



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 27, 2024 – 10:47 PM EDT

PDB ID : 1NF5  
Title : Crystal Structure of Lactose Synthase, Complex with Glucose  
Authors : Ramakrishnan, B.; Qasba, P.K.  
Deposited on : 2002-12-13  
Resolution : 2.00 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

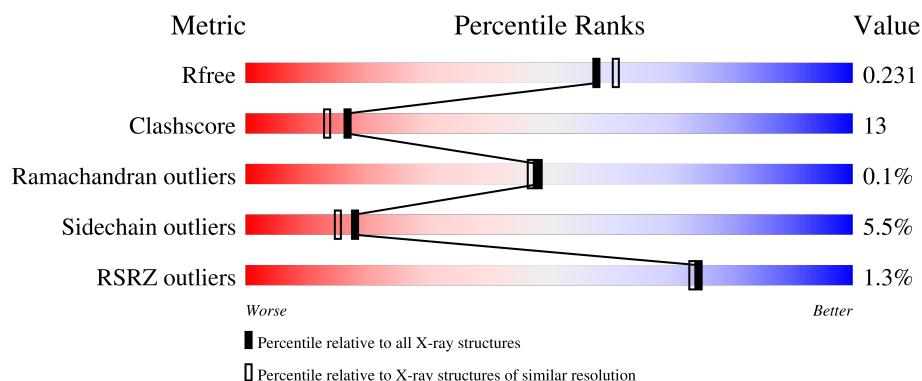
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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 of 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	123	<div> <div>%</div> <div> <div></div> <div>76%</div> <div>20%</div> <div>• •</div> </div> </div>
1	C	123	<div> <div>2%</div> <div> <div></div> <div>80%</div> <div>18%</div> <div>•</div> </div> </div>
2	B	286	<div> <div>2%</div> <div> <div></div> <div>71%</div> <div>20%</div> <div>• 5%</div> </div> </div>
2	D	286	<div> <div>%</div> <div> <div></div> <div>73%</div> <div>20%</div> <div>• 5%</div> </div> </div>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 7037 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 Alpha-lactalbumin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	123	Total	C	N	O	S	0	0	0
			980	620	156	195	9			
1	C	123	Total	C	N	O	S	0	0	0
			980	620	156	195	9			

- Molecule 2 is a protein called beta-1,4-galactosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	272	Total	C	N	O	S	0	0	0
			2218	1424	382	398	14			
2	D	272	Total	C	N	O	S	0	0	0
			2218	1424	382	398	14			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	117	ALA	-	see remark 999	UNP P08037
B	118	SER	-	see remark 999	UNP P08037
B	119	MET	-	see remark 999	UNP P08037
B	120	THR	-	see remark 999	UNP P08037
B	121	GLY	-	see remark 999	UNP P08037
B	122	GLY	-	see remark 999	UNP P08037
B	123	GLN	-	see remark 999	UNP P08037
B	124	GLN	-	see remark 999	UNP P08037
B	125	MET	-	see remark 999	UNP P08037
B	126	GLY	-	see remark 999	UNP P08037
B	127	ARG	-	see remark 999	UNP P08037
B	128	GLY	-	see remark 999	UNP P08037
B	129	SER	-	see remark 999	UNP P08037
D	117	ALA	-	see remark 999	UNP P08037
D	118	SER	-	see remark 999	UNP P08037
D	119	MET	-	see remark 999	UNP P08037

