



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 07:14 am BST

PDB ID : 5YK4
Title : Mismatch Repair Protein
Authors : Nirwal, S.; Nair, D.T.
Deposited on : 2017-10-12
Resolution : 2.97 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

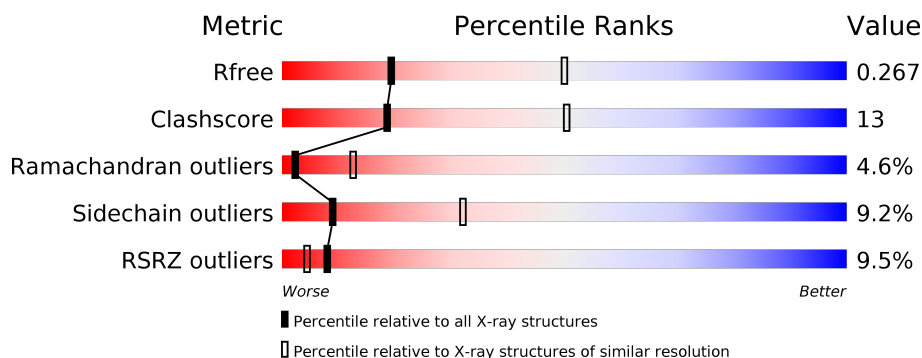
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.97 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}	130704	2754 (3.00-2.96)
Clashscore	141614	3103 (3.00-2.96)
Ramachandran outliers	138981	2993 (3.00-2.96)
Sidechain outliers	138945	2996 (3.00-2.96)
RSRZ outliers	127900	2644 (3.00-2.96)

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 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	819	<div> <div>4%</div> <div> <div></div> <div>70%</div> <div>21%</div> <div>• 6%</div> </div> </div>
1	B	819	<div> <div>14%</div> <div> <div></div> <div>57%</div> <div>29%</div> <div>7%</div> <div>• 6%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11983 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 MutS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	771	Total	C	N	O	S	0	0	0
			5906	3733	1046	1107	20			
1	B	767	Total	C	N	O	S	0	0	0
			5904	3735	1044	1106	19			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	expression tag	UNP Q5F5J4
A	-3	PRO	-	expression tag	UNP Q5F5J4
A	-2	LEU	-	expression tag	UNP Q5F5J4
A	-1	GLY	-	expression tag	UNP Q5F5J4
A	0	SER	-	expression tag	UNP Q5F5J4
B	-4	GLY	-	expression tag	UNP Q5F5J4
B	-3	PRO	-	expression tag	UNP Q5F5J4
B	-2	LEU	-	expression tag	UNP Q5F5J4
B	-1	GLY	-	expression tag	UNP Q5F5J4
B	0	SER	-	expression tag	UNP Q5F5J4

- Molecule 2 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: C₁₀H₁₅N₅O₁₀P₂) (labeled as "Ligand of Interest" by author).



