



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 02:09 pm BST

PDB ID : 2IH2  
Title : Crystal structure of the adenine-specific DNA methyltransferase M.TaqI complexed with the cofactor analog AETA and a 10 bp DNA containing 5-methylpyrimidin-2(1H)-one at the target base partner position  
Authors : Lenz, T.; Scheidig, A.J.; Weinhold, E.  
Deposited on : 2006-09-25  
Resolution : 1.61 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](mailto: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

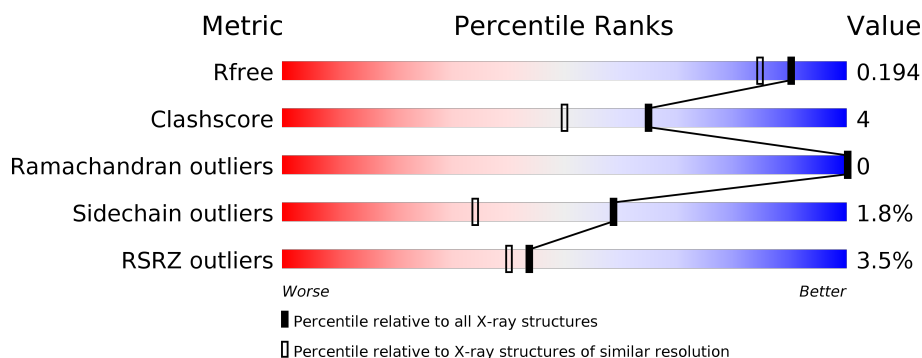
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.61 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	4693 (1.64-1.60)
Clashscore	141614	5002 (1.64-1.60)
Ramachandran outliers	138981	4888 (1.64-1.60)
Sidechain outliers	138945	4887 (1.64-1.60)
RSRZ outliers	127900	4609 (1.64-1.60)

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 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	B	10	<div> <div></div> <div>50%50%</div> </div>
1	E	10	<div> <div></div> <div>20%70%10%</div> </div>
2	C	10	<div> <div></div> <div>60%40%</div> </div>
2	F	10	<div> <div></div> <div>60%40%</div> </div>
3	A	421	<div> <div>3%</div> <div>86%7%7%</div> </div>
3	D	421	<div> <div>4%</div> <div>87%6%7%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8673 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 DNA chain called 5'-D(\*GP\*TP\*TP\*CP\*GP\*AP\*TP\*GP\*TP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	10	Total	C	N	O	P	0	0	0
			202	98	34	61	9			
1	E	10	Total	C	N	O	P	0	0	0
			202	98	34	61	9			

- Molecule 2 is a DNA chain called 5'-D(\*GP\*AP\*CP\*AP\*(5PY)P\*CP\*GP\*(6MA)P\*AP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	P	0	0	0
			202	98	41	54	9			
2	F	10	Total	C	N	O	P	0	0	0
			202	98	41	54	9			

- Molecule 3 is a protein called Modification methylase TaqI.

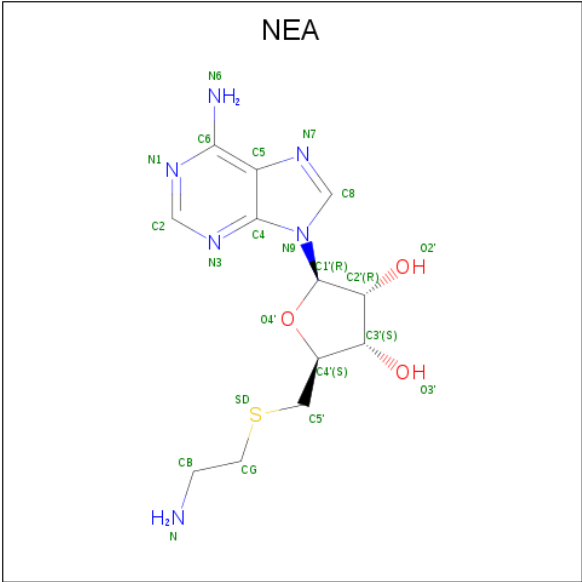
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	393	Total	C	N	O	S	0	0	0
			3189	2087	548	548	6			
3	D	393	Total	C	N	O	S	0	0	0
			3189	2087	548	548	6			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

