



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 2, 2025 – 10:06 am BST

PDB ID : 8RAV / pdb_00008rav
Title : The crystal structure of DNA-bound human MutSbeta (MSH2/MSH3) in the canonical mismatch bound conformation with ADP bound in MSH2
Authors : Thomsen, M.; Costanzi, E.
Deposited on : 2023-12-01
Resolution : 2.63 Å(reported)

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

MolProbity : 4-5-2 with Phenix2.0rc1
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 2.0rc1
EDS : 3.0
buster-report : 1.1.7 (2018)
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

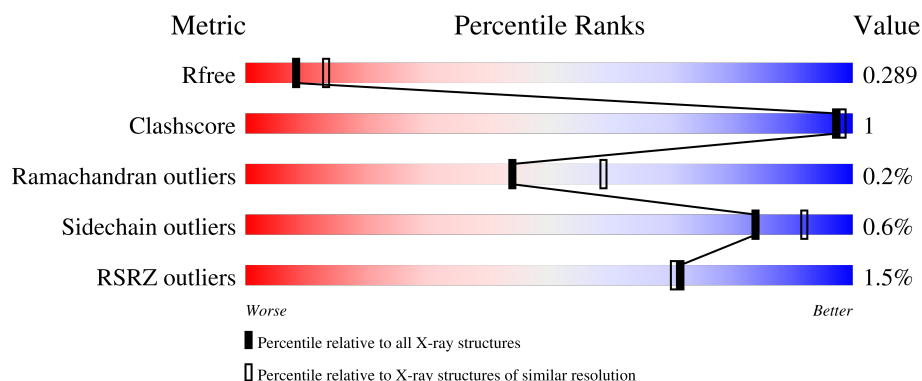
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.63 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	1851 (2.66-2.62)
Clashscore	180529	1953 (2.66-2.62)
Ramachandran outliers	177936	1929 (2.66-2.62)
Sidechain outliers	177891	1929 (2.66-2.62)
RSRZ outliers	164620	1850 (2.66-2.62)

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 $\leq 5\%$. 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	<div> <div>2%</div> <div> <div></div> <div>91%</div> <div>7%</div> </div> </div>
2	B	918	<div> <div>%</div> <div> <div></div> <div>89%</div> <div>8%</div> </div> </div>
3	C	24	<div> <div></div> <div> <div>83%</div> <div>17%</div> </div> </div>
4	D	24	<div> <div></div> <div> <div>92%</div> <div>8%</div> </div> </div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 14707 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	871	Total	C	N	O	S	339	1	0
			6916	4395	1176	1309	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	849	Total	C	N	O	S	203	1	0
			6801	4339	1162	1269	31			

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*CP*TP*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	20	Total	C	N	O	P	0	0	0
			410	195	75	120	20			

- 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*AP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			

- Molecule 5 is ADENOSINE-5'-DIPHOSPHATE (CCD ID: ADP) (formula: C₁₀H₁₅N₅O₁₀P₂) (labeled as "Ligand of Interest" by depositor).



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

- Molecule 6 is MAGNESIUM ION (CCD ID: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

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

- Molecule 7 is SODIUM ION (CCD ID: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	1	Total Na 1 1	0	0

- Molecule 8 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: $\text{C}_2\text{H}_6\text{O}_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	B	1	Total	C	O	0	0
			4	2	2		
8	B	1	Total	C	O	0	0
			4	2	2		
8	B	1	Total	C	O	0	0
			4	2	2		

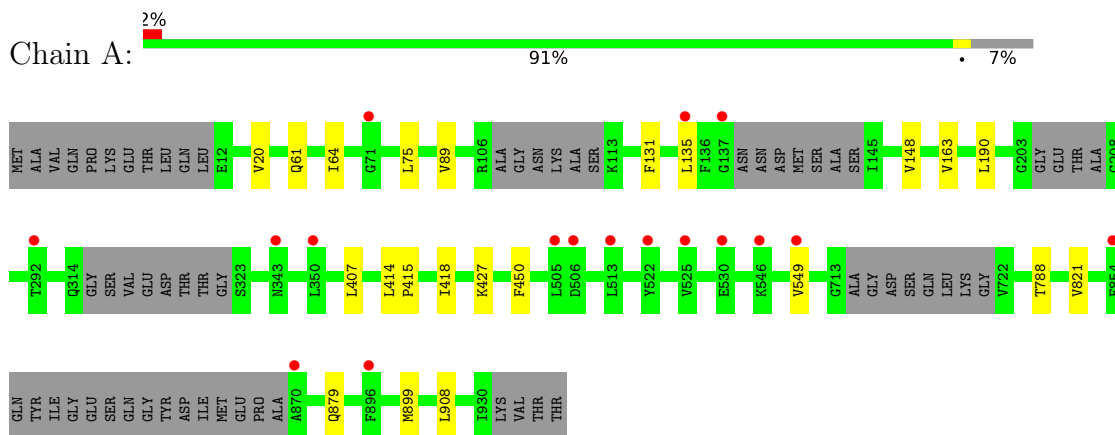
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	28	Total	O	0	0
			28	28		
9	B	16	Total	O	0	0
			16	16		
9	C	4	Total	O	0	0
			4	4		
9	D	2	Total	O	0	0
			2	2		

