



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 04:32 am GMT

PDB ID : 4AY5
Title : Human O-GlcNAc transferase (OGT) in complex with UDP and glycopeptide
Authors : Schimpl, M.; Zheng, X.; Blair, D.E.; Schuettelkopf, A.W.; Navratilova, I.; Aristotelous, T.; Ferenbach, A.T.; Macnaughtan, M.A.; Borodkin, V.S.; Van Aalten, D.M.F.
Deposited on : 2012-06-18
Resolution : 3.15 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

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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

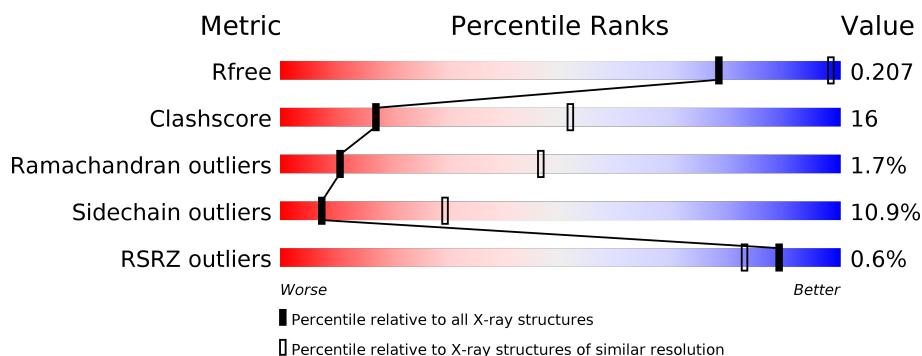
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.15 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}	100719	1259 (3.20-3.12)
Clashscore	112137	1397 (3.20-3.12)
Ramachandran outliers	110173	1368 (3.20-3.12)
Sidechain outliers	110143	1367 (3.20-3.12)
RSRZ outliers	101464	1264 (3.20-3.12)

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. 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	723	<div> <div></div> <div>63% 29% 5% .</div> </div>
1	B	723	<div> <div></div> <div>61% 29% 6% .</div> </div>
1	C	723	<div> <div></div> <div>63% 28% 5% .</div> </div>
1	D	723	<div> <div></div> <div>63% 29% . .</div> </div>
2	I	11	<div> <div></div> <div>27% 55% 18%</div> </div>
2	J	11	<div> <div>18%</div> <div>18% 55% 27%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	K	11	<div><div></div><div>18%27%45%27%</div></div>
2	L	11	<div><div></div><div>9%18%55%27%</div></div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 22524 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-ACETYLGLUCOSAMINYL TRANSFERASE 110 KDA SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	698	Total	C	N	O	S	0	0	0
			5514	3499	964	1013	38			
1	B	698	Total	C	N	O	S	0	0	0
			5514	3499	964	1013	38			
1	C	698	Total	C	N	O	S	0	0	0
			5514	3499	964	1013	38			
1	D	698	Total	C	N	O	S	0	0	0
			5514	3499	964	1013	38			

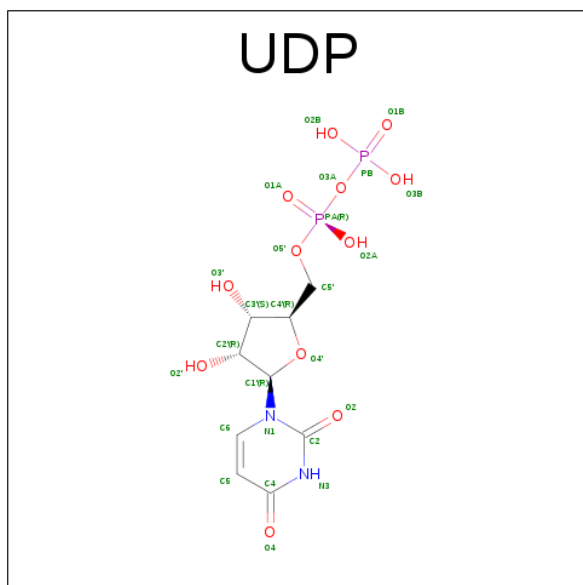
There are 16 discrepancies between the modelled and reference sequences:

