



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:57 AM GMT

PDB ID : 3GNV  
Title : HCV NS5B polymerase in complex with 1,5 benzodiazepine inhibitor 1b  
Authors : De Bondt, H.; Nyanguile, O.  
Deposited on : 2009-03-18  
Resolution : 2.75 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

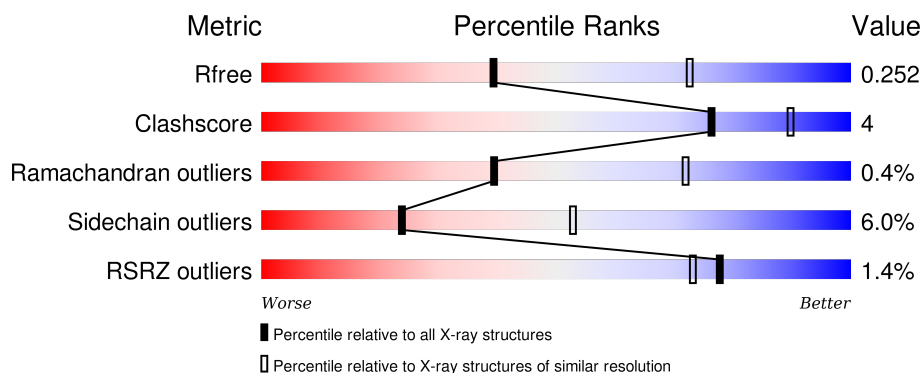
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	3340 (2.80-2.72)
Clashscore	102246	3829 (2.80-2.72)
Ramachandran outliers	100387	3767 (2.80-2.72)
Sidechain outliers	100360	3770 (2.80-2.72)
RSRZ outliers	91569	3352 (2.80-2.72)

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	581	 82% 14% • •
1	B	581	 81% 13% • 5%

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
3	GOL	B	580	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8957 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 RNA-directed RNA polymerase.

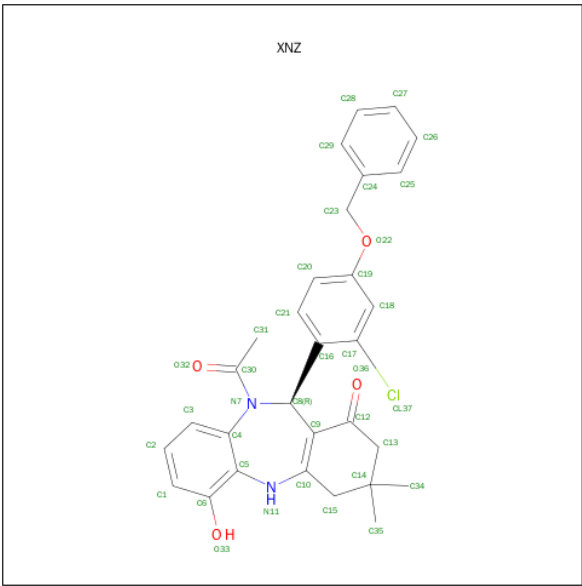
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	563	Total	C	N	O	S	20	0	0
			4384	2760	775	817	32			
1	B	553	Total	C	N	O	S	5	0	0
			4311	2716	762	801	32			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	EXPRESSION TAG	UNP O92972
A	-1	ALA	-	EXPRESSION TAG	UNP O92972
A	0	SER	-	EXPRESSION TAG	UNP O92972
A	571	LEU	-	EXPRESSION TAG	UNP O92972
A	572	GLU	-	EXPRESSION TAG	UNP O92972
A	573	HIS	-	EXPRESSION TAG	UNP O92972
A	574	HIS	-	EXPRESSION TAG	UNP O92972
A	575	HIS	-	EXPRESSION TAG	UNP O92972
A	576	HIS	-	EXPRESSION TAG	UNP O92972
A	577	HIS	-	EXPRESSION TAG	UNP O92972
A	578	HIS	-	EXPRESSION TAG	UNP O92972
B	-2	MET	-	EXPRESSION TAG	UNP O92972
B	-1	ALA	-	EXPRESSION TAG	UNP O92972
B	0	SER	-	EXPRESSION TAG	UNP O92972
B	571	LEU	-	EXPRESSION TAG	UNP O92972
B	572	GLU	-	EXPRESSION TAG	UNP O92972
B	573	HIS	-	EXPRESSION TAG	UNP O92972
B	574	HIS	-	EXPRESSION TAG	UNP O92972
B	575	HIS	-	EXPRESSION TAG	UNP O92972
B	576	HIS	-	EXPRESSION TAG	UNP O92972
B	577	HIS	-	EXPRESSION TAG	UNP O92972
B	578	HIS	-	EXPRESSION TAG	UNP O92972