*Continued on next page...*

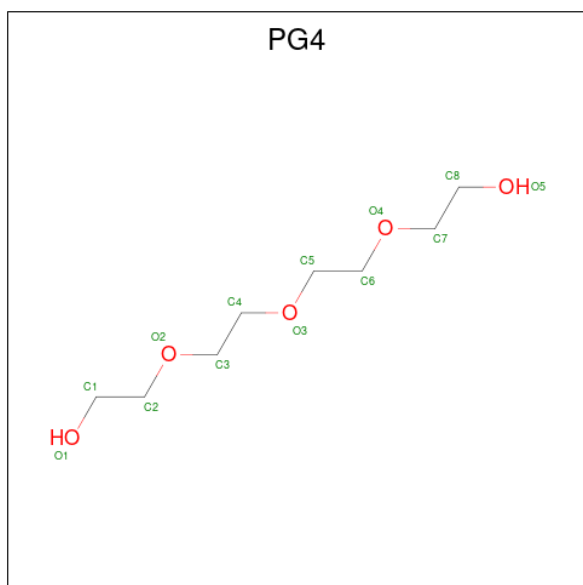
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	120	THR	-	see remark 999	UNP P08037
D	121	GLY	-	see remark 999	UNP P08037
D	122	GLY	-	see remark 999	UNP P08037
D	123	GLN	-	see remark 999	UNP P08037
D	124	GLN	-	see remark 999	UNP P08037
D	125	MET	-	see remark 999	UNP P08037
D	126	GLY	-	see remark 999	UNP P08037
D	127	ARG	-	see remark 999	UNP P08037
D	128	GLY	-	see remark 999	UNP P08037
D	129	SER	-	see remark 999	UNP P08037

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

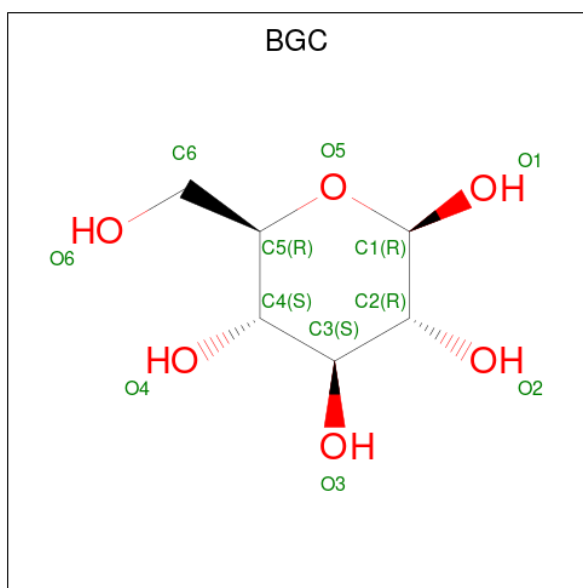
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Ca 1 1	0	0
3	C	1	Total Ca 1 1	0	0

- Molecule 4 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C<sub>8</sub>H<sub>18</sub>O<sub>5</sub>).



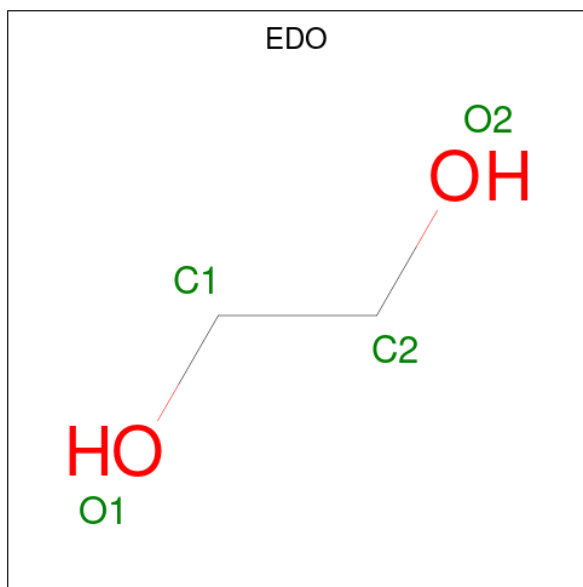
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C O 10 6 4	0	0
4	B	1	Total C O 10 6 4	0	0

- Molecule 5 is beta-D-glucopyranose (three-letter code: BGC) (formula:  $C_6H_{12}O_6$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			12	6	6		
5	D	1	Total	C	O	0	0
			12	6	6		

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

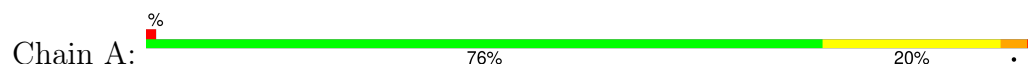
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	106	Total 106	O 106	0	0
7	B	181	Total 181	O 181	0	0
7	C	119	Total 119	O 119	0	0
7	D	185	Total 185	O 185	0	0

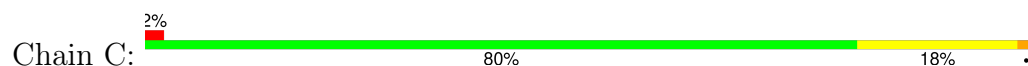
### 3 Residue-property plots [i](#)

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.