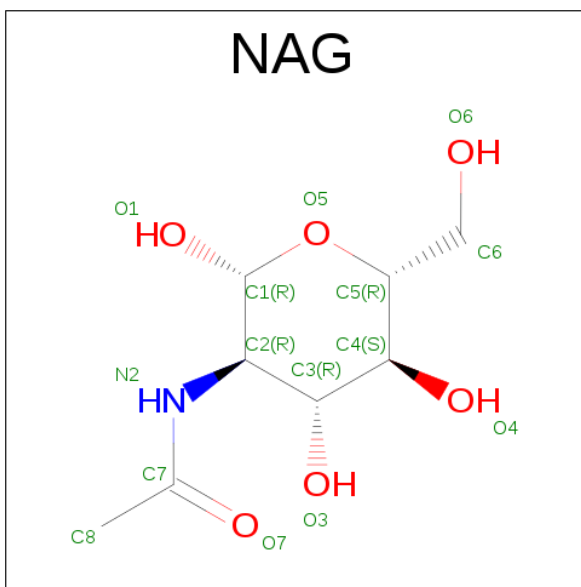
Chain	Residue	Modelled	Actual	Comment	Reference
A	309	GLY	-	EXPRESSION TAG	UNP O15294
A	310	PRO	-	EXPRESSION TAG	UNP O15294
A	311	GLY	-	EXPRESSION TAG	UNP O15294
A	312	SER	-	EXPRESSION TAG	UNP O15294
B	309	GLY	-	EXPRESSION TAG	UNP O15294
B	310	PRO	-	EXPRESSION TAG	UNP O15294
B	311	GLY	-	EXPRESSION TAG	UNP O15294
B	312	SER	-	EXPRESSION TAG	UNP O15294
C	309	GLY	-	EXPRESSION TAG	UNP O15294
C	310	PRO	-	EXPRESSION TAG	UNP O15294
C	311	GLY	-	EXPRESSION TAG	UNP O15294
C	312	SER	-	EXPRESSION TAG	UNP O15294
D	309	GLY	-	EXPRESSION TAG	UNP O15294
D	310	PRO	-	EXPRESSION TAG	UNP O15294
D	311	GLY	-	EXPRESSION TAG	UNP O15294
D	312	SER	-	EXPRESSION TAG	UNP O15294

- Molecule 2 is a protein called GTAB1TIDE.

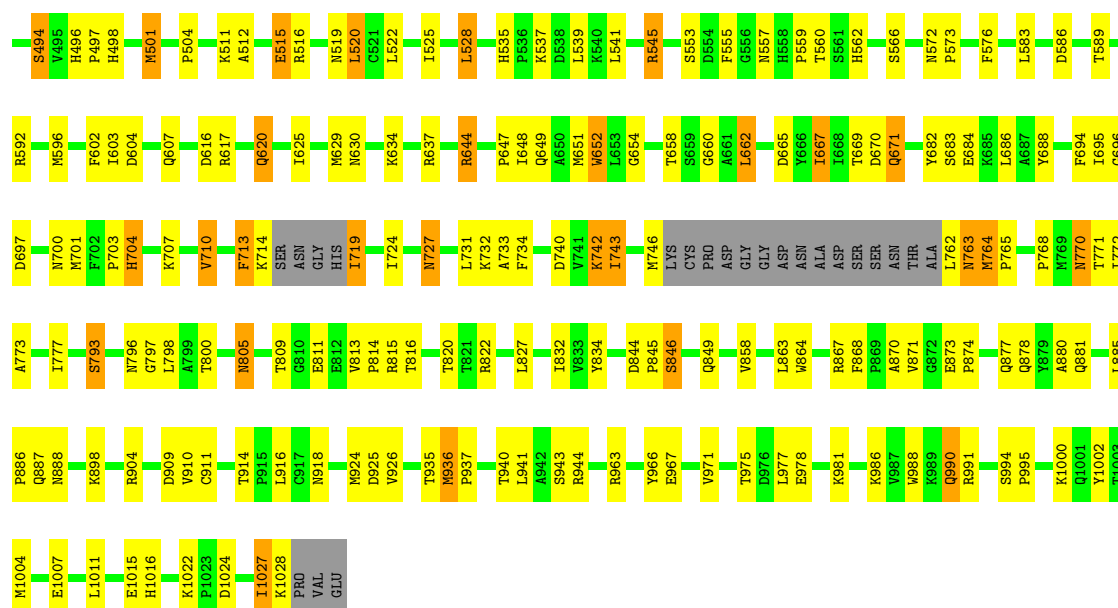
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	I	11	Total	C	N	O	0	0	0
			78	49	12	17			
2	J	11	Total	C	N	O	0	0	0
			78	49	12	17			
2	K	11	Total	C	N	O	0	0	0
			78	49	12	17			
2	L	11	Total	C	N	O	0	0	0
			78	49	12	17			

