



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 03:33 PM GMT

PDB ID : 4CJC  
Title : orthorhombic crystal form of Bogt6a E192Q in complex with UDP-GalNAc, UDP, GalNAc  
Authors : Pham, T.; Stinson, B.; Thiyagarajan, N.; Lizotte-Waniewski, M.; Brew, K.; Acharya, K.R.  
Deposited on : 2013-12-19  
Resolution : 3.42 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

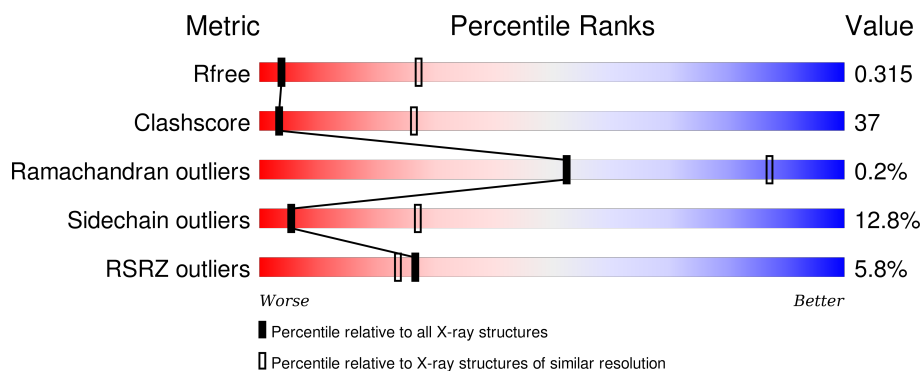
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.42 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	1049 (3.52-3.32)
Clashscore	102246	1032 (3.50-3.34)
Ramachandran outliers	100387	1002 (3.50-3.34)
Sidechain outliers	100360	1003 (3.50-3.34)
RSRZ outliers	91569	1054 (3.52-3.32)

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	246	<div> <div>6%</div> <div>48%</div> <div>41%</div> <div>9%</div> <div>.</div> </div>
1	B	246	<div> <div>51%</div> <div>39%</div> <div>7%</div> <div>.</div> </div>
1	C	246	<div> <div>6%</div> <div>51%</div> <div>35%</div> <div>12%</div> <div>.</div> </div>
1	D	246	<div> <div>15%</div> <div>40%</div> <div>48%</div> <div>7%</div> <div>.</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8120 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.

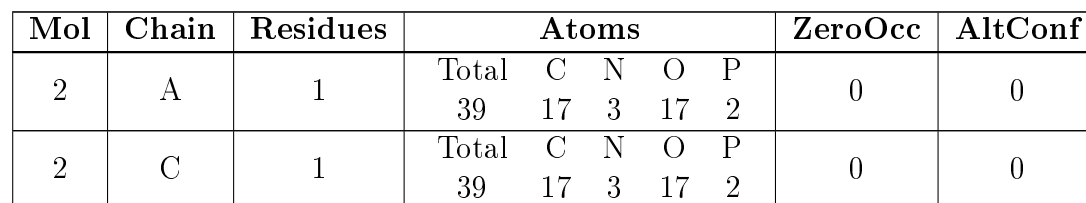
- Molecule 1 is a protein called GLYCOSYLTRANSFERASE FAMILY 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	241	Total	C	N	O	S	7	0	0
			2008	1312	327	362	7			
1	B	238	Total	C	N	O	S	10	0	0
			1981	1295	322	357	7			
1	C	240	Total	C	N	O	S	10	0	0
			2000	1306	326	361	7			
1	D	236	Total	C	N	O	S	10	0	0
			1973	1291	320	355	7			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	192	GLN	GLU	ENGINEERED MUTATION	UNP A7LVT2
B	192	GLN	GLU	ENGINEERED MUTATION	UNP A7LVT2
C	192	GLN	GLU	ENGINEERED MUTATION	UNP A7LVT2
D	192	GLN	GLU	ENGINEERED MUTATION	UNP A7LVT2

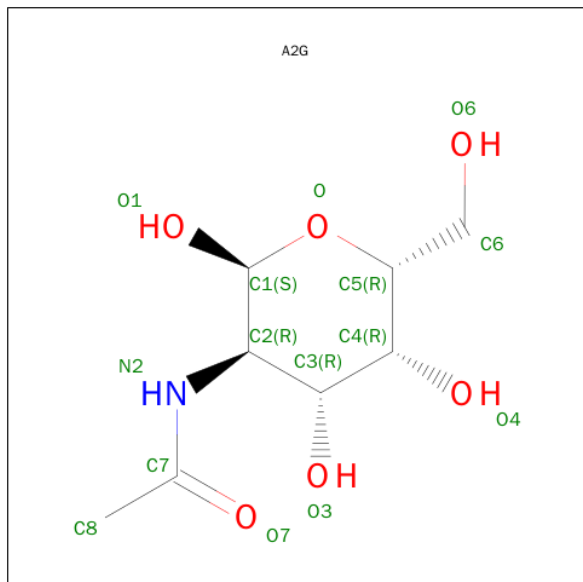
- Molecule 2 is URIDINE-DIPHOSPHATE-N-ACETYLGALACTOSAMINE (three-letter code: UD2) (formula: C<sub>17</sub>H<sub>27</sub>N<sub>3</sub>O<sub>17</sub>P<sub>2</sub>).