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

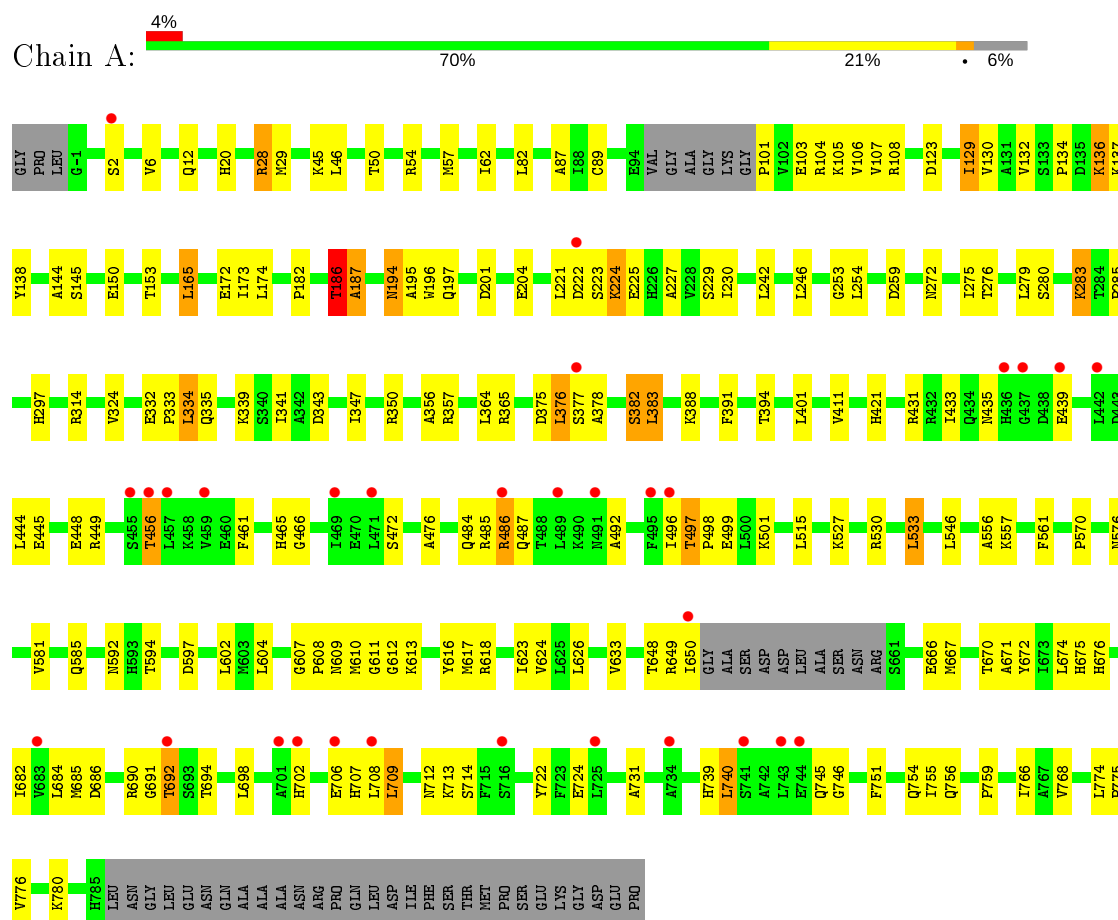
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	89	Total	O	0	0
			89	89		
3	B	30	Total	O	0	0
			30	30		

