WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 10, 2024 – 03:19 PM EST

PDB ID	:	2OZN
Title	:	The Cohesin-Dockerin Complex of NagJ and NagH from Clostridium perfringens
Authors	:	Adams, J.J.; Boraston, A.; Smith, S.P.
Deposited on	:	2007-02-26
Resolution	:	1.60 Å(reported)

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

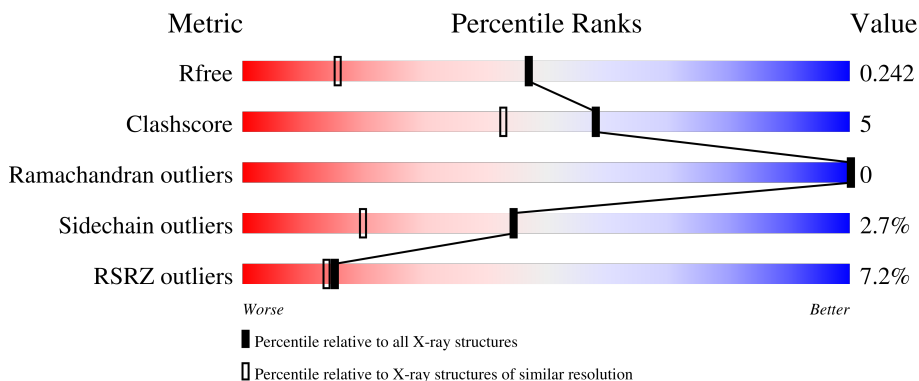|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

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 1.60 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		0.242
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		2.7%
RSRZ outliers		7.2%

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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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	165	3% / 73% / 7% / 19%
2	B	140	10% / 86% / 8% / 6%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 2313 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 O-GlcNAcase nagJ.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	133	991	620	157	214	0	0	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	745	MET	-	initiating methionine	UNP Q0TR53
A	746	GLY	-	expression tag	UNP Q0TR53
A	747	SER	-	expression tag	UNP Q0TR53
A	748	SER	-	expression tag	UNP Q0TR53
A	749	HIS	-	expression tag	UNP Q0TR53
A	750	HIS	-	expression tag	UNP Q0TR53
A	751	HIS	-	expression tag	UNP Q0TR53
A	752	HIS	-	expression tag	UNP Q0TR53
A	753	HIS	-	expression tag	UNP Q0TR53
A	754	HIS	-	expression tag	UNP Q0TR53
A	755	SER	-	expression tag	UNP Q0TR53
A	756	SER	-	expression tag	UNP Q0TR53
A	757	GLY	-	expression tag	UNP Q0TR53
A	758	LEU	-	expression tag	UNP Q0TR53
A	759	VAL	-	expression tag	UNP Q0TR53
A	760	PRO	-	expression tag	UNP Q0TR53
A	761	ARG	-	expression tag	UNP Q0TR53
A	762	GLY	-	expression tag	UNP Q0TR53
A	763	SER	-	expression tag	UNP Q0TR53
A	764	HIS	-	expression tag	UNP Q0TR53
A	765	MET	-	expression tag	UNP Q0TR53
A	766	ALA	-	expression tag	UNP Q0TR53
A	767	SER	-	expression tag	UNP Q0TR53

- Molecule 2 is a protein called Hyaluronoglucosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	131	1004	618	166	219	1	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1497	MET	-	expression tag	UNP P26831
B	1629	LEU	-	expression tag	UNP P26831
B	1630	GLU	-	expression tag	UNP P26831
B	1631	HIS	-	expression tag	UNP P26831
B	1632	HIS	-	expression tag	UNP P26831
B	1633	HIS	-	expression tag	UNP P26831
B	1634	HIS	-	expression tag	UNP P26831
B	1635	HIS	-	expression tag	UNP P26831
B	1636	HIS	-	expression tag	UNP P26831

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
3	A	1	1	1	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
4	B	2	2	2	0	0

