



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2020 – 04:08 AM BST

PDB ID : 4NZ5  
Title : Structure of Vibrio cholerae chitin de-N-acetylase in complex with 2-(ACETYLAMINO)-2-DEOXY-A-D-GLUCOPYRANOSE (NDG) and cadmium ion  
Authors : Albesa-Jove, D.; Andres, E.; Biarnes, X.; Planas, A.; Guerin, M.E.  
Deposited on : 2013-12-11  
Resolution : 1.87 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 : 2.13.1  
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) : 2.13.1

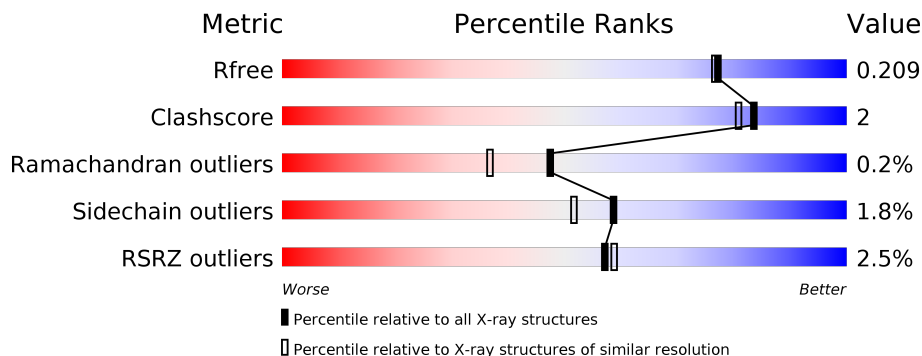
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.87 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	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)
RSRZ outliers	127900	9303 (1.90-1.86)

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	A	419	
1	B	419	

## 2 Entry composition i

There are 7 unique types of molecules in this entry. The entry contains 6992 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 Deacetylase DA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	404	3159	2008	529	611	11	0	3	1
1	B	404	3140	1998	528	604	10	0	0	1

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	MET	-	expression tag	UNP M7M1G8
A	432	GLU	-	expression tag	UNP M7M1G8
A	433	LEU	-	expression tag	UNP M7M1G8
A	434	ARG	-	expression tag	UNP M7M1G8
A	435	ARG	-	expression tag	UNP M7M1G8
A	436	GLN	-	expression tag	UNP M7M1G8
A	437	TRP	-	expression tag	UNP M7M1G8
A	438	SER	-	expression tag	UNP M7M1G8
A	439	HIS	-	expression tag	UNP M7M1G8
A	440	PRO	-	expression tag	UNP M7M1G8
A	441	GLN	-	expression tag	UNP M7M1G8
A	442	PHE	-	expression tag	UNP M7M1G8
A	443	GLU	-	expression tag	UNP M7M1G8
A	444	LYS	-	expression tag	UNP M7M1G8
B	26	MET	-	expression tag	UNP M7M1G8
B	432	GLU	-	expression tag	UNP M7M1G8
B	433	LEU	-	expression tag	UNP M7M1G8
B	434	ARG	-	expression tag	UNP M7M1G8
B	435	ARG	-	expression tag	UNP M7M1G8
B	436	GLN	-	expression tag	UNP M7M1G8
B	437	TRP	-	expression tag	UNP M7M1G8
B	438	SER	-	expression tag	UNP M7M1G8
B	439	HIS	-	expression tag	UNP M7M1G8
B	440	PRO	-	expression tag	UNP M7M1G8
B	441	GLN	-	expression tag	UNP M7M1G8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	442	PHE	-	expression tag	UNP M7M1G8
B	443	GLU	-	expression tag	UNP M7M1G8
B	444	LYS	-	expression tag	UNP M7M1G8

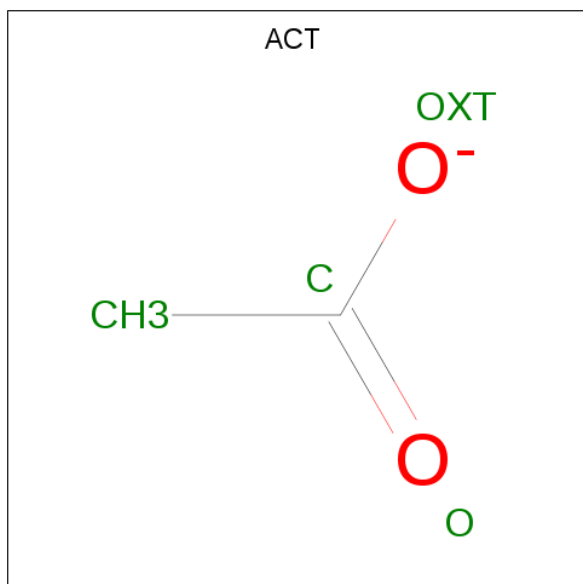
- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Na 1 1	0	0
2	A	1	Total Na 1 1	0	0

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: Cd).