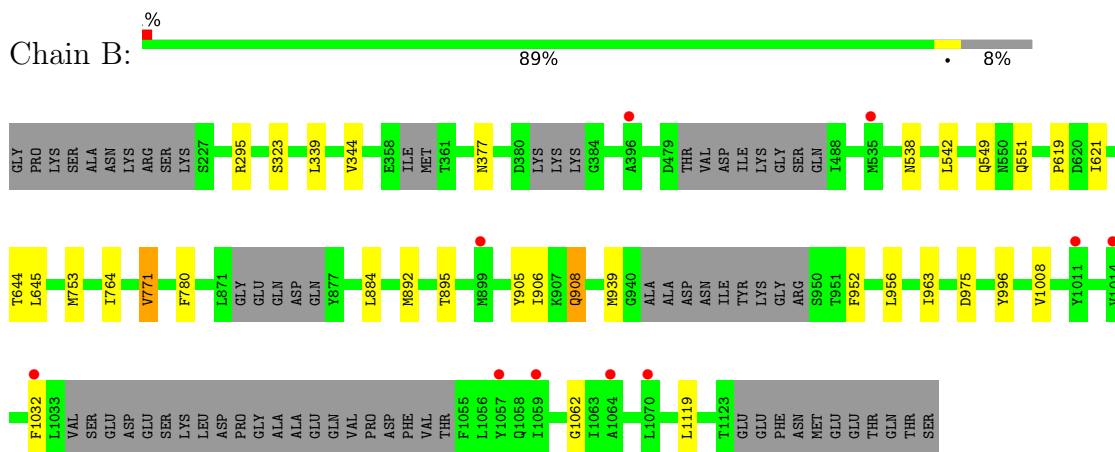
3 Residue-property plots [i](#)

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.

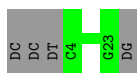
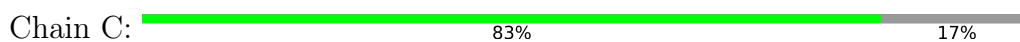
- Molecule 1: DNA mismatch repair protein Msh2



- Molecule 2: DNA mismatch repair protein Msh3

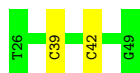


- Molecule 3: DNA (5'-D(P*CP*TP*AP*TP*CP*TP*GP*AP*AP*GP*CP*CP*GP*AP*TP*C
P*GP*AP*TP*G)-3')



- 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*AP*GP*G)-3')

Chain D:  92% 8%



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	66.60Å 91.12Å 96.48Å 67.56° 86.81° 74.64°	Depositor
Resolution (Å)	44.94 – 2.63 44.94 – 2.63	Depositor EDS
% Data completeness (in resolution range)	96.1 (44.94-2.63) 96.1 (44.94-2.63)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.03	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.29 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.239 , 0.284 0.240 , 0.289	Depositor DCC
R_{free} test set	857 reflections (1.44%)	wwPDB-VP
Wilson B-factor (Å ²)	74.4	Xtriage
Anisotropy	0.406	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 62.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	14707	wwPDB-VP
Average B, all atoms (Å ²)	102.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.58% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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: NA, EDO, 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	A	1.06	1/7023 (0.0%)	1.58	2/9457 (0.0%)
2	B	1.05	0/6923	1.57	2/9345 (0.0%)
3	C	0.29	0/459	0.68	0/706
4	D	0.30	0/548	0.71	2/844 (0.2%)
All	All	1.02	1/14953 (0.0%)	1.53	6/20352 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	418	ILE	CB-CG1	-5.87	1.41	1.53

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1032	PHE	CA-CB-CG	8.14	121.94	113.80
4	D	39	DC	C2'-C3'-O3'	6.50	121.25	111.50
4	D	42	DC	C4'-C3'-O3'	5.77	118.65	110.00
2	B	1062	GLY	CA-C-O	-5.37	118.58	122.45
1	A	879	GLN	CA-C-N	5.09	125.73	120.03

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	6916	0	6988	8	0
2	B	6801	0	6907	14	0
3	C	410	0	226	0	0
4	D	489	0	272	0	0
5	A	27	0	12	0	0
6	A	1	0	0	0	0
7	A	1	0	0	0	0
8	B	12	0	18	0	0
9	A	28	0	0	0	0
9	B	16	0	0	0	0
9	C	4	0	0	0	0
9	D	2	0	0	0	0
All	All	14707	0	14423	21	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 1.

