Full wwPDB X-ray Structure Validation Report (i)

Feb 21, 2024 – 02:21 AM EST

PDB ID	:	4NUZ
Title	:	Crystal structure of a glycosynthase mutant (D233Q) of EndoS, an endo-beta -N-acetyl-glucosaminidase from Streptococcus pyogenes
Authors	:	Trastoy, B.; Guenther, S.; Snyder, G.A.; Sundberg, E.J.
Deposited on	:	2013-12-04
Resolution	:	1.91 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 (1)) were used in the production of this report:

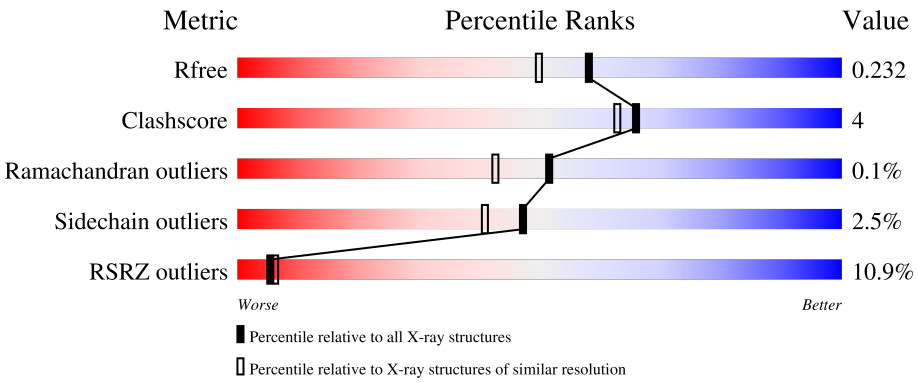
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.91 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 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	899	11% 88% 9% ••

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7896 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-beta-N-acetylglucosaminidase F2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	888	7047	4457	1189	1384	17	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	233	GLN	ASP	engineered mutation	UNP Q48WW7
A	996	LEU	-	expression tag	UNP Q48WW7

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

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

- Molecule 3 is water.

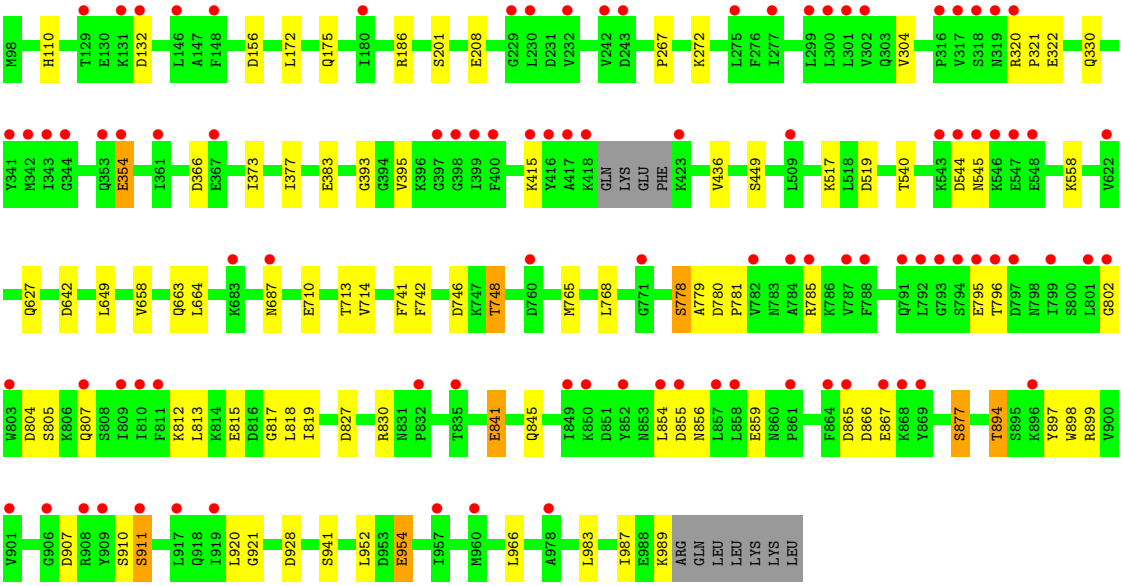
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	848	848	848	0	0

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: Endo-beta-N-acetylglucosaminidase F2

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	92.77Å 96.39Å 141.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.68 – 1.91 29.68 – 1.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (29.68-1.91) 98.8 (29.68-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.47 (at 1.91Å)	Xtriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, R_{free}	0.192 , 0.235 0.191 , 0.232	Depositor DCC
R_{free} test set	1987 reflections (2.01%)	wwPDB-VP
Wilson B-factor (Å^2)	31.4	Xtriage
Anisotropy	0.402	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 50.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	0.038 for k,h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7896	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.28% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.75	0/7189	0.74	0/9718

