



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 04:37 PM GMT

PDB ID : 1O23  
Title : CRYSTAL STRUCTURE OF LACTOSE SYNTHASE IN THE PRESENCE  
OF UDP-GLUCOSE  
Authors : Ramakrishnan, B.; Shah, P.S.; Qasba, P.K.  
Deposited on : 2003-01-29  
Resolution : 2.32 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

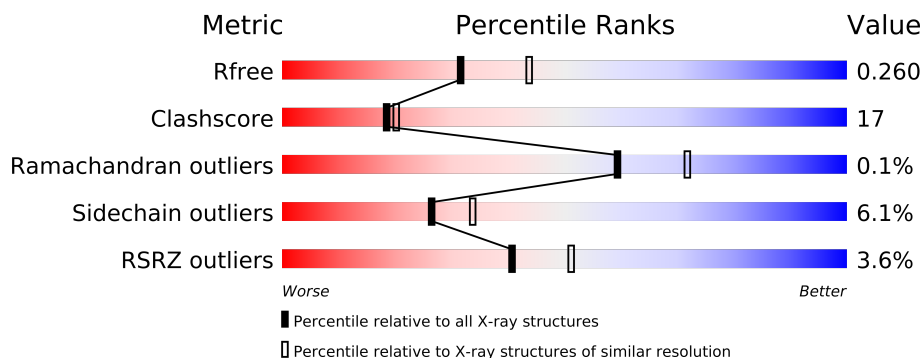
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.32 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))
$R_{free}$	66092	3293 (2.34-2.30)
Clashscore	79885	4097 (2.34-2.30)
Ramachandran outliers	78287	4055 (2.34-2.30)
Sidechain outliers	78261	4054 (2.34-2.30)
RSRZ outliers	66119	3294 (2.34-2.30)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	123	
1	C	123	
2	B	286	
2	D	286	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
6	UDP	B	811	-	X
6	UDP	D	812	-	X
7	PG4	A	813	-	X
7	PG4	C	814	-	X
8	MES	A	815	-	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 6972 atoms, of which 0 are hydrogen and 0 are deuterium.

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	519	ALA	-	SEE REMARK 999	UNP P08037
D	520	SER	-	SEE REMARK 999	UNP P08037
D	521	MET	-	SEE REMARK 999	UNP P08037

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	522	THR	-	SEE REMARK 999	UNP P08037
D	523	GLY	-	SEE REMARK 999	UNP P08037
D	524	GLY	-	SEE REMARK 999	UNP P08037
D	525	GLN	-	SEE REMARK 999	UNP P08037
D	526	GLN	-	SEE REMARK 999	UNP P08037
D	527	MET	-	SEE REMARK 999	UNP P08037
D	528	GLY	-	SEE REMARK 999	UNP P08037
D	529	ARG	-	SEE REMARK 999	UNP P08037
D	530	GLY	-	SEE REMARK 999	UNP P08037
D	531	SER	-	SEE REMARK 999	UNP P08037

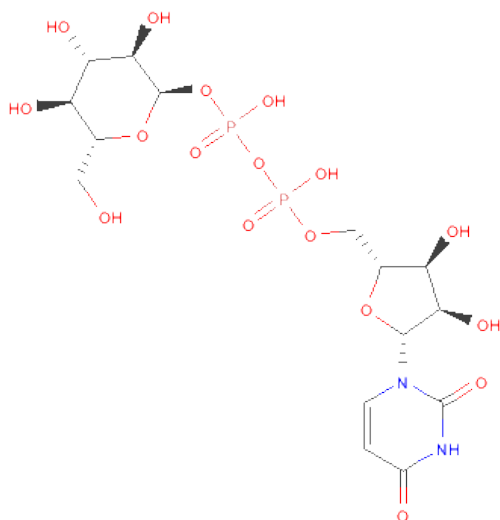
- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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 MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