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



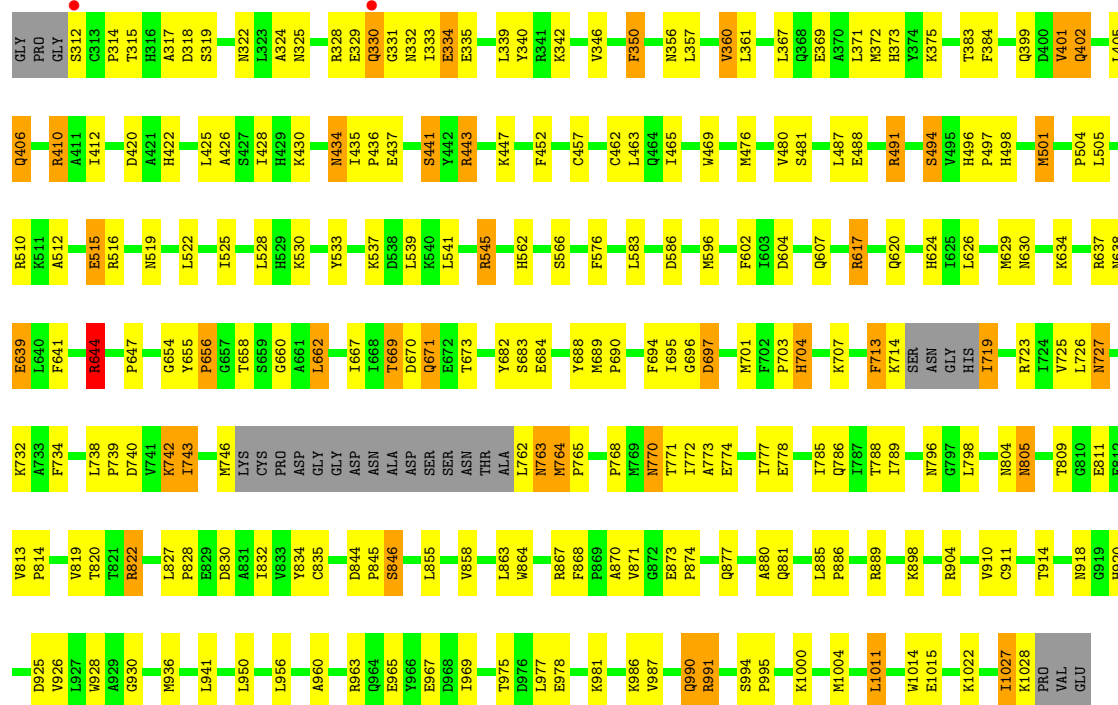


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	I	1	Total	C	N	O	0	0
			14	8	1	5		
4	J	1	Total	C	N	O	0	0
			14	8	1	5		
4	K	1	Total	C	N	O	0	0
			14	8	1	5		
4	L	1	Total	C	N	O	0	0
			14	8	1	5		



