W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Dec 3, 2023 – 09:21 pm GMT

PDB ID	:	2CBJ
Title	:	Structure of the Clostridium perfringens NagJ family 84 glycoside hydrolase, a homologue of human O-GlcNAcase in complex with PUGNAc
Authors	:	Rao, F.V.; Dorfmueller, H.C.; Villa, F.; Allwood, M.; Eggleston, I.M.; van Aalten, D.M.F.
Deposited on	:	2006-01-05
Resolution	:	2.35 Å(reported)

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

MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

PERCENTILES INFOmissingINFO

1   Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 9589 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.
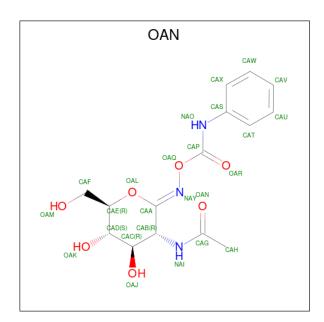
- Molecule 1 is a protein called HYALURONIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	585	Total 4628	C 2913	N 757	O 941	S 17	0	0	0
1	B	585	Total 4629	C 2913	N 758	O 941	S 17	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	196	GLN	LYS	conflict	UNP Q8XL08
A	234	SER	ASN	conflict	UNP Q8XL08
A	244	ASN	ASP	conflict	UNP Q8XL08
A	268	ASP	GLU	conflict	UNP Q8XL08
A	279	THR	ALA	conflict	UNP Q8XL08
A	348	ALA	THR	conflict	UNP Q8XL08
B	196	GLN	LYS	conflict	UNP Q8XL08
B	234	SER	ASN	conflict	UNP Q8XL08
B	244	ASN	ASP	conflict	UNP Q8XL08
B	268	ASP	GLU	conflict	UNP Q8XL08
B	279	THR	ALA	conflict	UNP Q8XL08
B	348	ALA	THR	conflict	UNP Q8XL08

- Molecule 2 is O-(2-ACETAMIDO-2-DEOXY D-GLUCOPYRANOSYLIDENE) AMINO-N-PHENYLCARBAMATE (three-letter code: OAN) (formula: $C_{15}H_{19}N_3O_7$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 25	C 15	N 3	O 7	0	0
2	B	1	Total 25	C 15	N 3	O 7	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total 2	Cl 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	143	Total 143	O 143	0	0
4	B	137	Total 137	O 137	0	0

SEQUENCE-PLOTS INFOmissingINFO

2 Data and refinement statistics ⓘ

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	129.61Å 145.75Å 152.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.35 19.99 – 2.35	Depositor EDS
% Data completeness (in resolution range)	93.5 (20.00-2.35) 93.3 (19.99-2.35)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 2.35Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.193 , 0.248 0.192 , 0.246	Depositor DCC
R_{free} test set	562 reflections (1.00%)	wwPDB-VP
Wilson B-factor (Å^2)	41.7	Xtriage
Anisotropy	0.291	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 29.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.011 for -h,-l,-k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9589	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.48% 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.

3 Model quality ⓘ

3.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: OAN, CL

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	0.66	0/4724	0.69	2/6413 (0.0%)
1	B	0.65	1/4725 (0.0%)	0.67	2/6415 (0.0%)
All	All	0.65	1/9449 (0.0%)	0.68	4/12828 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	267	GLY	C-O	5.49	1.32	1.23

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	502	ARG	NE-CZ-NH2	-6.81	116.90	120.30
1	B	502	ARG	NE-CZ-NH1	6.26	123.43	120.30
1	A	389	ARG	NE-CZ-NH1	5.71	123.15	120.30
1	A	502	ARG	NE-CZ-NH2	-5.34	117.63	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	97	ASN	Peptide
1	A	98	ASP	Peptide

3.2 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	4628	0	4418	62	0
1	B	4629	0	4420	52	0
2	A	25	0	19	1	0
2	B	25	0	19	1	0
3	B	2	0	0	0	0
4	A	143	0	0	2	0
4	B	137	0	0	0	0
All	All	9589	0	8876	109	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 6.

The worst 5 of 109 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:123:THR:HG21	1:A:172:ASN:HD21	1.05	1.09
1:A:123:THR:HG21	1:A:172:ASN:ND2	1.72	1.03
1:B:305:ALA:O	1:B:309:GLN:HG2	1.61	1.00
1:A:619:PHE:HE2	1:A:621:LEU:HD13	1.29	0.93
1:A:619:PHE:CE2	1:A:621:LEU:HD13	2.10	0.87

There are no symmetry-related clashes.

3.3 Torsion angles ⓘ

3.3.1 Protein backbone ⓘ

There are no protein backbone outliers to report in this entry.

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

3.3.2   Protein sidechains ⓘ

There are no protein residues with a non-rotameric sidechain to report in this entry.

3.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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

3.5   Carbohydrates ⓘ

There are no monosaccharides in this entry.