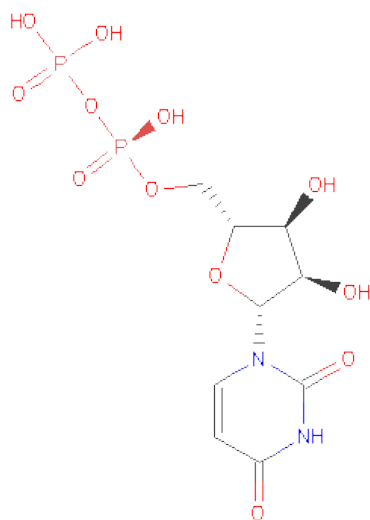
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Mn 1 1	0	0
4	D	1	Total Mn 1 1	0	0

- Molecule 5 is URIDINE-5'-DIPHOSPHATE-GLUCOSE (three-letter code: UPG) (formula: C<sub>15</sub>H<sub>24</sub>N<sub>2</sub>O<sub>17</sub>P<sub>2</sub>).



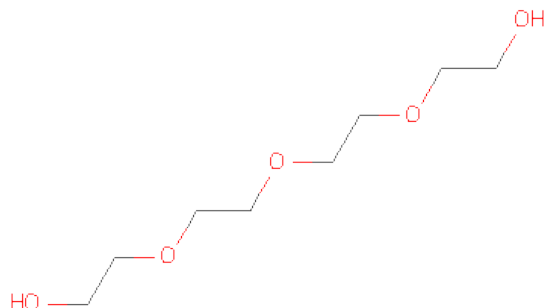
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total	C	N	O	P	0	0
			36	15	2	17	2		
5	D	1	Total	C	N	O	P	0	0
			36	15	2	17	2		

- Molecule 6 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula:  $C_9H_{14}N_2O_{12}P_2$ ).



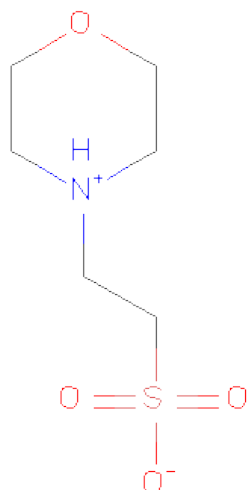
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
6	D	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 7 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula:  $C_8H_{18}O_5$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			13	8	5		
7	C	1	Total	C	O	0	0
			13	8	5		

- Molecule 8 is 2-(N-MORPHOLINO)-ETHANESULFONICACID (three-letter code: MES) (formula:  $C_6H_{13}NO_4S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
8	A	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 9 is water.

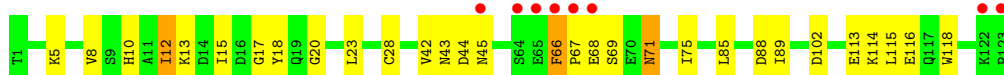
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	63	Total	O	0	0
			63	63		
9	B	138	Total	O	0	0
			138	138		
9	C	87	Total	O	0	0
			87	87		
9	D	124	Total	O	0	0
			124	124		

### 3 Residue-property plots

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.

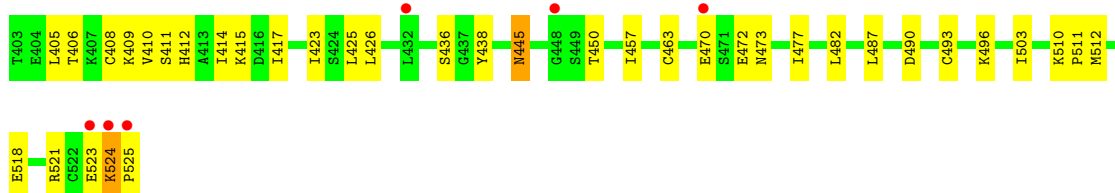
- Molecule 1: ALPHA-LACTALBUMIN

Chain A: 



- Molecule 1: ALPHA-LACTALBUMIN

Chain C: 