- 
- The chemical structure of UDP (Uridine Diphosphate) is shown. It consists of a uracil base (a six-membered ring with two nitrogen atoms, N1 and N3) attached to a ribose sugar (a five-membered ring with oxygen atom O4'). The ribose sugar is further attached to a diphosphate group (two phosphorus atoms, P1 and P2, connected by an oxygen atom O5'). The diphosphate group is shown with various oxygen atoms (O1A, O1B, O2A, O2B, O3A, O3B, O4, O5') and hydroxyl groups (OH). The structure is labeled with atom names and numbers, indicating the specific atoms involved in the reaction.

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

- Molecule 4 is SUGAR (N-ACETYL-2-DEOXY-2-AMINO-GALACTOSE) (three-letter code: A2G) (formula:  $C_8H_{15}NO_6$ ).

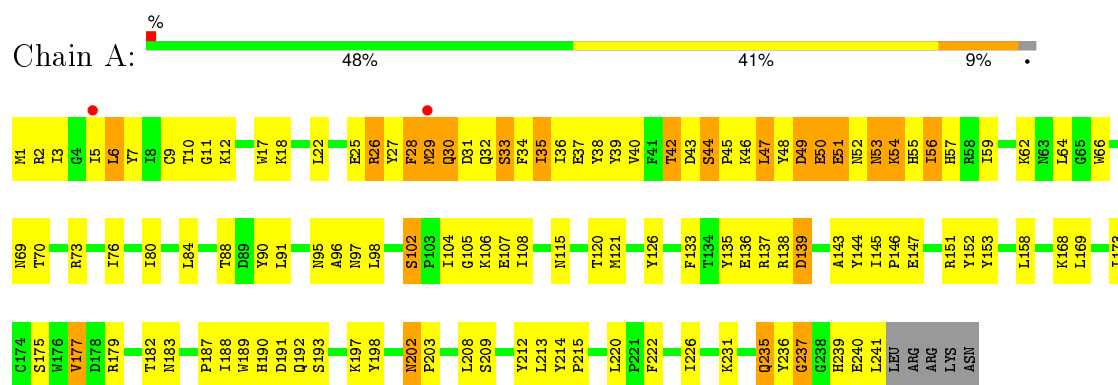


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	N	O	0	0
			15	8	1	6		
4	D	1	Total	C	N	O	0	0
			15	8	1	6		

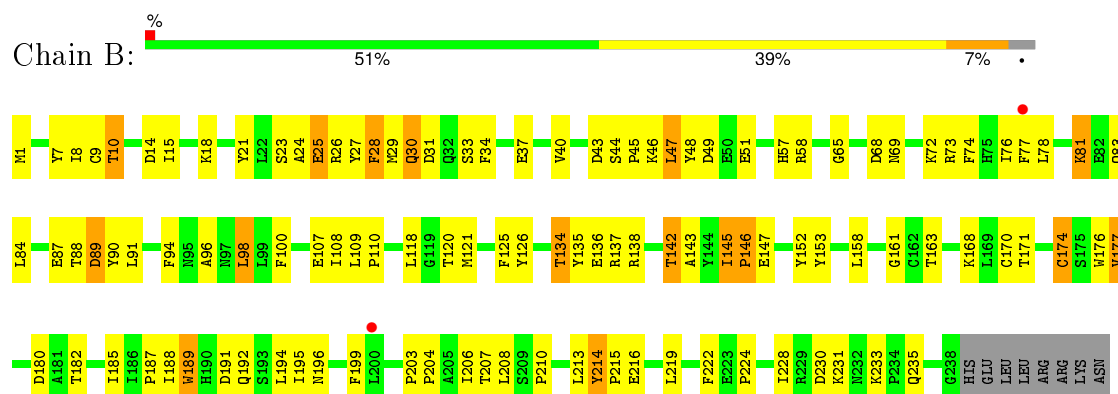
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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.

