



wwPDB X-ray Structure Validation Summary Report

Feb 15, 2017 – 04:45 am GMT

PDB ID : 3ECQ
Title : Endo-alpha-N-acetylgalactosaminidase from Streptococcus pneumoniae:
SeMet structure
Authors : Caines, M.E.C.; Zhu, H.; Vuckovic, M.; Strynadka, N.C.J.
Deposited on : 2008-09-01
Resolution : 2.90 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

<http://wwpdb.org/validation/2016/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
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : recalc28949

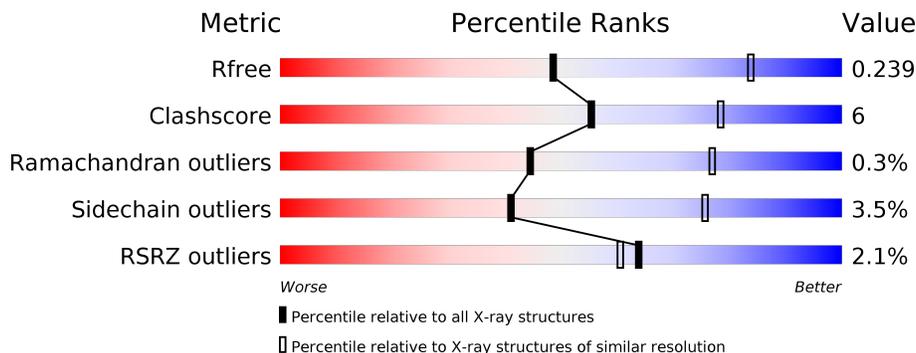
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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}	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	1531	 3% 74% 13% 13%
1	B	1531	 3% 74% 14% • 12%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	GOL	A	2003	-	-	-	X

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 21372 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 Endo-alpha-N-acetylgalactosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	1336	10586	6644	1831	2088	23	0	1	0
1	B	1347	10671	6695	1844	2108	24	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	37	GLY	-	EXPRESSION TAG	UNP Q8DR60
A	38	SER	-	EXPRESSION TAG	UNP Q8DR60
A	39	MSE	-	EXPRESSION TAG	UNP Q8DR60
B	37	GLY	-	EXPRESSION TAG	UNP Q8DR60
B	38	SER	-	EXPRESSION TAG	UNP Q8DR60
B	39	MSE	-	EXPRESSION TAG	UNP Q8DR60

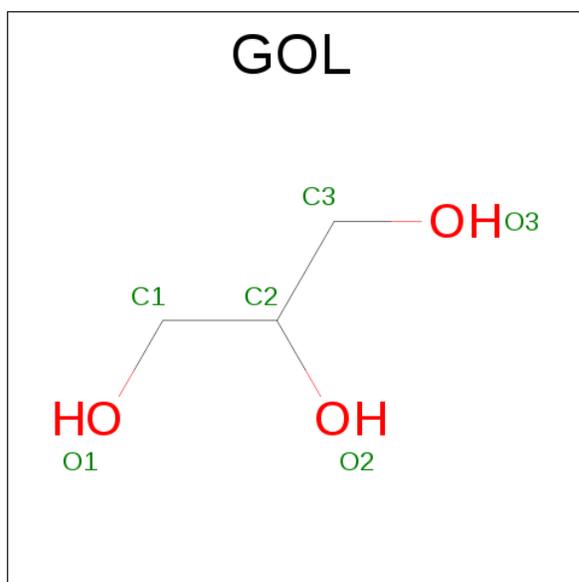
- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total 1	Na 1	0	0
3	A	1	Total 1	Na 1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	53	Total O 53 53	0	0
5	B	44	Total O 44 44	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	210.61Å 158.21Å 112.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	112.51 – 2.90 112.44 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.4 (112.51-2.90) 99.4 (112.44-2.90)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.30 (at 2.91Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.190 , 0.235 0.198 , 0.239	Depositor DCC
R_{free} test set	4164 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	56.7	Xtrriage
Anisotropy	0.136	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 43.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	21372	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.80% 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: GOL, CA, NA