The worst 5 of 21 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:788:THR:HG21	1:A:821:VAL:HG23	1.70	0.73
1:A:908:LEU:HD11	2:B:1119:LEU:HD11	1.87	0.56
1:A:61:GLN:HG3	1:A:64:ILE:HD12	1.88	0.54
2:B:621:ILE:HG21	2:B:645:LEU:HD21	1.92	0.52
2:B:952:PHE:CZ	2:B:956:LEU:HD11	2.47	0.50

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	858/934 (92%)	815 (95%)	40 (5%)	3 (0%)	37 51

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	836/918 (91%)	797 (95%)	39 (5%)	0	100	100
All	All	1694/1852 (92%)	1612 (95%)	79 (5%)	3 (0%)	44	59

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	427	LYS
1	A	899	MET
1	A	135	LEU

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	761/808 (94%)	759 (100%)	2 (0%)	91	96
2	B	760/818 (93%)	753 (99%)	7 (1%)	75	87
All	All	1521/1626 (94%)	1512 (99%)	9 (1%)	84	92

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	895	THR
2	B	908	GLN
2	B	377	ASN
2	B	538	ASN
2	B	542	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	791	HIS
2	B	1027	ASN
2	B	1071	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	817	HIS
2	B	288	HIS

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 ⓘ

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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	$\# Z > 2$
5	ADP	A	1001	6	24,29,29	0.64	0	29,45,45	0.73	1 (3%)
8	EDO	B	1201	-	3,3,3	0.10	0	2,2,2	0.31	0
8	EDO	B	1202	-	3,3,3	0.16	0	2,2,2	0.42	0
8	EDO	B	1203	-	3,3,3	0.09	0	2,2,2	0.27	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
5	ADP	A	1001	6	-	0/12/32/32	0/3/3/3
8	EDO	B	1201	-	-	1/1/1/1	-
8	EDO	B	1202	-	-	1/1/1/1	-
8	EDO	B	1203	-	-	0/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
5	A	1001	ADP	C5-C6-N6	2.19	123.68	120.35

There are no chirality outliers.