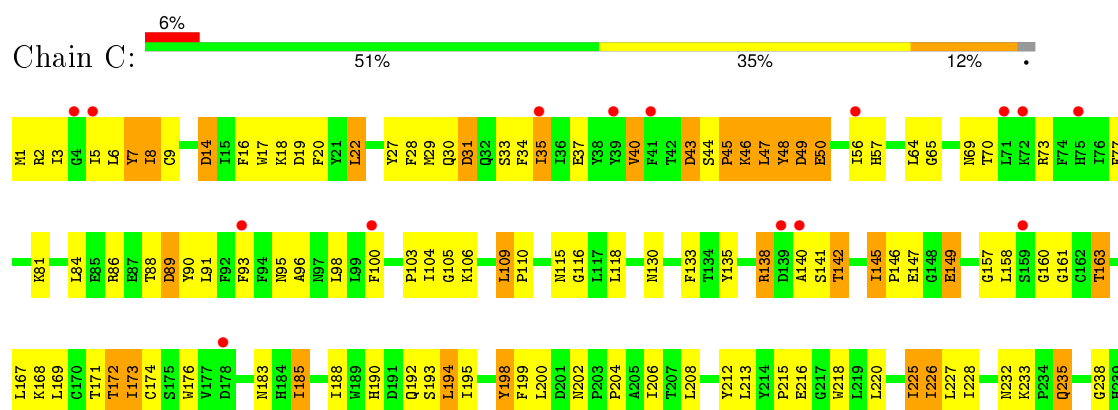
#### • Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6



#### • Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6

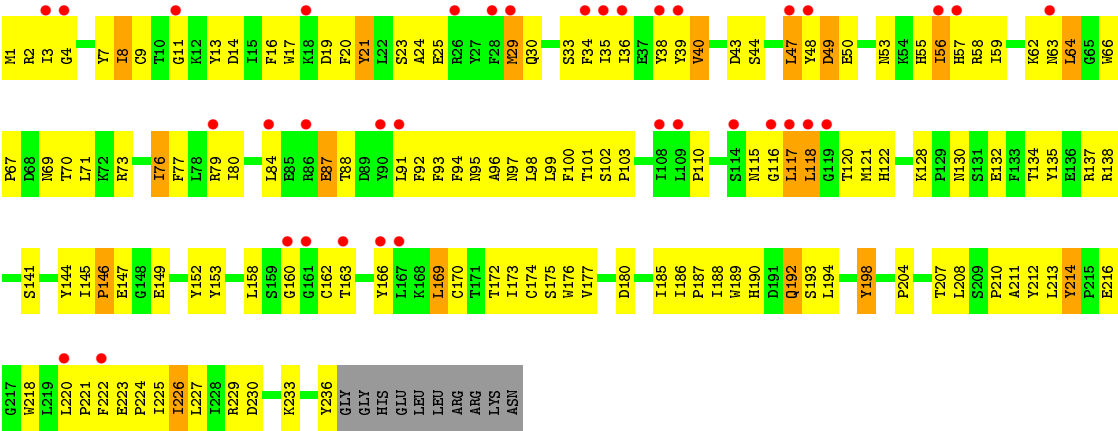


#### • Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6



E240
LEU
LEU
ARG
ARG
LYS
ASN

● Molecule 1: GLYCOSYLTRANSFERASE FAMILY 6



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	80.12Å 120.15Å 131.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	88.80 – 3.42 88.80 – 3.42	Depositor EDS
% Data completeness (in resolution range)	97.0 (88.80-3.42) 90.4 (88.80-3.42)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.12 (at 3.41Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.283 , 0.314 0.281 , 0.315	Depositor DCC
$R_{free}$ test set	878 reflections (5.74%)	DCC
Wilson B-factor (Å <sup>2</sup> )	89.6	Xtriage
Anisotropy	0.402	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 53.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	9 of 17344 reflections (0.052%)	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	8120	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.20% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: UD2, UDP, A2G

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	0/2074	0.65	2/2816 (0.1%)
1	B	0.47	0/2046	0.65	4/2778 (0.1%)
1	C	0.38	0/2066	0.62	0/2805
1	D	0.36	0/2038	0.67	2/2768 (0.1%)
All	All	0.43	0/8224	0.65	8/11167 (0.1%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	49	ASP	N-CA-C	-8.23	88.78	111.00
1	B	47	LEU	CA-CB-CG	-7.24	98.66	115.30
1	D	169	LEU	CA-CB-CG	6.82	130.98	115.30
1	A	237	GLY	N-CA-C	-6.36	97.20	113.10
1	B	33	SER	CB-CA-C	-6.03	98.64	110.10

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	2008	0	1944	201	0
1	B	1981	0	1920	100	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2000	0	1933	130	0
1	D	1973	0	1914	157	0
2	A	39	0	25	9	0
2	C	39	0	25	5	0
3	B	25	0	11	2	0
3	D	25	0	11	2	0
4	B	15	0	15	0	0
4	D	15	0	15	1	0
All	All	8120	0	7813	586	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 37.