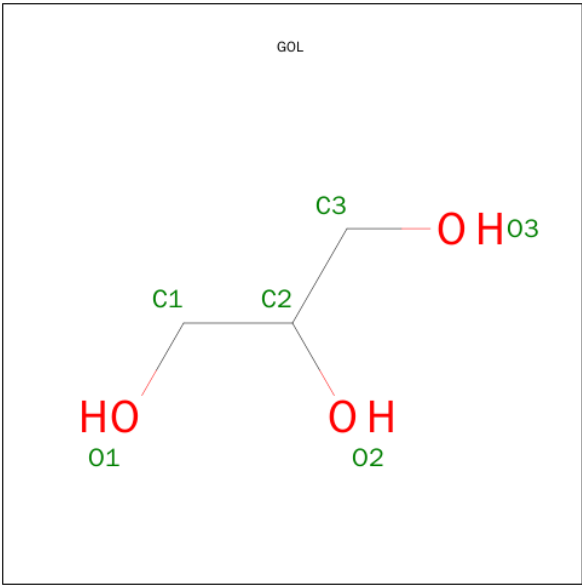
- Molecule 2 is (11R)-10-ACETYL-11-[4-(BENZYLOXY)-2-CHLOROPHENYL]-6-HYDROXY-3,3-DIMETHYL-2,3,4,5,10,11-HEXAHYDRO-1H-DIBENZO[B,E][1,4]DIAZEPIN-1-ON

E (three-letter code: XNZ) (formula: C<sub>30</sub>H<sub>29</sub>ClN<sub>2</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	0	0
			37	30	1	2	4		
2	B	1	Total	C	Cl	N	O	0	0
			37	30	1	2	4		

- Molecule 3 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
3	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

*Continued from previous page...*

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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	111	Total	O	0	0
			111	111		
4	B	65	Total	O	0	0
			65	65		

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.

Chain A:

Sequence logo for Chain A showing amino acid conservation across 400 positions. The y-axis represents information content in bits (0.00 to 0.25). The x-axis shows positions 1 to 400. A green bar at the top indicates 82% conservation, and a yellow bar indicates 14% conservation. A legend at the bottom shows amino acid codes: MET, ALA, SER, S1, L10, M24, S27, R32, R48, V52, R56, L57, L68, L82, E86, S96, K106, R114, L126, T130, E131, V144, Q148, P149, G152, L159, K172, M173, S180, P197, V201, K211, Y219, D220, T233, R222, C223, F224, D225, S226, R234, V235, E236, A249, L260, T261, I262, C274, R277, R280, V284, T287, N291, L314, D318, D319, L320, V321, A327, Q330, E331, Y346, S347, D352, Q355, T364, S365, S368, L384, W397, T403, P404, L416, P417, L425, Q438, L439, E440, K441, Q446, A450, L459, P460, Q461, L469, V485, L489, Q493, T499, R517, R531, P540, A541, A542, D546, L547, S548, Y555, S556, S563, LEU, SER, ARG, ALA, ARG, ARG, PRO, LEU, ARG, GLU, GLY, HIS, HIS, HIS, HIS, HIS.

Chain B:

Met	Ala	Ser	S1	T5	C14	E18	P22	Ile	Asn	Pro	Leu	Ser	Asn	Ser	L30	L31	R32	S46	Q49	V52	R56	L57	L68	T77	A80	R81	L82	L83	E86	R120	E124	D125	L126	L127	D128	D129	T130	Q148	P149	GLU	Lys	GLY	G153																																		
M173	P183	S189	Q194	P197	K198	V201	K211	P215	Y219	D220	C223	S226	E237	L245	A249	R254	C274	G275	R280	C289	G290	N291	R310	D319	L320	C324	R325	S326	D359	L362	T363	T364	S368	D387	T388	R401	A416	P417	T418	L419	W420	A421	L425	H428	L434	L459	P460	Q461	R465	G468	L469	S470	N483	S487	C488	L497	W507	R508	R517	Y524	A529	V530	R531	T532	K533	L534	K535	A541	A542	S543	Q544	L545	D546	L547	S548	S556	C562

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.29Å 107.97Å 133.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	83.33 – 2.75 83.23 – 2.75	Depositor EDS
% Data completeness (in resolution range)	100.0 (83.33-2.75) 100.0 (83.23-2.75)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.85 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.174 , 0.234 0.202 , 0.252	Depositor DCC
$R_{free}$ test set	2037 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.0	Xtriage
Anisotropy	0.549	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 36.7	EDS
Estimated twinning fraction	0.028 for k,h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 40722 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8957	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: GOL, XNZ

