



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2017 – 12:38 PM EDT

PDB ID : 5W0U  
Title : Crystal structure of MBP fused activation-induced cytidine deaminase (AID)  
in complex with dCMP  
Authors : Qiao, Q.; Wang, L.; Wu, H.  
Deposited on : unknown  
Resolution : 2.90 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20029824

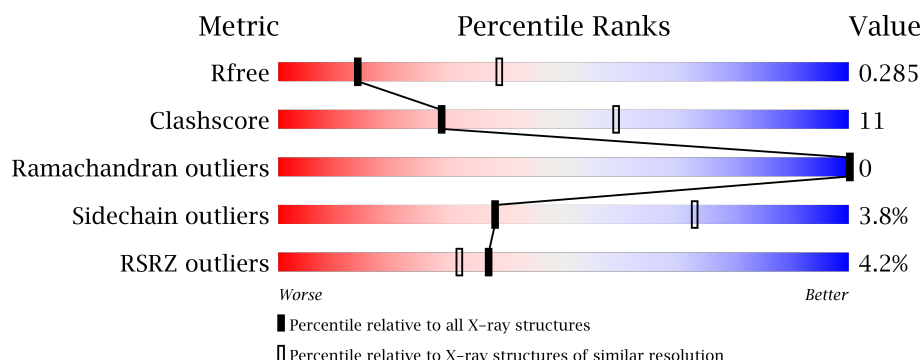
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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	549	<div> <div>5%</div> <div>77%</div> <div>19%</div> <div>..</div> </div>
1	B	549	<div> <div>3%</div> <div>78%</div> <div>19%</div> <div>..</div> </div>
2	D	12	<div> <div>17%</div> <div>42%</div> <div>58%</div> </div>
3	G	12	<div> <div>17%</div> <div>58%</div> <div>42%</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	DCM	A	2002	-	-	X	-
5	DCM	B	2002	-	-	X	-

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 9130 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 Maltose-binding periplasmic protein, Single-stranded DNA cytosine deaminase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	542	Total	C	N	O	S	0	0	0
			4298	2762	726	795	15			
1	A	542	Total	C	N	O	S	0	0	0
			4297	2762	726	794	15			

There are 58 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP P0AEY0
B	83	ALA	ASP	engineered mutation	UNP P0AEY0
B	84	ALA	LYS	engineered mutation	UNP P0AEY0
B	173	ALA	GLU	engineered mutation	UNP P0AEY0
B	174	ALA	ASN	engineered mutation	UNP P0AEY0
B	240	ALA	LYS	engineered mutation	UNP P0AEY0
B	360	ALA	GLU	engineered mutation	UNP P0AEY0
B	363	ALA	LYS	engineered mutation	UNP P0AEY0
B	364	ALA	ASP	engineered mutation	UNP P0AEY0
B	368	ASN	-	linker	UNP P0AEY0
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0
B	371	ALA	-	linker	UNP P0AEY0
B	372	GLU	-	linker	UNP P0AEY0
B	373	PHE	-	linker	UNP P0AEY0
B	1006	MET	-	linker	UNP P0AEY0
B	1007	ASP	-	linker	UNP P0AEY0
B	1008	PRO	-	linker	UNP P0AEY0
B	1009	ALA	-	linker	UNP P0AEY0
B	1010	THR	-	linker	UNP P0AEY0
B	1011	PHE	-	linker	UNP P0AEY0
B	1012	THR	-	linker	UNP P0AEY0
B	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1058	ALA	GLU	engineered mutation	UNP Q9GZX7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	1130	ALA	HIS	engineered mutation	UNP Q9GZX7
B	1131	GLU	ARG	engineered mutation	UNP Q9GZX7
B	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
B	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
B	1181	GLN	LEU	engineered mutation	UNP Q9GZX7
A	1	MET	-	initiating methionine	UNP P0AEY0
A	83	ALA	ASP	engineered mutation	UNP P0AEY0
A	84	ALA	LYS	engineered mutation	UNP P0AEY0
A	173	ALA	GLU	engineered mutation	UNP P0AEY0
A	174	ALA	ASN	engineered mutation	UNP P0AEY0
A	240	ALA	LYS	engineered mutation	UNP P0AEY0
A	360	ALA	GLU	engineered mutation	UNP P0AEY0
A	363	ALA	LYS	engineered mutation	UNP P0AEY0
A	364	ALA	ASP	engineered mutation	UNP P0AEY0
A	368	ASN	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
A	371	ALA	-	linker	UNP P0AEY0
A	372	GLU	-	linker	UNP P0AEY0
A	373	PHE	-	linker	UNP P0AEY0
A	1006	MET	-	linker	UNP P0AEY0
A	1007	ASP	-	linker	UNP P0AEY0
A	1008	PRO	-	linker	UNP P0AEY0
A	1009	ALA	-	linker	UNP P0AEY0
A	1010	THR	-	linker	UNP P0AEY0
A	1011	PHE	-	linker	UNP P0AEY0
A	1012	THR	-	linker	UNP P0AEY0
A	1042	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1058	ALA	GLU	engineered mutation	UNP Q9GZX7
A	1130	ALA	HIS	engineered mutation	UNP Q9GZX7
A	1131	GLU	ARG	engineered mutation	UNP Q9GZX7
A	1141	TYR	PHE	engineered mutation	UNP Q9GZX7
A	1145	GLU	PHE	engineered mutation	UNP Q9GZX7
A	1181	GLN	LEU	engineered mutation	UNP Q9GZX7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			243	116	48	68	11			

