



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:38 PM GMT

PDB ID : 3UBT  
Title : Crystal Structure of C71S Mutant of DNA Cytosine-5 Methyltransferase  
M.HaeIII Bound to DNA  
Authors : Verdine, G.L.; Didovyk, A.  
Deposited on : 2011-10-24  
Resolution : 2.50 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

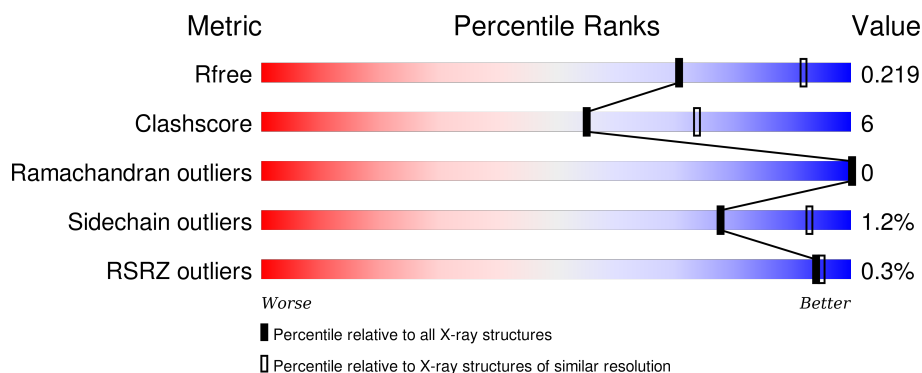
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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.



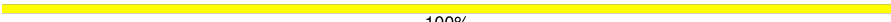



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	331	<div> <div>84%</div> <div>15%</div> <div>.</div> </div>
1	B	331	<div> <div>85%</div> <div>14%</div> <div>.</div> </div>
1	Y	331	<div> <div>85%</div> <div>15%</div> <div>.</div> </div>
2	C	6	<div> <div>67%</div> <div>33%</div> </div>
2	D	6	<div> <div>17%</div> <div>83%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	E	6	 100%
2	F	6	 67% 33%
2	G	6	 17% 83%
2	H	6	 17% 100%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	ATP	A	333	-	-	-	X
4	ATP	Y	334	-	-	-	X
5	2PE	B	333	-	-	-	X
5	2PE	Y	335	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 9258 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 Modification methylase HaeIII.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	Y	328	Total	C	N	O	S	0	0	0
			2615	1683	450	470	12			
1	A	328	Total	C	N	O	S	0	0	0
			2607	1678	447	470	12			
1	B	328	Total	C	N	O	S	0	2	0
			2623	1690	453	468	12			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Y	71	SER	CYS	ENGINEERED MUTATION	UNP P20589
Y	331	ASN	-	EXPRESSION TAG	UNP P20589
A	71	SER	CYS	ENGINEERED MUTATION	UNP P20589
A	331	ASN	-	EXPRESSION TAG	UNP P20589
B	71	SER	CYS	ENGINEERED MUTATION	UNP P20589
B	331	ASN	-	EXPRESSION TAG	UNP P20589

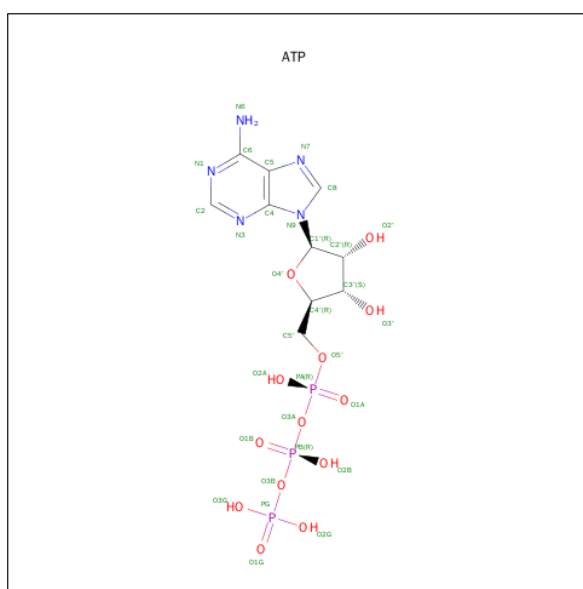
- Molecule 2 is a DNA chain called 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	6	Total	C	N	O	P	0	0	0
			120	58	23	34	5			
2	E	6	Total	C	N	O	P	0	0	0
			120	58	23	34	5			
2	F	6	Total	C	N	O	P	0	0	0
			120	58	23	34	5			
2	G	6	Total	C	N	O	P	0	0	0
			120	58	23	34	5			
2	H	6	Total	C	N	O	P	0	0	0
			111	53	19	34	5			
2	C	6	Total	C	N	O	P	0	0	0
			120	58	23	34	5			

