



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:47 PM GMT

PDB ID : 4G68
Title : Biochemical and structural insights into xylan utilization by the thermophilic bacterium *Caldanaerobius polysaccharolyticus*
Authors : Agarwal, V.; Nair, S.K.
Deposited on : 2012-07-18
Resolution : 1.80 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 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

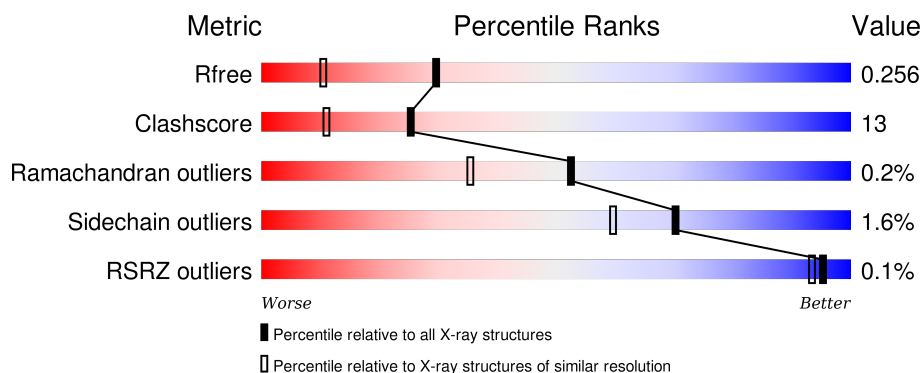
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION




The reported resolution of this entry is 1.80 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	4533 (1.80-1.80)
Clashscore	102246	5383 (1.80-1.80)
Ramachandran outliers	100387	5320 (1.80-1.80)
Sidechain outliers	100360	5319 (1.80-1.80)
RSRZ outliers	91569	4547 (1.80-1.80)

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	456	
2	B	432	
3	C	432	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10210 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 ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	392	Total	C	N	O	S	Se	0	2	0
			3058	1957	494	597	4	6			

- Molecule 2 is a protein called ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	392	Total	C	N	O	S	Se	0	0	0
			3050	1951	494	596	4	5			

- Molecule 3 is a protein called ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	392	Total	C	N	O	S	Se	0	0	0
			3050	1951	494	596	8	1			

- Molecule 4 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	3	Total	C	O	0	0
			28	15	13		
4	B	3	Total	C	O	0	0
			28	15	13		

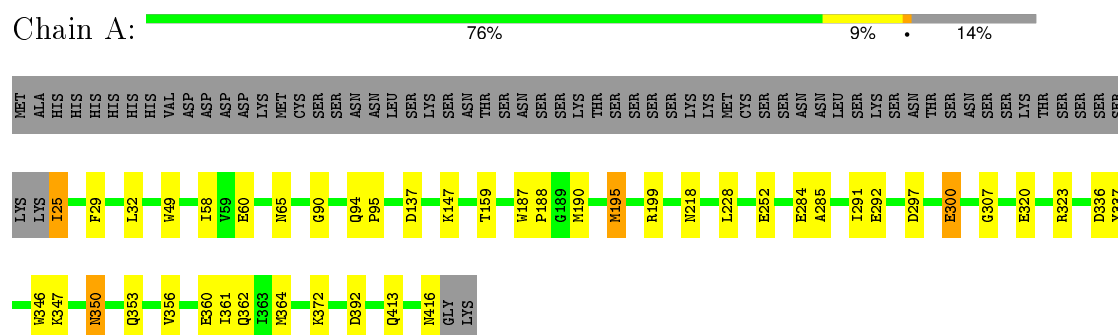
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	364	Total	O	0	0
			364	364		
5	B	363	Total	O	0	0
			363	363		
5	C	269	Total	O	0	0
			269	269		

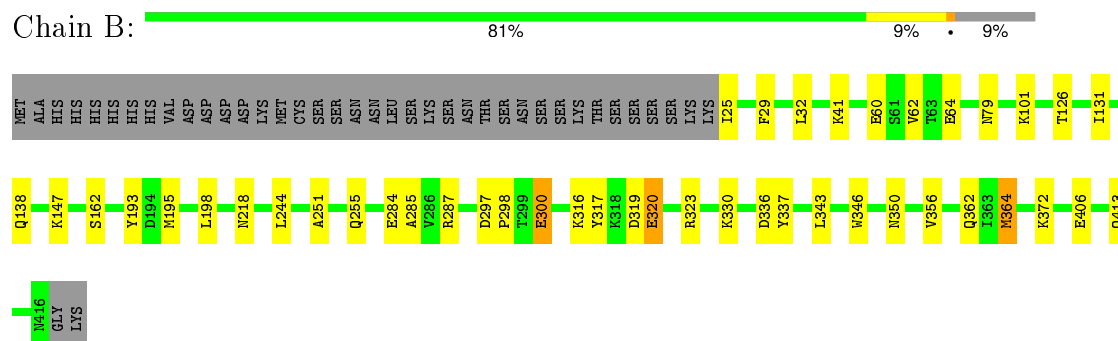
3 Residue-property plots

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.