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

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



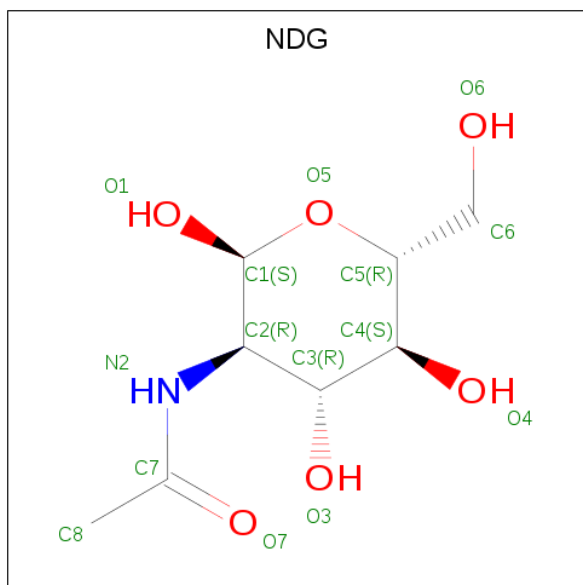
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 4 2 2	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 5 is 2-acetamido-2-deoxy- $\alpha$ -D-glucopyranose (three-letter code: NDG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			15	8	1	6		
5	B	1	Total	C	N	O	0	0
			15	8	1	6		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C O 12 6 6	0	1
6	B	1	Total C O 6 3 3	0	0
6	B	1	Total C O 6 3 3	0	0

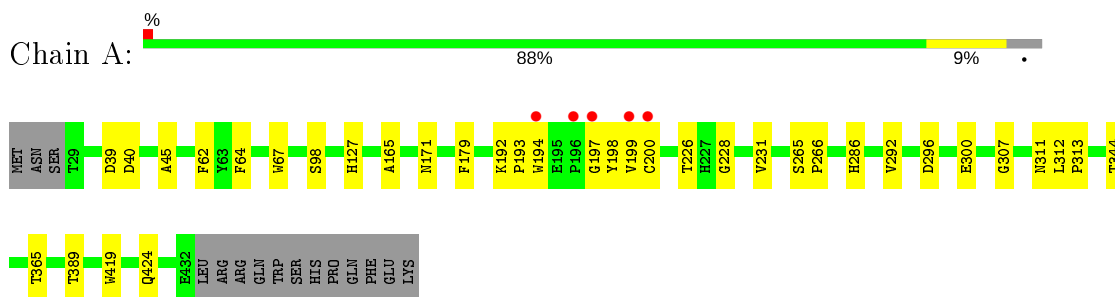
- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	311	Total O 316 316	0	4
7	B	306	Total O 311 311	0	5

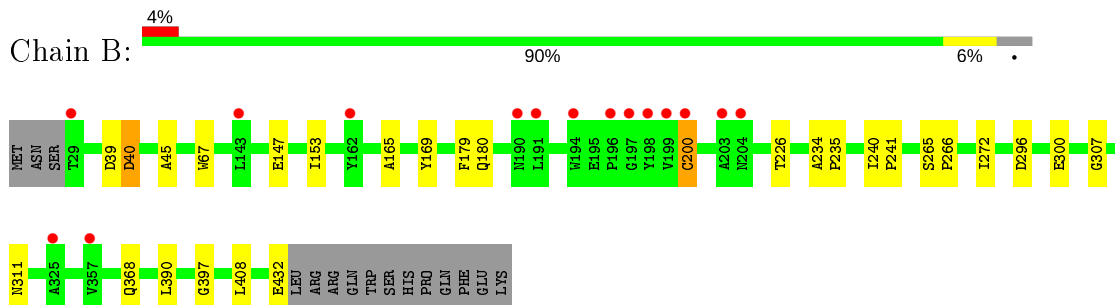
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Deacetylase DA1