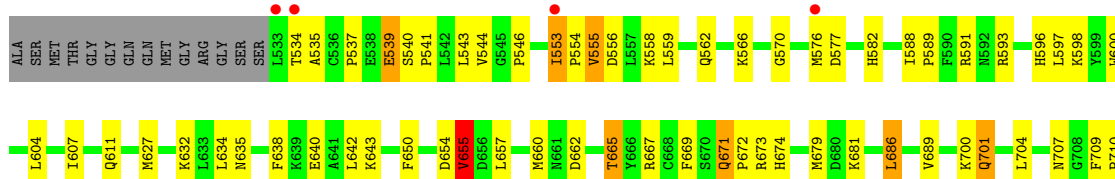
- Molecule 2: BETA-1,4-GALACTOSYLTRANSFERASE

Chain B: 

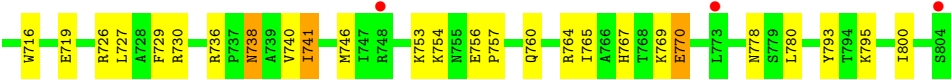


- Molecule 2: BETA-1,4-GALACTOSYLTRANSFERASE

Chain D: 







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	55.55Å 99.38Å 102.57Å 90.00° 104.09° 90.00°	Depositor
Resolution (Å)	19.96 – 2.32 19.96 – 2.30	Depositor EDS
% Data completeness (in resolution range)	98.8 (19.96-2.32) 98.5 (19.96-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.53 (at 2.30Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.201 , 0.265 0.198 , 0.260	Depositor DCC
$R_{free}$ test set	4654 reflections (10.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	41.9	Xtriage
Anisotropy	0.618	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 44.9	EDS
Estimated twinning fraction	0.033 for h,-k,-h-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 47286 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6972	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.22% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: UDP, UPG, CA, MN, PG4, MES

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	0.50	0/1001	0.71	0/1350
1	C	0.59	0/1001	0.77	0/1350
2	B	0.52	0/2278	0.74	1/3085 (0.0%)
2	D	0.51	0/2278	0.73	2/3085 (0.1%)
All	All	0.52	0/6558	0.74	3/8870 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	655	VAL	CB-CA-C	-6.28	99.46	111.40
2	B	253	VAL	CB-CA-C	-5.37	101.19	111.40
2	D	686	LEU	CA-CB-CG	5.19	127.24	115.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	980	0	936	33	0
1	C	980	0	933	32	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	2218	0	2185	62	0
2	D	2218	0	2185	89	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
5	B	36	0	21	2	0
5	D	36	0	21	6	0
6	B	25	0	11	1	0
6	D	25	0	11	1	0
7	A	13	0	18	0	0
7	C	13	0	18	3	0
8	A	12	0	13	11	0
9	A	63	0	0	2	0
9	B	138	0	0	3	0
9	C	87	0	0	1	0
9	D	124	0	0	5	0
All	All	6972	0	6352	217	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 17.

The worst 5 of 217 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:D:553:ILE:HD12	2:D:554:PRO:HD2	1.39	1.04
2:D:738:ASN:HD22	2:D:738:ASN:C	1.69	0.94
2:D:738:ASN:HD21	2:D:741:ILE:H	1.12	0.91
2:D:701:GLN:NE2	2:D:701:GLN:H	1.71	0.88
1:C:523:GLU:O	1:C:524:LYS:HG2	1.75	0.85

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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%)	113 (93%)	8 (7%)	0	100	100
1	C	121/123 (98%)	115 (95%)	5 (4%)	1 (1%)	27	30
2	B	270/286 (94%)	264 (98%)	6 (2%)	0	100	100
2	D	270/286 (94%)	256 (95%)	14 (5%)	0	100	100
All	All	782/818 (96%)	748 (96%)	33 (4%)	1 (0%)	59	72

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	524	LYS

### 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%)	103 (94%)	6 (6%)	30	39
1	C	109/109 (100%)	105 (96%)	4 (4%)	45	60
2	B	245/254 (96%)	230 (94%)	15 (6%)	26	34
2	D	245/254 (96%)	227 (93%)	18 (7%)	20	24
All	All	708/726 (98%)	665 (94%)	43 (6%)	26	34