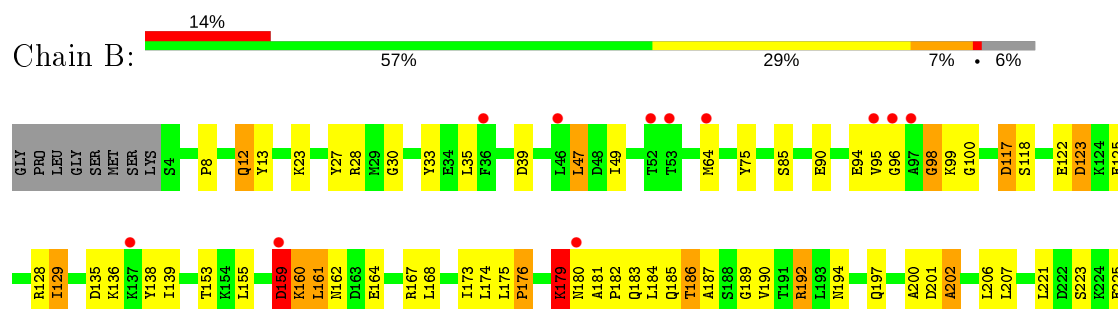
3 Residue-property plots

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



• Molecule 1: DNA mismatch repair protein MutS



ALA	ALA	K713	V643	E558	Y483	D415	Q385	H226
ALA	ALA	S714	D644	R559	Q484	Y415	C336	A227
ASN	ASN	F715	Q645	N560	R485	N420	E337	I230
ARG	ARG	S716		F561	V562	H424	I341	L236
PRO	PRO	L717	H649	R563	Q487	P425	E345	L237
GLN	GLN		ILE	P564	T488	E426	R346	R238
LEU	LEU	E723	GLY		L489	L427	I347	Y239
ASP	ASP	E724	ALA	V571	N491	D428	A348	T243
ILE	ILE	L725	SER	I574	A492	R431	A349	Q244
PHE	PHE	L726	ASP		E493	R432	R350	E250
SER	SER	L728	ASP	G577	R494	I433	A352	
THR	THR	P729	LEU	R577	F495	H436	V353	L254
MET	MET		ALA	R578	T497	Q437	A362	E257
PRO	PRO	A734	ALA	H579	L500	D438	S363	T258
SER	SER	A735	ASN	P580	K501	E439	L364	E259
GLU	GLU	Y736	ASN	V581	A502	F440	R365	S260
LYS	LYS	N737	ARG	V582	F503	L441	F369	Q261
GLY	GLY	N738	SER	H583	E504	L442	L263	Y262
ASP	ASP	H739	THR	Q584	L508	D443	E370	G264
GLU	GLU	L740	PHE	Q585	L509	L444	L371	M265
PRO	PRO		MET	R586	A510	E445	I374	
		L743	V665	H587	Q511	K446	D375	R270
		E744	B666	H588	Q513	R449	L376	S280
		Q745	M667		L517	E450	S377	
			S668		E518	R451	A378	T284
		D748	B689	N592	K519	R452	T379	P285
		L752	T670	H593	Q520	L453	S382	L286
		H753	I673	T594	L521	R454	L383	L287
		Q754	L674	D595	F522	E455	L384	
		I755	H675	R601	L526	T456	E385	I290
		Q756	H676	L602	K527	L457	T386	L291
			A677		N528	R458	L387	
		P759	S681	L605	L529	V459	K388	M298
		A760	I682	N609	E530	E460	A389	G299
		G761		M610	T531	N462	F391	
		K762		G611	A532	R463	L305	
		S763	M685	G612	L533	V464	H308	
		Y764	V688	K613	Q535	G465	L311	
		G765	G689		L536	F467	P396	
		I766	R690	Y616		Y468	Y397	
		A767	G691	Q619		T469	R314	
		Y768	T692		A542	E470	I317	
		A769	F695	I622	A543	L471	R318	
				I623	A544	S472	A319	
		V776	L698	V624		K473	R320	
		L779	A699	L625		Q475	Q321	
			L700				E322	
		Q783		T629			A323	
		K784	I704	G630			V324	
		H785	A705	C631				
		LEU	E706	F632			L412	
		ASN	H707	V633			K413	
		GLY	I708	P634			F414	
		LEU	L709	A635				
		GLU	Q710	D636				
		ASN	K711	A637				
		GLN	N712	A638				
		ALA						

4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2 ₁	Depositor
Cell constants a, b, c, α , β , γ	89.94Å 102.40Å 236.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.97 58.65 – 2.97	Depositor EDS
% Data completeness (in resolution range)	99.4 (50.00-2.97) 99.4 (58.65-2.97)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.40 (at 2.96Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.246 , 0.269 0.246 , 0.267	Depositor DCC
R_{free} test set	2301 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	78.6	Xtriage
Anisotropy	0.445	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 62.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11983	wwPDB-VP
Average B, all atoms (Å ²)	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.98% 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: 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	0.37	0/6015	0.50	2/8154 (0.0%)
1	B	0.30	0/6014	0.51	1/8154 (0.0%)
All	All	0.34	0/12029	0.51	3/16308 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	674	LEU	CA-CB-CG	5.79	128.62	115.30
1	A	186	THR	C-N-CA	5.43	135.29	121.70
1	A	775	PRO	N-CA-CB	5.22	109.56	103.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	159	ASP	Peptide
1	B	669	GLU	Peptide

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	5906	0	5863	115	0
1	B	5904	0	5915	201	0
2	A	27	0	12	0	0
2	B	27	0	12	1	0
3	A	89	0	0	9	0
3	B	30	0	0	3	0
All	All	11983	0	11802	312	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 13.

The worst 5 of 312 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:B:441:LEU:O	1:B:445:GLU:HB3	1.52	1.09
1:A:332:GLU:HG2	1:A:333:PRO:HD3	1.50	0.93
1:B:395:LEU:HB2	1:B:396:PRO:HD2	1.49	0.92
1:B:180:ASN:HB3	1:B:181:ALA:HA	1.50	0.92
1:B:762:LYS:HG2	1:B:763:SER:HA	1.60	0.82

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	765/819 (93%)	688 (90%)	55 (7%)	22 (3%)	4 22

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	763/819 (93%)	631 (83%)	83 (11%)	49 (6%)	1	6
All	All	1528/1638 (93%)	1319 (86%)	138 (9%)	71 (5%)	2	12

5 of 71 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	187	ALA
1	A	285	PRO
1	A	377	SER
1	A	713	LYS
1	A	774	LEU

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	
1	A	608/669 (91%)	572 (94%)	36 (6%)	19	52
1	B	615/669 (92%)	539 (88%)	76 (12%)	4	19
All	All	1223/1338 (91%)	1111 (91%)	112 (9%)	9	32

5 of 112 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	263	ILE
1	B	371	LEU
1	B	710	GLN
1	B	284	THR
1	B	317	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	162	ASN
1	B	261	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	721	HIS
1	B	250	HIS
1	B	308	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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

2 ligands are modelled in this entry.

