Full wwPDB X-ray Structure Validation Report ⓘ

Apr 29, 2025 – 09:55 AM EDT

PDB ID	:	4N3A / pdb_00004n3a
Title	:	Crystal Structure of human O-GlcNAc transferase bound to a peptide from HCF-1 pro-repeat 2 (1-26)E10A
Authors	:	Lazarus, M.B.; Herr, W.; Walker, S.
Deposited on	:	2013-10-06
Resolution	:	1.88 Å(reported)

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

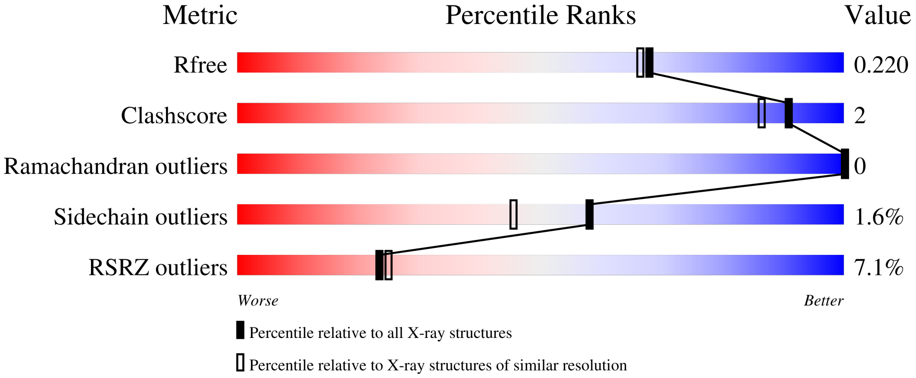
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

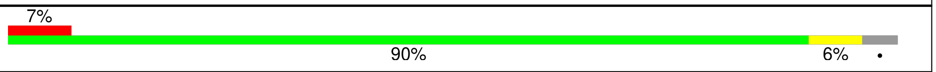
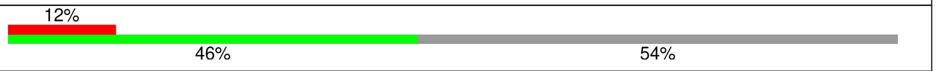
The reported resolution of this entry is 1.88 Å.

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	1090 (1.88-1.88)
Clashscore	180529	1144 (1.88-1.88)
Ramachandran outliers	177936	1135 (1.88-1.88)
Sidechain outliers	177891	1135 (1.88-1.88)
RSRZ outliers	164620	1090 (1.88-1.88)

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 >=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	723	
2	B	26	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5783 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 UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	697	5530	3512	966	1014	38	0	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	309	GLY	-	expression tag	UNP O15294
A	310	PRO	-	expression tag	UNP O15294
A	311	GLY	-	expression tag	UNP O15294
A	312	SER	-	expression tag	UNP O15294

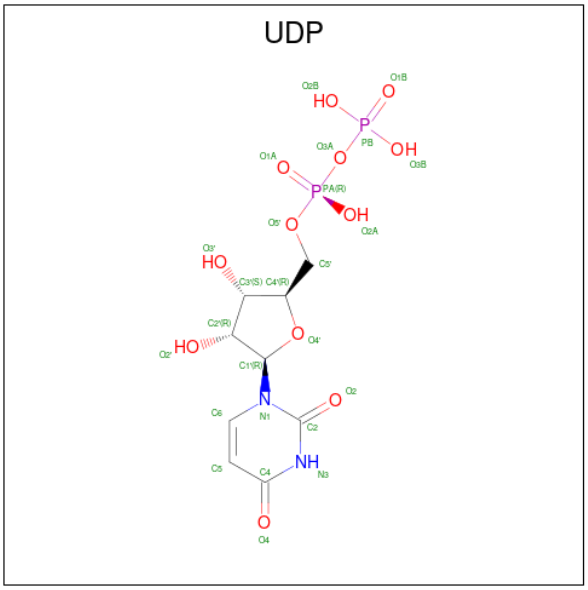
- Molecule 2 is a protein called Host cell factor 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	12	80	45	13	22	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	10	ALA	GLU	SEE REMARK 999	UNP P51610

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (CCD ID: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$).



