



Full wwPDB X-ray Structure Validation Report ⓘ

Jun 2, 2025 – 10:06 am BST

PDB ID : 8RAZ / pdb_00008raz
Title : The crystal structure of DNA-bound human MutSbeta (MSH2_E749A/MSH3_E976A) in the canonical mismatch bound conformation with ADP bound in MSH2 and MSH3
Authors : Thomsen, M.; Costanzi, E.
Deposited on : 2023-12-01
Resolution : 2.84 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 ⓘ](#)) 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

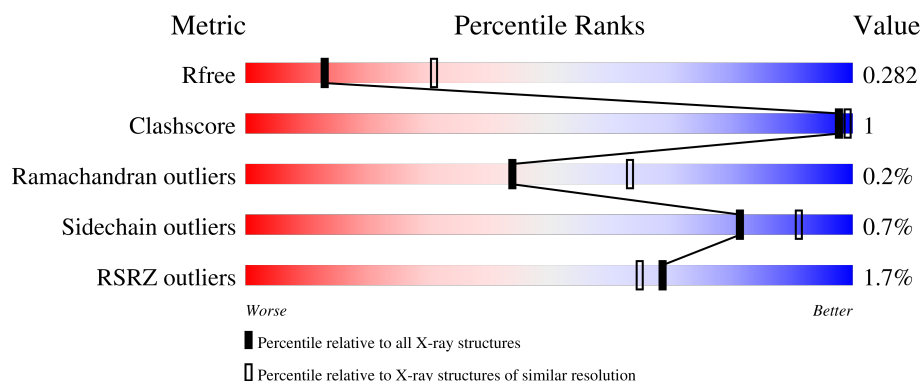
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.84 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	1367 (2.86-2.82)
Clashscore	180529	1455 (2.86-2.82)
Ramachandran outliers	177936	1422 (2.86-2.82)
Sidechain outliers	177891	1423 (2.86-2.82)
RSRZ outliers	164620	1368 (2.86-2.82)

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>93%</div> <div>5%</div> </div>
2	B	1137	<div> <div>%</div> <div>72%</div> <div>25%</div> </div>
3	C	24	<div> <div>71%</div> <div>8%</div> <div>21%</div> </div>
4	D	24	<div> <div>79%</div> <div>21%</div> </div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 14782 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	889	Total	C	N	O	S	545	1	0
			7029	4462	1193	1336	38			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	749	ALA	GLU	engineered mutation	UNP P43246

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	854	Total	C	N	O	S	336	1	0
			6842	4368	1172	1272	30			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	79	VAL	ILE	conflict	UNP P20585
B	949	ARG	GLN	conflict	UNP P20585
B	976	ALA	GLU	engineered mutation	UNP P20585
B	1045	THR	ALA	conflict	UNP P20585

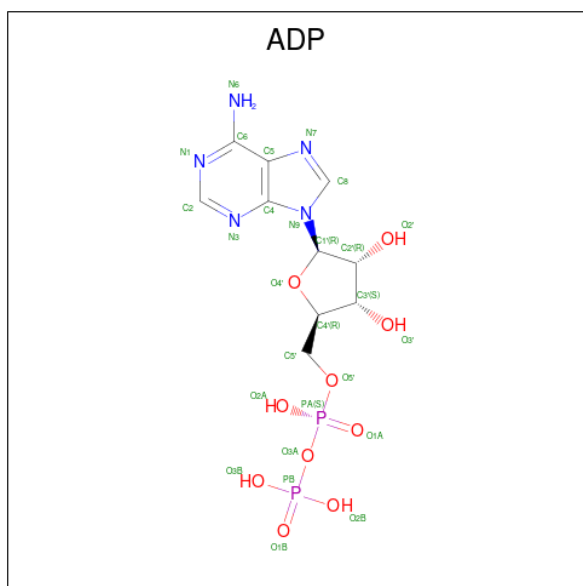
- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*TP*CP*TP*GP*AP*AP*GP*CP*CP*GP*AP*TP*CP*GP*AP*TP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	19	Total	C	N	O	P	0	0	0
			393	186	75	113	19			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	19	Total	C	N	O	P	0	0	0
			387	184	71	113	19			

- Molecule 5 is ADENOSINE-5'-DIPHOSPHATE (CCD ID: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$) (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		
5	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 6 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).