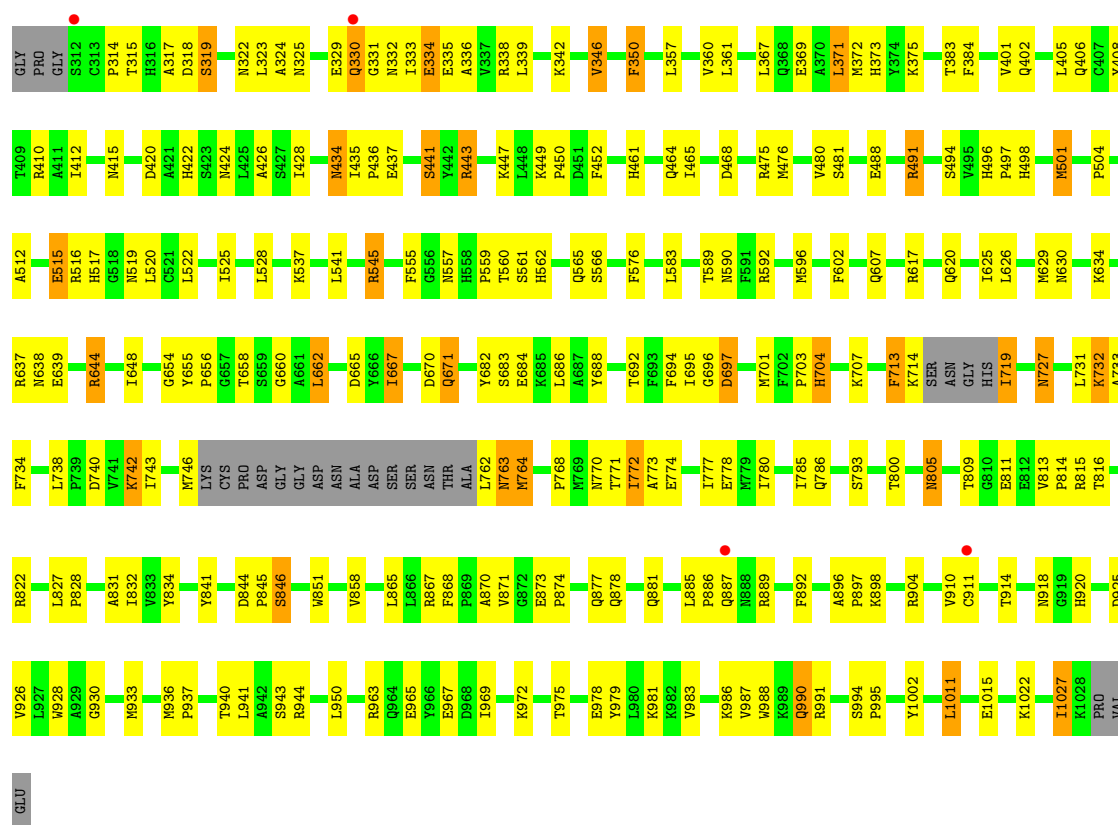
• Molecule 1: UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYL TRANSFERASE 110 KDA SUBUNIT

Chain C: 63% 28% 5% •

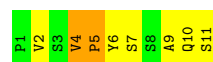
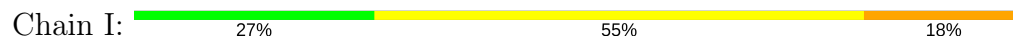


• Molecule 1: UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYL TRANSFERASE 110 KDA SUBUNIT

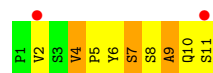
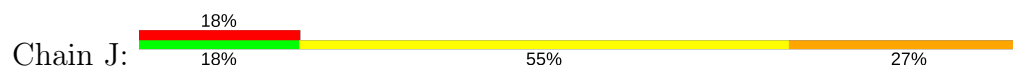
Chain D: 63% 29% • •



• Molecule 2: GTAB1TIDE



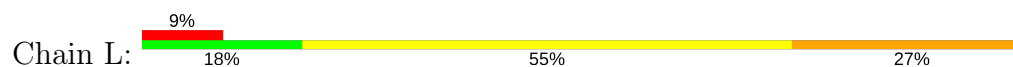
• Molecule 2: GTAB1TIDE



• Molecule 2: GTAB1TIDE



• Molecule 2: GTAB1TIDE



4 Data and refinement statistics

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α , β , γ	274.17Å 274.17Å 142.15Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	237.44 – 3.15 29.91 – 3.15	Depositor EDS
% Data completeness (in resolution range)	99.6 (237.44-3.15) 99.7 (29.91-3.15)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.06 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.174 , 0.204 0.173 , 0.207	Depositor DCC
R_{free} test set	501 reflections (0.48%)	DCC
Wilson B-factor (Å ²)	67.1	Xtriage
Anisotropy	0.006	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 58.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.036 for -h,-k,l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	22524	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.68% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: UDP, NAG