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	$\# Z > 2$
2	ADP	A	901	-	24,29,29	1.03	2 (8%)	29,45,45	1.44	4 (13%)
2	ADP	B	901	-	24,29,29	1.08	1 (4%)	29,45,45	1.51	4 (13%)

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
2	ADP	A	901	-	-	3/12/32/32	0/3/3/3

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	ADP	B	901	-	-	3/12/32/32	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	901	ADP	C5-C4	2.73	1.48	1.40
2	A	901	ADP	C5-C4	2.73	1.48	1.40
2	A	901	ADP	C2-N3	2.09	1.35	1.32

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	901	ADP	PA-O3A-PB	-4.06	118.89	132.83
2	A	901	ADP	PA-O3A-PB	-3.75	119.97	132.83
2	B	901	ADP	C3'-C2'-C1'	3.43	106.15	100.98
2	B	901	ADP	N3-C2-N1	-3.28	123.55	128.68
2	A	901	ADP	N3-C2-N1	-2.88	124.18	128.68

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	901	ADP	C5'-O5'-PA-O3A
2	B	901	ADP	C5'-O5'-PA-O3A
2	B	901	ADP	PA-O3A-PB-O1B
2	A	901	ADP	C5'-O5'-PA-O1A
2	A	901	ADP	C5'-O5'-PA-O2A

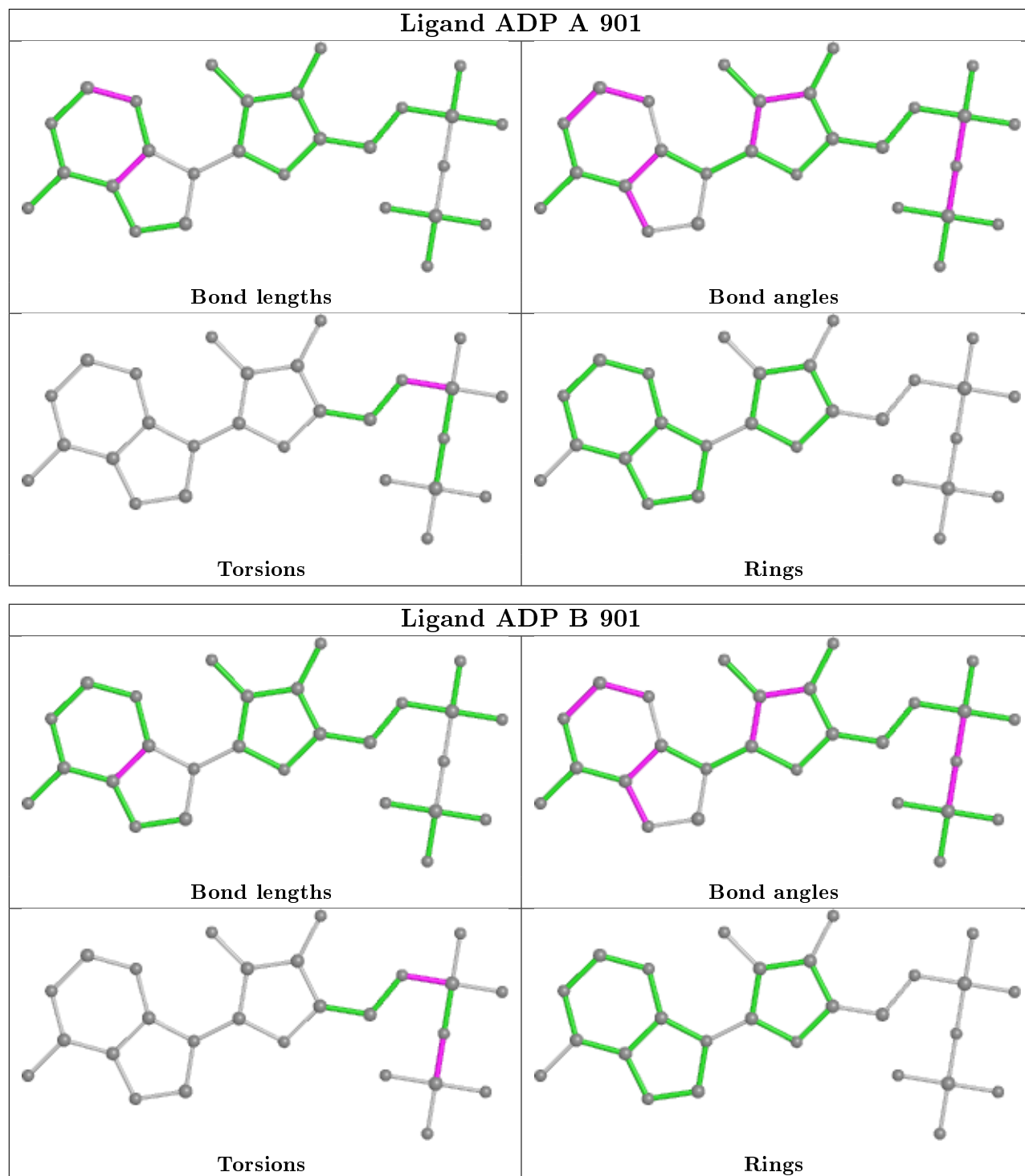
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	901	ADP	1	0