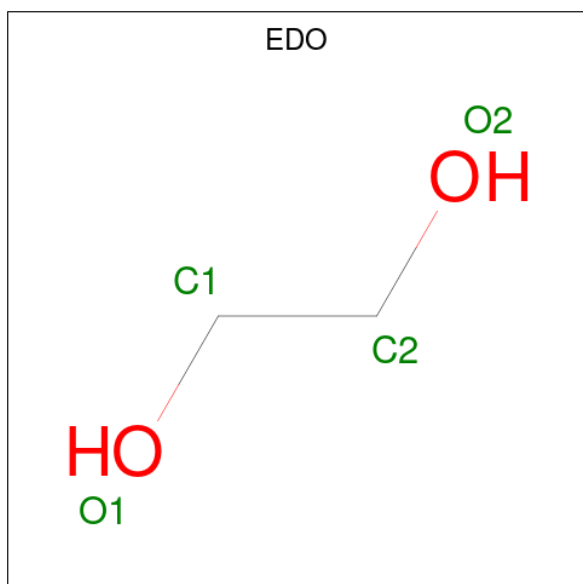


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 7 is CHLORIDE ION (CCD ID: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	1	Total	Cl	0	0
			1	1		
7	B	1	Total	Cl	0	0
			1	1		

- Molecule 8 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂).



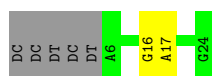
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	B	1	Total	C	O	0	0
			4	2	2		
8	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 9 is water.

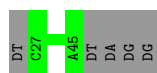
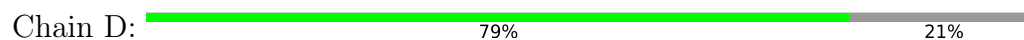
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	31	Total	O	0	0
			31	31		
9	B	29	Total	O	0	0
			29	29		
9	D	1	Total	O	0	0
			1	1		



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	70.32Å 92.41Å 94.22Å 67.13° 86.69° 75.82°	Depositor
Resolution (Å)	56.39 – 2.84 56.39 – 2.84	Depositor EDS
% Data completeness (in resolution range)	93.0 (56.39-2.84) 93.1 (56.39-2.84)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.76 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.214 , 0.279 0.216 , 0.282	Depositor DCC
R_{free} test set	1351 reflections (2.70%)	wwPDB-VP
Wilson B-factor (Å ²)	65.1	Xtriage
Anisotropy	0.216	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 52.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	14782	wwPDB-VP
Average B, all atoms (Å ²)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% 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

5.1 Standard geometry

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

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	0/7138	1.58	5/9614 (0.1%)
2	B	1.05	0/6963	1.55	3/9392 (0.0%)
3	C	0.32	0/441	0.75	0/679
4	D	0.32	0/433	0.73	0/665
All	All	1.03	0/14975	1.52	8/20350 (0.0%)

There are no bond length outliers.

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	379	ASP	CA-CB-CG	6.42	119.02	112.60
1	A	163	VAL	CA-C-N	5.33	125.28	121.65
1	A	163	VAL	C-N-CA	5.33	125.28	121.65
1	A	717	SER	CA-C-N	5.21	128.64	120.82
1	A	717	SER	C-N-CA	5.21	128.64	120.82
2	B	538	ASN	CA-C-N	5.12	125.76	120.03
2	B	538	ASN	C-N-CA	5.12	125.76	120.03
2	B	1012	PRO	N-CA-C	5.08	116.90	110.70

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	7029	0	7097	5	0
2	B	6842	0	6960	15	0
3	C	393	0	214	1	0
4	D	387	0	214	0	0
5	A	27	0	12	0	0
5	B	27	0	12	0	0
6	A	6	0	8	0	0
7	A	1	0	0	0	0
7	B	1	0	0	0	0
8	B	8	0	12	0	0
9	A	31	0	0	0	0
9	B	29	0	0	0	0
9	D	1	0	0	0	0
All	All	14782	0	14529	20	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.