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	0.93	3/5641 (0.1%)	1.03	11/7650 (0.1%)
1	B	0.91	2/5641 (0.0%)	1.01	10/7650 (0.1%)
1	C	0.91	2/5641 (0.0%)	1.01	10/7650 (0.1%)
1	D	0.80	3/5641 (0.1%)	0.92	5/7650 (0.1%)
2	I	1.59	2/80 (2.5%)	1.63	0/109
2	J	1.67	1/80 (1.2%)	1.59	0/109
2	K	1.57	1/80 (1.2%)	1.61	0/109
2	L	1.50	1/80 (1.2%)	1.55	0/109
All	All	0.90	15/22884 (0.1%)	1.01	36/31036 (0.1%)

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	L	6	TYR	CB-CG	-6.71	1.41	1.51
1	A	988	TRP	CD2-CE2	6.64	1.49	1.41
1	B	652	TRP	CD2-CE2	6.36	1.49	1.41
2	J	6	TYR	CB-CG	-6.16	1.42	1.51
1	A	598	GLU	CD-OE1	6.13	1.32	1.25

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	904	ARG	NE-CZ-NH2	-9.25	115.68	120.30
1	A	904	ARG	NE-CZ-NH2	-7.90	116.35	120.30
1	B	925	ASP	CB-CG-OD1	7.52	125.07	118.30
1	A	855	LEU	CB-CG-CD1	-7.24	98.70	111.00
1	B	904	ARG	NE-CZ-NH2	-6.63	116.98	120.30

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	5514	0	5489	169	0
1	B	5514	0	5489	191	0
1	C	5514	0	5489	167	0
1	D	5514	0	5489	165	0
2	I	78	0	75	4	0
2	J	78	0	75	5	0
2	K	78	0	75	4	0
2	L	78	0	75	5	0
3	A	25	0	11	2	0
3	B	25	0	11	2	0
3	C	25	0	11	2	0
3	D	25	0	11	3	0
4	I	14	0	13	0	0
4	J	14	0	13	0	0
4	K	14	0	13	0	0
4	L	14	0	13	0	0
All	All	22524	0	22352	701	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 16.

The worst 5 of 701 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:443:ARG:HH11	1:A:443:ARG:HG2	1.26	1.01
1:A:773:ALA:O	1:A:777:ILE:HG22	1.61	1.00
1:D:713:PHE:H	1:D:713:PHE:HD1	1.03	0.99
1:A:713:PHE:HD1	1:A:713:PHE:H	1.02	0.96
1:B:662:LEU:HD12	1:B:662:LEU:H	1.30	0.95

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	692/723 (96%)	622 (90%)	63 (9%)	7 (1%)	18	60
1	B	692/723 (96%)	623 (90%)	59 (8%)	10 (1%)	13	51
1	C	692/723 (96%)	618 (89%)	65 (9%)	9 (1%)	14	53
1	D	692/723 (96%)	624 (90%)	58 (8%)	10 (1%)	13	51
2	I	9/11 (82%)	6 (67%)	0	3 (33%)	0	0
2	J	9/11 (82%)	6 (67%)	0	3 (33%)	0	0
2	K	9/11 (82%)	6 (67%)	0	3 (33%)	0	0
2	L	9/11 (82%)	6 (67%)	0	3 (33%)	0	0
All	All	2804/2936 (96%)	2511 (90%)	245 (9%)	48 (2%)	11	46

5 of 48 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	2	VAL
2	I	9	ALA
2	J	2	VAL
2	J	9	ALA
2	K	2	VAL

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	600/618 (97%)	535 (89%)	65 (11%)	7	30

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	600/618 (97%)	535 (89%)	65 (11%)	7	30
1	C	600/618 (97%)	537 (90%)	63 (10%)	8	31
1	D	600/618 (97%)	537 (90%)	63 (10%)	8	31
2	I	10/10 (100%)	8 (80%)	2 (20%)	1	7
2	J	10/10 (100%)	7 (70%)	3 (30%)	0	1
2	K	10/10 (100%)	7 (70%)	3 (30%)	0	1
2	L	10/10 (100%)	7 (70%)	3 (30%)	0	1
All	All	2440/2512 (97%)	2173 (89%)	267 (11%)	7	29

5 of 267 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	871	VAL
1	C	488	GLU
1	D	871	VAL
1	B	981	LYS
1	C	371	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 44 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	881	GLN
1	C	406	GLN
1	D	805	ASN
1	B	990	GLN
1	C	322	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 [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