- Molecule 1: Deacetylase DA1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.66Å 112.53Å 69.81Å 90.00° 94.35° 90.00°	Depositor
Resolution (Å)	44.44 – 1.87 44.44 – 1.87	Depositor EDS
% Data completeness (in resolution range)	97.2 (44.44-1.87) 97.2 (44.44-1.87)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.44 (at 1.87Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.168 , 0.201 0.178 , 0.209	Depositor DCC
$R_{free}$ test set	3803 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.6	Xtrriage
Anisotropy	0.106	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 45.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6992	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.91% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT, GOL, CD, NDG, NA

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	0.31	0/3258	0.48	0/4460
1	B	0.29	0/3233	0.47	0/4428
All	All	0.30	0/6491	0.47	0/8888

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	3159	0	2966	17	0
1	B	3140	0	2951	12	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	4	0	3	0	0
4	B	4	0	3	0	0
5	A	15	0	12	0	0
5	B	15	0	12	0	0
6	A	12	0	16	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	12	0	16	2	0
7	A	316	0	0	3	0
7	B	311	0	0	2	0
All	All	6992	0	5979	30	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 2.

All (30) 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:193:PRO:HA	1:A:198:TYR:CD1	2.35	0.62
1:A:127:HIS:HD2	7:A:895:HOH:O	1.89	0.54
1:A:265:SER:N	1:A:266:PRO:CD	2.72	0.53
1:B:432:GLU:N	7:B:835:HOH:O	2.44	0.50
1:B:45:ALA:HB2	1:B:300:GLU:HB2	1.95	0.49
1:B:169:TYR:HA	6:B:506:GOL:H11	1.95	0.48
1:B:200:CYS:SG	1:B:272:ILE:HD11	2.54	0.48
1:B:234:ALA:HB1	1:B:235:PRO:HD2	1.96	0.48
1:B:39:ASP:O	1:B:40:ASP:HB2	2.14	0.47
1:A:197:GLY:O	1:A:199:VAL:HG23	2.15	0.46
1:A:45:ALA:HB2	1:A:300:GLU:HB3	1.97	0.46
1:B:147:GLU:HG2	1:B:153:ILE:CG2	2.46	0.46
1:A:286:HIS:CD2	7:A:624:HOH:O	2.69	0.45
1:A:389:THR:HG23	7:A:801:HOH:O	2.16	0.45
1:A:419:TRP:HA	1:A:424:GLN:HG3	1.99	0.45
1:A:171:ASN:OD1	1:A:228:GLY:HA3	2.17	0.44
1:A:312:LEU:N	1:A:313:PRO:CD	2.80	0.44
1:A:344:THR:HA	1:A:365:THR:HA	2.00	0.44
1:B:240:ILE:HB	1:B:241:PRO:HD2	1.98	0.44
1:A:307:GLY:O	1:A:311:ASN:HB2	2.18	0.43
1:A:39:ASP:O	1:A:40:ASP:HB2	2.17	0.43
1:B:265:SER:N	1:B:266:PRO:CD	2.81	0.43
1:B:307:GLY:O	1:B:311:ASN:HB2	2.19	0.41
1:A:231:VAL:CG2	1:A:292:VAL:HG22	2.51	0.41
1:B:397:GLY:HA2	1:B:408:LEU:HD11	2.02	0.41
1:A:192:LYS:HA	1:A:193:PRO:HD3	1.82	0.41
1:B:165:ALA:O	1:B:226:THR:HA	2.21	0.41
1:A:165:ALA:O	1:A:226:THR:HA	2.21	0.40
1:A:62:PHE:HB3	1:A:64:PHE:CE1	2.56	0.40
6:B:506:GOL:H12	7:B:766:HOH:O	2.21	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	405/419 (97%)	392 (97%)	12 (3%)	1 (0%)	47	37
1	B	402/419 (96%)	391 (97%)	10 (2%)	1 (0%)	47	37
All	All	807/838 (96%)	783 (97%)	22 (3%)	2 (0%)	47	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	98	SER
1	B	40	ASP

### 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	337/358 (94%)	332 (98%)	5 (2%)	65	59
1	B	333/358 (93%)	326 (98%)	7 (2%)	53	45
All	All	670/716 (94%)	658 (98%)	12 (2%)	59	52

