W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 08:09 pm BST

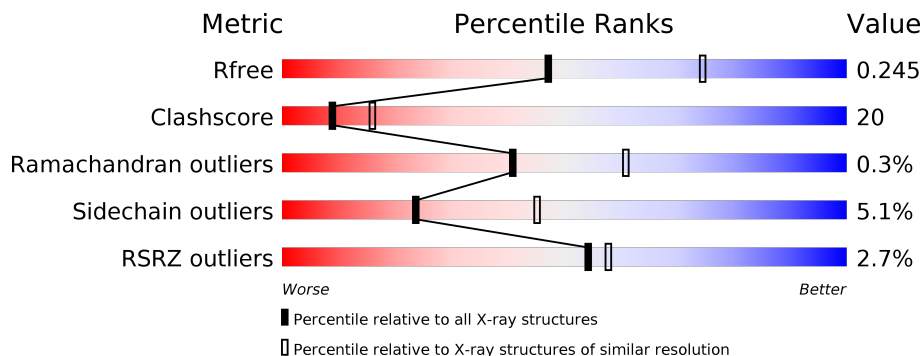|		
PDB ID	:	1NHE
Title	:	Crystal structure of Lactose synthase complex with UDP
Authors	:	Ramakrishnan, B.; Qasba, P.K.
Deposited on	:	2002-12-19
Resolution	:	2.50 Å(reported)

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 2.50 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	Percentile Ranks	Value
Rfree		0.245
Clashscore		20
Ramachandran outliers		0.3%
Sidechain outliers		5.1%
RSRZ outliers		2.7%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 >=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 <=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	4% 72% 28% •
1	C	123	3% 69% 30% •
2	B	286	2% 59% 33% • 5%
2	D	286	2% 59% 32% • 5%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6803 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 979	C 620	N 156	O 194	S 9	0	0	0
1	C	123	Total 979	C 620	N 156	O 194	S 9	0	0	0

- Molecule 2 is a protein called BETA-1,4-GALACTOSYLTRANSFERASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	272	Total 2218	C 1424	N 382	O 398	S 5	Se 9	0	0	0
2	D	272	Total 2218	C 1424	N 382	O 398	S 5	Se 9	0	0	0

There are 44 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	MSE	-	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	MSE	-	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
B	145	MSE	MET	CLONING ARTIFACT	UNP P08037
B	174	MSE	MET	CLONING ARTIFACT	UNP P08037
B	225	MSE	MET	CLONING ARTIFACT	UNP P08037

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	258	MSE	MET	CLONING ARTIFACT	UNP P08037
B	277	MSE	MET	CLONING ARTIFACT	UNP P08037
B	330	MSE	MET	CLONING ARTIFACT	UNP P08037
B	344	MSE	MET	CLONING ARTIFACT	UNP P08037
B	370	MSE	MET	CLONING ARTIFACT	UNP P08037
B	381	MSE	MET	CLONING ARTIFACT	UNP P08037
D	117	ALA	-	SEE REMARK 999	UNP P08037
D	118	SER	-	SEE REMARK 999	UNP P08037
D	119	MSE	-	SEE REMARK 999	UNP P08037
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	MSE	-	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
D	145	MSE	MET	CLONING ARTIFACT	UNP P08037
D	174	MSE	MET	CLONING ARTIFACT	UNP P08037
D	225	MSE	MET	CLONING ARTIFACT	UNP P08037
D	258	MSE	MET	CLONING ARTIFACT	UNP P08037
D	277	MSE	MET	CLONING ARTIFACT	UNP P08037
D	330	MSE	MET	CLONING ARTIFACT	UNP P08037
D	344	MSE	MET	CLONING ARTIFACT	UNP P08037
D	370	MSE	MET	CLONING ARTIFACT	UNP P08037
D	381	MSE	MET	CLONING ARTIFACT	UNP P08037

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

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

- Molecule 4 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).