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

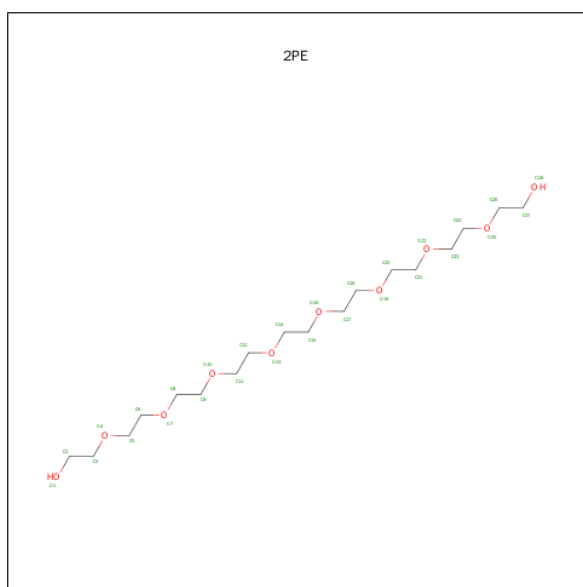
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Cl	0	0
			1	1		
3	B	1	Total	Cl	0	0
			1	1		
3	Y	1	Total	Cl	0	0
			1	1		

- Molecule 4 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	Y	1	Total	C	N	O	P	0	0
			23	10	5	7	1		
4	Y	1	Total	C	N	O	P	0	0
			31	10	5	13	3		
4	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

- Molecule 5 is NONAETHYLENE GLYCOL (three-letter code: 2PE) (formula: C<sub>18</sub>H<sub>38</sub>O<sub>10</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	Y	1	Total	C	O	0	0
			13	8	5		
5	A	1	Total	C	O	0	0
			8	6	2		
5	B	1	Total	C	O	0	0
			13	8	5		

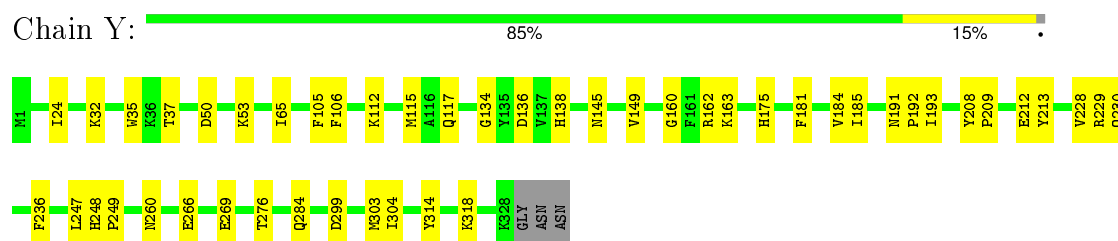
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	Y	171	Total	O	0	4
			175	175		
6	A	171	Total	O	0	2
			173	173		
6	B	173	Total	O	0	4
			177	177		
6	D	4	Total	O	0	0
			4	4		
6	E	7	Total	O	0	0
			7	7		
6	F	11	Total	O	0	1
			12	12		
6	G	5	Total	O	0	0
			5	5		
6	H	11	Total	O	0	1
			12	12		
6	C	15	Total	O	0	0
			15	15		

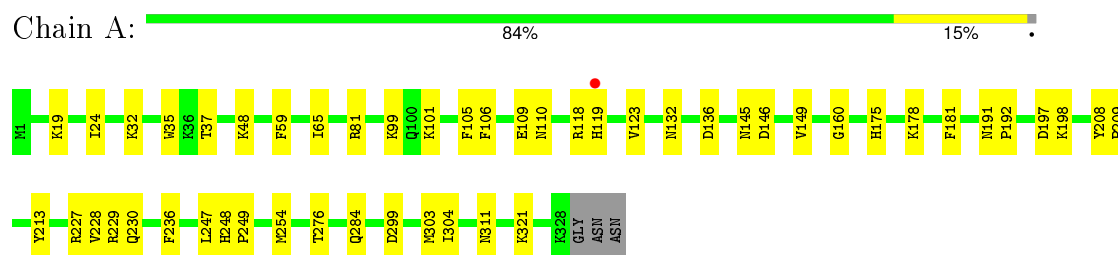
### 3 Residue-property plots

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.