- Molecule 5 is 5'-DEOXY-5'-[2-(AMINO)ETHYLTHIO]ADENOSINE (three-letter code: NEA) (formula: C<sub>12</sub>H<sub>18</sub>N<sub>6</sub>O<sub>3</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	S	0	0
			22	12	6	3	1		
5	D	1	Total	C	N	O	S	0	0
			22	12	6	3	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	70	Total	O	0	0
			70	70		
6	C	63	Total	O	0	0
			63	63		
6	E	56	Total	O	0	0
			56	56		
6	F	47	Total	O	0	0
			47	47		
6	A	610	Total	O	0	0
			610	610		
6	D	567	Total	O	0	0
			567	567		

### 3 Residue-property plots [i](#)

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: 5'-D(\*GP\*TP\*TP\*CP\*GP\*AP\*TP\*GP\*TP\*C)-3'



- Molecule 1: 5'-D(\*GP\*TP\*TP\*CP\*GP\*AP\*TP\*GP\*TP\*C)-3'



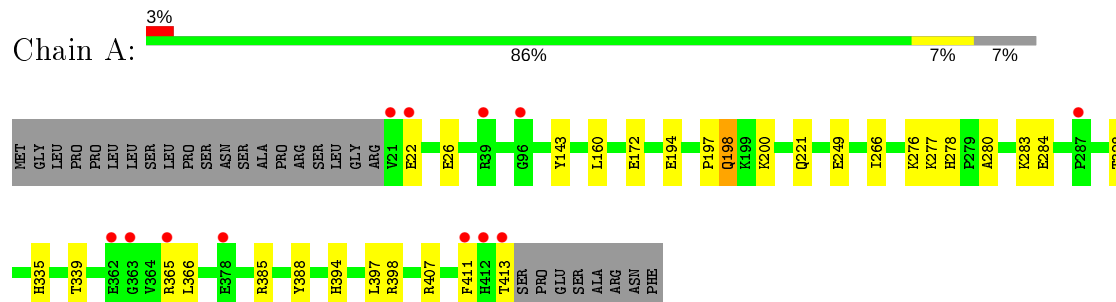
- Molecule 2: 5'-D(\*GP\*AP\*CP\*AP\*(5PY)P\*CP\*GP\*(6MA)P\*AP\*C)-3'



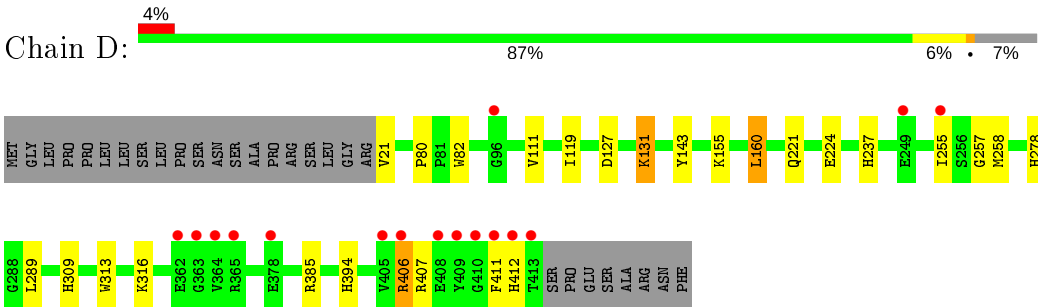
- Molecule 2: 5'-D(\*GP\*AP\*CP\*AP\*(5PY)P\*CP\*GP\*(6MA)P\*AP\*C)-3'