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

- Molecule 4 is water.

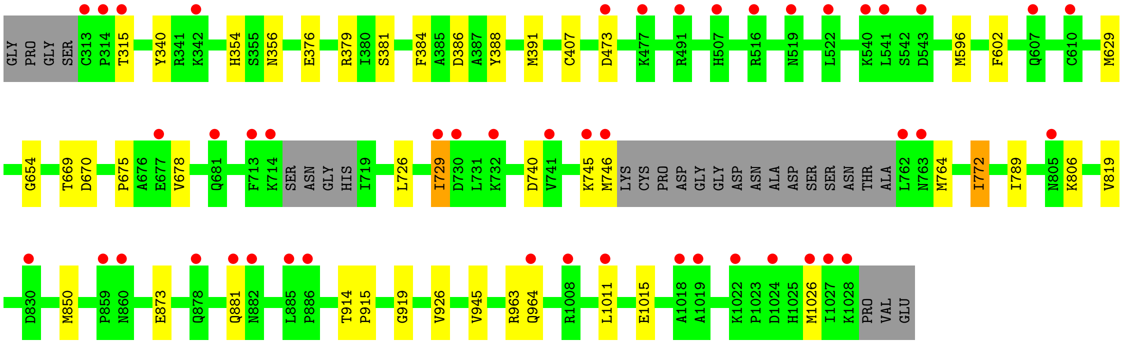
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	140	140	140	0	0
4	B	8	8	8	0	0

3 Residue-property plots ⓘ

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.

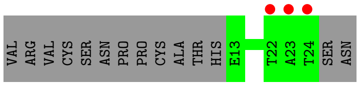
- Molecule 1: UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit

Chain A:



- Molecule 2: Host cell factor 1

Chain B:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α, β, γ	98.93Å 98.93Å 367.01Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.86 – 1.88 45.86 – 1.88	Depositor EDS
% Data completeness (in resolution range)	94.3 (45.86-1.88) 94.3 (45.86-1.88)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	1.87 (at 1.88Å)	Xtriage
Refinement program	PHENIX 1.7.3_928	Depositor
R, R_{free}	0.200 , 0.222 0.198 , 0.220	Depositor DCC
R_{free} test set	4137 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	30.4	Xtriage
Anisotropy	0.011	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 42.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5783	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% 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: UDP

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	0.48	0/5667	0.78	5/7687 (0.1%)
2	B	0.44	0/79	0.63	0/109
All	All	0.48	0/5746	0.78	5/7796 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	945	VAL	N-CA-C	5.76	115.95	110.42
1	A	919	GLY	N-CA-C	-5.51	104.92	111.36
1	A	381	SER	CA-C-N	5.29	124.96	119.56
1	A	381	SER	C-N-CA	5.29	124.96	119.56
1	A	740	ASP	N-CA-C	5.13	116.47	110.41

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	5530	0	5513	18	1
2	B	80	0	73	0	0
3	A	25	0	11	0	0
4	A	140	0	0	1	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	8	0	0	0	0
All	All	5783	0	5597	18	1

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 2.

