wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 07:43 am BST

PDB ID	:	3QXF
Title	:	Structure of the bacterial cellulose synthase subunit Z
Authors	:	Zimmer, J.
Deposited on	:	2011-03-01
Resolution	:	1.85 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:

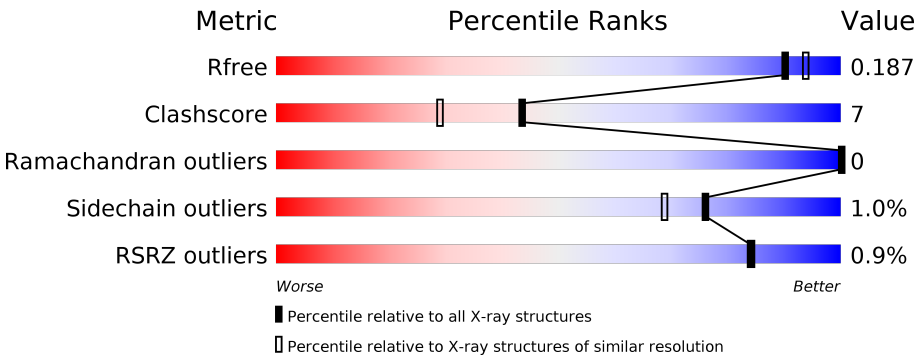
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	355	% 85% 10% 5%
1	B	355	% 85% 10% 5%
1	C	355	% 83% 11% • 5%
1	D	355	% 82% 13% 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 12415 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 Endoglucanase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	337	2722	1742	471	500	9	0	0	0
1	B	337	2727	1745	471	502	9	0	1	0
1	C	337	2727	1745	471	502	9	0	1	0
1	D	337	2722	1742	471	500	9	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	369	LEU	-	EXPRESSION TAG	UNP P37651
A	370	GLU	-	EXPRESSION TAG	UNP P37651
A	371	HIS	-	EXPRESSION TAG	UNP P37651
A	372	HIS	-	EXPRESSION TAG	UNP P37651
A	373	HIS	-	EXPRESSION TAG	UNP P37651
A	374	HIS	-	EXPRESSION TAG	UNP P37651
A	375	HIS	-	EXPRESSION TAG	UNP P37651
A	376	HIS	-	EXPRESSION TAG	UNP P37651
B	369	LEU	-	EXPRESSION TAG	UNP P37651
B	370	GLU	-	EXPRESSION TAG	UNP P37651
B	371	HIS	-	EXPRESSION TAG	UNP P37651
B	372	HIS	-	EXPRESSION TAG	UNP P37651
B	373	HIS	-	EXPRESSION TAG	UNP P37651
B	374	HIS	-	EXPRESSION TAG	UNP P37651
B	375	HIS	-	EXPRESSION TAG	UNP P37651
B	376	HIS	-	EXPRESSION TAG	UNP P37651
C	369	LEU	-	EXPRESSION TAG	UNP P37651
C	370	GLU	-	EXPRESSION TAG	UNP P37651
C	371	HIS	-	EXPRESSION TAG	UNP P37651
C	372	HIS	-	EXPRESSION TAG	UNP P37651
C	373	HIS	-	EXPRESSION TAG	UNP P37651

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	374	HIS	-	EXPRESSION TAG	UNP P37651
C	375	HIS	-	EXPRESSION TAG	UNP P37651
C	376	HIS	-	EXPRESSION TAG	UNP P37651
D	369	LEU	-	EXPRESSION TAG	UNP P37651
D	370	GLU	-	EXPRESSION TAG	UNP P37651
D	371	HIS	-	EXPRESSION TAG	UNP P37651
D	372	HIS	-	EXPRESSION TAG	UNP P37651
D	373	HIS	-	EXPRESSION TAG	UNP P37651
D	374	HIS	-	EXPRESSION TAG	UNP P37651
D	375	HIS	-	EXPRESSION TAG	UNP P37651
D	376	HIS	-	EXPRESSION TAG	UNP P37651

- Molecule 2 is water.

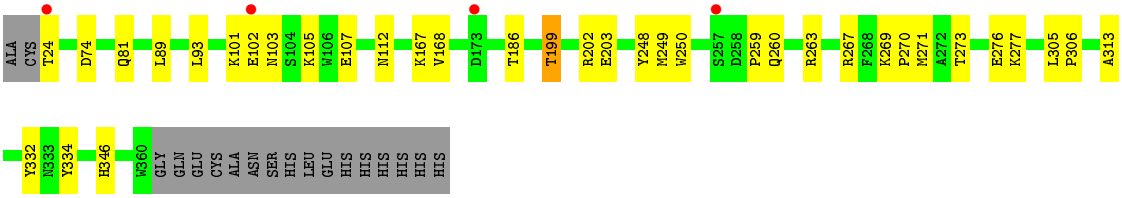
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	377	377	377	0	0
			Total	O		
2	B	324	324	324	0	0
			Total	O		
2	C	423	423	423	0	0
			Total	O		
2	D	393	393	393	0	0

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.