PG4
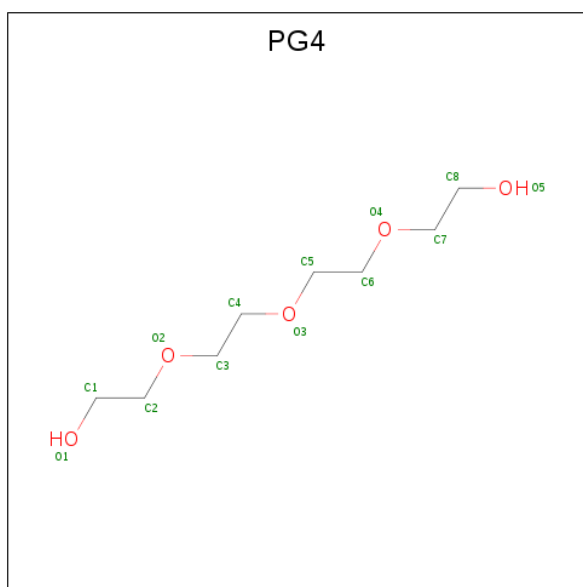
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 13	C 8	O 5	0	0
4	C	1	Total 13	C 8	O 5	0	0

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

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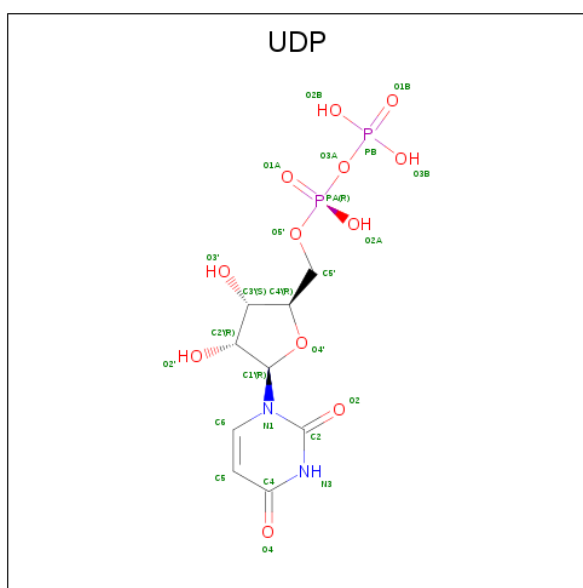
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total 25	C 9	N 2	O 12	P 2	0	0
5	D	1	Total 25	C 9	N 2	O 12	P 2	0	0