8 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
3	UDP	A	1201	-	21,26,26	1.12	2 (9%)	22,40,40	2.13	4 (18%)
3	UDP	B	1201	-	21,26,26	0.90	0	22,40,40	2.39	4 (18%)
3	UDP	C	1201	-	21,26,26	1.10	1 (4%)	22,40,40	2.65	2 (9%)
3	UDP	D	1201	-	21,26,26	0.82	1 (4%)	22,40,40	2.37	4 (18%)
4	NAG	I	1202	2	14,14,15	1.82	3 (21%)	15,19,21	2.37	4 (26%)
4	NAG	J	1202	2	14,14,15	1.74	4 (28%)	15,19,21	2.25	4 (26%)
4	NAG	K	1202	2	14,14,15	1.74	4 (28%)	15,19,21	2.28	4 (26%)
4	NAG	L	1202	2	14,14,15	1.60	4 (28%)	15,19,21	2.05	4 (26%)

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	-	-	0/12/32/32	0/2/2/2
3	UDP	B	1201	-	-	0/12/32/32	0/2/2/2
3	UDP	C	1201	-	-	0/12/32/32	0/2/2/2
3	UDP	D	1201	-	-	0/12/32/32	0/2/2/2
4	NAG	I	1202	2	-	0/6/23/26	0/1/1/1
4	NAG	J	1202	2	-	0/6/23/26	0/1/1/1
4	NAG	K	1202	2	-	0/6/23/26	0/1/1/1
4	NAG	L	1202	2	-	0/6/23/26	0/1/1/1

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	J	1202	NAG	O5-C1	-3.84	1.37	1.43
4	L	1202	NAG	O5-C1	-2.41	1.39	1.43
4	K	1202	NAG	C4-C5	-2.30	1.48	1.53
4	I	1202	NAG	O5-C1	-2.27	1.40	1.43
3	A	1201	UDP	C2-N3	-2.26	1.33	1.38

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	J	1202	NAG	C2-N2-C7	-3.44	117.92	122.94
4	K	1202	NAG	O5-C1-C2	-2.66	107.78	111.47
3	D	1201	UDP	C4'-O4'-C1'	-2.61	106.99	109.77
4	I	1202	NAG	C2-N2-C7	-2.56	119.22	122.94
4	I	1202	NAG	C6-C5-C4	-2.55	107.03	113.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1201	UDP	2	0
3	B	1201	UDP	2	0
3	C	1201	UDP	2	0
3	D	1201	UDP	3	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, 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	698/723 (96%)	-0.53	4 (0%) 89 83	30, 53, 95, 148	0
1	B	698/723 (96%)	-0.54	2 (0%) 93 92	27, 54, 91, 137	0
1	C	698/723 (96%)	-0.59	2 (0%) 93 92	30, 51, 88, 122	0
1	D	698/723 (96%)	-0.41	4 (0%) 89 83	36, 66, 97, 134	0
2	I	11/11 (100%)	0.09	0 100 100	67, 82, 106, 108	0
2	J	11/11 (100%)	0.52	2 (18%) 1 1	74, 85, 121, 140	0
2	K	11/11 (100%)	0.51	2 (18%) 1 1	66, 86, 111, 142	0
2	L	11/11 (100%)	0.01	1 (9%) 10 5	82, 101, 121, 129	0
All	All	2836/2936 (96%)	-0.50	17 (0%) 89 83	27, 56, 97, 148	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	K	11	SER	4.9
1	B	312	SER	4.2
2	J	11	SER	4.1
1	D	911	CYS	3.2
2	L	11	SER	3.1

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, 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	LLDF	B-factors(Å ²)	Q<0.9
4	NAG	J	1202	14/15	0.96	0.23	1.65	56,69,83,85	0
4	NAG	K	1202	14/15	0.98	0.23	1.12	54,59,76,81	0
4	NAG	L	1202	14/15	0.95	0.19	0.73	63,79,93,94	0
4	NAG	I	1202	14/15	0.97	0.17	0.44	57,67,77,81	0
3	UDP	B	1201	25/25	0.97	0.17	-0.22	35,62,72,76	0
3	UDP	C	1201	25/25	0.98	0.17	-0.24	44,57,63,64	0
3	UDP	D	1201	25/25	0.98	0.17	-0.31	60,76,83,85	0
3	UDP	A	1201	25/25	0.97	0.16	-0.44	32,55,61,64	0

6.5 Other polymers [i](#)

There are no such residues in this entry.