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	0.52	0/10800	0.64	4/14588 (0.0%)
1	B	0.50	0/10880	0.62	1/14691 (0.0%)
All	All	0.51	0/21680	0.63	5/29279 (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	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	607	LYS	CB-CA-C	-5.33	99.74	110.40
1	B	1307	ARG	NE-CZ-NH1	5.25	122.92	120.30
1	A	870	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	A	1307	ARG	NE-CZ-NH1	5.08	122.84	120.30
1	A	613	ARG	NE-CZ-NH1	5.07	122.83	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1377	GLY	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	1377	GLY	Peptide

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	10586	0	10172	123	0
1	B	10671	0	10253	124	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	6	0	8	0	0
4	B	6	0	8	0	0
5	A	53	0	0	0	0
5	B	44	0	0	0	0
All	All	21372	0	20441	247	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 247 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:472:MSE:HE2	1:A:498:MSE:HE1	1.31	1.09
1:B:472:MSE:HE2	1:B:498:MSE:HE1	1.10	1.07
1:A:1419:THR:O	1:A:1424:MSE:HE3	1.73	0.87
1:B:219:TRP:CZ2	1:B:269:VAL:HG21	2.14	0.81
1:B:795:ILE:HD11	1:B:808:HIS:CD2	2.19	0.78

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	1331/1531 (87%)	1272 (96%)	55 (4%)	4 (0%)	44	77
1	B	1341/1531 (88%)	1276 (95%)	60 (4%)	5 (0%)	38	72
All	All	2672/3062 (87%)	2548 (95%)	115 (4%)	9 (0%)	44	77

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1254	ASP
1	B	1254	ASP
1	B	662	LEU
1	B	816	TYR
1	A	662	LEU

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	1132/1257 (90%)	1094 (97%)	38 (3%)	42	76
1	B	1140/1257 (91%)	1098 (96%)	42 (4%)	39	74
All	All	2272/2514 (90%)	2192 (96%)	80 (4%)	41	75

5 of 80 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1342	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	288	LYS
1	B	1324	SER
1	A	1363	LYS
1	B	194	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	657	HIS
1	B	694	HIS
1	B	1300	GLN
1	B	413	GLN
1	B	1428	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 6 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
4	GOL	A	2003	-	5,5,5	0.42	0	5,5,5	0.56	0
4	GOL	B	2003	-	5,5,5	0.42	0	5,5,5	0.32	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
4	GOL	A	2003	-	-	0/4/4/4	0/0/0/0
4	GOL	B	2003	-	-	0/4/4/4	0/0/0/0

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	1313/1531 (85%)	0.00	4 (0%) 93 93	29, 49, 83, 128	0
1	B	1323/1531 (86%)	0.21	52 (3%) 40 35	32, 55, 103, 146	0
All	All	2636/3062 (86%)	0.11	56 (2%) 64 60	29, 52, 96, 146	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	281	ASN	5.4
1	B	271	LEU	4.9
1	B	202	LYS	4.1
1	B	275	VAL	3.9
1	B	282	GLU	3.8

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
4	GOL	A	2003	6/6	0.73	0.30	2.23	81,82,82,82	0
2	CA	A	2000	1/1	0.84	0.15	-0.52	63,63,63,63	0
4	GOL	B	2003	6/6	0.79	0.21	-0.68	60,64,65,66	0
2	CA	B	2000	1/1	0.98	0.13	-0.75	54,54,54,54	0
2	CA	B	2001	1/1	0.99	0.14	-1.52	36,36,36,36	0
3	NA	B	2002	1/1	0.79	0.15	-1.61	41,41,41,41	0
3	NA	A	2002	1/1	0.95	0.14	-2.01	20,20,20,20	0
2	CA	A	2001	1/1	0.99	0.14	-2.08	42,42,42,42	0

6.5 Other polymers [i](#)

There are no such residues in this entry.