3.6   Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	#$|Z| > 2$
2	OAN	A	1625	-	24,26,26	1.34	3 (12%)	24,35,35	1.18	2 (8%)
2	OAN	B	1625	-	24,26,26	1.37	3 (12%)	24,35,35	1.70	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
2	OAN	A	1625	-	-	3/12/35/35	0/2/2/2
2	OAN	B	1625	-	-	3/12/35/35	0/2/2/2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1625	OAN	CAS-NAO	-3.48	1.34	1.41
2	B	1625	OAN	OAQ-NAY	-3.00	1.37	1.44
2	B	1625	OAN	CAH-CAG	2.95	1.56	1.50
2	A	1625	OAN	CAA-NAY	2.93	1.34	1.27
2	A	1625	OAN	CAC-CAB	2.69	1.56	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1625	OAN	CAS-NAO-CAP	-5.75	116.91	126.36
2	B	1625	OAN	CAC-CAB-NAI	-3.89	106.18	112.28
2	A	1625	OAN	CAF-CAE-CAD	2.66	119.23	113.00
2	B	1625	OAN	OAR-CAP-NAO	-2.38	120.92	126.11
2	B	1625	OAN	OAL-CAA-CAB	2.28	124.54	118.10

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1625	OAN	OAR-CAP-OAQ-NAY
2	A	1625	OAN	NAO-CAP-OAQ-NAY
2	B	1625	OAN	CAT-CAS-NAO-CAP
2	B	1625	OAN	CAX-CAS-NAO-CAP
2	A	1625	OAN	CAA-CAB-NAI-CAG

There are no ring outliers.

2 monomers are involved in 2 short contacts:

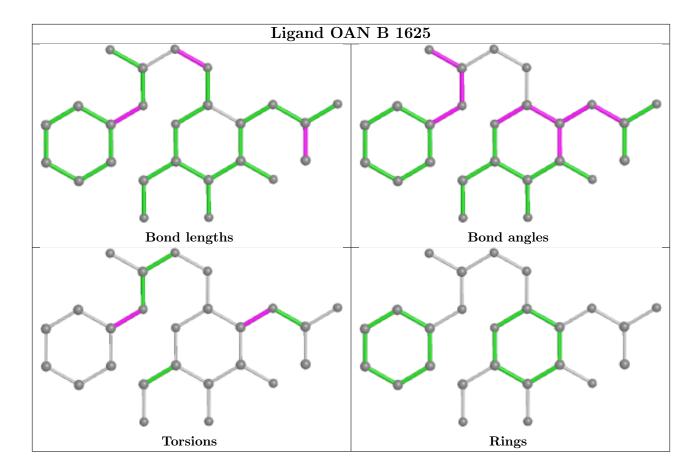| Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1625	OAN	1	0
2	B	1625	OAN	1	0

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 OAN A 1625

Bond lengths

Bond angles

Torsions

Rings

Ligand OAN B 1625

Bond lengths

Bond angles

Torsions

Rings

3.7 Other polymers ⓘ

There are no such residues in this entry.

3.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

4 Fit of model and data ⓘ

4.1 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(Å²)	Q<0.9
1	A	585/594 (98%)	-0.27	6 (1%)	82	88	27, 45, 74, 90	14 (2%)
1	B	585/594 (98%)	-0.23	11 (1%)	66	76	26, 46, 80, 98	10 (1%)
All	All	1170/1188 (98%)	-0.25	17 (1%)	73	81	26, 45, 78, 98	24 (2%)

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	99	PRO	4.1
1	B	122	GLY	4.0
1	B	343	ASN	3.5
1	A	40	GLN	3.3
1	A	99	PRO	2.8

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

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

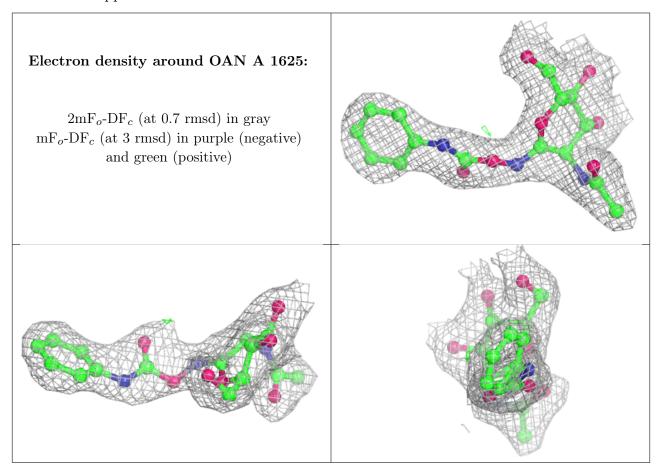## 4.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

4.4 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.
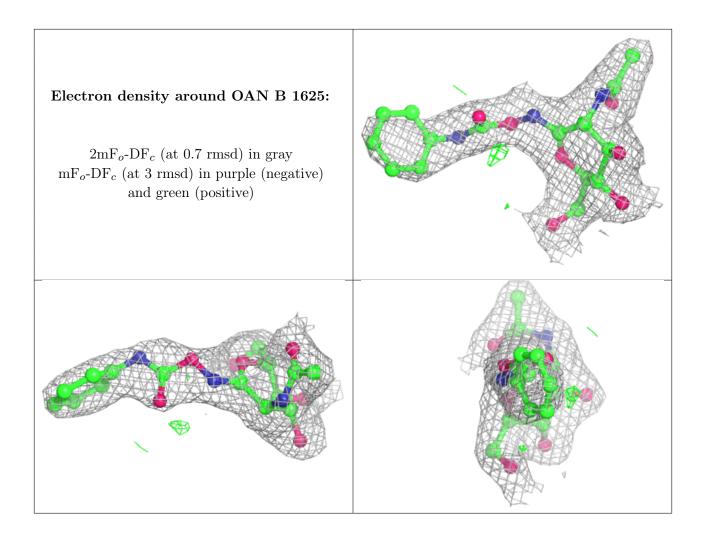
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	CL	B	1627	1/1	0.89	0.34	59,59,59,59	1
3	CL	B	1626	1/1	0.93	0.12	63,63,63,63	0
2	OAN	A	1625	25/25	0.95	0.12	25,37,49,54	0
2	OAN	B	1625	25/25	0.95	0.14	22,35,68,70	0

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 OAN A 1625:

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

Electron density around OAN B 1625:

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

4.5 Other polymers (i)

There are no such residues in this entry.