wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 11:58 pm BST

PDB ID	:	1H1N
Title	:	Atomic resolution structure of the major endoglucanase from Thermoascus aurantiacus
Authors	:	Van Petegem, F.; Vandenberghe, I.; Bhat, M.K.; Van Beeumen, J.
Deposited on	:	2002-07-19
Resolution	:	1.12 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
https://www.wwpdb.org/validation/2017/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:

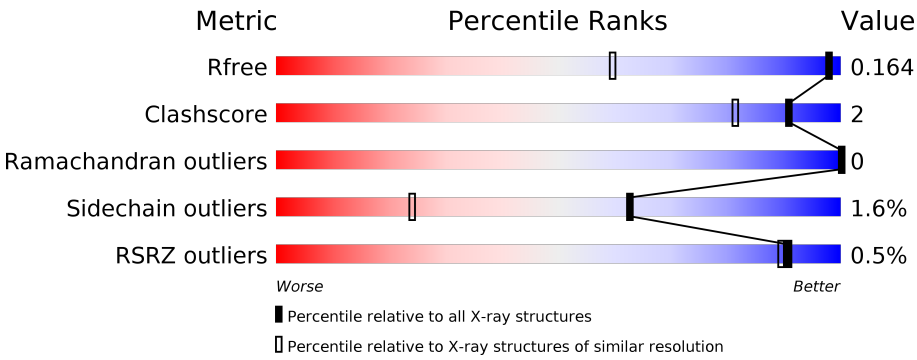
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.12 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	1168 (1.14-1.10)
Clashscore	141614	1205 (1.14-1.10)
Ramachandran outliers	138981	1168 (1.14-1.10)
Sidechain outliers	138945	1165 (1.14-1.10)
RSRZ outliers	127900	1146 (1.14-1.10)

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 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	305	91% 9%
1	B	305	% 92% 7% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5739 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 TYPE CELLULASE ENGI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	305	2384	1518	381	472	13	0	8	0
1	B	304	2379	1515	380	471	13	0	9	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	110	GLU	GLN	conflict	UNP Q8TG26
A	131	ASP	ASN	conflict	UNP Q8TG26
A	246	ASP	ASN	conflict	UNP Q8TG26
A	247	ASN	ASP	conflict	UNP Q8TG26
B	110	GLU	GLN	conflict	UNP Q8TG26
B	131	ASP	ASN	conflict	UNP Q8TG26
B	246	ASP	ASN	conflict	UNP Q8TG26
B	247	ASN	ASP	conflict	UNP Q8TG26

- Molecule 2 is water.

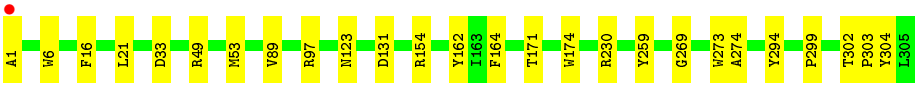
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	495	495	495	0	0
2	B	481	481	481	0	0

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 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.

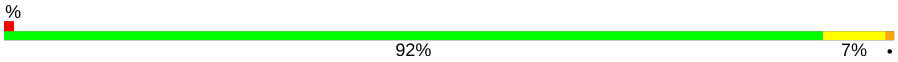
- Molecule 1: ENDO TYPE CELLULASE ENGI

Chain A:



- Molecule 1: ENDO TYPE CELLULASE ENGI

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	75.76Å 84.73Å 89.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 1.12 14.94 – 1.10	Depositor EDS
% Data completeness (in resolution range)	94.1 (15.00-1.12) 90.0 (14.94-1.10)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.86 (at 1.10Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.133 , 0.171 0.133 , 0.164	Depositor DCC
R_{free} test set	11469 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	10.3	Xtriage
Anisotropy	0.432	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 68.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.013 for -h,l,k	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	5739	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.01% 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)

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	0.78	0/2485	1.22	14/3397 (0.4%)
1	B	0.77	0/2485	1.21	20/3396 (0.6%)
All	All	0.78	0/4970	1.21	34/6793 (0.5%)

There are no bond length outliers.

The worst 5 of 34 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	230	ARG	NE-CZ-NH2	-13.68	113.46	120.30
1	A	259	TYR	CZ-CE2-CD2	11.93	130.53	119.80
1	A	259	TYR	CB-CG-CD2	-10.14	114.92	121.00
1	A	259	TYR	CB-CG-CD1	9.65	126.79	121.00
1	B	230	ARG	NE-CZ-NH1	8.30	124.45	120.30

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	2384	0	2229	8	0
1	B	2379	0	2216	8	0
2	A	495	0	0	1	0
2	B	481	0	0	3	0
All	All	5739	0	4445	16	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 2.

The worst 5 of 16 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:302:THR:OG1	1:A:303:PRO:HD3	2.01	0.61
1:A:21:LEU:HD21	1:A:53[A]:MET:CE	2.33	0.59
1:B:203:SER:HB2	2:B:2349:HOH:O	2.04	0.57
1:A:6:TRP:O	1:A:269:GLY:HA3	2.08	0.54
1:B:2:LYS:HE2	1:B:4:PHE:O	2.13	0.49

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	311/305 (102%)	302 (97%)	9 (3%)	0	100	100
1	B	311/305 (102%)	304 (98%)	7 (2%)	0	100	100
All	All	622/610 (102%)	606 (97%)	16 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	259/254 (102%)	255 (98%)	4 (2%)	65	26
1	B	258/254 (102%)	254 (98%)	4 (2%)	62	24
All	All	517/508 (102%)	509 (98%)	8 (2%)	62	26

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	174	TRP
1	B	174	TRP
1	B	131	ASP
1	A	171	THR
1	B	2	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	79	ASN
1	A	123	ASN
1	A	139	GLN
1	A	247	ASN
1	B	263	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)

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 (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	305/305 (100%)	-0.30	1 (0%)	94	91	8, 13, 21, 43	0
1	B	304/305 (99%)	-0.37	2 (0%)	87	85	9, 13, 24, 37	0
All	All	609/610 (99%)	-0.34	3 (0%)	91	89	8, 13, 23, 43	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1	ALA	6.5
1	B	19	GLN	2.6
1	B	302	THR	2.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.