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 (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	7047	0	6993	56	0
2	A	1	0	0	0	0
3	A	848	0	0	14	0
All	All	7896	0	6993	56	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 (56) 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:395:VAL:HG11	1:A:449:SER:HB3	1.75	0.69

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:517:LYS:HE2	1:A:519:ASP:OD2	2.01	0.60
1:A:687:ASN:HB3	3:A:1419:HOH:O	2.01	0.59
1:A:894:THR:HG21	1:A:928:ASP:HA	1.83	0.59
1:A:765:MET:HE2	1:A:920:LEU:C	2.23	0.58
1:A:845:GLN:OE1	1:A:899:ARG:NH1	2.37	0.56
1:A:907:ASP:O	1:A:910:SER:OG	2.23	0.56
1:A:856:ASN:ND2	1:A:859:GLU:OE1	2.34	0.54
1:A:540:THR:OG1	1:A:545:ASN:HA	2.08	0.53
1:A:841:GLU:OE2	1:A:877:SER:HB3	2.07	0.53
1:A:866:ASP:OD1	1:A:867:GLU:N	2.42	0.52
1:A:320:ARG:HG3	1:A:321:PRO:HD2	1.92	0.52
1:A:322:GLU:HG2	3:A:1915:HOH:O	2.08	0.51
1:A:132:ASP:OD1	1:A:132:ASP:N	2.42	0.50
1:A:778:SER:OG	1:A:865:ASP:OD2	2.25	0.50
1:A:172:LEU:O	1:A:175:GLN:HG2	2.11	0.49
1:A:393:GLY:HA2	3:A:1442:HOH:O	2.13	0.48
1:A:815:GLU:HG2	3:A:1819:HOH:O	2.14	0.48
1:A:804:ASP:CG	1:A:805:SER:H	2.17	0.48
1:A:746:ASP:OD1	1:A:748:THR:HG22	2.13	0.48
1:A:373:ILE:HD12	3:A:1849:HOH:O	2.14	0.47
1:A:713:THR:OG1	1:A:742:PHE:HB2	2.14	0.47
1:A:354:GLU:H	1:A:354:GLU:HG2	1.40	0.47
1:A:827:ASP:CG	1:A:830:ARG:HH21	2.19	0.47
1:A:954:GLU:HG2	3:A:1639:HOH:O	2.15	0.46
1:A:987:ILE:O	1:A:989:LYS:HA	2.15	0.46
1:A:765:MET:HE3	1:A:921:GLY:HA2	1.98	0.46
1:A:941:SER:HA	1:A:952:LEU:HD22	1.98	0.46
1:A:804:ASP:OD1	1:A:805:SER:N	2.45	0.45
1:A:854:LEU:HD11	1:A:897:TYR:CE1	2.52	0.45
1:A:658:VAL:CG1	1:A:663:GLN:HG3	2.47	0.44
1:A:983:LEU:O	1:A:987:ILE:HG13	2.16	0.44
1:A:415:LYS:N	3:A:1580:HOH:O	2.23	0.44
1:A:377:ILE:HG22	1:A:383:GLU:HG3	2.00	0.43
1:A:714:VAL:HG12	1:A:741:PHE:CD1	2.53	0.43
1:A:819:ILE:HD13	1:A:898:TRP:HZ2	1.83	0.43
1:A:267:PRO:O	1:A:272:LYS:HE3	2.19	0.43
1:A:812:LYS:HB2	1:A:897:TYR:CE2	2.54	0.43
1:A:802:GLY:O	1:A:911:SER:HB3	2.18	0.43
1:A:856:ASN:HA	1:A:859:GLU:OE1	2.19	0.43
1:A:208:GLU:HG2	3:A:1691:HOH:O	2.18	0.42
1:A:855:ASP:OD1	1:A:856:ASN:N	2.52	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:558:LYS:HE2	3:A:1601:HOH:O	2.18	0.42
1:A:807:GLN:NE2	3:A:1858:HOH:O	2.51	0.42
1:A:866:ASP:OD1	1:A:866:ASP:C	2.58	0.42
1:A:156:ASP:HB3	3:A:1588:HOH:O	2.19	0.42
1:A:110:HIS:HE1	3:A:1651:HOH:O	2.02	0.41
1:A:186:ARG:NH1	3:A:1307:HOH:O	2.45	0.41
1:A:818:LEU:HD22	1:A:966:LEU:HD22	2.02	0.41
1:A:330:GLN:HG3	3:A:1910:HOH:O	2.20	0.41
1:A:649:LEU:HB2	1:A:664:LEU:HD21	2.03	0.40
1:A:377:ILE:CD1	1:A:436:VAL:HG13	2.51	0.40
1:A:768:LEU:HD13	1:A:817:GLY:HA3	2.03	0.40
1:A:856:ASN:O	1:A:859:GLU:HG2	2.20	0.40
1:A:894:THR:O	1:A:894:THR:HG22	2.21	0.40
1:A:780:ASP:HA	1:A:781:PRO:HD3	1.76	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	884/899 (98%)	859 (97%)	24 (3%)	1 (0%)	51 42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	779	ALA