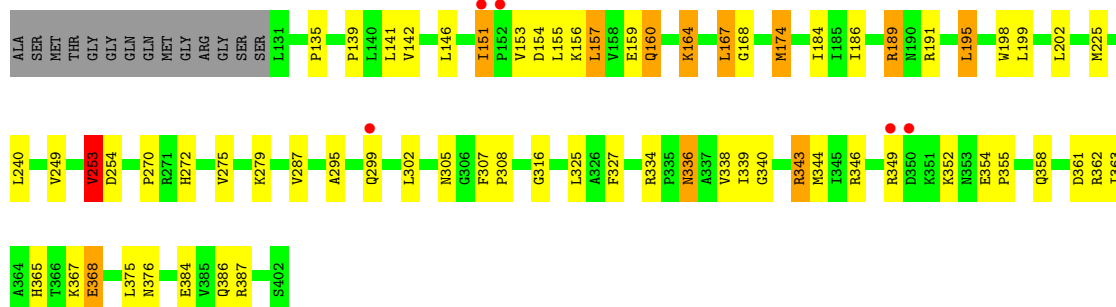
- Molecule 1: Alpha-lactalbumin



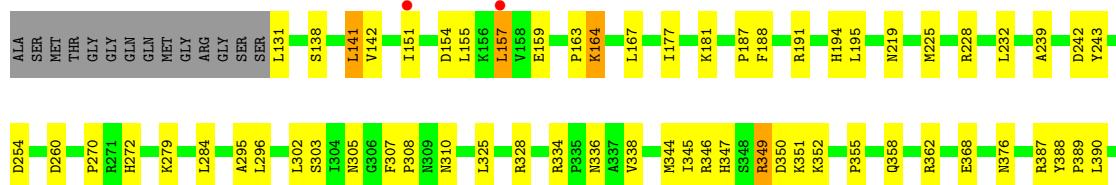
- Molecule 1: Alpha-lactalbumin



- Molecule 2: beta-1,4-galactosyltransferase



- Molecule 2: beta-1,4-galactosyltransferase



8402

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.22Å 93.90Å 99.97Å 90.00° 101.50° 90.00°	Depositor
Resolution (Å)	24.73 – 2.00 24.73 – 2.00	Depositor EDS
% Data completeness (in resolution range)	81.7 (24.73-2.00) 81.7 (24.73-2.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 1.99Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.190 , 0.241 0.182 , 0.231	Depositor DCC
$R_{free}$ test set	2893 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.8	Xtriage
Anisotropy	0.236	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 49.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7037	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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: BGC, CA, PG4, EDO

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	0.57	0/1001	0.78	1/1350 (0.1%)
1	C	0.59	0/1001	0.79	1/1350 (0.1%)
2	B	0.53	0/2278	0.75	1/3085 (0.0%)
2	D	0.53	0/2278	0.77	1/3085 (0.0%)
All	All	0.54	0/6558	0.76	4/8870 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	253	VAL	CB-CA-C	-5.83	100.32	111.40
2	D	284	LEU	CA-CB-CG	5.72	128.47	115.30
1	C	87	ASP	CB-CG-OD1	5.14	122.93	118.30
1	A	45	ASN	O-C-N	-5.10	114.53	123.20

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	980	0	936	28	0
1	C	980	0	936	25	0
2	B	2218	0	2185	60	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	2218	0	2185	49	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	A	10	0	13	0	0
4	B	10	0	13	0	0
5	B	12	0	12	0	0
5	D	12	0	12	2	0
6	C	4	0	6	0	0
7	A	106	0	0	3	0
7	B	181	0	0	1	0
7	C	119	0	0	2	0
7	D	185	0	0	1	0
All	All	7037	0	6298	160	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 13.