- Molecule 3: Modification methylase TaqI



- Molecule 3: Modification methylase TaqI



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.44Å 68.88Å 114.35Å 90.00° 92.27° 90.00°	Depositor
Resolution (Å)	19.64 – 1.61 19.66 – 1.61	Depositor EDS
% Data completeness (in resolution range)	95.9 (19.64-1.61) 98.1 (19.66-1.61)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.02 (at 1.61Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.160 , 0.195 0.160 , 0.194	Depositor DCC
$R_{free}$ test set	5827 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	10.2	Xtriage
Anisotropy	0.162	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 55.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.021 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	8673	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.52% of the height of the origin peak. No significant pseudotranslation is detected.*

<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, 0.333 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: NEA, GOL, 5PY, 6MA

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	B	1.00	0/225	1.89	4/346 (1.2%)
1	E	0.95	0/225	1.89	8/346 (2.3%)
2	C	1.02	0/179	1.72	5/269 (1.9%)
2	F	0.88	0/179	1.70	3/269 (1.1%)
3	A	0.50	0/3293	0.63	0/4475
3	D	0.48	0/3293	0.62	1/4475 (0.0%)
All	All	0.56	0/7394	0.86	21/10180 (0.2%)

There are no bond length outliers.

All (21) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	E	8	DG	O4'-C1'-N9	-13.48	98.56	108.00
1	E	6	DA	O4'-C1'-N9	-11.98	99.61	108.00
1	B	6	DA	O4'-C1'-N9	-10.03	100.98	108.00
2	F	20	DC	O4'-C1'-N1	9.67	114.77	108.00
2	F	12	DA	O4'-C1'-N9	-9.45	101.38	108.00
1	E	4	DC	O4'-C1'-N1	-7.64	102.65	108.00
2	C	12	DA	O4'-C1'-N9	-7.06	103.06	108.00
1	B	8	DG	O4'-C1'-N9	-7.03	103.08	108.00
3	D	160	LEU	CA-CB-CG	6.98	131.35	115.30
1	E	3	DT	O4'-C1'-N1	-6.68	103.32	108.00
1	E	1	DG	P-O3'-C3'	6.20	127.14	119.70
1	B	4	DC	O4'-C1'-N1	-6.18	103.67	108.00
1	B	3	DT	O4'-C1'-N1	-6.01	103.80	108.00
2	C	13	DC	O4'-C1'-N1	5.74	112.02	108.00
1	E	10	DC	O4'-C1'-N1	5.71	112.00	108.00
2	C	20	DC	O4'-C1'-N1	5.64	111.95	108.00
2	C	13	DC	C1'-O4'-C4'	-5.33	104.77	110.10
1	E	5	DG	C5-C6-N1	5.24	114.12	111.50

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	E	9	DT	C6-C5-C7	-5.22	119.77	122.90
2	C	17	DG	N9-C4-C5	5.17	107.47	105.40
2	F	13	DC	O4'-C1'-C2'	-5.06	101.85	105.90

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	B	202	0	116	1	0
1	E	202	0	116	1	0
2	C	202	0	115	0	0
2	F	202	0	115	1	0
3	A	3189	0	3184	30	0
3	D	3189	0	3184	27	0
4	A	6	0	8	0	0
4	C	6	0	8	0	0
4	D	18	0	24	2	0
5	A	22	0	18	0	0
5	D	22	0	18	0	0
6	A	610	0	0	15	0
6	B	70	0	0	2	0
6	C	63	0	0	0	0
6	D	567	0	0	15	0
6	E	56	0	0	0	0
6	F	47	0	0	1	0
All	All	8673	0	6906	58	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 4.