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	771/819 (94%)	0.45	31 (4%) 38 23	42, 74, 133, 184	0
1	B	767/819 (93%)	0.94	115 (14%) 2 1	60, 113, 191, 254	0
All	All	1538/1638 (93%)	0.69	146 (9%) 8 4	42, 93, 169, 254	0

The worst 5 of 146 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	452	THR	11.5
1	B	681	SER	8.2
1	B	531	THR	7.8
1	B	459	VAL	7.7
1	B	470	GLU	7.0

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](#)

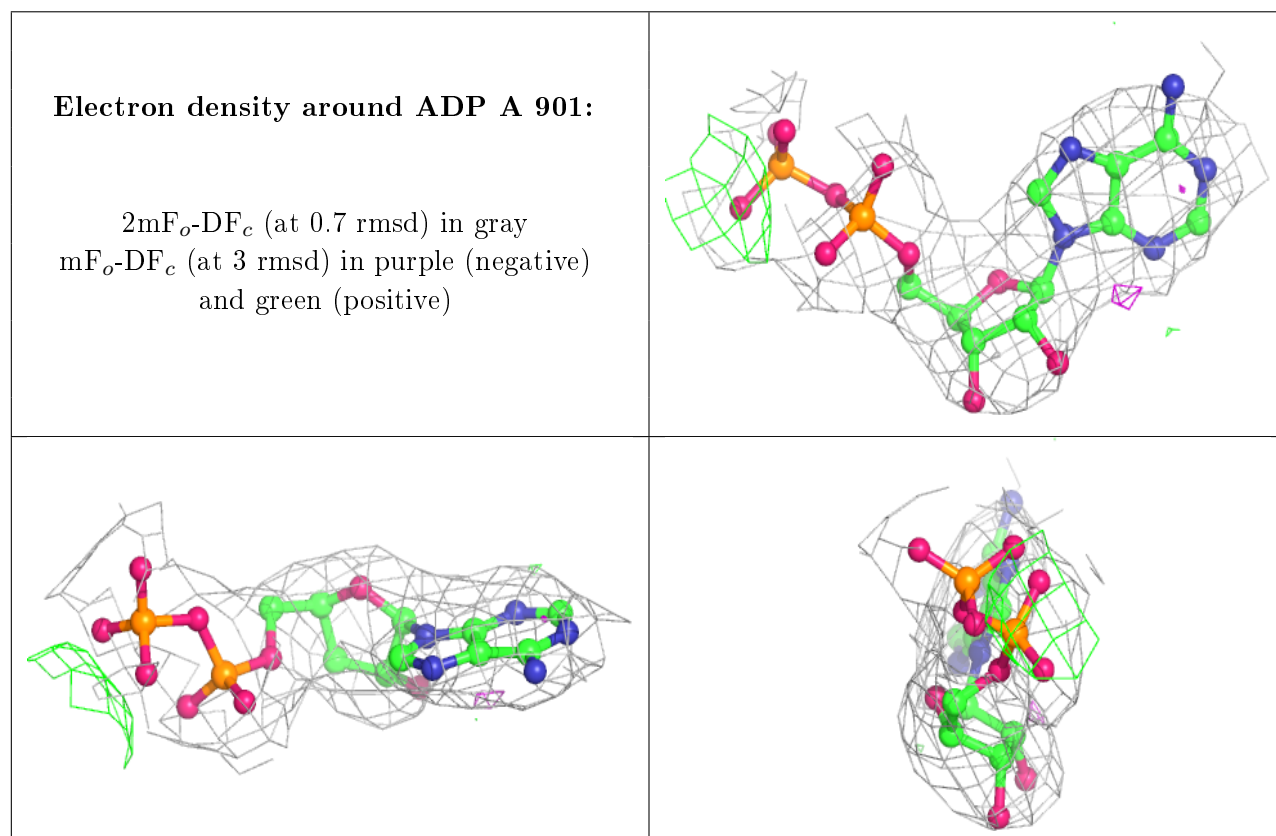
There are no carbohydrates in this entry.