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	55	Total 55	O 55	0	0
6	B	134	Total 134	O 134	0	0
6	C	56	Total 56	O 56	0	0
6	D	86	Total 86	O 86	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.
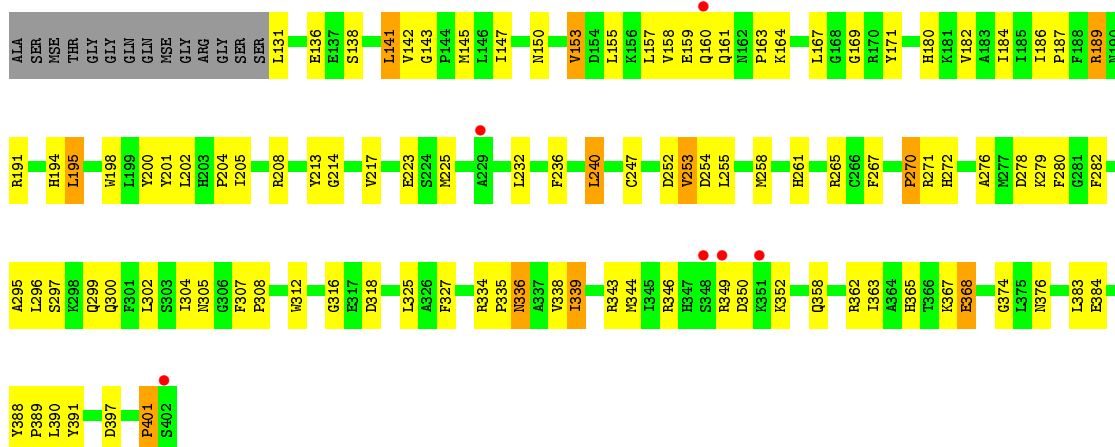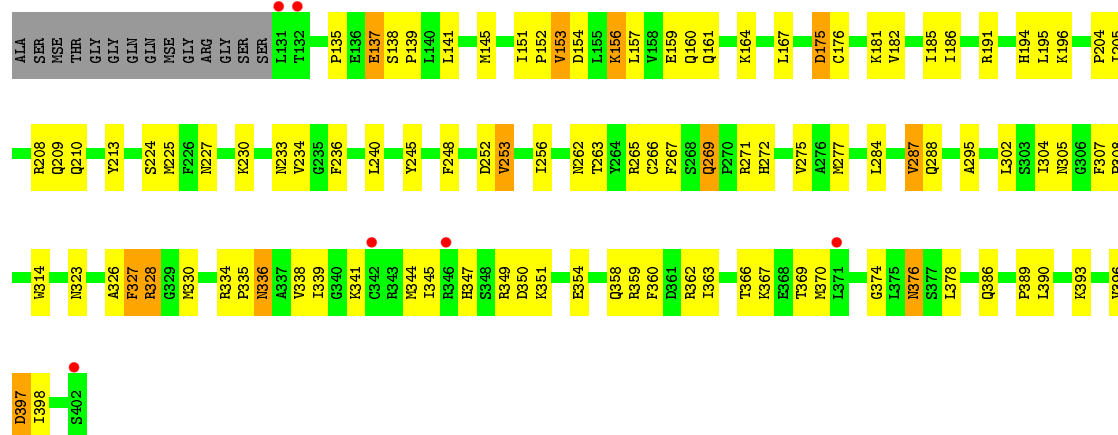
- Molecule 1: ALPHA-LACTALBUMIN
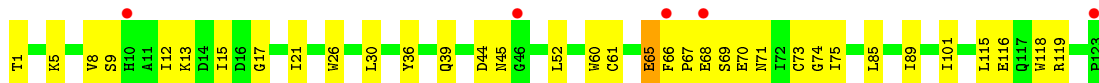
- Molecule 1: ALPHA-LACTALBUMIN
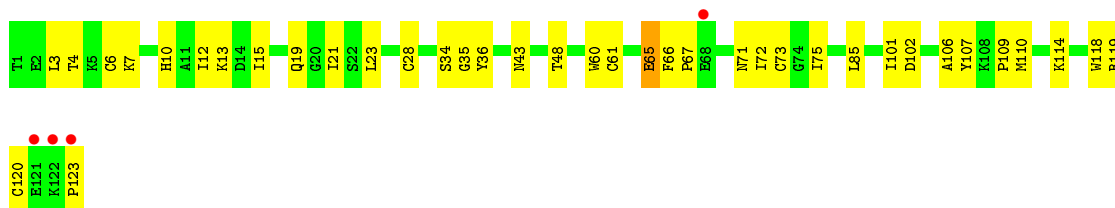
- Molecule 2: BETA-1,4-GALACTOSYLTRANSFERASE

- Molecule 2: BETA-1,4-GALACTOSYLTRANSFERASE

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.17Å 99.01Å 102.38Å 90.00° 103.92° 90.00°	Depositor
Resolution (Å)	19.90 – 2.50 19.90 – 2.49	Depositor EDS
% Data completeness (in resolution range)	99.4 (19.90-2.50) 99.0 (19.90-2.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 2.50Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.197 , 0.255 0.187 , 0.245	Depositor DCC
R_{free} test set	3686 reflections (9.94%)	wwPDB-VP
Wilson B-factor (Å^2)	43.6	Xtriage
Anisotropy	0.503	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 45.9	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.033 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6803	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.34% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: PG4, UDP, 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	0.51	0/1000	0.73	0/1350
1	C	0.57	0/1000	0.72	0/1350
2	B	0.54	0/2269	0.75	0/3058
2	D	0.49	0/2269	0.74	1/3058 (0.0%)
All	All	0.52	0/6538	0.74	1/8816 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	167	LEU	CA-CB-CG	5.15	127.15	115.30