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	0.56	2/4480 (0.0%)	0.69	3/6080 (0.0%)
1	B	0.50	2/4404 (0.0%)	0.68	3/5974 (0.1%)
All	All	0.53	4/8884 (0.0%)	0.68	6/12054 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	223	CYS	CB-SG	-6.11	1.71	1.82
1	A	274	CYS	CB-SG	-6.05	1.72	1.82
1	B	223	CYS	CB-SG	-5.59	1.72	1.81
1	B	274	CYS	CB-SG	-5.31	1.73	1.81

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	56	ARG	NE-CZ-NH2	-6.67	116.96	120.30
1	B	280	ARG	NE-CZ-NH1	5.85	123.22	120.30
1	B	280	ARG	NE-CZ-NH2	-5.65	117.47	120.30
1	A	280	ARG	NE-CZ-NH1	5.20	122.90	120.30
1	A	56	ARG	NE-CZ-NH1	5.18	122.89	120.30

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	A	4384	0	4397	35	0
1	B	4311	0	4322	33	0
2	A	37	0	28	0	0
2	B	37	0	28	0	0
3	A	6	0	8	0	0
3	B	6	0	8	0	0
4	A	111	0	0	3	0
4	B	65	0	0	4	0
All	All	8957	0	8791	67	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.

The worst 5 of 67 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:198:LYS:HD2	4:B:631:HOH:O	1.85	0.74
1:B:30:LEU:HD12	1:B:31:LEU:HB2	1.76	0.67
1:A:32:ARG:HD3	1:A:493:GLY:O	1.95	0.67
1:B:530:VAL:O	1:B:533:LYS:NZ	2.28	0.67
1:A:236:GLU:OE1	1:A:280:ARG:NH2	2.27	0.66

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	561/581 (97%)	538 (96%)	21 (4%)	2 (0%)	39	72
1	B	547/581 (94%)	528 (96%)	17 (3%)	2 (0%)	39	72
All	All	1108/1162 (95%)	1066 (96%)	38 (3%)	4 (0%)	39	72

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	PRO
1	B	223	CYS
1	A	152	GLY
1	B	532	THR

### 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	479/495 (97%)	453 (95%)	26 (5%)	27	58
1	B	470/495 (95%)	439 (93%)	31 (7%)	21	48
All	All	949/990 (96%)	892 (94%)	57 (6%)	24	53

5 of 57 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	556	SER
1	B	56	ARG
1	B	487	SER
1	B	14	CYS
1	B	31	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	446	GLN
1	A	514	GLN
1	B	206	ASN
1	A	355	GLN
1	A	438	GLN

### 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 ⓘ

4 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	XNZ	A	579	-	39,41,41	0.97	3 (7%)	52,61,61	1.80	6 (11%)
3	GOL	A	580	-	5,5,5	0.37	0	5,5,5	0.29	0
2	XNZ	B	579	-	39,41,41	0.91	3 (7%)	52,61,61	2.02	11 (21%)
3	GOL	B	580	-	5,5,5	0.40	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
2	XNZ	A	579	-	-	0/13/47/47	0/4/5/5
3	GOL	A	580	-	-	0/4/4/4	0/0/0/0
2	XNZ	B	579	-	-	0/13/47/47	0/4/5/5
3	GOL	B	580	-	-	0/4/4/4	0/0/0/0

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	579	XNZ	C5-N11	-2.69	1.39	1.42
2	B	579	XNZ	C12-C9	2.13	1.49	1.45
2	A	579	XNZ	C12-C9	2.20	1.49	1.45
2	B	579	XNZ	C30-N7	2.22	1.41	1.37
2	A	579	XNZ	C15-C10	2.98	1.54	1.50

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	579	XNZ	C4-N7-C30	-5.77	115.80	122.67
2	A	579	XNZ	O32-C30-N7	-4.14	115.50	120.72
2	A	579	XNZ	C4-N7-C30	-4.04	117.85	122.67
2	B	579	XNZ	O32-C30-N7	-3.76	115.98	120.72
2	B	579	XNZ	O36-C12-C9	-2.52	118.79	121.32

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	563/581 (96%)	0.13	2 (0%) 93 92	7, 19, 34, 64	5 (0%)
1	B	553/581 (95%)	0.28	14 (2%) 61 54	6, 19, 33, 48	1 (0%)
All	All	1116/1162 (96%)	0.20	16 (1%) 78 73	6, 19, 33, 64	6 (0%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	531	ARG	3.3
1	B	542	ALA	2.9
1	B	535	LYS	2.9
1	B	468	GLY	2.9
1	B	153	GLY	2.5

### 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
3	GOL	B	580	6/6	0.81	0.25	2.66	43,51,54,56	0
3	GOL	A	580	6/6	0.93	0.21	1.18	14,22,22,29	0
2	XNZ	B	579	37/37	0.96	0.19	0.18	14,22,32,34	0
2	XNZ	A	579	37/37	0.97	0.20	0.04	4,17,23,24	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.