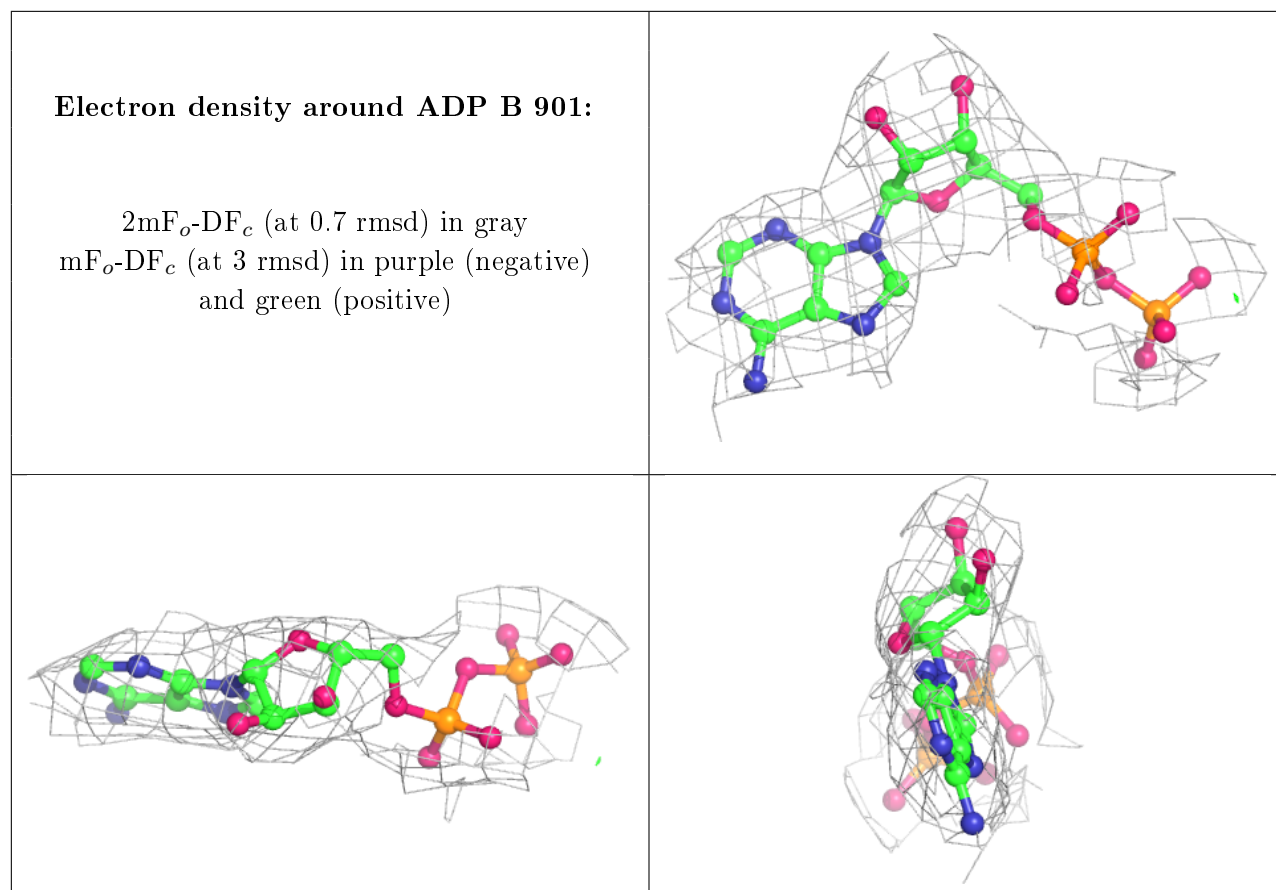
6.4 Ligands [i](#)

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
2	ADP	A	901	27/27	0.93	0.25	59,64,70,73	0
2	ADP	B	901	27/27	0.93	0.20	79,88,102,106	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.





6.5 Other polymers ⓘ

There are no such residues in this entry.