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	979	0	936	32	0
1	C	979	0	936	35	0
2	B	2218	0	2185	83	0
2	D	2218	0	2185	114	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	13	0	18	3	0
4	C	13	0	18	3	0
5	B	25	0	11	2	0
5	D	25	0	11	5	0
6	A	55	0	0	2	0
6	B	134	0	0	6	0
6	C	56	0	0	2	0
6	D	86	0	0	3	0
All	All	6803	0	6300	259	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 20.

The worst 5 of 259 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)
2:D:336:ASN:HD21	2:D:339:ILE:H	1.04	0.98
2:D:336:ASN:ND2	2:D:339:ILE:H	1.66	0.93
2:D:335:PRO:HB2	2:D:339:ILE:HD11	1.53	0.89
2:B:155:LEU:O	2:B:159:GLU:HG3	1.73	0.88
2:D:336:ASN:C	2:D:336:ASN:HD22	1.81	0.84

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%)	110 (91%)	11 (9%)	0	100	100
1	C	121/123 (98%)	113 (93%)	8 (7%)	0	100	100
2	B	270/286 (94%)	255 (94%)	14 (5%)	1 (0%)	34	54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	270/286 (94%)	247 (92%)	22 (8%)	1 (0%)	34	54
All	All	782/818 (96%)	725 (93%)	55 (7%)	2 (0%)	41	61

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	327	PHE
2	B	401	PRO

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	109/109 (100%)	107 (98%)	2 (2%)	59	81
1	C	109/109 (100%)	107 (98%)	2 (2%)	59	81
2	B	245/243 (101%)	229 (94%)	16 (6%)	17	33
2	D	245/243 (101%)	229 (94%)	16 (6%)	17	33
All	All	708/704 (101%)	672 (95%)	36 (5%)	24	45

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	339	ILE
2	D	137	GLU
2	D	351	LYS
1	C	65	GLU
2	D	153	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	71	ASN
2	D	160	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	305	ASN
1	C	43	ASN
2	D	336	ASN

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 carbohydrates in this entry.

5.6   Ligand geometry (i)

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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	UDP	D	810	-	20,26,26	2.12	5 (25%)	25,40,40	1.59	5 (20%)
4	PG4	A	813	-	12,12,12	0.63	0	11,11,11	0.30	0
4	PG4	C	814	-	12,12,12	0.51	0	11,11,11	0.22	0
5	UDP	B	809	-	20,26,26	2.19	5 (25%)	25,40,40	1.60	5 (20%)

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	UDP	D	810	-	-	1/14/32/32	0/2/2/2
4	PG4	A	813	-	-	<mark>5/10/10/10</mark>	-
4	PG4	C	814	-	-	<mark>4/10/10/10</mark>	-
5	UDP	B	809	-	-	<mark>5/14/32/32</mark>	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	809	UDP	C4-N3	5.49	1.42	1.33
5	B	809	UDP	C6-N1	5.40	1.42	1.35
5	D	810	UDP	C6-N1	5.21	1.42	1.35
5	D	810	UDP	C4-N3	4.87	1.41	1.33
5	D	810	UDP	PB-O1B	3.81	1.62	1.50

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	809	UDP	C5-C4-N3	-4.05	114.39	123.31
5	D	810	UDP	C5-C4-N3	-3.93	114.67	123.31
5	B	809	UDP	O3B-PB-O1B	3.71	125.22	110.68
5	D	810	UDP	O3B-PB-O1B	3.70	125.17	110.68
5	B	809	UDP	O5'-C5'-C4'	3.16	119.88	108.99

There are no chirality outliers.