All (18) 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:726:LEU:HD22	1:A:819:VAL:HG22	1.74	0.69
1:A:746:MET:HE3	1:A:764:MET:HE2	1.83	0.59
1:A:629:MET:O	1:A:654:GLY:HA3	2.09	0.53
1:A:729:ILE:HD13	1:A:729:ILE:HA	1.80	0.47
1:A:806:LYS:HE2	4:A:1379:HOH:O	2.15	0.47
1:A:1011:LEU:O	1:A:1015:GLU:HG2	2.15	0.47
1:A:675:PRO:O	1:A:678:VAL:HG22	2.15	0.46
1:A:850:MET:HE1	1:A:915:PRO:HG2	1.98	0.45
1:A:596:MET:HG2	1:A:602:PHE:CD1	2.51	0.45
1:A:669:THR:OG1	1:A:670:ASP:N	2.49	0.44
1:A:388:TYR:O	1:A:407:CYS:HB3	2.19	0.43
1:A:384:PHE:CE1	1:A:386:ASP:HB2	2.54	0.43
1:A:772:ILE:HG12	1:A:789:ILE:HG21	2.01	0.43
1:A:850:MET:HE2	1:A:963:ARG:NH1	2.33	0.43
1:A:391:MET:HE3	1:A:407:CYS:SG	2.59	0.43
1:A:354:HIS:CE1	1:A:376:GLU:HG2	2.55	0.42
1:A:376:GLU:OE1	1:A:379:ARG:NH2	2.38	0.40
1:A:340:TYR:CZ	1:A:356:ASN:HB3	2.56	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:873:GLU:OE1	4:A:1434:HOH:O[12_545]	2.06	0.14

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	695/723 (96%)	678 (98%)	17 (2%)	0	100	100
2	B	10/26 (38%)	10 (100%)	0	0	100	100
All	All	705/749 (94%)	688 (98%)	17 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	603/618 (98%)	593 (98%)	10 (2%)	56	43
2	B	9/22 (41%)	9 (100%)	0	100	100
All	All	612/640 (96%)	602 (98%)	10 (2%)	58	45

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	315	THR
1	A	473	ASP
1	A	729	ILE
1	A	745	LYS
1	A	772	ILE
1	A	881	GLN
1	A	914	THR
1	A	926	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	964	GLN
1	A	1026	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	332	ASN
1	A	490	ASN
1	A	681	GLN
1	A	691	HIS
1	A	700	ASN
1	A	802	GLN
1	A	839	GLN
1	A	1021	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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

1 ligand is 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	$\#\|Z\| > 2$
3	UDP	A	1201	-	25,26,26	0.97	1 (4%)	38,40,40	1.66	5 (13%)

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
3	UDP	A	1201	-	-	2/16/32/32	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1201	UDP	C6-C5	2.06	1.39	1.35

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1201	UDP	C4-N3-C2	-5.78	119.43	126.61
3	A	1201	UDP	C5-C4-N3	4.32	120.85	114.80
3	A	1201	UDP	N3-C2-N1	4.08	120.21	114.89
3	A	1201	UDP	O2A-PA-O3A	2.80	114.83	107.27
3	A	1201	UDP	O4-C4-C5	-2.56	120.74	125.16

There are no chirality outliers.

All (2) torsion outliers are listed below:

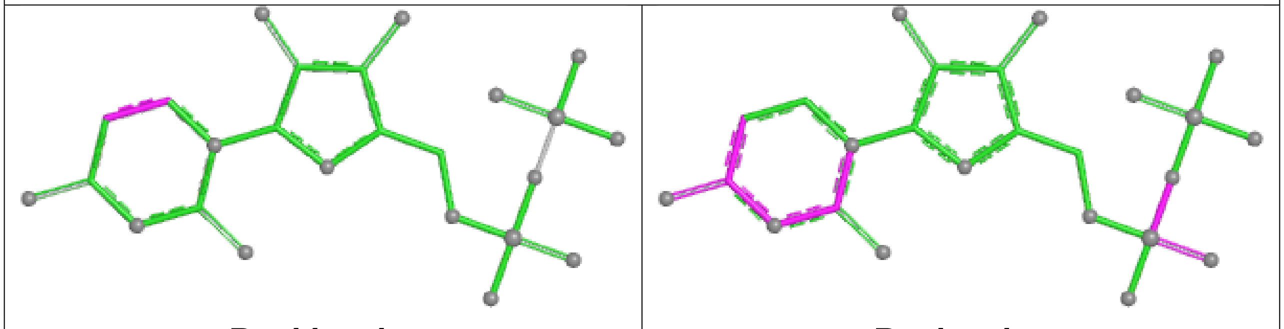
Mol	Chain	Res	Type	Atoms
3	A	1201	UDP	O4'-C4'-C5'-O5'
3	A	1201	UDP	C3'-C4'-C5'-O5'

There are no ring outliers.