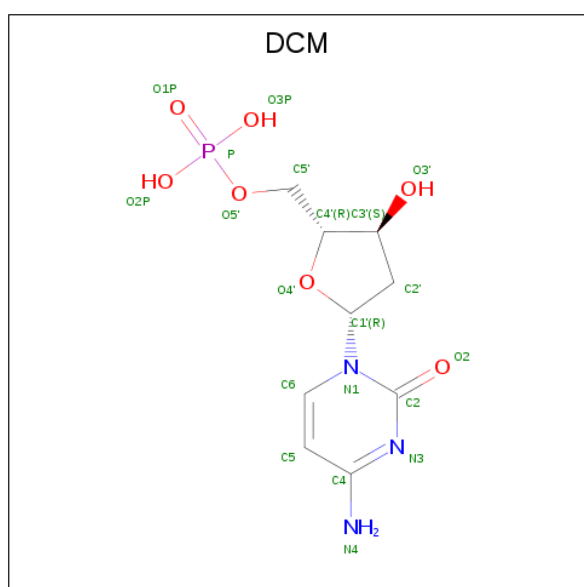
- Molecule 3 is a DNA chain called DNA (5'-D(\*CP\*TP\*GP\*GP\*CP\*CP\*TP\*TP\*GP\*AP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	12	Total	C	N	O	P	0	0	0
			240	115	43	71	11			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Zn	0	0
			1	1		
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is 2'-DEOXYCYTIDINE-5'-MONOPHOSPHATE (three-letter code: DCM) (formula: C<sub>9</sub>H<sub>14</sub>N<sub>3</sub>O<sub>7</sub>P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total	C	N	O	P	0	0
			20	9	3	7	1		
5	A	1	Total	C	N	O	P	0	0
			20	9	3	7	1		

- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	B	1	Total Ca 1 1	0	0
6	A	1	Total Ca 1 1	0	0

- Molecule 7 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
7	B	1	Total C O 6 3 3	0	0

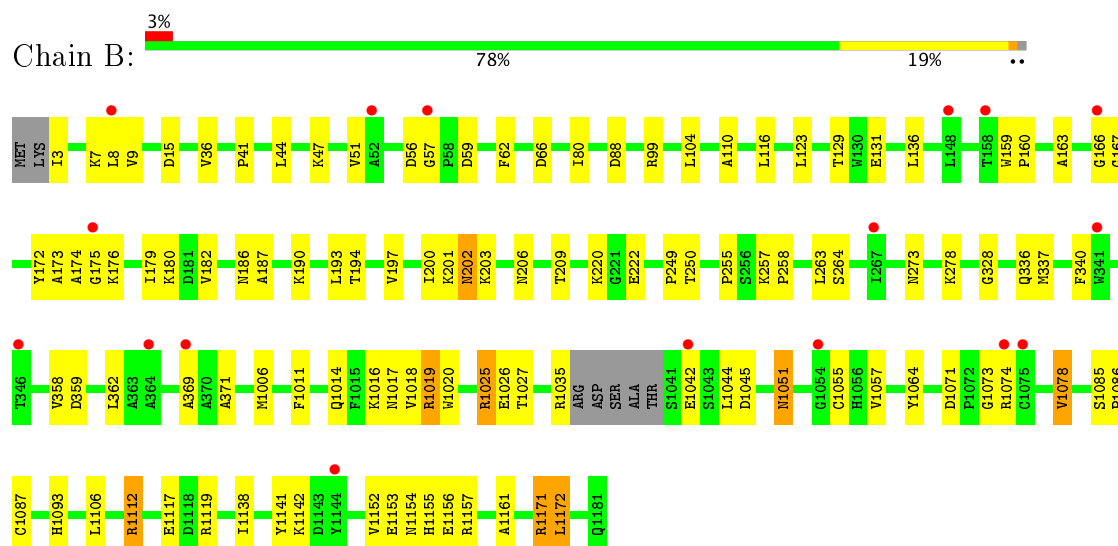
- Molecule 8 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	B	1	Total O 1 1	0	0
8	A	1	Total O 1 1	0	0