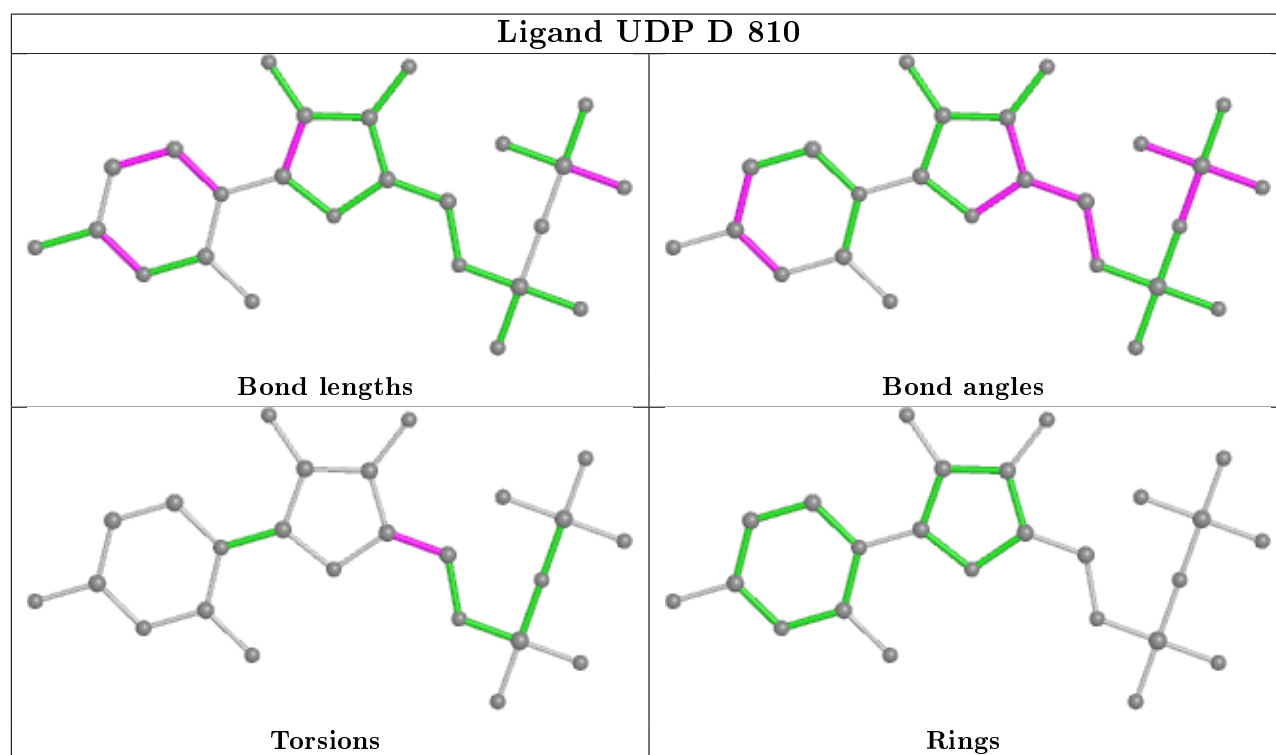5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	809	UDP	C5'-O5'-PA-O1A
5	B	809	UDP	C5'-O5'-PA-O2A
5	B	809	UDP	O4'-C4'-C5'-O5'
4	A	813	PG4	O2-C3-C4-O3
4	A	813	PG4	O3-C5-C6-O4

There are no ring outliers.