The worst 5 of 586 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:D:29:MET:HG3	1:D:34:PHE:CZ	1.33	1.63
1:C:48:TYR:CD1	1:C:49:ASP:HB3	1.41	1.52
1:A:3:ILE:HD12	1:A:29:MET:CE	1.41	1.46
1:A:3:ILE:CD1	1:A:29:MET:HE3	1.42	1.44
1:A:188:ILE:HD13	1:A:189:TRP:CD1	1.49	1.44

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	239/246 (97%)	222 (93%)	17 (7%)	0	100	100
1	B	236/246 (96%)	219 (93%)	16 (7%)	1 (0%)	39	79
1	C	238/246 (97%)	219 (92%)	19 (8%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	234/246 (95%)	218 (93%)	15 (6%)	1 (0%)	39	79
All	All	947/984 (96%)	878 (93%)	67 (7%)	2 (0%)	52	87

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	146	PRO
1	B	146	PRO

### 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	216/221 (98%)	184 (85%)	32 (15%)	4	21
1	B	213/221 (96%)	194 (91%)	19 (9%)	12	47
1	C	215/221 (97%)	177 (82%)	38 (18%)	2	12
1	D	213/221 (96%)	192 (90%)	21 (10%)	10	41
All	All	857/884 (97%)	747 (87%)	110 (13%)	5	27

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	206	ILE
1	C	45	PRO
1	D	117	LEU
1	C	1	MET
1	C	22	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	53	ASN
1	D	196	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	69	ASN
1	A	202	ASN
1	D	130	ASN

### 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 ⓘ

6 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	UD2	A	1242	-	32,41,41	0.56	0	46,62,62	1.56	5 (10%)
3	UDP	B	1239	-	18,26,26	0.71	0	26,40,40	1.77	2 (7%)
4	A2G	B	1240	-	15,15,15	0.39	0	17,21,21	0.72	0
2	UD2	C	1241	-	32,41,41	0.55	0	46,62,62	1.71	7 (15%)
3	UDP	D	1237	-	18,26,26	0.66	0	26,40,40	1.65	2 (7%)
4	A2G	D	1238	-	15,15,15	0.40	0	17,21,21	0.42	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	UD2	A	1242	-	-	0/22/63/63	0/3/3/3
3	UDP	B	1239	-	-	0/12/32/32	0/2/2/2
4	A2G	B	1240	-	-	0/6/26/26	0/1/1/1
2	UD2	C	1241	-	-	0/22/63/63	0/3/3/3
3	UDP	D	1237	-	-	0/12/32/32	0/2/2/2
4	A2G	D	1238	-	-	2/6/26/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	1237	UDP	PA-O3A-PB	-3.56	120.73	132.67
2	C	1241	UD2	PB-O3A-PA	-3.55	122.75	132.73
3	B	1239	UDP	PA-O3A-PB	-3.30	121.61	132.67
2	A	1242	UD2	PB-O3A-PA	-3.11	123.99	132.73
2	C	1241	UD2	C2'-N2'-C7'	-2.68	116.24	123.10

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	1238	A2G	C8-C7-N2-C2
4	D	1238	A2G	O7-C7-N2-C2

There are no ring outliers.

5 monomers are involved in 19 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1242	UD2	9	0
3	B	1239	UDP	2	0
2	C	1241	UD2	5	0
3	D	1237	UDP	2	0
4	D	1238	A2G	1	0

## 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 [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	241/246 (97%)	-0.10	2 (0%) 87 82	23, 51, 79, 94	2 (0%)
1	B	238/246 (96%)	-0.02	2 (0%) 87 82	33, 64, 84, 95	3 (1%)
1	C	240/246 (97%)	0.28	15 (6%) 23 21	25, 97, 148, 184	3 (1%)
1	D	236/246 (95%)	0.73	36 (15%) 3 3	46, 119, 199, 358	3 (1%)
All	All	955/984 (97%)	0.22	55 (5%) 26 24	23, 74, 160, 358	11 (1%)

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	116	GLY	7.8
1	D	166	TYR	5.1
1	D	108	ILE	4.7
1	D	161	GLY	4.4
1	D	84	LEU	4.1

### 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
4	A2G	B	1240	15/15	0.83	0.23	0.50	62,79,86,89	0
3	UDP	B	1239	25/25	0.84	0.24	0.09	65,78,105,116	0
2	UD2	C	1241	39/39	0.84	0.21	-0.46	69,86,97,98	0
3	UDP	D	1237	25/25	0.80	0.20	-0.55	97,127,136,141	0
2	UD2	A	1242	39/39	0.91	0.19	-0.70	33,49,62,65	0
4	A2G	D	1238	15/15	0.88	0.19	-2.82	50,59,71,72	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.