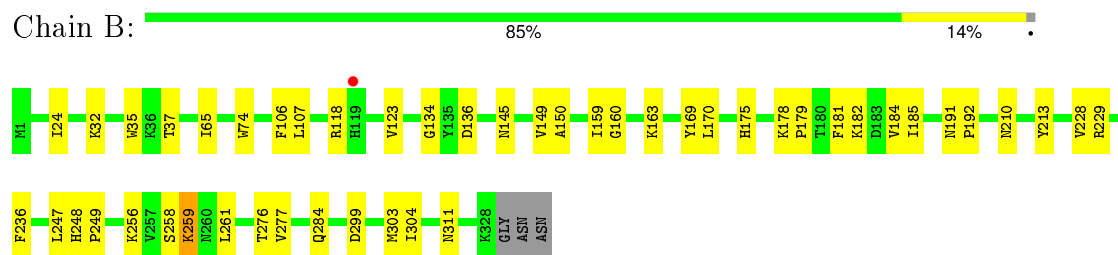
#### • Molecule 1: Modification methylase HaeIII



#### • Molecule 1: Modification methylase HaeIII




#### • Molecule 1: Modification methylase HaeIII



#### • Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'



#### • Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'

Chain E:  100%

T4  
G5  
G6  
C7  
C8  
A9

- Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'

Chain F:  67% 33%

T7  
C11  
A12

- Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'

Chain G:  17% 83%

T4  
G5  
G6  
C7  
C8  
A9

- Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'

Chain H:  17% 100%

T7  
A12

- Molecule 2: 5'-D(\*TP\*GP\*GP\*CP\*CP\*A)-3'

Chain C:  67% 33%

T7  
G8  
A12



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.56 Å 129.59 Å 132.38 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	16.98 – 2.50 16.98 – 2.50	Depositor EDS
% Data completeness (in resolution range)	90.9 (16.98-2.50) 91.1 (16.98-2.50)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.43 (at 2.49 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, $R_{free}$	0.182 , 0.221 0.181 , 0.219	Depositor DCC
$R_{free}$ test set	1538 reflections (5.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.1	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 48.9	EDS
Estimated twinning fraction	0.000 for -h,l,k	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	3 of 31721 reflections (0.009%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9258	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 62.18 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.1719e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: 2PE, ATP, 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	0.22	0/2673	0.37	0/3614
1	B	0.21	0/2696	0.36	0/3641
1	Y	0.21	0/2680	0.37	0/3620
2	C	0.50	0/134	0.91	0/205
2	D	0.41	0/134	1.07	0/205
2	E	0.41	0/134	0.92	0/205
2	F	0.43	0/134	0.99	0/205
2	G	0.40	0/134	1.03	0/205
2	H	0.41	0/123	0.89	0/188
All	All	0.24	0/8842	0.46	0/12088

There are no bond length outliers.

There are no bond angle outliers.

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	2607	0	2549	33	0
1	B	2623	0	2591	26	0
1	Y	2615	0	2575	29	0
2	C	120	0	69	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	120	0	69	5	0
2	E	120	0	69	6	0
2	F	120	0	69	2	0
2	G	120	0	69	4	0
2	H	111	0	65	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	Y	1	0	0	0	0
4	A	31	0	12	2	0
4	Y	54	0	24	2	0
5	A	8	0	8	0	0
5	B	13	0	17	1	0
5	Y	13	0	17	0	0
6	A	173	0	0	4	0
6	B	177	0	0	1	0
6	C	15	0	0	0	0
6	D	4	0	0	0	0
6	E	7	0	0	0	0
6	F	12	0	0	0	0
6	G	5	0	0	0	0
6	H	12	0	0	0	0
6	Y	175	0	0	4	0
All	All	9258	0	8203	107	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 6.