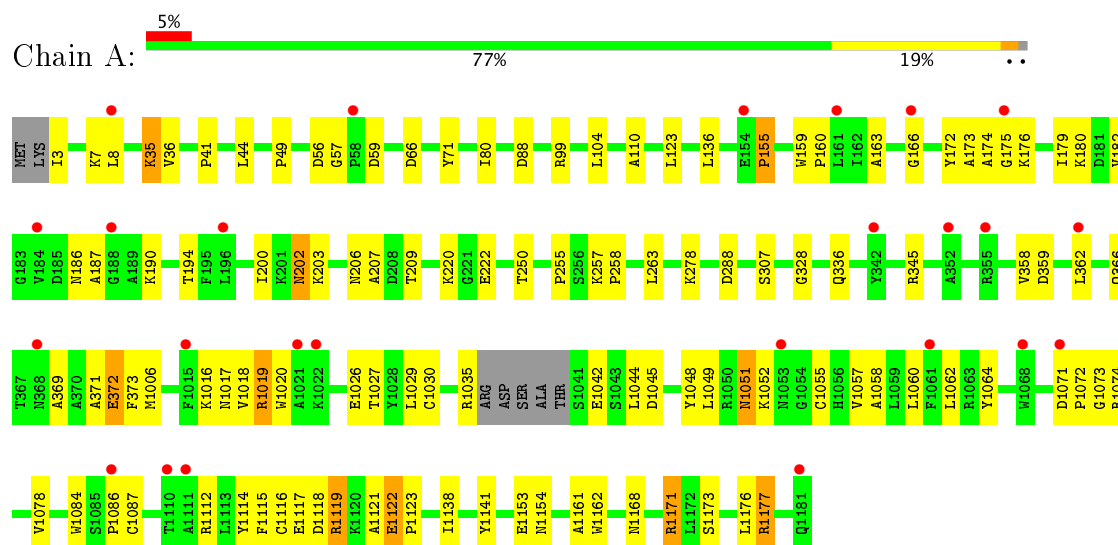
### 3 Residue-property plots [i](#)

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.

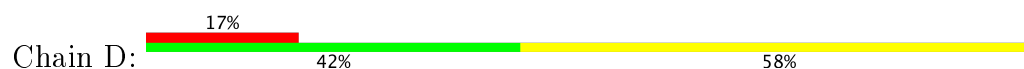
- Molecule 1: Maltose-binding periplasmic protein,Single-stranded DNA cytosine deaminase



- Molecule 1: Maltose-binding periplasmic protein,Single-stranded DNA cytosine deaminase



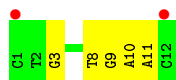
- Molecule 2: DNA (5'-D(\*GP\*TP\*TP\*CP\*AP\*AP\*GP\*GP\*CP\*CP\*AP\*G)-3')







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



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.00Å 39.78Å 153.94Å 90.00° 90.14° 90.00°	Depositor
Resolution (Å)	153.94 – 2.90 153.93 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (153.94-2.90) 98.6 (153.93-2.90)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.62 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, $R_{free}$	0.251 , 0.281 0.250 , 0.285	Depositor DCC
$R_{free}$ test set	1344 reflections (3.93%)	DCC
Wilson B-factor (Å <sup>2</sup> )	82.8	Xtriage
Anisotropy	0.697	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 60.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9130	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	100.0	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 58.26 % 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 2.0998e-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, 0.333 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: GOL, ZN, DCM, CA