All (58) 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)
3:D:411:PHE:HB2	6:D:3565:HOH:O	1.53	1.06
6:B:79:HOH:O	3:A:200:LYS:HD2	1.62	0.97
3:D:131:LYS:HE3	6:D:3570:HOH:O	1.80	0.80
3:D:221:GLN:HA	4:D:3005:GOL:H12	1.63	0.80
3:A:339:THR:HG22	6:A:3610:HOH:O	1.83	0.78
3:D:127:ASP:O	3:D:131:LYS:HE2	1.82	0.77
1:B:5:DG:N7	3:A:394:HIS:HE1	1.86	0.73
1:E:5:DG:N7	3:D:394:HIS:HE1	1.86	0.72
3:A:278:HIS:HD2	3:A:280:ALA:H	1.38	0.71
3:D:155:LYS:HE3	6:D:3533:HOH:O	1.92	0.70
3:A:365:ARG:HH21	3:A:411:PHE:HB3	1.57	0.69
3:A:26:GLU:HG2	6:A:3241:HOH:O	1.91	0.69
3:D:258:MET:HE1	3:D:411:PHE:HA	1.78	0.65
3:A:194:GLU:HB2	6:A:3609:HOH:O	1.97	0.65
3:D:278:HIS:CD2	3:D:280:ALA:H	2.15	0.65
3:A:278:HIS:CD2	3:A:280:ALA:H	2.15	0.64
3:A:328:THR:HB	6:A:3384:HOH:O	1.99	0.61
3:A:26:GLU:CG	6:A:3241:HOH:O	2.48	0.61
3:D:278:HIS:HD2	3:D:280:ALA:H	1.48	0.59
2:F:11:DG:N7	6:F:66:HOH:O	2.32	0.58
3:A:365:ARG:HH12	3:A:413:THR:HA	1.68	0.58
3:D:21:VAL:N	6:D:3369:HOH:O	2.35	0.58
3:A:385:ARG:HA	6:A:3610:HOH:O	2.05	0.57
3:D:258:MET:HE1	3:D:412:HIS:H	1.71	0.55
3:A:266:ILE:HG21	3:A:397:LEU:HD13	1.89	0.55
3:D:309:HIS:HE1	6:D:3192:HOH:O	1.89	0.55
3:A:385:ARG:HG3	6:A:3227:HOH:O	2.07	0.54
3:D:257:GLY:HA3	6:D:3566:HOH:O	2.07	0.54
3:D:385:ARG:HG3	6:D:3557:HOH:O	2.06	0.54
3:D:224:GLU:HG2	6:D:3567:HOH:O	2.07	0.54
3:A:276:LYS:HE3	6:A:3586:HOH:O	2.07	0.53
3:D:406:ARG:HD3	6:D:3216:HOH:O	2.10	0.51
3:D:407:ARG:HD2	6:D:3034:HOH:O	2.11	0.51
3:D:255:ILE:HG13	6:D:3246:HOH:O	2.10	0.50
3:A:385:ARG:HG2	6:A:3319:HOH:O	2.12	0.50
3:A:249:GLU:HG2	6:A:3216:HOH:O	2.12	0.48
3:D:131:LYS:NZ	6:D:3443:HOH:O	2.44	0.47
3:D:221:GLN:HA	4:D:3005:GOL:C1	2.40	0.47
3:A:198:GLN:NE2	6:A:3064:HOH:O	2.47	0.47
3:D:309:HIS:HD2	6:D:3363:HOH:O	1.96	0.47
3:A:197:PRO:O	3:A:198:GLN:HG2	2.14	0.46
6:B:79:HOH:O	3:A:200:LYS:CD	2.40	0.46

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:385:ARG:HD3	6:A:3612:HOH:O	2.15	0.45
3:D:316:LYS:HE2	6:D:3539:HOH:O	2.15	0.45
3:D:289:LEU:HB3	3:D:313:TRP:HB3	1.98	0.45
3:D:80:PRO:HG2	3:D:82:TRP:CZ2	2.52	0.45
3:D:258:MET:CE	3:D:411:PHE:HA	2.47	0.45
3:A:276:LYS:CE	6:A:3586:HOH:O	2.63	0.45
3:A:365:ARG:NH2	3:A:411:PHE:CD2	2.85	0.45
3:A:277:LYS:NZ	6:A:3244:HOH:O	2.50	0.44
3:A:335:HIS:O	3:A:394:HIS:HD2	2.01	0.43
3:A:407:ARG:HB2	3:A:411:PHE:CD1	2.54	0.43
3:D:237:HIS:HE1	6:D:3440:HOH:O	2.02	0.43
3:A:339:THR:HG21	3:A:388:TYR:HB2	2.01	0.42
3:A:26:GLU:CD	3:A:26:GLU:H	2.24	0.41
3:A:172:GLU:OE2	3:A:398:ARG:NE	2.46	0.41
3:D:111:VAL:HA	3:D:119:ILE:HB	2.03	0.40
3:A:398:ARG:NH1	6:A:3604:HOH:O	2.54	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	
3	A	391/421 (93%)	379 (97%)	12 (3%)	0	100	100
3	D	391/421 (93%)	379 (97%)	12 (3%)	0	100	100
All	All	782/842 (93%)	758 (97%)	24 (3%)	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	
3	A	332/356 (93%)	324 (98%)	8 (2%)	49	22
3	D	332/356 (93%)	328 (99%)	4 (1%)	71	52
All	All	664/712 (93%)	652 (98%)	12 (2%)	59	34