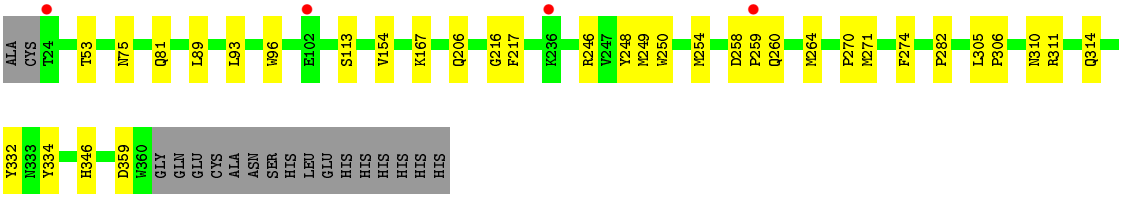
- Molecule 1: Endoglucanase

Chain A:



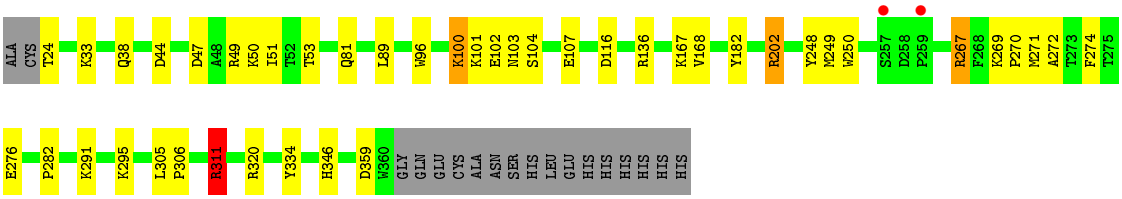
- Molecule 1: Endoglucanase

Chain B:



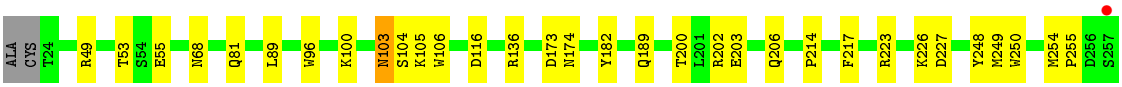
- Molecule 1: Endoglucanase

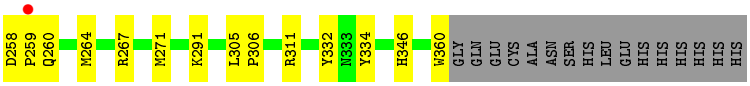
Chain C:



- Molecule 1: Endoglucanase

Chain D:




D258
P259
Q260
M264
R267
M271
K291
L305
P306
R311
Y332
N333
Y334
H346
W360
GLY
GLN
GLU
CYS
ALA
ASN
SER
HIS
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	54.82Å 87.93Å 91.75Å 69.96° 74.36° 78.22°	Depositor
Resolution (Å)	24.10 – 1.85 24.10 – 1.85	Depositor EDS
% Data completeness (in resolution range)	97.7 (24.10-1.85) 97.7 (24.10-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.55 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.136 , 0.178 0.149 , 0.187	Depositor DCC
R_{free} test set	6480 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	13.0	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 46.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	12415	wwPDB-VP
Average B, all atoms (Å^2)	13.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.31% 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	0.40	0/2793	0.46	0/3776
1	B	0.37	0/2801	0.46	0/3787
1	C	0.42	0/2801	0.48	0/3787
1	D	0.39	0/2793	0.47	0/3776
All	All	0.40	0/11188	0.47	0/15126

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	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	311	ARG	Sidechain

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	2722	0	2637	30	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2727	0	2641	33	0
1	C	2727	0	2641	40	0
1	D	2722	0	2637	41	0
2	A	377	0	0	14	0
2	B	324	0	0	9	0
2	C	423	0	0	13	0
2	D	393	0	0	14	0
All	All	12415	0	10556	143	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 7.

The worst 5 of 143 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:C:47[A]:ASP:OD2	1:C:51:ILE:HG12	1.30	1.25
1:C:167:LYS:HE2	2:C:1305:HOH:O	1.42	1.18
1:A:186:THR:HB	2:A:1458:HOH:O	1.54	1.05
1:A:202:ARG:NH2	1:A:203:GLU:OE2	2.00	0.95
1:A:24:THR:HG23	2:A:1294:HOH:O	1.67	0.95

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	335/355 (94%)	326 (97%)	9 (3%)	0	100	100
1	B	336/355 (95%)	329 (98%)	7 (2%)	0	100	100
1	C	336/355 (95%)	329 (98%)	7 (2%)	0	100	100
1	D	335/355 (94%)	327 (98%)	8 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	1342/1420 (94%)	1311 (98%)	31 (2%)	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	282/288 (98%)	280 (99%)	2 (1%)	84	79
1	B	283/288 (98%)	282 (100%)	1 (0%)	91	89
1	C	283/288 (98%)	278 (98%)	5 (2%)	59	45
1	D	282/288 (98%)	279 (99%)	3 (1%)	73	65
All	All	1130/1152 (98%)	1119 (99%)	11 (1%)	76	69

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	202	ARG
1	C	267	ARG
1	D	103	ASN
1	C	100	LYS
1	C	334	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	346	HIS
1	C	81	GLN
1	D	206	GLN
1	B	260	GLN
1	B	317	GLN

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	328/355 (92%)	-0.51	4 (1%)	79	79	4, 10, 23, 42	0
1	B	328/355 (92%)	-0.49	4 (1%)	79	79	5, 12, 24, 32	0
1	C	328/355 (92%)	-0.62	2 (0%)	89	89	3, 8, 20, 31	0
1	D	328/355 (92%)	-0.51	2 (0%)	89	89	4, 10, 22, 34	0
All	All	1312/1420 (92%)	-0.54	12 (0%)	84	84	3, 10, 23, 42	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	102	GLU	4.9
1	D	259	PRO	2.9
1	B	24	THR	2.8
1	D	257	SER	2.5
1	C	257	SER	2.4

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.