



wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 16, 2021 – 06:08 PM JST

PDB ID : 7VI6
Title : Crystal structure of GH3 beta-N-acetylhexosaminidase Amuc_2109 from Akkermansia muciniphila
Authors : Qian, K.; Yang, W.; Chen, X.; Wang, Y.; Zhang, M.; Wang, M.
Deposited on : 2021-09-26
Resolution : 2.00 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.24
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)	:	2.24

i

X-RAY DIFFRACTION

A.

the following graphic. The table shows the number of entries on which the scores are based.

 R_{free}

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.

1

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5740 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 Beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	342	Total	C	N	O	S	0	3	0
			2620	1654	469	480	17			
1	B	340	Total	C	N	O	S	0	2	0
			2605	1646	464	478	17			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	354	LEU	-	expression tag	UNP B2UPP0
A	355	GLU	-	expression tag	UNP B2UPP0
A	356	HIS	-	expression tag	UNP B2UPP0
A	357	HIS	-	expression tag	UNP B2UPP0
A	358	HIS	-	expression tag	UNP B2UPP0
A	359	HIS	-	expression tag	UNP B2UPP0
A	360	HIS	-	expression tag	UNP B2UPP0
A	361	HIS	-	expression tag	UNP B2UPP0
B	354	LEU	-	expression tag	UNP B2UPP0
B	355	GLU	-	expression tag	UNP B2UPP0
B	356	HIS	-	expression tag	UNP B2UPP0
B	357	HIS	-	expression tag	UNP B2UPP0
B	358	HIS	-	expression tag	UNP B2UPP0
B	359	HIS	-	expression tag	UNP B2UPP0
B	360	HIS	-	expression tag	UNP B2UPP0
B	361	HIS	-	expression tag	UNP B2UPP0

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Mg	0	0
			2	2		
2	B	3	Total	Mg	0	0
			3	3		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	5	Total 5	Cl 5	0	0
3	B	5	Total 5	Cl 5	0	0

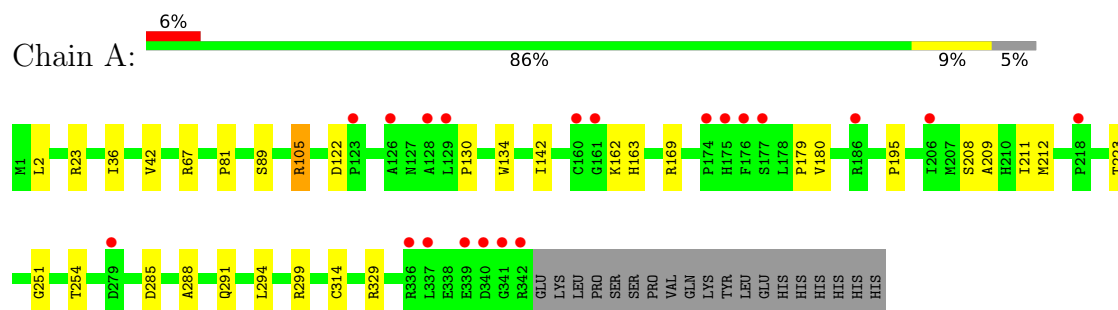
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	255	Total 255	O 255	0	0
4	B	245	Total 245	O 245	0	0

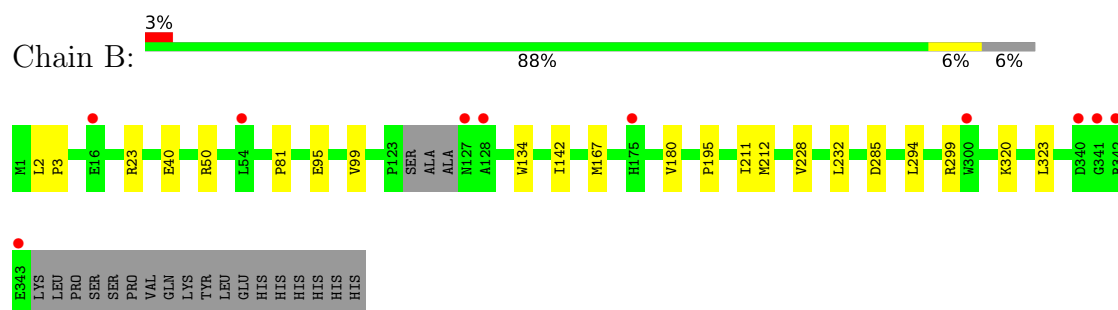
3 Residue-property plots [i](#)