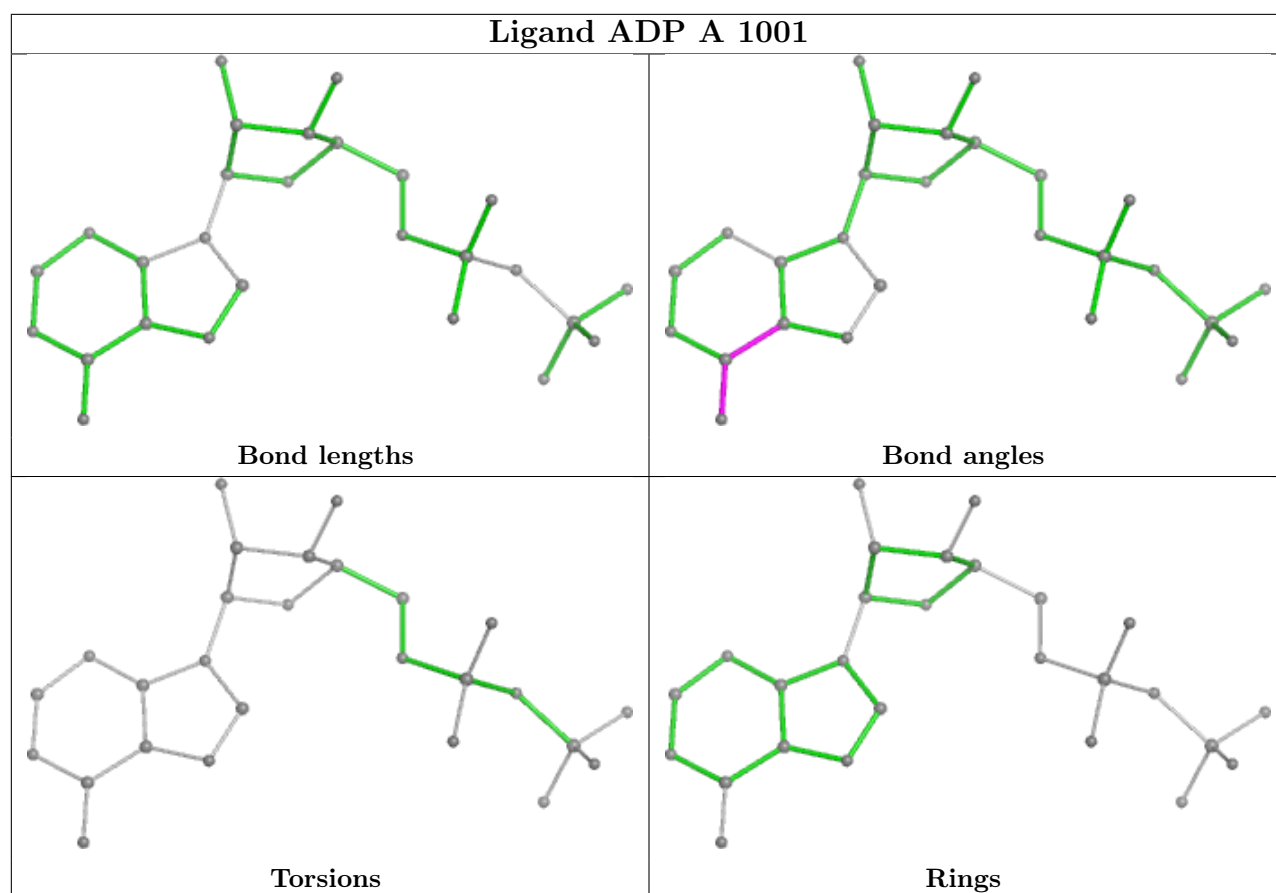
All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	B	1202	EDO	O1-C1-C2-O2
8	B	1201	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



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](#)

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	871/934 (93%)	0.20	17 (1%) 64 63	37, 96, 142, 191	107 (12%)
2	B	849/918 (92%)	0.13	10 (1%) 76 75	45, 86, 139, 192	67 (7%)
3	C	20/24 (83%)	-0.03	0 100 100	74, 125, 308, 336	0
4	D	24/24 (100%)	-0.08	0 100 100	73, 113, 236, 294	0
All	All	1764/1900 (92%)	0.16	27 (1%) 71 70	37, 91, 143, 336	174 (9%)

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	899[A]	MET	5.2
2	B	1070	LEU	3.5
1	A	549	VAL	3.1
1	A	71	GLY	3.1
1	A	854	PHE	3.0

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

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](#)

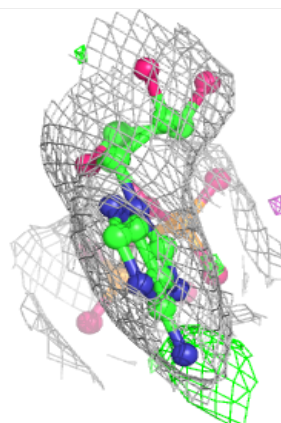
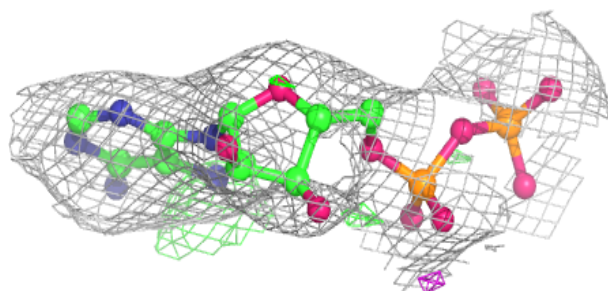
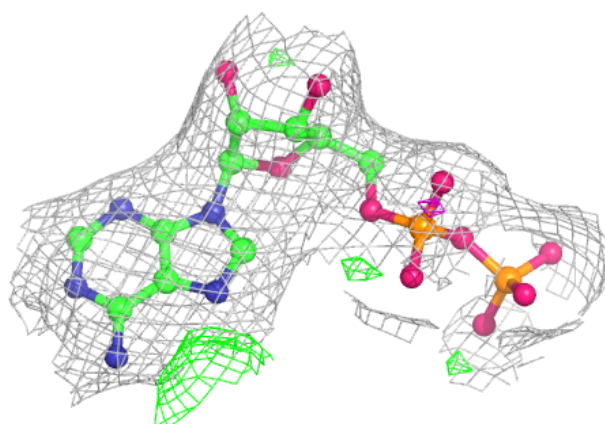
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
8	EDO	B	1201	4/4	0.86	0.13	93,94,94,94	0
8	EDO	B	1202	4/4	0.86	0.18	92,93,95,95	0
8	EDO	B	1203	4/4	0.86	0.10	81,81,81,81	0
5	ADP	A	1001	27/27	0.90	0.08	73,76,79,83	0
7	NA	A	1003	1/1	0.94	0.15	49,49,49,49	0
6	MG	A	1002	1/1	0.98	0.06	63,63,63,63	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