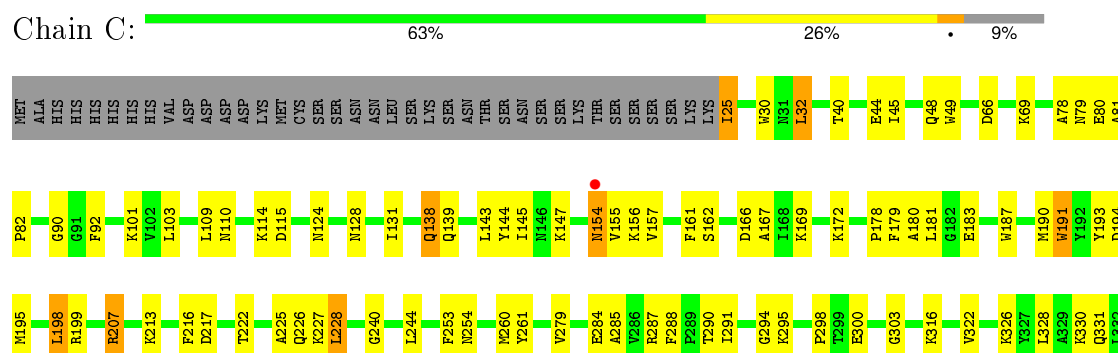
• Molecule 1: ABC transporter



• Molecule 2: ABC transporter



• Molecule 3: ABC transporter



S333	D334	L343	K346	K347	Y348	D349	N350	L351	D352	Q353	S354	K355	Y356	L359	Q362	I363	K364	N365	N370	K371	K372	P376	K377	K378	D379	I380	Y381	L382	S383	K384	D385	K386	K387	Q388	K391	K395	Q396	L397	F398	K399	K400	Q401	I402	S408	Q412	K416	GLY	LYS
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	59.33Å 150.86Å 150.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 1.80 38.36 – 1.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (25.00-1.80) 99.7 (38.36-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.42 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.5.0056	Depositor
R, R_{free}	0.230 , 0.261 0.215 , 0.256	Depositor DCC
R_{free} test set	6312 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	13.8	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 19.4	EDS
Estimated twinning fraction	0.478 for -h,l,k	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	2 of 125762 reflections (0.002%)	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	10210	wwPDB-VP
Average B, all atoms (Å ²)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.92% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: XYP, XYS

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	1.11	4/3125 (0.1%)	0.99	6/4224 (0.1%)
2	B	1.08	5/3112 (0.2%)	0.97	9/4209 (0.2%)
3	C	0.91	5/3116 (0.2%)	0.87	3/4221 (0.1%)
All	All	1.04	14/9353 (0.1%)	0.94	18/12654 (0.1%)

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	320	GLU	CD-OE1	6.94	1.33	1.25
2	B	406	GLU	CD-OE2	-6.76	1.18	1.25
2	B	346	TRP	CD2-CE2	6.33	1.49	1.41
2	B	320	GLU	CD-OE1	6.33	1.32	1.25
2	B	64	GLU	CD-OE1	6.33	1.32	1.25

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	195	MET	CG-SD-CE	-11.49	81.82	100.20
1	A	323	ARG	NE-CZ-NH1	8.62	124.61	120.30
1	A	323	ARG	NE-CZ-NH2	-7.25	116.67	120.30
3	C	207	ARG	NE-CZ-NH2	-6.72	116.94	120.30
1	A	195	MSE	CG-SE-CE	-6.64	84.30	98.90

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	3058	0	2995	32	1
2	B	3050	0	2981	34	1
3	C	3050	0	2981	165	2
4	A	28	0	24	0	0
4	B	28	0	24	0	0
5	A	364	0	0	14	0
5	B	363	0	0	14	0
5	C	269	0	0	126	0
All	All	10210	0	9005	229	2

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 13.