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: Beta-N-acetylhexosaminidase



• Molecule 1: Beta-N-acetylhexosaminidase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	78.93Å 81.11Å 126.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.46 – 2.00 39.46 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.7 (39.46-2.00) 99.8 (39.46-2.00)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.11 (at 2.00Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.183 , 0.226 0.183 , 0.225	Depositor DCC
R_{free} test set	2778 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	36.5	Xtriage
Anisotropy	0.543	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 46.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.035 for k,h,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5740	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 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	0.29	0/2684	0.48	0/3643
1	B	0.30	0/2665	0.48	0/3616
All	All	0.29	0/5349	0.48	0/7259

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	2620	0	2638	18	0
1	B	2605	0	2617	12	0
2	A	2	0	0	0	0
2	B	3	0	0	0	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0
4	A	255	0	0	3	0
4	B	245	0	0	2	1
All	All	5740	0	5255	30	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 3.

The worst 5 of 30 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:291:GLN:NE2	4:A:503:HOH:O	2.25	0.69
1:A:105:ARG:HD2	1:A:314[B]:CYS:SG	2.33	0.68
1:B:320:LYS:HD2	1:B:323:LEU:HB2	1.79	0.64
1:B:294:LEU:HB2	1:B:299:ARG:HD2	1.87	0.57
1:B:23:ARG:NH1	1:B:285:ASP:OD1	2.31	0.56

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)
4:B:697:HOH:O	4:B:725:HOH:O[4_566]	2.15	0.05

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

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	343/361 (95%)	334 (97%)	9 (3%)	0	100	100
1	B	338/361 (94%)	328 (97%)	10 (3%)	0	100	100
All	All	681/722 (94%)	662 (97%)	19 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

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	282/298 (95%)	280 (99%)	2 (1%)	84	88
1	B	281/298 (94%)	281 (100%)	0	100	100
All	All	563/596 (94%)	561 (100%)	2 (0%)	91	93

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	105	ARG
1	A	169	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 15 ligands modelled in this entry, 15 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 [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	342/361 (94%)	0.23	20 (5%)	23 22	28, 40, 64, 86	0
1	B	340/361 (94%)	0.10	10 (2%)	51 50	29, 40, 57, 83	0
All	All	682/722 (94%)	0.16	30 (4%)	34 33	28, 40, 61, 86	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	343	GLU	5.9
1	B	128	ALA	5.4
1	A	175	HIS	5.1
1	B	340	ASP	4.6
1	A	341	GLY	4.3

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 monosaccharides 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. 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(\AA^2)	Q<0.9
3	CL	B	407	1/1	0.91	0.10	64,64,64,64	0
2	MG	B	403	1/1	0.92	0.27	56,56,56,56	0
2	MG	B	401	1/1	0.92	0.16	46,46,46,46	0
2	MG	A	401	1/1	0.96	0.14	51,51,51,51	0
2	MG	B	402	1/1	0.96	0.19	53,53,53,53	0
3	CL	A	406	1/1	0.97	0.08	45,45,45,45	0
2	MG	A	402	1/1	0.97	0.16	53,53,53,53	0
3	CL	A	403	1/1	0.98	0.16	55,55,55,55	0
3	CL	B	405	1/1	0.98	0.06	46,46,46,46	0
3	CL	B	406	1/1	0.98	0.19	68,68,68,68	0
3	CL	A	404	1/1	0.98	0.07	43,43,43,43	0
3	CL	B	408	1/1	0.98	0.09	47,47,47,47	0
3	CL	A	407	1/1	0.99	0.05	42,42,42,42	0
3	CL	B	404	1/1	0.99	0.17	55,55,55,55	0
3	CL	A	405	1/1	0.99	0.06	49,49,49,49	0

6.5 Other polymers [i](#)

There are no such residues in this entry.