The worst 5 of 107 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:119:HIS:HB2	6:A:353:HOH:O	1.75	0.84
2:D:6:DG:H2"	2:D:7:DC:H5"	1.61	0.81
2:F:11:DC:H2"	2:F:12:DA:H5'	1.69	0.74
1:Y:314:TYR:CZ	1:Y:318:LYS:HD2	2.24	0.73
1:Y:162:ARG:HD3	6:Y:416:HOH:O	1.88	0.71

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	326/331 (98%)	319 (98%)	7 (2%)	0	100	100
1	B	328/331 (99%)	321 (98%)	7 (2%)	0	100	100
1	Y	326/331 (98%)	317 (97%)	9 (3%)	0	100	100
All	All	980/993 (99%)	957 (98%)	23 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	273/287 (95%)	270 (99%)	3 (1%)	80	94
1	B	277/287 (96%)	271 (98%)	6 (2%)	60	84
1	Y	275/287 (96%)	273 (99%)	2 (1%)	88	97
All	All	825/861 (96%)	814 (99%)	11 (1%)	78	92

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	311	ASN
1	B	175	HIS
1	B	259[B]	LYS
1	A	181	PHE
1	B	259[A]	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	138	HIS
1	B	262	ASN
1	B	145	ASN
1	A	132	ASN
1	B	202	ASN

### 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 ⓘ

Of 9 ligands modelled in this entry, 3 are monoatomic - leaving 6 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	$\# Z  > 2$
4	ATP	A	333	-	24,33,33	0.96	1 (4%)	31,52,52	1.90	5 (16%)
5	2PE	A	334	-	7,7,27	0.50	0	6,6,26	0.24	0
5	2PE	B	333	-	12,12,27	0.41	0	11,11,26	0.30	0
4	ATP	Y	333	-	20,25,33	1.10	1 (5%)	22,38,52	1.95	4 (18%)
4	ATP	Y	334	-	24,33,33	0.99	1 (4%)	31,52,52	1.86	5 (16%)
5	2PE	Y	335	-	12,12,27	0.40	0	11,11,26	0.33	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
4	ATP	A	333	-	-	0/18/38/38	0/3/3/3
5	2PE	A	334	-	-	0/5/5/25	0/0/0/0
5	2PE	B	333	-	-	0/10/10/25	0/0/0/0
4	ATP	Y	333	-	-	0/6/26/38	0/3/3/3
4	ATP	Y	334	-	-	0/18/38/38	0/3/3/3
5	2PE	Y	335	-	-	0/10/10/25	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	333	ATP	C5-C4	3.08	1.47	1.40
4	Y	333	ATP	C5-C4	3.09	1.47	1.40
4	Y	334	ATP	C5-C4	3.12	1.47	1.40

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	333	ATP	N3-C2-N1	-6.89	123.62	128.89
4	Y	333	ATP	N3-C2-N1	-6.74	123.73	128.89
4	Y	334	ATP	N3-C2-N1	-6.42	123.98	128.89
4	Y	334	ATP	PA-O3A-PB	-3.98	121.55	132.73
4	A	333	ATP	PA-O3A-PB	-3.61	122.61	132.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	333	ATP	2	0
5	B	333	2PE	1	0
4	Y	334	ATP	2	0

## 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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	A	328/331 (99%)	-0.44	1 (0%) 94 95	8, 19, 35, 50	0
1	B	328/331 (99%)	-0.39	1 (0%) 94 95	8, 19, 37, 57	0
1	Y	328/331 (99%)	-0.49	0 100 100	8, 17, 32, 40	0
2	C	6/6 (100%)	0.11	0 100 100	22, 33, 39, 46	0
2	D	6/6 (100%)	0.32	0 100 100	27, 31, 38, 45	0
2	E	6/6 (100%)	0.60	0 100 100	37, 43, 47, 48	0
2	F	6/6 (100%)	0.17	0 100 100	23, 35, 44, 44	0
2	G	6/6 (100%)	0.26	0 100 100	34, 40, 45, 52	0
2	H	6/6 (100%)	0.11	1 (16%) 2 2	19, 32, 50, 73	0
All	All	1020/1029 (99%)	-0.41	3 (0%) 94 95	8, 19, 38, 73	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	119	HIS	3.1
1	B	119	HIS	2.9
2	H	12	DA	2.1

### 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. 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, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	2PE	B	333	13/28	0.90	0.18	3.71	23,35,45,45	0
4	ATP	Y	334	31/31	0.88	0.24	3.45	20,44,85,106	0
5	2PE	Y	335	13/28	0.93	0.17	2.03	28,35,56,72	0
4	ATP	A	333	31/31	0.76	0.32	2.02	32,77,109,118	0
5	2PE	A	334	8/28	0.88	0.22	1.98	33,38,45,52	0
4	ATP	Y	333	23/31	0.92	0.16	0.33	15,36,64,76	0
3	CL	A	332	1/1	0.98	0.18	-	23,23,23,23	0
3	CL	Y	332	1/1	0.98	0.12	-	33,33,33,33	0
3	CL	B	332	1/1	0.97	0.20	-	36,36,36,36	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.