All (20) 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)
2:B:884:LEU:HB3	2:B:1004:LEU:HD22	1.81	0.60
1:A:769:TYR:CG	2:B:1082:ILE:HD11	2.43	0.53
2:B:564:LEU:HA	2:B:840:VAL:HG21	1.90	0.53
1:A:86:GLU:HG2	1:A:134:ILE:HG23	1.94	0.48
2:B:340:ILE:HD11	2:B:395:PRO:CB	2.45	0.46
2:B:764:ILE:HG23	2:B:768:TRP:CE3	2.51	0.46
2:B:621:ILE:HG21	2:B:645:LEU:HD21	1.97	0.45
1:A:438:THR:HB	1:A:439:PRO:HD3	1.99	0.44
2:B:601:LEU:HD11	2:B:841:ALA:CB	2.48	0.44
2:B:952:PHE:CE2	2:B:956:LEU:HD11	2.53	0.44
2:B:750:GLN:HB3	2:B:753:MET:HE3	2.00	0.43
1:A:788:THR:HG21	1:A:821:VAL:HG13	2.01	0.43
2:B:456:GLU:HG2	2:B:509:TYR:CE1	2.54	0.43
2:B:506:ILE:HG22	2:B:519:LEU:HD11	2.01	0.43
2:B:584:LEU:HD11	2:B:593:ARG:CZ	2.51	0.41
2:B:401:VAL:HA	2:B:526:LYS:O	2.21	0.41
2:B:680:SER:N	2:B:681:PRO:CD	2.84	0.41
2:B:656:ILE:N	2:B:657:PRO:HD2	2.35	0.41
1:A:414:LEU:N	1:A:415:PRO:CD	2.84	0.40
3:C:16:DG:H2'	3:C:17:DA:C8	2.57	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	878/934 (94%)	835 (95%)	41 (5%)	2 (0%)	44	63
2	B	837/1137 (74%)	801 (96%)	35 (4%)	1 (0%)	48	69
All	All	1715/2071 (83%)	1636 (95%)	76 (4%)	3 (0%)	44	63

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	466	HIS
2	B	620	ASP
1	A	648	ILE

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	772/807 (96%)	770 (100%)	2 (0%)	91	96
2	B	763/998 (76%)	755 (99%)	8 (1%)	73	87
All	All	1535/1805 (85%)	1525 (99%)	10 (1%)	81	91

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	577	ILE
1	A	638	ARG
2	B	249	CYS
2	B	404	SER
2	B	540	THR
2	B	660	ASN
2	B	780	PHE
2	B	781	HIS
2	B	782	SER
2	B	903	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	74	ASN
1	A	105	ASN
1	A	388	ASN
1	A	397	GLN
1	A	593	GLN
1	A	903	ASN
2	B	791	HIS
2	B	819	HIS
2	B	843	GLN
2	B	861	ASN

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	GOL	A	1002	-	5,5,5	0.10	0	5,5,5	0.30	0
8	EDO	B	1202	-	3,3,3	0.05	0	2,2,2	0.26	0
5	ADP	B	1201	-	24,29,29	0.64	0	29,45,45	0.76	1 (3%)
8	EDO	B	1203	-	3,3,3	0.16	0	2,2,2	0.36	0
5	ADP	A	1001	-	24,29,29	0.65	0	29,45,45	0.75	1 (3%)

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	GOL	A	1002	-	-	0/4/4/4	-
8	EDO	B	1202	-	-	0/1/1/1	-
5	ADP	B	1201	-	-	0/12/32/32	0/3/3/3
8	EDO	B	1203	-	-	1/1/1/1	-
5	ADP	A	1001	-	-	2/12/32/32	0/3/3/3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1001	ADP	C5-C6-N6	2.09	123.53	120.35
5	B	1201	ADP	C5-C6-N6	2.01	123.40	120.35

There are no chirality outliers.

