



Full wwPDB X-ray Structure Validation Report ⓘ

Jun 20, 2024 – 07:39 pm BST

PDB ID : 8QK1
Title : Crystal structure of Trichuris suis beta-N-acetyl-D-hexosaminidase - HEX-2
in apo form
Authors : Dutkiewicz, Z.; Varrot, A.
Deposited on : 2023-09-14
Resolution : 2.55 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.37.1
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) : 2.37.1

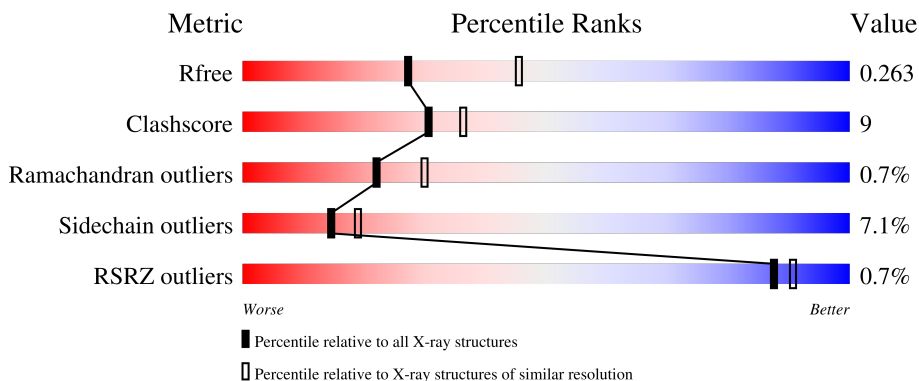
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 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}	130704	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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 $\leq 5\%$. 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	545	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3734 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
			Total	C	N	O	S			
1	A	456	3688	2385	623	661	19	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	537	ALA	-	expression tag	UNP A0A085NCI8
A	538	VAL	-	expression tag	UNP A0A085NCI8
A	539	ASP	-	expression tag	UNP A0A085NCI8
A	540	HIS	-	expression tag	UNP A0A085NCI8
A	541	HIS	-	expression tag	UNP A0A085NCI8
A	542	HIS	-	expression tag	UNP A0A085NCI8
A	543	HIS	-	expression tag	UNP A0A085NCI8
A	544	HIS	-	expression tag	UNP A0A085NCI8
A	545	HIS	-	expression tag	UNP A0A085NCI8

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	44	Total	O	0	1
			45	45		

4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	96.12Å 59.04Å 105.98Å 90.00° 114.49° 90.00°	Depositor
Resolution (Å)	43.74 – 2.55 43.74 – 2.55	Depositor EDS
% Data completeness (in resolution range)	96.7 (43.74-2.55) 96.8 (43.74-2.55)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.63 (at 2.54Å)	Xtrriage
Refinement program	REFMAC 5.8.0419	Depositor
R, R_{free}	0.189 , 0.259 0.196 , 0.263	Depositor DCC
R_{free} test set	836 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å ²)	44.1	Xtrriage
Anisotropy	0.096	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 39.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.025 for h,-k,-h-l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3734	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.16% 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:
ZN

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	0.64	2/3788 (0.1%)	1.02	0/5145

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	101	GLU	CD-OE1	5.37	1.31	1.25
1	A	259	GLU	CD-OE2	5.20	1.31	1.25

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	3688	0	3581	67	0
2	A	1	0	0	0	0
3	A	45	0	0	0	0
All	All	3734	0	3581	67	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 9.

