Full wwPDB X-ray Structure Validation Report (i)

Aug 19, 2023 – 06:17 PM EDT

PDB ID	:	2G9G
Title	:	Crystal structure of His-tagged mouse PNGase C-terminal domain
Authors	:	Zhou, X.; Zhao, G.; Wang, L.; Li, G.; Lennarz, W.J.; Schindelin, H.
Deposited on	:	2006-03-06
Resolution	:	2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

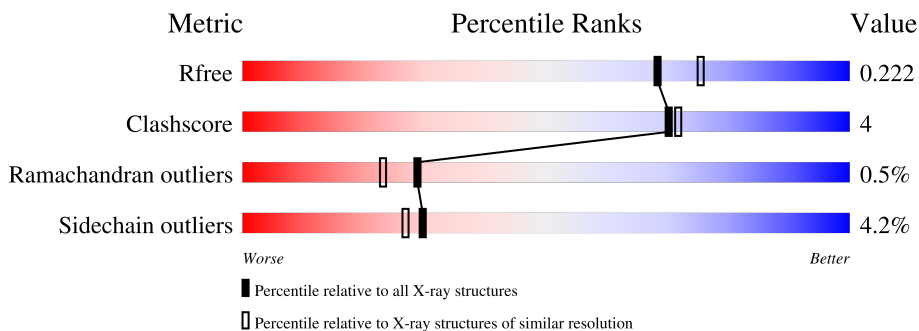|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.00 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		0.222
Clashscore		4
Ramachandran outliers		0.5%
Sidechain outliers		4.2%

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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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%

Mol	Chain	Length	Quality of chain
1	A	221	75% 10% 14%

2　Entry composition　ⓘ

There are 4 unique types of molecules in this entry. The entry contains 1685 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 peptide N-glycanase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	189	1543	969	274	297	3	0	2	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	431	MET	-	cloning artifact	GB 30517852
A	432	GLY	-	cloning artifact	GB 30517852
A	433	SER	-	cloning artifact	GB 30517852
A	434	SER	-	cloning artifact	GB 30517852
A	435	HIS	-	expression tag	GB 30517852
A	436	HIS	-	expression tag	GB 30517852
A	437	HIS	-	expression tag	GB 30517852
A	438	HIS	-	expression tag	GB 30517852
A	439	HIS	-	expression tag	GB 30517852
A	440	HIS	-	expression tag	GB 30517852
A	441	SER	-	cloning artifact	GB 30517852
A	442	SER	-	cloning artifact	GB 30517852
A	443	GLY	-	cloning artifact	GB 30517852
A	444	LEU	-	cloning artifact	GB 30517852
A	445	VAL	-	cloning artifact	GB 30517852
A	446	PRO	-	cloning artifact	GB 30517852
A	447	ARG	-	cloning artifact	GB 30517852
A	448	GLY	-	cloning artifact	GB 30517852
A	449	SER	-	cloning artifact	GB 30517852
A	450	HIS	-	cloning artifact	GB 30517852

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O3 O⁻ — S — O⁻ O4

S

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0
2	A	1	Total 5	O 4	S 1	0	0

• Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

C3

OH O3

C1   C2

HO   OH

O1   O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0

Continued on next page...

W O R L D W I D E

PDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	115	Total 115	O 115	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. 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. 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: peptide N-glycanase

Chain A:
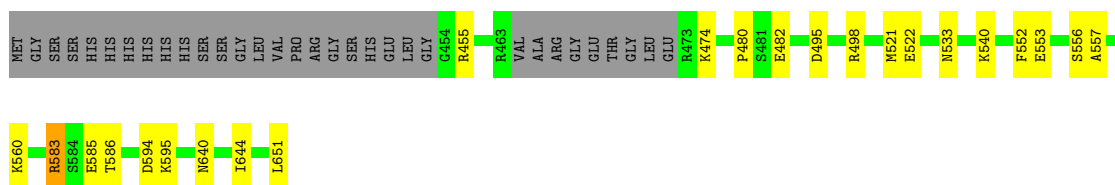
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	71.73Å 40.75Å 66.51Å 90.00° 95.34° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 35.71 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.7 (20.00-2.00) 99.4 (35.71-2.00)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	3.64 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.150 , 0.207 0.163 , 0.222	Depositor DCC
R_{free} test set	648 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	21.1	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 49.1	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.49, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1685	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.43% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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: GOL, SO4

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	0.71	1/1582 (0.1%)	0.77	0/2130

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	482	GLU	CD-OE2	5.88	1.32	1.25

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	1543	0	1496	10	0
2	A	15	0	0	0	0
3	A	12	0	16	2	0
4	A	115	0	0	2	0
All	All	1685	0	1512	11	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 4.

All (11) 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:585:GLU:HG2	1:A:586:THR:HG23	1.67	0.74
1:A:521:MET:SD	4:A:103:HOH:O	2.50	0.68
1:A:552:PHE:CZ	1:A:644:ILE:HD13	2.40	0.55
1:A:560:LYS:HB3	1:A:651:LEU:HD13	1.90	0.53
1:A:556:SER:OG	1:A:556:SER:O	2.28	0.51
1:A:552:PHE:HZ	1:A:644:ILE:HD13	1.78	0.47
1:A:495:ASP:OD2	1:A:498:ARG:NE	2.50	0.44
1:A:583:ARG:HH22	3:A:121:GOL:H12	1.84	0.43
1:A:480:PRO:HD3	1:A:553:GLU:O	2.19	0.43
3:A:122:GOL:H12	4:A:112:HOH:O	2.18	0.42
1:A:594:ASP:O	1:A:595:LYS:HB2	2.21	0.40

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	187/221 (85%)	178 (95%)	8 (4%)	1 (0%)	29	23

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	557	ALA

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	169/192 (88%)	162 (96%)	7 (4%)	30 28

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	455	ARG
1	A	474	LYS
1	A	522	GLU
1	A	533	ASN
1	A	540	LYS
1	A	583	ARG
1	A	640	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	601	ASN
1	A	640	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 ⓘ

5 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	SO4	A	123	-	4,4,4	0.11	0	6,6,6	0.26	0
3	GOL	A	122	-	5,5,5	0.59	0	5,5,5	1.10	0
2	SO4	A	125	-	4,4,4	0.25	0	6,6,6	0.39	0
2	SO4	A	124	-	4,4,4	0.30	0	6,6,6	0.60	0
3	GOL	A	121	-	5,5,5	0.56	0	5,5,5	0.46	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
3	GOL	A	122	-	-	2/4/4/4	-
3	GOL	A	121	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	121	GOL	O2-C2-C3-O3
3	A	121	GOL	O1-C1-C2-C3
3	A	121	GOL	C1-C2-C3-O3
3	A	122	GOL	C1-C2-C3-O3
3	A	121	GOL	O1-C1-C2-O2
3	A	122	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	122	GOL	1	0
3	A	121	GOL	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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.