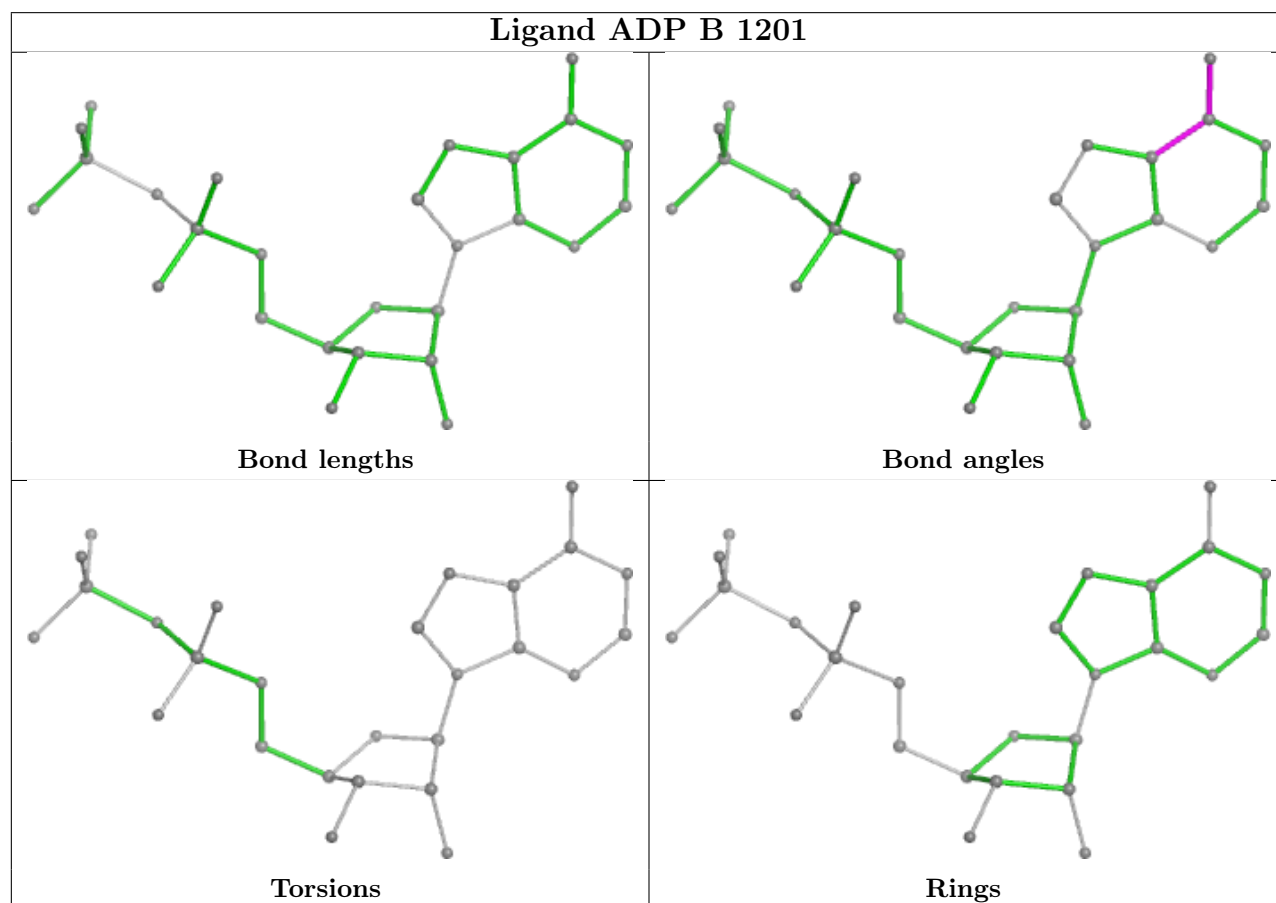
All (3) torsion outliers are listed below:

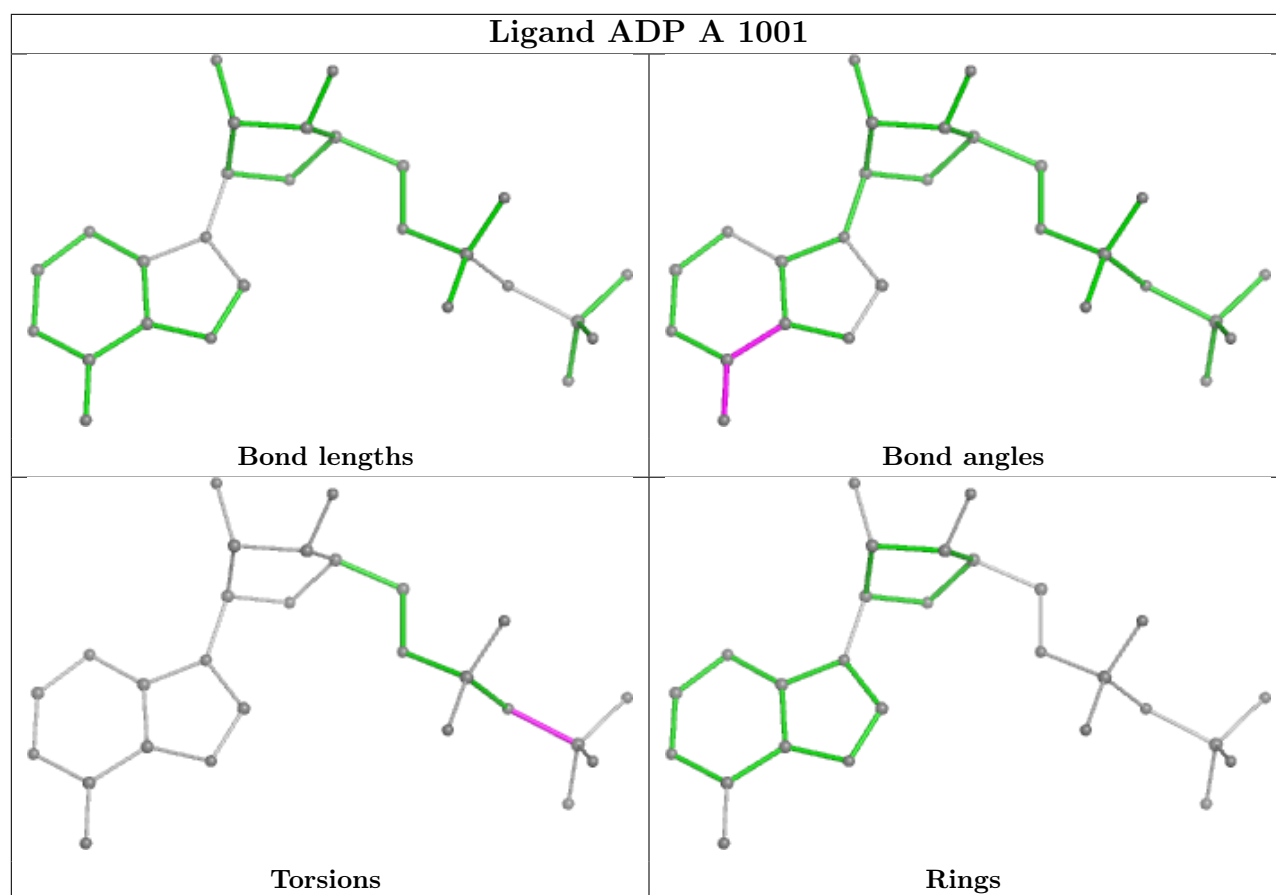
Mol	Chain	Res	Type	Atoms
5	A	1001	ADP	PA-O3A-PB-O3B
8	B	1203	EDO	O1-C1-C2-O2
5	A	1001	ADP	PA-O3A-PB-O1B

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

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	A	889/934 (95%)	0.03	22 (2%) 58 55	20, 65, 112, 155	191 (21%)
2	B	854/1137 (75%)	-0.24	9 (1%) 77 76	25, 56, 91, 129	119 (13%)
3	C	19/24 (79%)	-0.23	0 100 100	45, 80, 170, 179	0
4	D	19/24 (79%)	-0.42	0 100 100	54, 69, 129, 150	0
All	All	1781/2119 (84%)	-0.11	31 (1%) 69 65	20, 61, 106, 179	310 (17%)

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	533	LEU	4.5
1	A	507	PRO	4.0
1	A	510	GLN	3.2
1	A	543	ASP	2.9
1	A	113	LYS	2.8
1	A	511	ILE	2.8
1	A	529	GLU	2.6
1	A	68	GLY	2.6
1	A	41	ASP	2.6
1	A	467	GLU	2.5
2	B	778	SER	2.5
2	B	359	ILE	2.5
1	A	719	LEU	2.4
2	B	356	VAL	2.4
2	B	354	VAL	2.3
1	A	497	ILE	2.3
1	A	664	PHE	2.3
1	A	294[A]	PHE	2.3
2	B	1011	TYR	2.2
1	A	139	ASN	2.2
1	A	542	VAL	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	513	LEU	2.2
2	B	776	ALA	2.2
1	A	505	LEU	2.2
2	B	447	SER	2.1
1	A	517	ALA	2.1
1	A	714	ALA	2.1
1	A	550	LYS	2.1
2	B	450	ASP	2.0
2	B	1124	GLU	2.0
1	A	466	HIS	2.0

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](#)