All (67) 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:381:ASP:OD1	1:A:383:SER:HB3	1.66	0.95
1:A:64:ASP:OD1	1:A:66:LYS:HG3	1.82	0.80
1:A:198:ASP:HB3	1:A:246:TRP:CD1	2.19	0.78
1:A:389:ILE:HG23	1:A:394:CYS:HB2	1.70	0.71
1:A:287:TRP:HB3	1:A:336:LYS:HD3	1.80	0.64
1:A:198:ASP:HB3	1:A:246:TRP:CG	2.34	0.62
1:A:67:GLY:O	1:A:349:GLN:HG2	1.99	0.61
1:A:361:LEU:HB3	1:A:362:PRO:HD3	1.84	0.58
1:A:409:PHE:H	1:A:412:SER:HB3	1.67	0.58
1:A:232:SER:HA	1:A:243:VAL:HG21	1.87	0.56
1:A:414:ILE:HG13	1:A:480:ALA:HB1	1.89	0.55
1:A:372:MET:O	1:A:376:ARG:HG3	2.07	0.54
1:A:323:HIS:CD2	1:A:371:ASN:HB3	2.44	0.53
1:A:247:HIS:CG	1:A:248:ASP:N	2.77	0.52
1:A:291:SER:HB2	1:A:336:LYS:O	2.10	0.52
1:A:271:LEU:HB2	1:A:297:MET:CE	2.40	0.52
1:A:305:GLY:HA3	1:A:348:TRP:HB2	1.91	0.51
1:A:100:TYR:N	1:A:108:ARG:HD3	2.26	0.51
1:A:258:ALA:HA	1:A:263:LEU:HD12	1.93	0.50
1:A:98:PHE:CG	1:A:99:PRO:HD2	2.47	0.50
1:A:414:ILE:HD11	1:A:481:MET:HA	1.93	0.50
1:A:396:THR:HG22	1:A:397:GLY:H	1.77	0.50
1:A:198:ASP:CB	1:A:246:TRP:CD1	2.93	0.49
1:A:346:THR:CG2	1:A:348:TRP:CE2	2.95	0.49
1:A:304:LYS:HB2	1:A:359:GLU:HG3	1.94	0.49
1:A:167:SER:OG	1:A:223:LEU:HD23	2.14	0.48
1:A:227:HIS:O	1:A:231:VAL:HG23	2.14	0.48
1:A:443:ASN:OD1	1:A:443:ASN:N	2.45	0.48
1:A:91:LEU:HD12	1:A:133:ILE:O	2.14	0.47
1:A:366:PRO:O	1:A:370:VAL:HG23	2.13	0.47
1:A:109:ASN:O	1:A:112:ALA:HB2	2.15	0.47
1:A:303:PHE:CD2	1:A:368:LEU:HB2	2.51	0.46
1:A:346:THR:HG21	1:A:348:TRP:CE2	2.50	0.46
1:A:168:ASN:OD1	1:A:168:ASN:C	2.54	0.45
1:A:409:PHE:HB2	1:A:410:PRO:CD	2.46	0.45
1:A:144:TRP:HA	1:A:147:LYS:HE2	1.99	0.45
1:A:229:VAL:O	1:A:230:ASN:C	2.54	0.45
1:A:198:ASP:HB3	1:A:246:TRP:CB	2.47	0.45
1:A:336:LYS:HB3	1:A:336:LYS:HE3	1.49	0.45
1:A:139:ILE:HG12	1:A:227:HIS:CD2	2.52	0.44
1:A:529:LYS:HD3	1:A:529:LYS:HA	1.75	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:403:ASP:HB3	1:A:405:TYR:CE2	2.53	0.44
1:A:358:CYS:HB2	1:A:359:GLU:H	1.67	0.43
1:A:346:THR:HG21	1:A:348:TRP:NE1	2.33	0.43
1:A:409:PHE:O	1:A:412:SER:HB3	2.18	0.43
1:A:397:GLY:O	1:A:398:SER:O	2.37	0.43
1:A:346:THR:HG21	1:A:348:TRP:CZ2	2.54	0.43
1:A:460:VAL:N	1:A:461:PRO:CD	2.82	0.43
1:A:359:GLU:H	1:A:359:GLU:HG2	1.09	0.42
1:A:256:SER:OG	1:A:257:LEU:N	2.52	0.42
1:A:528:ILE:O	1:A:529:LYS:HB2	2.20	0.42
1:A:273:ASN:HB3	1:A:278:LEU:HD22	2.01	0.42
1:A:367:SER:O	1:A:371:ASN:ND2	2.52	0.42
1:A:74:HIS:NE2	1:A:483:GLU:O	2.53	0.41
1:A:396:THR:HG22	1:A:397:GLY:N	2.34	0.41
1:A:409:PHE:HB2	1:A:410:PRO:HD2	2.02	0.41
1:A:346:THR:CG2	1:A:348:TRP:NE1	2.83	0.41
1:A:385:ASN:O	1:A:388:ALA:HB3	2.20	0.41
1:A:481:MET:HB3	1:A:490:VAL:HG22	2.03	0.41
1:A:499:ASP:N	1:A:500:PRO:CD	2.84	0.41
1:A:62:HIS:O	1:A:63:LEU:HD12	2.21	0.41
1:A:303:PHE:CD1	1:A:368:LEU:HD13	2.55	0.41
1:A:100:TYR:H	1:A:108:ARG:HD3	1.85	0.41
1:A:384:VAL:O	1:A:385:ASN:C	2.59	0.40
1:A:381:ASP:OD1	1:A:383:SER:N	2.54	0.40
1:A:381:ASP:OD1	1:A:383:SER:CB	2.53	0.40
1:A:414:ILE:CD1	1:A:481:MET:HA	2.51	0.40

There are no symmetry-related clashes.