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

Mol	Chain	Res	Type
1	A	67	TRP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	179	PHE
1	A	194	TRP
1	A	200	CYS
1	A	296	ASP
1	B	67	TRP
1	B	179	PHE
1	B	180	GLN
1	B	200	CYS
1	B	296	ASP
1	B	368	GLN
1	B	390	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	127	HIS
1	B	178	HIS
1	B	396	GLN

### 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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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	NDG	B	504	-	15,15,15	1.23	1 (6%)	21,21,21	1.01	0
4	ACT	B	503	3	1,3,3	0.81	0	0,3,3	0.00	-
4	ACT	A	503	3	1,3,3	0.44	0	0,3,3	0.00	-
6	GOL	A	505[A]	-	5,5,5	0.37	0	5,5,5	0.39	0
6	GOL	B	506	-	5,5,5	0.32	0	5,5,5	0.37	0
6	GOL	B	505	-	5,5,5	0.40	0	5,5,5	0.27	0
6	GOL	A	505[B]	-	5,5,5	0.39	0	5,5,5	0.32	0
5	NDG	A	504	-	15,15,15	1.32	1 (6%)	21,21,21	1.21	2 (9%)

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	NDG	B	504	-	-	0/6/26/26	0/1/1/1
6	GOL	A	505[A]	-	-	4/4/4/4	-
6	GOL	B	506	-	-	0/4/4/4	-
6	GOL	B	505	-	-	0/4/4/4	-
6	GOL	A	505[B]	-	-	4/4/4/4	-
5	NDG	A	504	-	-	0/6/26/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	504	NDG	C1-C2	4.11	1.57	1.52
5	A	504	NDG	C1-C2	4.06	1.57	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	504	NDG	C1-C2-N2	2.89	114.08	110.73
5	A	504	NDG	O7-C7-N2	2.06	125.74	121.95

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	505[B]	GOL	O1-C1-C2-C3
6	A	505[B]	GOL	C1-C2-C3-O3
6	A	505[A]	GOL	O1-C1-C2-C3
6	A	505[B]	GOL	O1-C1-C2-O2
6	A	505[B]	GOL	O2-C2-C3-O3
6	A	505[A]	GOL	O1-C1-C2-O2
6	A	505[A]	GOL	C1-C2-C3-O3
6	A	505[A]	GOL	O2-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	506	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

### 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	A	404/419 (96%)	-0.06	5 (1%) 79 80	10, 19, 41, 69	0
1	B	404/419 (96%)	0.07	15 (3%) 41 43	10, 21, 37, 62	0
All	All	808/838 (96%)	0.00	20 (2%) 57 59	10, 20, 39, 69	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	199	VAL	4.1
1	A	196	PRO	3.7
1	B	194	TRP	3.6
1	B	197	GLY	3.5
1	B	191	LEU	3.3
1	A	200	CYS	3.2
1	A	194	TRP	3.1
1	A	197	GLY	2.8
1	B	198	TYR	2.8
1	B	29	THR	2.7
1	B	196	PRO	2.6
1	B	200	CYS	2.6
1	B	325	ALA	2.5
1	B	199	VAL	2.4
1	B	203	ALA	2.3
1	B	357	VAL	2.1
1	B	162	TYR	2.1
1	B	204	ASN	2.1
1	B	190	ASN	2.0
1	B	143	LEU	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, 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
6	GOL	B	506	6/6	0.68	0.24	47,47,47,47	0
5	NDG	B	504	15/15	0.81	0.17	28,39,46,49	0
6	GOL	B	505	6/6	0.84	0.23	26,28,36,37	0
5	NDG	A	504	15/15	0.85	0.21	29,35,44,44	0
4	ACT	A	503	4/4	0.95	0.09	11,15,16,21	0
2	NA	B	501	1/1	0.96	0.14	10,10,10,10	0
6	GOL	A	505[B]	6/6	0.96	0.25	22,22,22,22	6
6	GOL	A	505[A]	6/6	0.96	0.25	22,26,31,35	6
4	ACT	B	503	4/4	0.97	0.07	18,18,18,19	0
2	NA	A	501	1/1	0.99	0.14	14,14,14,14	0
3	CD	B	502	1/1	1.00	0.08	19,19,19,19	0
3	CD	A	502	1/1	1.00	0.08	19,19,19,19	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.