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	22	GLU
3	A	143	TYR
3	A	160	LEU
3	A	198	GLN
3	A	221	GLN
3	A	283	LYS
3	A	284	GLU
3	A	366	LEU
3	D	131	LYS
3	D	143	TYR
3	D	160	LEU
3	D	406	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	198	GLN
3	A	221	GLN
3	A	278	HIS
3	A	394	HIS
3	D	237	HIS
3	D	278	HIS
3	D	309	HIS
3	D	394	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

4 non-standard protein/DNA/RNA residues 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	6MA	F	18	1,2	18,24,25	0.78	0	15,34,37	1.31	1 (6%)
2	5PY	C	15	2	14,20,21	1.07	1 (7%)	18,28,31	2.54	2 (11%)
2	5PY	F	15	2	14,20,21	1.57	3 (21%)	18,28,31	2.45	3 (16%)
2	6MA	C	18	1,2	18,24,25	0.77	0	15,34,37	1.29	1 (6%)

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	6MA	F	18	1,2	-	0/5/23/24	0/3/3/3
2	5PY	C	15	2	-	4/4/21/22	0/2/2/2
2	5PY	F	15	2	-	4/4/21/22	0/2/2/2
2	6MA	C	18	1,2	-	0/5/23/24	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	15	5PY	O5'-C5'	-4.25	1.34	1.44
2	F	15	5PY	C4-C5	2.75	1.45	1.38
2	C	15	5PY	C4-C5	2.31	1.44	1.38
2	F	15	5PY	C6-C5	-2.03	1.35	1.39

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	15	5PY	C4-N3-C2	7.42	127.60	115.36
2	C	15	5PY	C5-C4-N3	-7.27	116.26	124.81
2	F	15	5PY	C4-N3-C2	6.79	126.56	115.36
2	F	15	5PY	C5-C4-N3	-6.60	117.05	124.81
2	F	18	6MA	C2-N1-C6	4.59	120.52	116.59
2	C	18	6MA	C2-N1-C6	4.59	120.52	116.59
2	F	15	5PY	O4'-C4'-C5'	-2.84	100.03	109.37

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	15	5PY	O4'-C1'-N1-C6
2	F	15	5PY	O4'-C4'-C5'-O5'
2	F	15	5PY	O4'-C1'-N1-C6
2	C	15	5PY	O4'-C4'-C5'-O5'
2	F	15	5PY	C3'-C4'-C5'-O5'
2	C	15	5PY	C3'-C4'-C5'-O5'
2	C	15	5PY	C4'-C5'-O5'-P
2	F	15	5PY	C4'-C5'-O5'-P