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	452/545 (83%)	422 (93%)	27 (6%)	3 (1%)	22 30

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	396	THR
1	A	398	SER
1	A	442	SER

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	394/487 (81%)	366 (93%)	28 (7%)	14 19

All (28) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	105	SER
1	A	127	GLN
1	A	157	ASN
1	A	159	ARG
1	A	191	GLN
1	A	199	GLU
1	A	236	LYS
1	A	250	ILE
1	A	256	SER
1	A	260	LYS
1	A	267	VAL
1	A	284	MET
1	A	332	THR
1	A	336	LYS
1	A	341	ARG
1	A	352	ASP
1	A	358	CYS
1	A	359	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	379	ARG
1	A	408	ARG
1	A	420	LEU
1	A	442	SER
1	A	444	ILE
1	A	450	MET
1	A	458	LEU
1	A	511	ARG
1	A	523	VAL
1	A	529	LYS

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	127	GLN
1	A	227	HIS

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 1 ligands modelled in this entry, 1 is 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 [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	456/545 (83%)	-0.31	3 (0%) 87 90	21, 41, 73, 106	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	397	GLY	3.3
1	A	253	ILE	2.9
1	A	250	ILE	2.8

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.

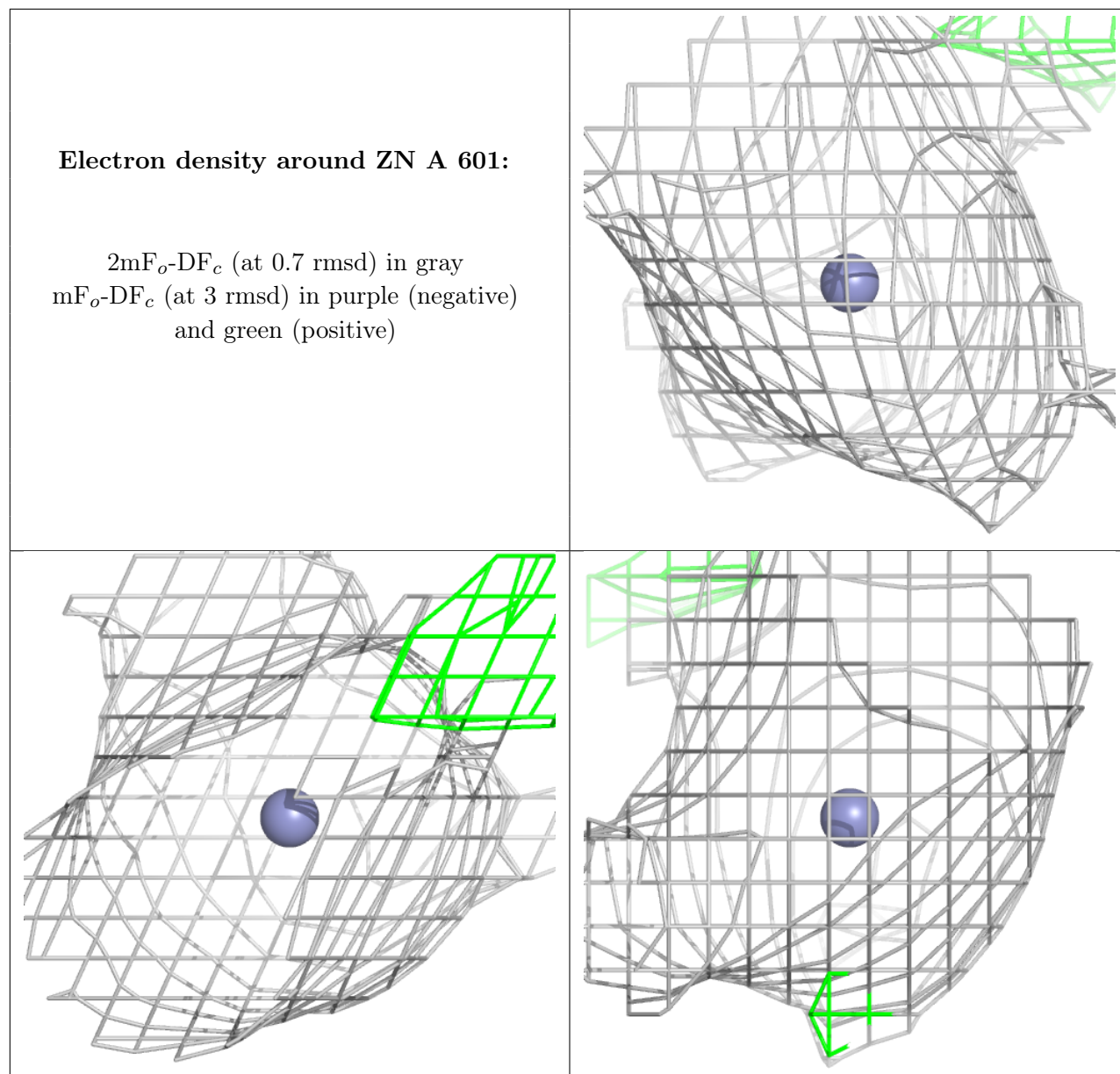
6.4 Ligands [i](#)

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(Å ²)	Q<0.9
2	ZN	A	601	1/1	0.97	0.04	81,81,81,81	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.



6.5 Other polymers [i](#)

There are no such residues in this entry.