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	0.41	0/4408	0.46	0/5982
1	B	0.39	0/4409	0.47	0/5983
2	D	0.63	0/273	0.94	0/420
3	G	0.65	0/267	0.98	0/408
All	All	0.41	0/9357	0.51	0/12793

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	4297	0	4207	98	0
1	B	4298	0	4210	87	0
2	D	243	0	132	6	0
3	G	240	0	133	4	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	20	0	12	12	0
5	B	20	0	12	8	0
6	A	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	1	0	0	0	0
7	B	6	0	8	2	0
8	A	1	0	0	2	0
8	B	1	0	0	1	0
All	All	9130	0	8714	196	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 11.

The worst 5 of 196 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:1086:PRO:HA	5:A:2002:DCM:N4	1.69	1.08
1:B:1027:THR:OG1	5:B:2002:DCM:O2	1.90	0.89
5:A:2002:DCM:C4	8:A:4001:HOH:O	2.21	0.88
1:A:1086:PRO:HA	5:A:2002:DCM:HN42	1.42	0.83
1:B:1017:ASN:O	1:B:1017:ASN:OD1	1.97	0.82

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	538/549 (98%)	521 (97%)	17 (3%)	0	100	100
1	B	538/549 (98%)	519 (96%)	19 (4%)	0	100	100
All	All	1076/1098 (98%)	1040 (97%)	36 (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	
1	A	438 / 445 (98%)	419 (96%)	19 (4%)	33	68
1	B	439 / 445 (99%)	425 (97%)	14 (3%)	44	78
All	All	877 / 890 (98%)	844 (96%)	33 (4%)	38	73

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	35	LYS
1	A	176	LYS
1	A	1154	ASN
1	A	66	ASP
1	A	88	ASP

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	A	1017	ASN
1	A	1175	GLN
1	A	1051	ASN
1	B	1051	ASN
1	A	1168	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 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	DCM	A	2002	-	18,21,21	0.60	0	24,31,31	1.17	1 (4%)
5	DCM	B	2002	-	18,21,21	0.61	0	24,31,31	1.18	2 (8%)
7	GOL	B	2004	-	5,5,5	0.35	0	5,5,5	0.16	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	DCM	A	2002	-	-	0/6/22/22	0/2/2/2
5	DCM	B	2002	-	-	0/6/22/22	0/2/2/2
7	GOL	B	2004	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	2002	DCM	N4-C4-N3	2.02	120.05	116.64
5	A	2002	DCM	O4'-C1'-N1	2.24	111.56	107.78
5	B	2002	DCM	O4'-C1'-N1	2.26	111.59	107.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	2002	DCM	12	0
5	B	2002	DCM	8	0
7	B	2004	GOL	2	0

## 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, 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	542/549 (98%)	0.47	25 (4%) 33 28	64, 94, 123, 148	0
1	B	542/549 (98%)	0.42	17 (3%) 49 43	68, 99, 124, 149	0
2	D	12/12 (100%)	1.32	2 (16%) 2 1	134, 153, 165, 167	0
3	G	12/12 (100%)	1.43	2 (16%) 2 1	144, 159, 168, 174	0
All	All	1108/1122 (98%)	0.47	46 (4%) 37 32	64, 97, 131, 174	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1181	GLN	4.4
1	A	1021	ALA	4.3
1	B	1054	GLY	3.8
1	B	1074	ARG	3.8
1	A	362	LEU	3.8

### 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( $\text{\AA}^2$ )	Q<0.9
7	GOL	B	2004	6/6	0.74	0.26	1.32	85,100,103,105	0
5	DCM	B	2002	20/20	0.80	0.24	0.12	77,122,152,179	0
6	CA	B	2003	1/1	0.56	0.20	0.06	110,110,110,110	0
6	CA	A	2003	1/1	0.94	0.22	-0.24	97,97,97,97	0
5	DCM	A	2002	20/20	0.86	0.23	-0.28	92,114,181,189	0
4	ZN	A	2001	1/1	0.46	0.28	-	166,166,166,166	0
4	ZN	B	2001	1/1	0.90	0.22	-	115,115,115,115	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.