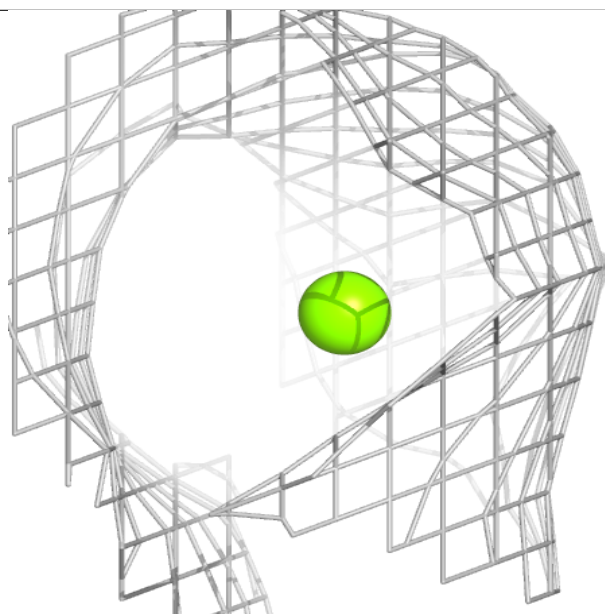
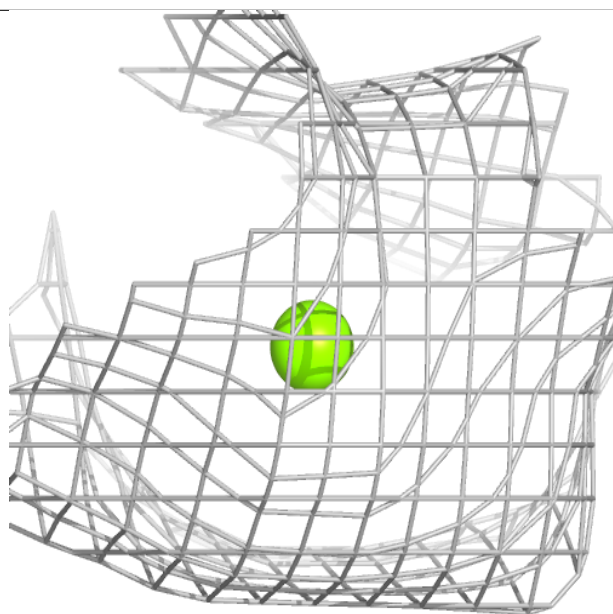
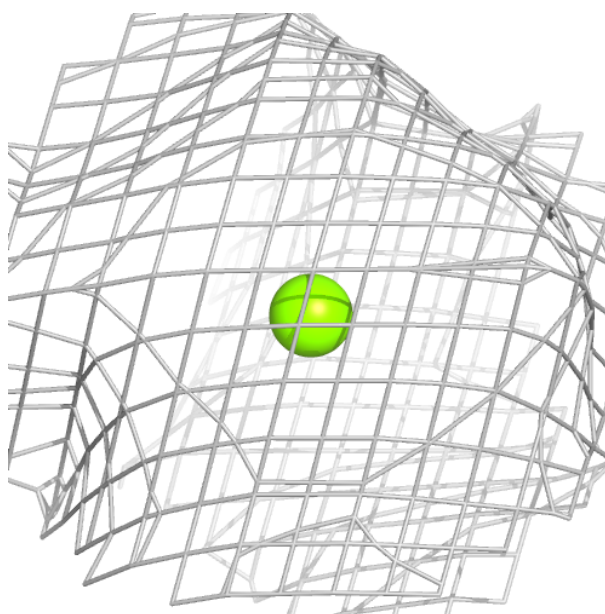
Electron density around ADP A 1001:

2mF_o-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



Electron density around MG A 1002:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers [i](#)

There are no such residues in this entry.