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	775/786 (99%)	756 (98%)	19 (2%)	47 41

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	201	SER
1	A	304	VAL
1	A	354	GLU
1	A	366	ASP
1	A	544	ASP
1	A	627	GLN
1	A	642	ASP
1	A	710	GLU
1	A	748	THR
1	A	778	SER
1	A	785	ARG
1	A	795	GLU
1	A	796	THR
1	A	813	LEU
1	A	841	GLU
1	A	877	SER
1	A	894	THR
1	A	911	SER
1	A	954	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	110	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 (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	888/899 (98%)	0.53	97 (10%) 5 6	15, 35, 69, 103	0

All (97) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	803	TRP	8.0
1	A	796	THR	7.3
1	A	544	ASP	7.1
1	A	909	TYR	6.8
1	A	864	PHE	5.7
1	A	546	LYS	5.3
1	A	242	VAL	4.7
1	A	317	VAL	4.7
1	A	545	ASN	4.6
1	A	850	LYS	4.4
1	A	301	LEU	4.2
1	A	417	ALA	4.1
1	A	799	ILE	4.1
1	A	908	ARG	4.1
1	A	316	PRO	3.9
1	A	687	ASN	3.8
1	A	855	ASP	3.7
1	A	852	TYR	3.7
1	A	793	GLY	3.6
1	A	795	GLU	3.6
1	A	785	ARG	3.6
1	A	415	LYS	3.5
1	A	277	ILE	3.5
1	A	320	ARG	3.5
1	A	849	ILE	3.4
1	A	543	LYS	3.4
1	A	180	ILE	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	367	GLU	3.3
1	A	857	LEU	3.3
1	A	794	SER	3.3
1	A	801	LEU	3.3
1	A	299	LEU	3.3
1	A	129	THR	3.2
1	A	423	LYS	3.1
1	A	400	PHE	3.1
1	A	835	THR	3.0
1	A	867	GLU	2.9
1	A	342	MET	2.9
1	A	341	TYR	2.9
1	A	547	GLU	2.9
1	A	354	GLU	2.8
1	A	243	ASP	2.8
1	A	906	GLY	2.8
1	A	416	TYR	2.8
1	A	399	ILE	2.7
1	A	319	ASN	2.7
1	A	343	ILE	2.7
1	A	418	LYS	2.7
1	A	869	TYR	2.7
1	A	978	ALA	2.7
1	A	854	LEU	2.6
1	A	810	ILE	2.6
1	A	782	VAL	2.6
1	A	807	GLN	2.6
1	A	901	VAL	2.6
1	A	760	ASP	2.6
1	A	809	ILE	2.6
1	A	797	ASP	2.6
1	A	398	GLY	2.6
1	A	787	VAL	2.5
1	A	548	GLU	2.5
1	A	917	LEU	2.5
1	A	683	LYS	2.5
1	A	911	SER	2.5
1	A	318	SER	2.4
1	A	146	LEU	2.4
1	A	802	GLY	2.4
1	A	622	VAL	2.4
1	A	300	LEU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	960	MET	2.3
1	A	771	GLY	2.3
1	A	148	PHE	2.3
1	A	361	ILE	2.3
1	A	275	LEU	2.3
1	A	509	LEU	2.3
1	A	788	PHE	2.2
1	A	791	GLN	2.2
1	A	784	ALA	2.2
1	A	861	PRO	2.2
1	A	232	VAL	2.2
1	A	792	LEU	2.2
1	A	131	LYS	2.2
1	A	811	PHE	2.2
1	A	919	ILE	2.2
1	A	896	LYS	2.2
1	A	858	LEU	2.1
1	A	865	ASP	2.1
1	A	302	VAL	2.1
1	A	229	GLY	2.1
1	A	868	LYS	2.1
1	A	832	PRO	2.1
1	A	132	ASP	2.1
1	A	957	ILE	2.1
1	A	344	GLY	2.1
1	A	353	GLN	2.1
1	A	397	GLY	2.0
1	A	230	LEU	2.0

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	CA	A	1001	1/1	0.99	0.08	42,42,42,42	0

6.5 Other polymers (i)

There are no such residues in this entry.