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

7 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
4	GOL	D	3004	-	5,5,5	0.40	0	5,5,5	0.42	0
4	GOL	D	3003	-	5,5,5	0.39	0	5,5,5	0.42	0
5	NEA	D	2002	-	21,24,24	1.18	2 (9%)	21,34,34	1.60	2 (9%)
5	NEA	A	2001	-	21,24,24	1.07	2 (9%)	21,34,34	1.65	2 (9%)
4	GOL	A	3002	-	5,5,5	0.42	0	5,5,5	0.24	0
4	GOL	C	3001	-	5,5,5	0.51	0	5,5,5	1.14	0
4	GOL	D	3005	-	5,5,5	0.54	0	5,5,5	0.77	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	GOL	D	3004	-	-	2/4/4/4	-
4	GOL	D	3003	-	-	0/4/4/4	-
5	NEA	D	2002	-	-	0/5/25/25	0/3/3/3
5	NEA	A	2001	-	-	0/5/25/25	0/3/3/3
4	GOL	A	3002	-	-	0/4/4/4	-
4	GOL	C	3001	-	-	2/4/4/4	-
4	GOL	D	3005	-	-	4/4/4/4	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	2002	NEA	C2-N3	3.83	1.38	1.32
5	A	2001	NEA	C2-N3	3.43	1.37	1.32
5	A	2001	NEA	C2-N1	2.35	1.38	1.33
5	D	2002	NEA	C2-N1	2.30	1.38	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	2001	NEA	N3-C2-N1	-6.29	118.85	128.68
5	D	2002	NEA	N3-C2-N1	-5.12	120.68	128.68
5	D	2002	NEA	O4'-C1'-C2'	-3.21	102.24	106.93
5	A	2001	NEA	O4'-C1'-C2'	-2.19	103.72	106.93

There are no chirality outliers.

All (8) torsion outliers are listed below:

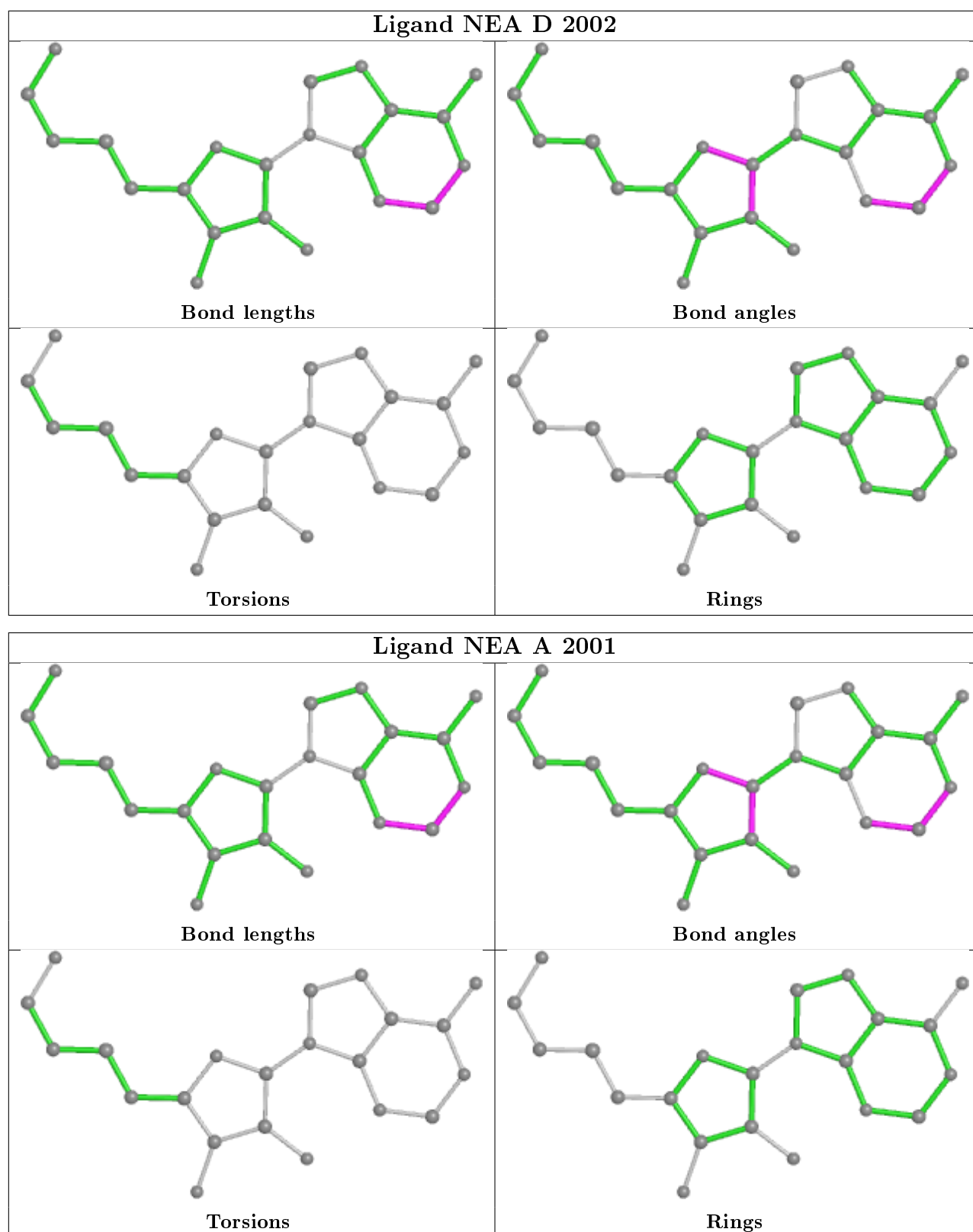
Mol	Chain	Res	Type	Atoms
4	D	3004	GOL	O1-C1-C2-C3
4	C	3001	GOL	C1-C2-C3-O3
4	D	3005	GOL	C1-C2-C3-O3
4	D	3005	GOL	O2-C2-C3-O3
4	D	3005	GOL	O1-C1-C2-C3
4	C	3001	GOL	O2-C2-C3-O3
4	D	3004	GOL	O1-C1-C2-O2
4	D	3005	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	3005	GOL	2	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 ⓘ