The worst 5 of 160 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:44:ASP:C	1:A:45:ASN:HD22	1.52	1.13
2:B:343:ARG:HG2	2:B:343:ARG:HH11	1.16	1.09
2:B:151:ILE:HD12	2:B:151:ILE:H	1.28	0.99
2:B:336:ASN:HD22	2:B:338:VAL:H	1.20	0.89
2:B:343:ARG:HH11	2:B:343:ARG:CG	1.87	0.88

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	121/123 (98%)	119 (98%)	1 (1%)	1 (1%)	16	12
1	C	121/123 (98%)	117 (97%)	4 (3%)	0	100	100
2	B	270/286 (94%)	259 (96%)	11 (4%)	0	100	100
2	D	270/286 (94%)	263 (97%)	7 (3%)	0	100	100
All	All	782/818 (96%)	758 (97%)	23 (3%)	1 (0%)	48	47

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	45	ASN

### 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	109/109 (100%)	102 (94%)	7 (6%)	14	11
1	C	109/109 (100%)	106 (97%)	3 (3%)	38	40
2	B	245/254 (96%)	224 (91%)	21 (9%)	8	5
2	D	245/254 (96%)	237 (97%)	8 (3%)	33	33
All	All	708/726 (98%)	669 (94%)	39 (6%)	18	15

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	71	ASN
2	D	302	LEU
1	C	85	LEU
2	D	157	LEU
2	D	355	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	310	ASN
2	D	305	ASN
1	C	39	GLN
2	D	194	HIS
1	C	32	HIS

### 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 oligosaccharides in this entry.

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

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	BGC	B	403	-	12,12,12	1.04	1 (8%)	17,17,17	1.84	3 (17%)
6	EDO	C	128	-	3,3,3	0.39	0	2,2,2	0.14	0
5	BGC	D	527	-	12,12,12	1.15	1 (8%)	17,17,17	2.32	9 (52%)
4	PG4	A	126	-	9,9,12	0.49	0	8,8,11	0.26	0
4	PG4	B	404	-	9,9,12	0.44	0	8,8,11	0.30	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	BGC	B	403	-	-	1/2/22/22	0/1/1/1
6	EDO	C	128	-	-	1/1/1/1	-
5	BGC	D	527	-	-	0/2/22/22	0/1/1/1
4	PG4	A	126	-	-	3/7/7/10	-
4	PG4	B	404	-	-	4/7/7/10	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	D	527	BGC	O5-C5	2.42	1.50	1.44
5	B	403	BGC	O5-C5	2.35	1.50	1.44

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	403	BGC	O5-C5-C6	-4.77	94.62	106.44
5	D	527	BGC	O5-C5-C6	-4.30	95.79	106.44
5	D	527	BGC	O5-C1-C2	4.24	117.75	110.30
5	D	527	BGC	O3-C3-C4	-3.28	102.65	110.38
5	D	527	BGC	O1-C1-C2	3.27	118.48	108.98

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	404	PG4	O3-C5-C6-O4
4	A	126	PG4	O1-C1-C2-O2
4	B	404	PG4	O1-C1-C2-O2
5	B	403	BGC	C4-C5-C6-O6
4	B	404	PG4	C6-C5-O3-C4

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	D	527	BGC	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	123/123 (100%)	-0.36	1 (0%) 82 82	18, 26, 51, 63	0
1	C	123/123 (100%)	-0.44	2 (1%) 70 69	15, 23, 43, 70	0
2	B	272/286 (95%)	-0.18	5 (1%) 67 66	18, 28, 52, 73	0
2	D	272/286 (95%)	-0.36	2 (0%) 84 83	17, 27, 51, 61	0
All	All	790/818 (96%)	-0.31	10 (1%) 74 73	15, 27, 51, 73	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	123	PRO	3.6
2	D	151	ILE	3.4
1	A	123	PRO	2.6
2	B	349	ARG	2.6
2	B	350	ASP	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 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( $\text{\AA}^2$ )	Q<0.9
6	EDO	C	128	4/4	0.83	0.11	50,50,51,52	0
4	PG4	B	404	10/13	0.86	0.13	50,52,53,54	0
4	PG4	A	126	10/13	0.93	0.11	44,47,50,50	0
5	BGC	B	403	12/12	0.96	0.06	19,21,24,24	0
5	BGC	D	527	12/12	0.97	0.05	17,19,20,21	0
3	CA	A	124	1/1	0.99	0.03	24,24,24,24	0
3	CA	C	526	1/1	1.00	0.02	24,24,24,24	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.