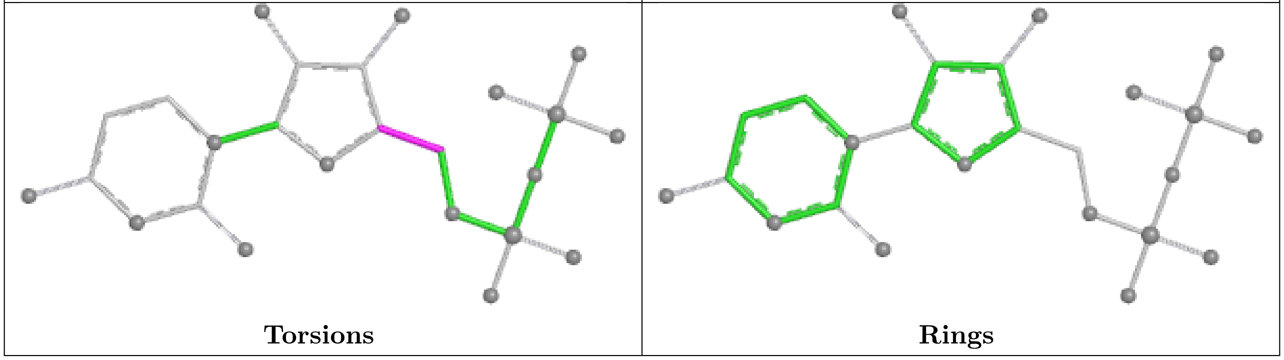
No monomer is involved in short contacts.

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 A 1201



Bond lengths **Bond angles**



Torsions **Rings**

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^{th} 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	697/723 (96%)	0.24	47 (6%)	25	27	14, 35, 68, 94	4 (0%)
2	B	12/26 (46%)	0.92	3 (25%)	2	1	33, 39, 75, 75	0
All	All	709/749 (94%)	0.25	50 (7%)	23	25	14, 35, 68, 94	4 (0%)

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	762	LEU	9.3
1	A	314	PRO	8.8
1	A	882	ASN	6.3
1	A	1028	LYS	5.4
1	A	763	ASN	5.3
1	A	313	CYS	5.2
1	A	714	LYS	5.1
1	A	830	ASP	4.3
1	A	964	GLN	4.1
1	A	860	ASN	3.8
2	B	24	THR	3.5
1	A	859	PRO	3.4
1	A	881	GLN	3.3
1	A	730	ASP	3.2
1	A	713	PHE	3.1
1	A	342	LYS	3.1
1	A	491	ARG	2.9
1	A	746	MET	2.9
1	A	805	ASN	2.9
1	A	729	ILE	2.8
1	A	681	GLN	2.7
1	A	1022	LYS	2.7
1	A	1027	ILE	2.7
1	A	522	LEU	2.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	1011	LEU	2.6
1	A	1018	ALA	2.6
1	A	886	PRO	2.6
1	A	315	THR	2.6
1	A	745	LYS	2.5
1	A	507	HIS	2.5
1	A	540	LYS	2.5
1	A	1019	ALA	2.4
1	A	878	GLN	2.4
1	A	885	LEU	2.4
1	A	610	CYS	2.4
1	A	741	VAL	2.3
2	B	22	THR	2.3
2	B	23	ALA	2.2
1	A	516	ARG	2.2
1	A	677	GLU	2.2
1	A	543	ASP	2.2
1	A	519	ASN	2.2
1	A	607	GLN	2.2
1	A	1008	ARG	2.2
1	A	1024	ASP	2.1
1	A	541	LEU	2.1
1	A	477	LYS	2.1
1	A	473	ASP	2.0
1	A	1026	MET	2.0
1	A	732	LYS	2.0