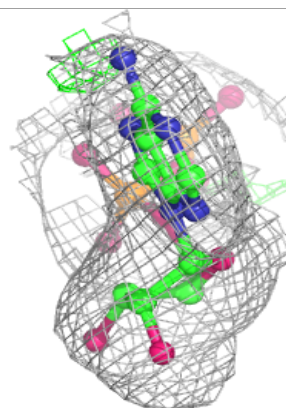
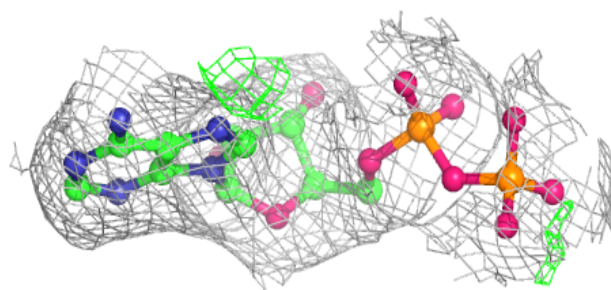
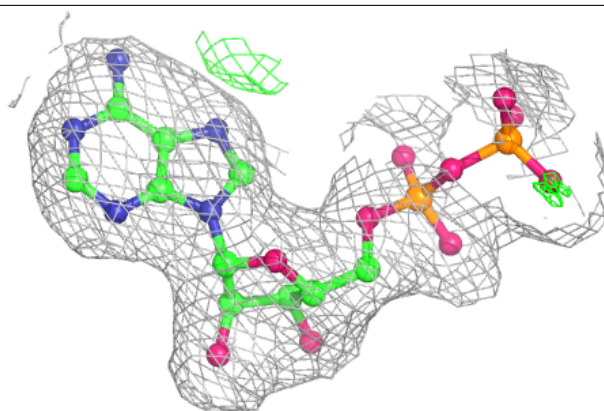
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
7	CL	B	1204	1/1	0.80	0.10	100,100,100,100	0
6	GOL	A	1002	6/6	0.90	0.13	74,79,80,81	0
8	EDO	B	1203	4/4	0.91	0.10	46,47,47,48	0
8	EDO	B	1202	4/4	0.92	0.09	53,54,55,55	0
7	CL	A	1003	1/1	0.94	0.06	90,90,90,90	0
5	ADP	B	1201	27/27	0.96	0.06	42,46,48,50	0
5	ADP	A	1001	27/27	0.96	0.06	53,56,61,64	0

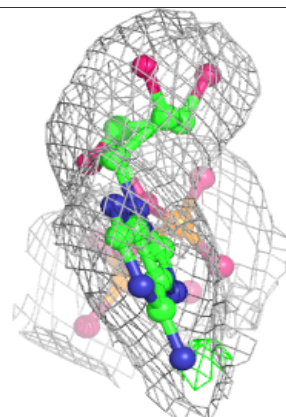
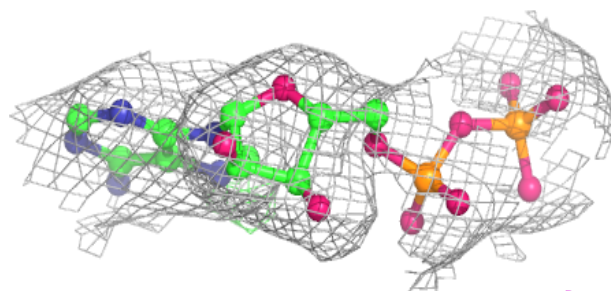
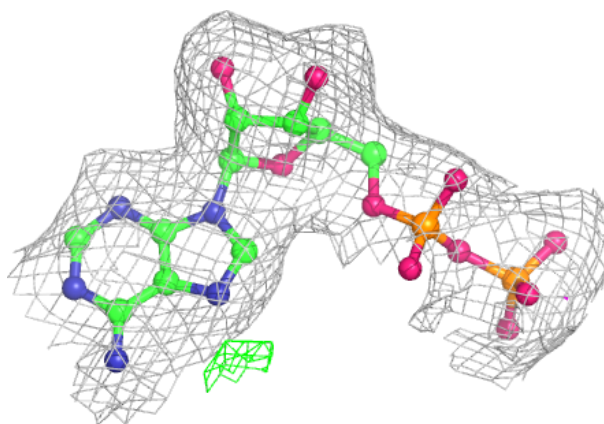
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around ADP B 1201:

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

**Electron density around 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)



6.5 Other polymers [i](#)

There are no such residues in this entry.