Full wwPDB X-ray Structure Validation Report (i)

Feb 19, 2016 – 08:33 PM GMT

PDB ID	:	4YHG
Title	:	NATIVE BACTEROIDETES-AFFILIATED GH5 CELLULASE LINKED WITH A POLYSACCHARIDE UTILIZATION LOCUS
Authors	:	Naas, A.E.; MacKenzie, A.K.; Dalhus, B.; Eijsink, V.G.H.; Pope, P.B.
Deposited on	:	2015-02-27
Resolution	:	2.40 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

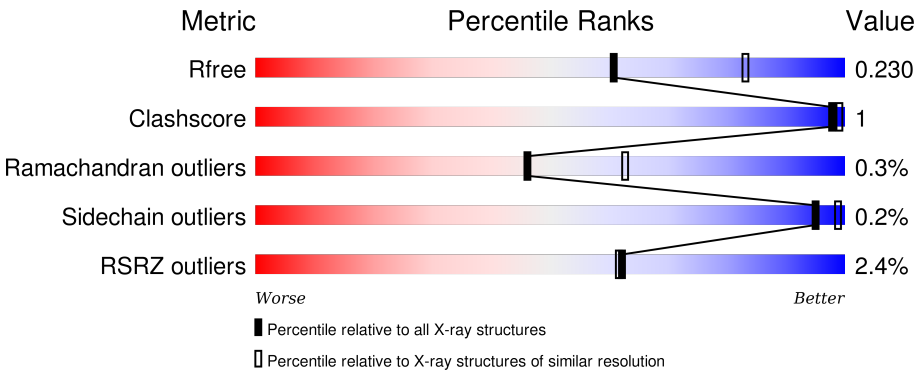
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 2.40 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}	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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 <=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	389	% 96% •
1	B	389	4% 91% • 7%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6207 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.

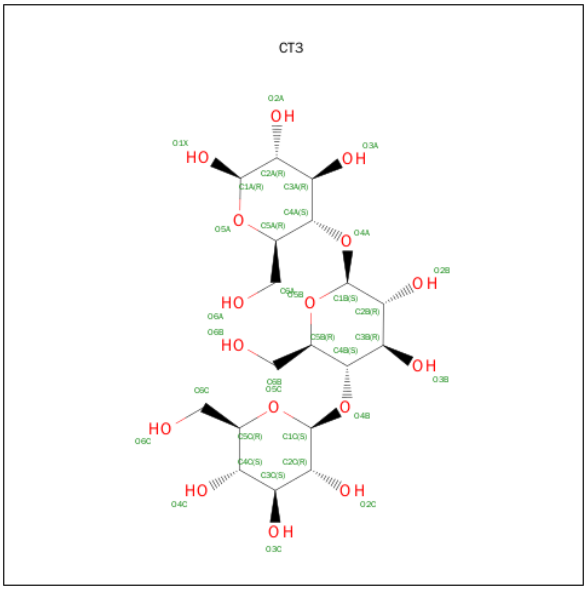
- Molecule 1 is a protein called GH5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	389	3073	1931	553	573	16	0	0	0
1	B	361	2868	1806	517	530	15	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	172	ALA	GLU	engineered mutation	UNP A0A076MPD7
B	172	ALA	GLU	engineered mutation	UNP A0A076MPD7

- Molecule 2 is beta-D-glucopyranosyl-(1->4)-beta-D-glucopyranosyl-(1->4)-beta-D-glucopyranose (three-letter code: CT3) (formula: $C_{18}H_{32}O_{16}$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	34	18	16	0	0

- Molecule 3 is water.

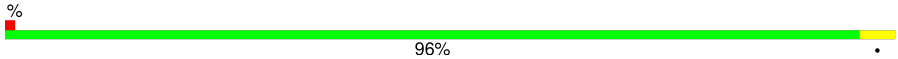
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	132	132	132	0	0
			Total	O		
3	B	100	100	100	0	0

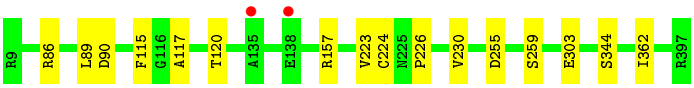
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: GH5

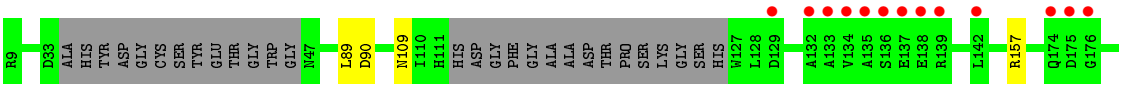
Chain A:

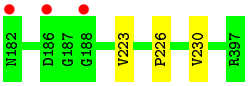




- Molecule 1: GH5

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	84.33Å 99.55Å 102.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	71.40 – 2.40 49.78 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.6 (71.40-2.40) 99.6 (49.78-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	2.38 (at 2.39Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.185 , 0.228 0.191 , 0.230	Depositor DCC
R_{free} test set	1733 reflections (5.33%)	DCC
Wilson B-factor (Å^2)	36.4	Xtriage
Anisotropy	0.096	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 24.6	EDS
Estimated twinning fraction	0.020 for -h,l,k	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 34276 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6207	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: CT3

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	0.54	0/3153	0.66	0/4278
1	B	0.55	0/2937	0.69	0/3981
All	All	0.54	0/6090	0.67	0/8259

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	3073	0	2904	9	0
1	B	2868	0	2739	4	0
2	A	34	0	32	1	0
3	A	132	0	0	1	0
3	B	100	0	0	0	0
All	All	6207	0	5675	13	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 1.

All (13) 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:90:ASP:OD1	1:A:157:ARG:NH2	2.33	0.62
1:B:90:ASP:OD1	1:B:157:ARG:NH2	2.43	0.51
1:A:226:PRO:O	1:A:230:VAL:HG13	2.11	0.50
1:A:115:PHE:CZ	1:A:117:ALA:HB3	2.49	0.48
1:A:89:LEU:HD23	1:A:157:ARG:NH1	2.32	0.45
1:B:89:LEU:HD23	1:B:157:ARG:NH1	2.32	0.45
1:A:224:CYS:HA	3:A:541:HOH:O	2.18	0.44
1:B:226:PRO:O	1:B:230:VAL:HG13	2.18	0.43
1:A:344:SER:HB3	1:A:362:ILE:HG21	2.01	0.43
1:A:303:GLU:OE1	2:A:401:CT3:H1A	2.20	0.41
1:B:109:ASN:OD1	1:B:109:ASN:C	2.58	0.41
1:A:255:ASP:HA	1:A:259:SER:OG	2.20	0.41
1:A:86:ARG:HB3	1:A:86:ARG:CZ	2.51	0.41

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	387/389 (100%)	376 (97%)	10 (3%)	1 (0%)	46 63
1	B	355/389 (91%)	345 (97%)	9 (2%)	1 (0%)	46 63
All	All	742/778 (95%)	721 (97%)	19 (3%)	2 (0%)	46 63

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	223	VAL
1	B	223	VAL

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	311/311 (100%)	310 (100%)	1 (0%)	94 98
1	B	292/311 (94%)	292 (100%)	0	100 100
All	All	603/622 (97%)	602 (100%)	1 (0%)	95 99

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	120	THR

Some 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 carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand is 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	CT3	A	401	-	36,36,36	0.92	0	53,53,53	1.14	5 (9%)

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
2	CT3	A	401	-	-	0/14/74/74	0/3/3/3

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401	CT3	O5A-C5A-C6A	2.00	111.56	106.38
2	A	401	CT3	O5B-C5B-C6B	2.14	111.93	106.38
2	A	401	CT3	O3B-C3B-C2B	2.49	115.97	110.36
2	A	401	CT3	O4A-C4A-C3A	2.84	114.58	107.18
2	A	401	CT3	O5C-C5C-C6C	2.84	113.73	106.38

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	CT3	1	0

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	389/389 (100%)	-0.35	2 (0%)	91	91	23, 38, 67, 96	0
1	B	361/389 (92%)	-0.15	16 (4%)	38	39	21, 40, 83, 106	0
All	All	750/778 (96%)	-0.25	18 (2%)	62	61	21, 39, 76, 106	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	134	VAL	4.4
1	B	135	ALA	4.0
1	B	175	ASP	4.0
1	B	142	LEU	3.6
1	B	133	ALA	3.5
1	B	139	ARG	3.4
1	A	138	GLU	3.3
1	B	138	GLU	3.0
1	B	176	GLY	2.9
1	B	174	GLN	2.9
1	B	132	ALA	2.7
1	B	182	ASN	2.6
1	B	137	GLU	2.5
1	B	186	ASP	2.4
1	B	136	SER	2.4
1	A	135	ALA	2.3
1	B	188	GLY	2.3
1	B	129	ASP	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 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(Å^2)	Q<0.9
2	CT3	A	401	34/34	0.85	0.17	1.13	38,55,65,69	0

6.5 Other polymers (i)

There are no such residues in this entry.