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 ⓘ

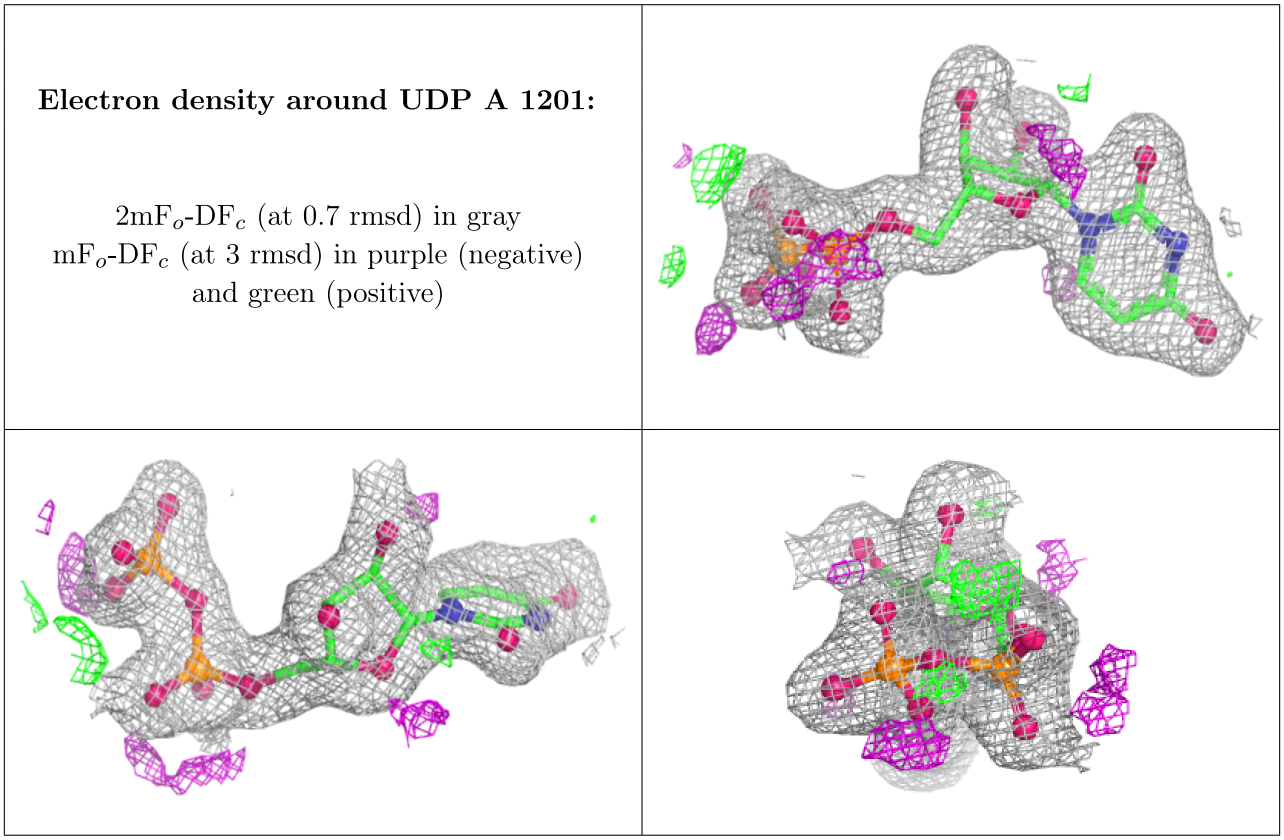
There are no monosaccharides in this entry.

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, 95^{th} 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($\mathrm{\AA}^2$)	Q<0.9
3	UDP	A	1201	25/25	0.98	0.04	17,22,26,27	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 A 1201:

$2mF_o$-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.