4 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	D	810	UDP	5	0
4	A	813	PG4	3	0
4	C	814	PG4	3	0
5	B	809	UDP	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand UDP D 810

Bond lengths

Bond angles

Torsions

Rings

Ligand UDP B 809

Bond lengths

Bond angles

Torsions

Rings
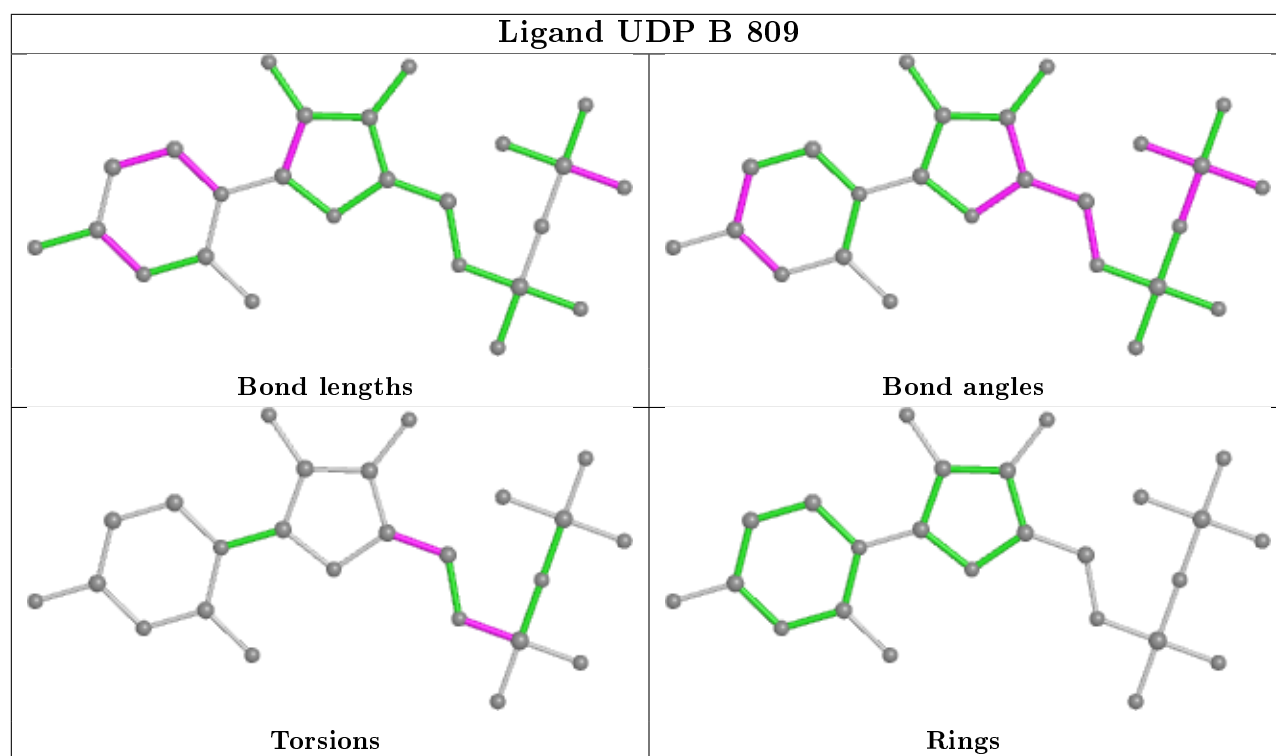
5.7 Other polymers (i)

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, 95th 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^2)	Q<0.9
1	A	123/123 (100%)	-0.03	5 (4%)	37	40	24, 42, 70, 85	0
1	C	123/123 (100%)	-0.06	4 (3%)	46	50	23, 37, 70, 101	0
2	B	263/286 (91%)	-0.18	6 (2%)	60	63	20, 38, 66, 82	0
2	D	263/286 (91%)	-0.13	6 (2%)	60	63	30, 44, 68, 81	0
All	All	772/818 (94%)	-0.12	21 (2%)	54	58	20, 41, 69, 101	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	123	PRO	8.4
1	C	122	LYS	4.4
2	D	371	LEU	4.3
2	D	346	ARG	4.2
2	B	402	SER	3.9

