V1130	Q949	L396	
GLU	GLU	M1136	L952	Y397	
CYS	CYS	M1137	E953	N404	
GLU	GLU	L1149	Y954	R411	
ASP	ASP	L1150	V968	K412	
PRO	PRO	M1153	Y969	I416	
SER	SER	A1154	W970	M438	
GLN	GLN	Q1155	L990	E487	
GLU	GLU	P1161	P991	R495	
THR	THR		GLU	C496	
THR	THR			M499	
ILE	ILE			I516	
T1284	T1284			T518	
C1294	C1294			S550	
P1295	P1295			GLY	
K1296	K1296			H552	
S1297	S1297			T553	
L1308	L1308				
P1309	P1309				

4 Data and refinement statistics

Property	Value	Source
Space group	P 43 3 2	Depositor
Cell constants a, b, c, α , β , γ	260.38Å 260.38Å 260.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.00 48.35 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-3.00) 99.8 (48.35-3.00)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.39 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.239 , 0.278 0.243 , 0.277	Depositor DCC
R_{free} test set	3088 reflections (5.38%)	DCC
Wilson B-factor (Å ²)	83.6	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 86.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 60518 reflections (0.002%)	Xtriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	14593	wwPDB-VP
Average B, all atoms (Å ²)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.47% 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	0.65	0/345	1.39	5/531 (0.9%)
2	F	0.66	0/337	1.39	4/518 (0.8%)
3	A	0.33	0/6539	0.61	0/8828
4	B	0.39	0/7587	0.66	1/10226 (0.0%)
All	All	0.38	0/14808	0.70	10/20103 (0.0%)

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	4	DC	O4'-C4'-C3'	-9.00	100.60	106.00
1	E	4	DC	O4'-C1'-N1	8.03	113.62	108.00
2	F	19	DA	P-O3'-C3'	6.97	128.06	119.70
2	F	20	DG	O4'-C1'-N9	6.77	112.74	108.00
1	E	4	DC	C4'-C3'-C2'	-6.12	97.59	103.10

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	307	0	168	1	0
2	F	302	0	169	1	0
3	A	6439	0	6409	65	0
4	B	7443	0	7414	72	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	A	27	0	12	0	0
6	B	27	0	12	0	0
7	A	2	0	0	0	0
7	B	40	0	0	0	0
7	E	1	0	0	0	0
7	F	3	0	0	0	0
All	All	14593	0	14184	137	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 5.

The worst 5 of 137 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:B:897:ASN:HB3	4:B:901:ARG:HE	1.35	0.90
3:A:39:ARG:HE	3:A:44:THR:HG21	1.48	0.77
4:B:897:ASN:HB3	4:B:901:ARG:NE	2.02	0.75
4:B:586:VAL:HG11	4:B:613:LEU:HD11	1.69	0.75
4:B:380:ASP:HB2	4:B:384:ARG:H	1.52	0.74

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	816/934 (87%)	732 (90%)	76 (9%)	8 (1%)	22 70
4	B	912/1022 (89%)	836 (92%)	66 (7%)	10 (1%)	21 67
All	All	1728/1956 (88%)	1568 (91%)	142 (8%)	18 (1%)	22 70

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	B	665	GLU
4	B	747	LEU
4	B	1297	SER
3	A	138	ASN
4	B	668	SER

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	693/808 (86%)	692 (100%)	1 (0%)	96	99
4	B	819/899 (91%)	817 (100%)	2 (0%)	96	99
All	All	1512/1707 (89%)	1509 (100%)	3 (0%)	96	99

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

Mol	Chain	Res	Type
3	A	711	ARG
4	B	937	TYR
4	B	1007	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	560	ASN
4	B	1327	ASN
4	B	751	ASN
3	A	388	ASN
4	B	1124	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 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	1.09	2 (6%)	45,45,45	1.78	8 (17%)
6	ADP	B	202	5	29,29,29	1.05	2 (6%)	45,45,45	1.81	7 (15%)

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
6	ADP	B	202	5	-	0/16/32/32	0/1/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	B	202	ADP	C5-C4	3.27	1.47	1.40
6	A	936	ADP	C5-C4	3.23	1.47	1.40
6	A	936	ADP	C4-N9	-2.63	1.33	1.37
6	B	202	ADP	C4-N9	-2.52	1.34	1.37

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
6	B	202	ADP	N3-C2-N1	-6.48	123.29	128.71
6	A	936	ADP	N3-C2-N1	-6.19	123.53	128.71
6	B	202	ADP	N3-C4-N9	5.22	134.85	125.43
6	A	936	ADP	N3-C4-N9	5.10	134.65	125.43
6	A	936	ADP	C4-C5-N7	-3.33	106.67	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	15/15 (100%)	1.16	4 (26%) 1 1	85, 93, 111, 114	0
2	F	15/15 (100%)	0.28	0 100 100	86, 90, 123, 126	0
3	A	830/934 (88%)	0.21	7 (0%) 83 26	20, 93, 98, 107	0
4	B	932/1022 (91%)	0.24	5 (0%) 88 36	34, 93, 107, 116	0
All	All	1792/1986 (90%)	0.23	16 (0%) 81 24	20, 93, 103, 126	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	15	DG	5.5
4	B	747	LEU	2.6
3	A	467	GLU	2.6
3	A	541	THR	2.6
4	B	954	TYR	2.5

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	0.43	11.92	92,92,92,92	0
5	MG	B	102	1/1	0.28	1.54	83,83,83,83	0
6	ADP	A	936	27/27	0.19	-0.96	89,89,91,91	0
6	ADP	B	202	27/27	0.16	-1.45	99,100,101,101	0

6.5 Other polymers ⓘ

There are no such residues in this entry.