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	181	181	181	0	0
5	B	134	134	134	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: O-GlcNAcase nagJ
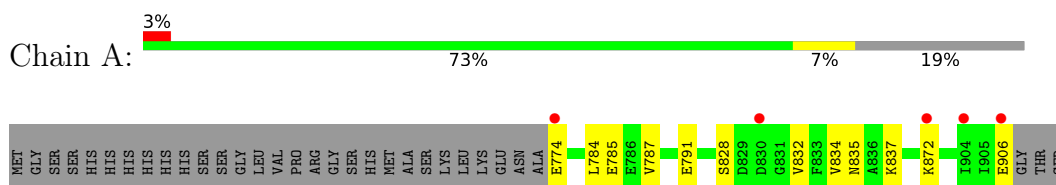
• Molecule 2: Hyaluronoglucosaminidase
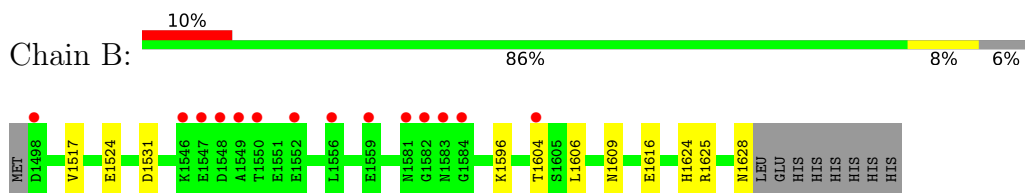
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	35.49Å 74.59Å 94.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	13.88 – 1.60 13.88 – 1.60	Depositor EDS
% Data completeness (in resolution range)	98.8 (13.88-1.60) 98.8 (13.88-1.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.83 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.204 , 0.246 0.202 , 0.242	Depositor DCC
R_{free} test set	1689 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	21.4	Xtriage
Anisotropy	0.033	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 61.9	EDS
L-test for twinning[2]	$<|L|> = 0.50$, $<L^2> = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2313	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.75% 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: CL, 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	0.56	0/1000	0.69	0/1357
2	B	0.46	0/1012	0.60	0/1360
All	All	0.51	0/2012	0.64	0/2717

There are no bond length outliers.

There are no bond angle outliers.

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	991	0	970	13	0
2	B	1004	0	967	7	0
3	A	1	0	0	0	0
4	B	2	0	0	0	0
5	A	181	0	0	2	0
5	B	134	0	0	5	0
All	All	2313	0	1937	19	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 5.

The worst 5 of 19 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:B:1531:ASP:HB2	5:B:1695:HOH:O	1.65	0.95
2:B:1604:THR:C	5:B:1720:HOH:O	2.18	0.82
2:B:1624:HIS:HD2	5:B:1722:HOH:O	1.65	0.79
2:B:1625:ARG:HG2	5:B:1739:HOH:O	1.87	0.73
1:A:828:SER:HB2	1:A:832:VAL:CG2	2.26	0.65

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	131/165 (79%)	130 (99%)	1 (1%)	0	100	100
2	B	129/140 (92%)	124 (96%)	5 (4%)	0	100	100
All	All	260/305 (85%)	254 (98%)	6 (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	108/134 (81%)	105 (97%)	3 (3%)	43	18
2	B	111/120 (92%)	108 (97%)	3 (3%)	44	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	219/254 (86%)	213 (97%)	6 (3%)	44	20

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1596	LYS
2	B	1606	LEU
2	B	1609	ASN
1	A	872	LYS
1	A	835	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	877	ASN
2	B	1624	HIS
2	B	1628	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	133/165 (80%)	-0.02	5 (3%)	40	37	12, 18, 27, 37	0
2	B	131/140 (93%)	0.54	14 (10%)	6	5	15, 22, 37, 47	0
All	All	264/305 (86%)	0.26	19 (7%)	15	14	12, 20, 34, 47	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1546	LYS	8.9
2	B	1582	GLY	6.9
2	B	1604	THR	6.4
2	B	1547	GLU	5.5
2	B	1581	ASN	5.0

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, 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
4	CA	B	402	1/1	0.97	0.07	27,27,27,27	0
4	CA	B	401	1/1	0.99	0.06	20,20,20,20	0
3	CL	A	403	1/1	1.00	0.14	26,26,26,26	0

6.5   Other polymers ⓘ

There are no such residues in this entry.