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

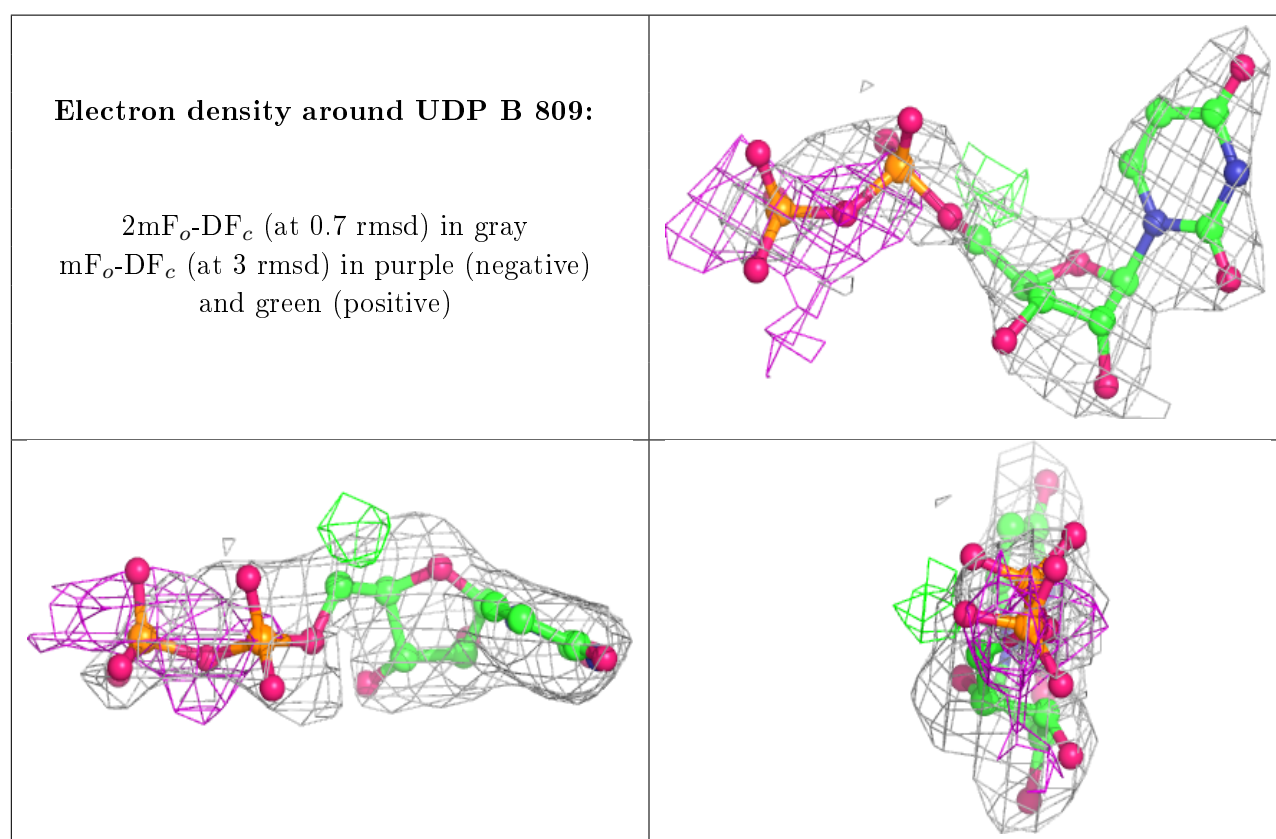There are no non-standard protein/DNA/RNA residues in this entry.