The worst 5 of 229 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)
2:B:218:ASN:HB3	5:B:1245:HOH:O	1.21	1.29
3:C:69:LYS:HD2	5:C:701:HOH:O	1.25	1.27
3:C:363:ILE:HG22	5:C:764:HOH:O	1.13	1.27
3:C:198:LEU:HG	5:C:765:HOH:O	1.28	1.27
3:C:279:VAL:HG21	5:C:584:HOH:O	1.26	1.26

All (2) 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)
2:B:162:SER:O	3:C:154:ASN:ND2[1_455]	2.02	0.18
1:A:347:LYS:NZ	3:C:295:LYS:NZ[1_455]	2.13	0.07

5.3 Torsion angles

5.3.1 Protein backbone

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	392/456 (86%)	386 (98%)	5 (1%)	1 (0%)	46	29
2	B	390/432 (90%)	382 (98%)	8 (2%)	0	100	100
3	C	390/432 (90%)	379 (97%)	10 (3%)	1 (0%)	46	29
All	All	1172/1320 (89%)	1147 (98%)	23 (2%)	2 (0%)	52	35

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	154	ASN
1	A	65	ASN

5.3.2 Protein sidechains

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	321/375 (86%)	317 (99%)	4 (1%)	78	71
2	B	319/352 (91%)	315 (99%)	4 (1%)	76	68
3	C	319/356 (90%)	312 (98%)	7 (2%)	60	45
All	All	959/1083 (89%)	944 (98%)	15 (2%)	70	59

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	356	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	372	LYS
3	C	228	LEU
2	B	244	LEU
3	C	198	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	79	ASN
2	B	219	GLN
3	C	79	ASN
2	B	57	GLN
3	C	229	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

5.5 Carbohydrates ⓘ

6 carbohydrates are 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$
4	XYS	A	800	4	10,10,10	2.05	3 (30%)	12,14,14	2.45	6 (50%)
4	XYP	A	801	4	9,9,10	0.99	0	12,12,14	2.32	4 (33%)
4	XYP	A	802	4	9,9,10	1.05	1 (11%)	12,12,14	1.36	1 (8%)
4	XYS	B	800	4	10,10,10	1.98	3 (30%)	12,14,14	2.10	7 (58%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	XYP	B	801	4	9,9,10	1.27	1 (11%)	12,12,14	2.24	3 (25%)
4	XYP	B	802	4	9,9,10	0.98	1 (11%)	12,12,14	1.19	1 (8%)

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	XYS	A	800	4	-	0/0/17/17	0/1/1/1
4	XYP	A	801	4	-	0/0/14/17	0/1/1/1
4	XYP	A	802	4	-	0/0/14/17	0/1/1/1
4	XYS	B	800	4	-	0/0/17/17	0/1/1/1
4	XYP	B	801	4	-	0/0/14/17	0/1/1/1
4	XYP	B	802	4	-	0/0/14/17	0/1/1/1

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	800	XYS	O5-C5	-4.90	1.35	1.43
4	B	800	XYS	O5-C5	-4.33	1.36	1.43
4	B	800	XYS	O5-C1	-2.92	1.38	1.43
4	B	800	XYS	O2-C2	-2.40	1.37	1.43
4	A	800	XYS	O5-C1	-2.38	1.39	1.43

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	801	XYP	C4B-C3B-C2B	-6.05	105.26	111.24
4	A	801	XYP	C4B-C3B-C2B	-5.61	105.70	111.24
4	A	801	XYP	O3B-C3B-C2B	-3.38	103.89	110.00
4	A	801	XYP	C1B-C2B-C3B	-3.06	105.93	109.54
4	B	802	XYP	O3B-C3B-C2B	-3.05	104.49	110.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	386/456 (84%)	-0.79	0 100 100	4, 9, 19, 36	0
2	B	387/432 (89%)	-0.80	0 100 100	4, 9, 18, 35	0
3	C	391/432 (90%)	-0.08	1 (0%) 94 92	15, 29, 41, 54	0
All	All	1164/1320 (88%)	-0.56	1 (0%) 95 93	4, 13, 36, 54	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	154	ASN	3.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.3 Carbohydrates ⓘ

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
4	XYS	A	800	10/10	0.96	0.07	1.66	8,10,15,17	0
4	XYS	B	800	10/10	0.97	0.08	1.64	9,11,14,19	0
4	XYP	A	801	9/10	0.99	0.06	0.18	5,5,6,6	0
4	XYP	B	801	9/10	0.99	0.05	-0.46	4,5,5,5	0
4	XYP	B	802	9/10	0.98	0.05	-0.74	3,3,4,4	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
4	XYP	A	802	9/10	0.99	0.05	-0.89	4,4,4,4	0

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.