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, 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	B	10/10 (100%)	-0.83	0 100 100	4, 6, 11, 12	0
1	E	10/10 (100%)	-0.76	0 100 100	5, 8, 14, 18	0
2	C	8/10 (80%)	-0.66	0 100 100	5, 7, 10, 15	0
2	F	8/10 (80%)	-0.36	0 100 100	7, 9, 20, 22	0
3	A	393/421 (93%)	-0.21	12 (3%) 49 46	4, 10, 23, 45	0
3	D	393/421 (93%)	-0.10	17 (4%) 35 31	5, 11, 24, 38	0
All	All	822/882 (93%)	-0.18	29 (3%) 44 40	4, 11, 24, 45	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	413	THR	8.7
3	D	413	THR	8.1
3	A	21	VAL	4.6
3	D	363	GLY	4.0
3	D	365	ARG	4.0
3	A	411	PHE	4.0
3	A	412	HIS	3.6
3	D	255	ILE	3.5
3	A	362	GLU	3.4
3	A	287	PRO	3.2
3	D	411	PHE	3.1
3	A	365	ARG	3.0
3	A	363	GLY	2.9
3	A	39	ARG	2.8
3	D	412	HIS	2.8
3	D	362	GLU	2.6
3	A	96	GLY	2.5
3	D	408	GLU	2.5
3	D	287	PRO	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
3	D	378	GLU	2.5
3	D	405	VAL	2.5
3	D	364	VAL	2.4
3	A	378	GLU	2.4
3	D	406	ARG	2.3
3	D	410	GLY	2.2
3	D	96	GLY	2.2
3	D	249	GLU	2.2
3	D	409	TYR	2.1
3	A	22	GLU	2.1

## 6.2 Non-standard residues in protein, DNA, RNA chains [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, 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	6MA	F	18	22/23	0.98	0.05	4,6,11,13	0
2	5PY	F	15	19/20	0.98	0.07	5,7,11,14	0
2	5PY	C	15	19/20	0.99	0.06	3,5,7,7	0
2	6MA	C	18	22/23	0.99	0.06	3,5,8,10	0