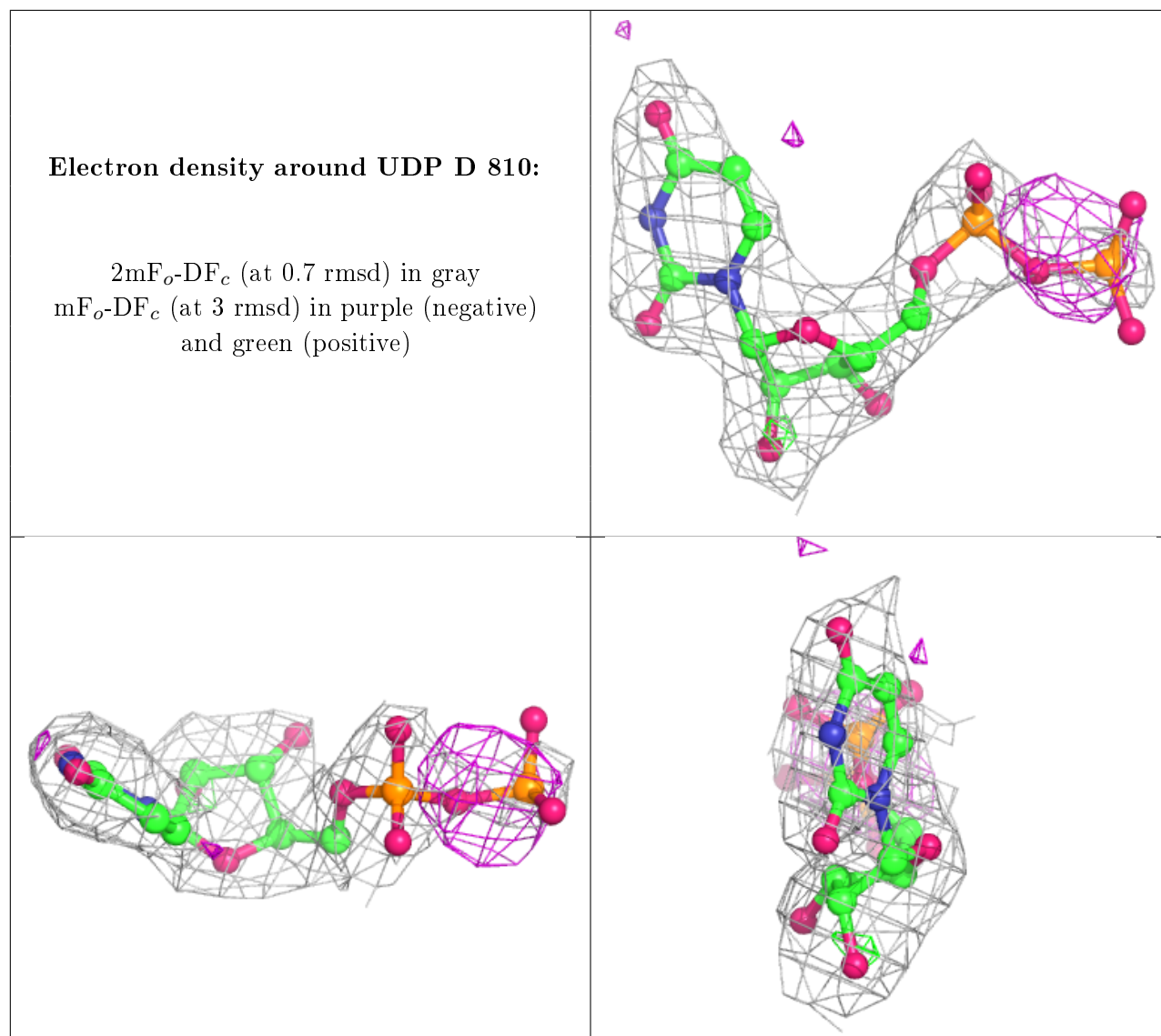## 6.3 Carbohydrates ⓘ

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. The B-factors column lists the minimum, median, 95th 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^2)	Q<0.9
5	UDP	B	809	25/25	0.73	0.31	71,79,110,110	0
5	UDP	D	810	25/25	0.80	0.27	58,68,104,104	0
4	PG4	A	813	13/13	0.87	0.21	51,53,58,58	0
4	PG4	C	814	13/13	0.90	0.17	64,66,67,67	0
3	CA	A	805	1/1	0.97	0.05	40,40,40,40	0
3	CA	C	806	1/1	0.98	0.08	35,35,35,35	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around UDP B 809:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative) and green (positive)

Electron density around UDP D 810:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.