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	336	ASN
1	C	490	ASP
2	D	730	ARG
2	B	339	ILE
1	C	445	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	445	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	D	562	GLN
2	D	738	ASN
1	C	473	ASN
2	D	563	GLN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 11 ligands modelled in this entry, 4 are monoatomic - leaving 7 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	$\# Z  > 2$
7	PG4	A	813	-	12,12,12	0.55	0	11,11,11	0.29	0
8	MES	A	815	-	12,12,12	1.25	1 (8%)	16,16,16	1.02	1 (6%)
5	UPG	B	809	4	38,38,38	1.68	12 (31%)	54,58,58	2.22	12 (22%)
6	UDP	B	811	-	26,26,26	1.81	6 (23%)	36,40,40	1.53	4 (11%)
7	PG4	C	814	-	12,12,12	0.52	0	11,11,11	0.19	0
5	UPG	D	810	4	38,38,38	1.91	13 (34%)	54,58,58	2.25	14 (25%)
6	UDP	D	812	-	26,26,26	1.84	5 (19%)	36,40,40	1.61	6 (16%)

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
7	PG4	A	813	-	-	0/10/10/10	0/0/0/0
8	MES	A	815	-	-	0/6/14/14	0/1/1/1
5	UPG	B	809	4	-	0/21/59/59	0/3/3/3
6	UDP	B	811	-	-	0/14/32/32	0/2/2/2
7	PG4	C	814	-	-	0/10/10/10	0/0/0/0
5	UPG	D	810	4	-	0/21/59/59	0/3/3/3
6	UDP	D	812	-	-	0/14/32/32	0/2/2/2

The worst 5 of 37 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	D	812	UDP	PA-O3A	4.53	1.68	1.59
5	D	810	UPG	O5'-C1'	4.28	1.52	1.41
5	D	810	UPG	PA-O3A	4.11	1.67	1.59
6	B	811	UDP	PA-O3A	4.08	1.67	1.59
5	D	810	UPG	C1'-C2'	3.88	1.64	1.52

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	D	810	UPG	O4'-C4'-C3'	7.17	126.42	110.35
5	B	809	UPG	O5'-C5'-C6'	-6.91	89.36	106.34
5	B	809	UPG	O4'-C4'-C3'	6.17	124.19	110.35
5	D	810	UPG	C4'-C3'-C2'	-5.91	99.89	110.82
5	D	810	UPG	O5'-C5'-C4'	-5.44	99.69	109.76

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

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	A	123/123 (100%)	0.08	8 (6%)	18 27	31, 45, 71, 91	0
1	C	123/123 (100%)	-0.01	6 (4%)	28 39	29, 40, 66, 107	0
2	B	272/286 (95%)	-0.05	7 (2%)	53 64	29, 44, 69, 81	0
2	D	272/286 (95%)	-0.03	7 (2%)	53 64	33, 46, 67, 83	0
All	All	790/818 (96%)	-0.02	28 (3%)	41 52	29, 45, 69, 107	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	525	PRO	8.9
2	D	804	SER	5.2
2	B	402	SER	4.6
1	A	68	GLU	4.0
2	D	533	LEU	3.9

### 6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates

There are no carbohydrates in this entry.

### 6.4 Ligands

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	MES	A	815	12/12	0.38	6.14	75,79,93,94	0
7	PG4	A	813	13/13	0.33	4.31	67,75,82,83	0
6	UDP	D	812	25/25	0.29	4.20	80,95,111,112	0
7	PG4	C	814	13/13	0.32	4.02	79,82,83,83	0
6	UDP	B	811	25/25	0.23	2.00	82,96,116,117	0
5	UPG	D	810	36/36	0.12	-0.11	37,44,55,58	0
5	UPG	B	809	36/36	0.11	-0.56	38,44,47,48	0
3	CA	A	805	1/1	0.07	-1.50	42,42,42,42	0
3	CA	C	806	1/1	0.05	-3.62	36,36,36,36	0
4	MN	D	808	1/1	0.07	-7.01	37,37,37,37	0
4	MN	B	807	1/1	0.04	-10.88	40,40,40,40	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.