## 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, 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	B-factors(Å <sup>2</sup> )	Q<0.9
4	GOL	D	3004	6/6	0.60	0.23	27,29,30,30	0
4	GOL	D	3005	6/6	0.71	0.30	24,27,29,31	0
4	GOL	D	3003	6/6	0.89	0.12	14,18,21,23	0
4	GOL	C	3001	6/6	0.92	0.11	13,16,17,17	0

*Continued on next page...*

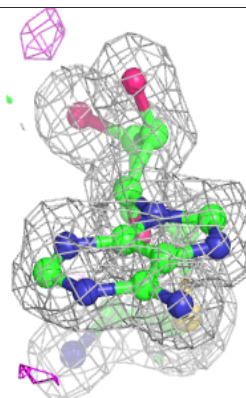
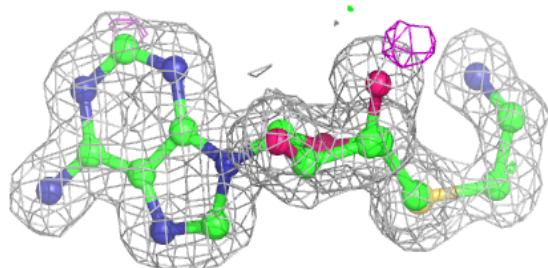
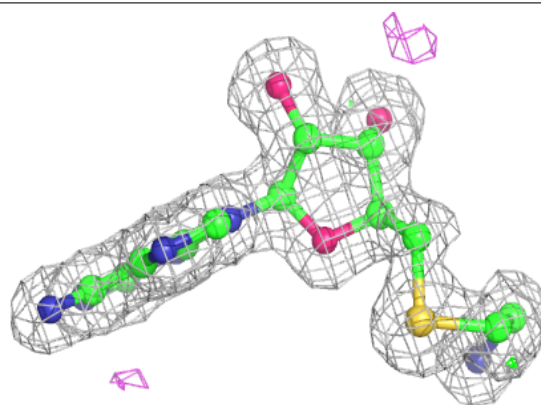
*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	A	3002	6/6	0.94	0.10	12,16,18,19	0
5	NEA	D	2002	22/22	0.98	0.05	6,7,9,10	0
5	NEA	A	2001	22/22	0.98	0.05	6,7,9,9	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.

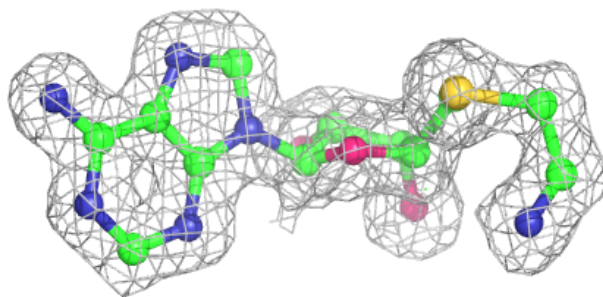
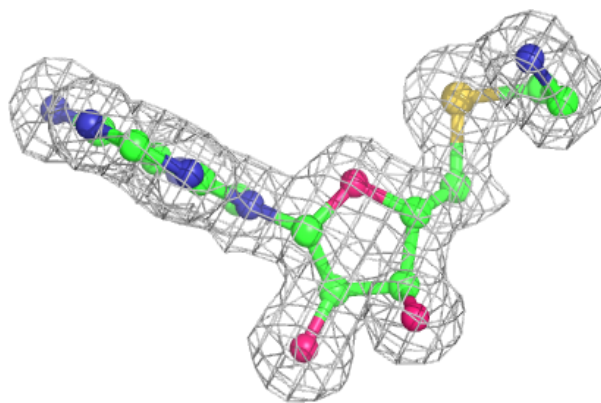
**Electron density around NEA D 2002:**

2mF<sub>o</sub>-DF<sub>c</sub> (at 0.7 rmsd) in gray  
mF<sub>o</sub>-DF<sub>c</sub> (at 3 rmsd) in purple (negative)  
and green (positive)



**Electron density around NEA A 2001:**

$2mF_o - DF_c$  (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.