wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 07:22 AM GMT

PDB ID	:	3AJ2
Title	:	The structure of AxCeSD octamer (C-terminal HIS-tag) from Acetobacter xylinum
Authors	:	Hu, S.Q.; Tajima, K.; Zhou, Y.; Tanaka, I.; Yao, M.
Deposited on	:	2010-05-20
Resolution	:	2.70 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 (i) symbol.

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

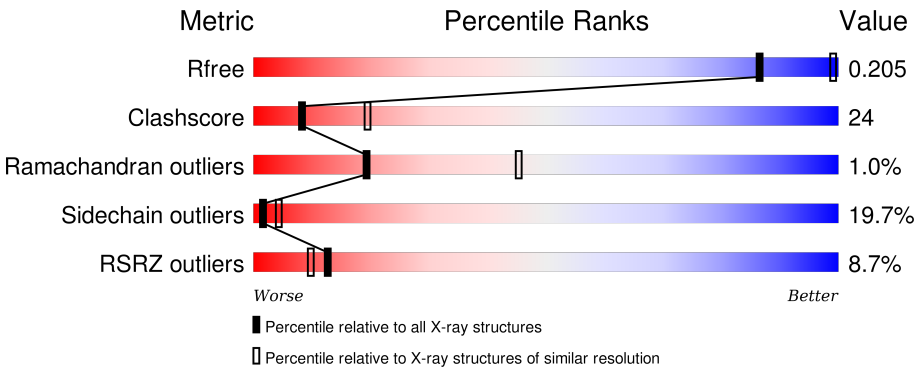
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.70 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	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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	162	10% 51% 33% 7% 9%
1	B	162	9% 53% 34% 10% •
1	C	162	7% 52% 28% 9% 10%
1	D	162	7% 59% 27% 11% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5046 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 Cellulose synthase operon protein D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	148	1155	732	198	220	5	0	0	0
1	B	158	1247	786	221	235	5	0	0	0
1	C	145	1137	725	193	214	5	0	0	0
1	D	158	1246	786	221	234	5	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	157	HIS	-	EXPRESSION TAG	UNP P37719
A	158	HIS	-	EXPRESSION TAG	UNP P37719
A	159	HIS	-	EXPRESSION TAG	UNP P37719
A	160	HIS	-	EXPRESSION TAG	UNP P37719
A	161	HIS	-	EXPRESSION TAG	UNP P37719
A	162	HIS	-	EXPRESSION TAG	UNP P37719
B	157	HIS	-	EXPRESSION TAG	UNP P37719
B	158	HIS	-	EXPRESSION TAG	UNP P37719
B	159	HIS	-	EXPRESSION TAG	UNP P37719
B	160	HIS	-	EXPRESSION TAG	UNP P37719
B	161	HIS	-	EXPRESSION TAG	UNP P37719
B	162	HIS	-	EXPRESSION TAG	UNP P37719
C	157	HIS	-	EXPRESSION TAG	UNP P37719
C	158	HIS	-	EXPRESSION TAG	UNP P37719
C	159	HIS	-	EXPRESSION TAG	UNP P37719
C	160	HIS	-	EXPRESSION TAG	UNP P37719
C	161	HIS	-	EXPRESSION TAG	UNP P37719
C	162	HIS	-	EXPRESSION TAG	UNP P37719
D	157	HIS	-	EXPRESSION TAG	UNP P37719
D	158	HIS	-	EXPRESSION TAG	UNP P37719
D	159	HIS	-	EXPRESSION TAG	UNP P37719

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	160	HIS	-	EXPRESSION TAG	UNP P37719
D	161	HIS	-	EXPRESSION TAG	UNP P37719
D	162	HIS	-	EXPRESSION TAG	UNP P37719

- Molecule 2 is water.

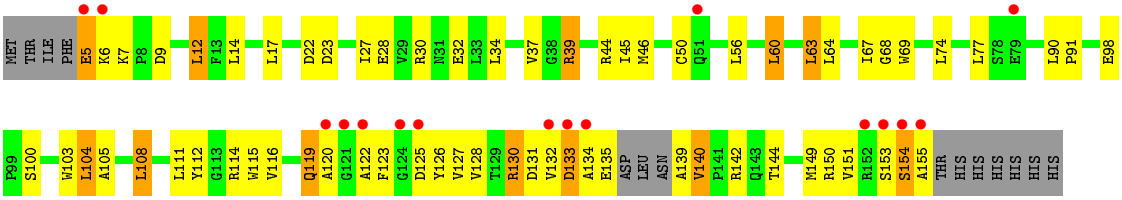
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	65	Total O 65 65	0	0
2	B	52	Total O 52 52	0	0
2	C	65	Total O 65 65	0	0
2	D	79	Total O 79 79	0	0

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.

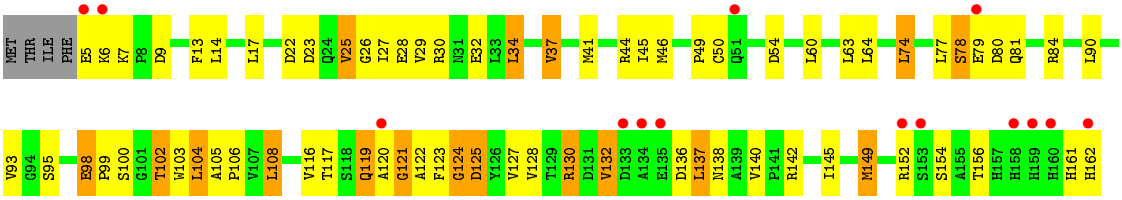
- Molecule 1: Cellulose synthase operon protein D

Chain A:



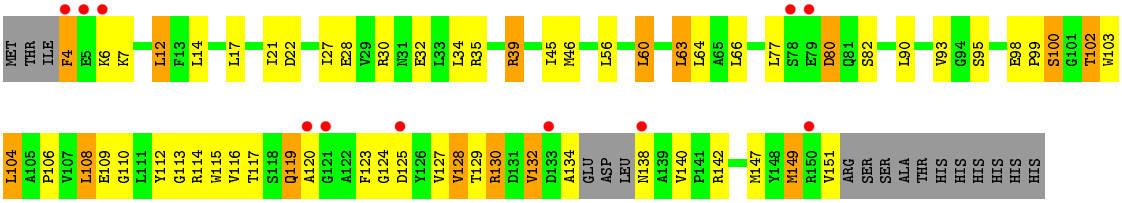
- Molecule 1: Cellulose synthase operon protein D

Chain B:



- Molecule 1: Cellulose synthase operon protein D

Chain C:



- Molecule 1: Cellulose synthase operon protein D

Chain D:

V116
Q119
A120
G121
A122
D125
V128
T129
R130
D131
V132
D133
A134
E135
N138
A139
V140
P141
R142
Q143
T144
M147
Y148
M149
R150
V151
R152
S153
S154
A155
T156
H157
H158
H159
H160
H161
H162

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	133.35Å 133.35Å 217.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.80 – 2.70 19.80 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (19.80-2.70) 99.7 (19.80-2.70)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.05 (at 2.71Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.209 , 0.276 0.213 , 0.205	Depositor DCC
R_{free} test set	2715 reflections (11.12%)	DCC
Wilson B-factor (Å^2)	57.3	Xtriage
Anisotropy	0.174	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 43.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.56$, $< L^2 > = 0.40$	Xtriage
Outliers	3 of 27126 reflections (0.011%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5046	wwPDB-VP
Average B, all atoms (Å^2)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 50.29 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 6.6114e-05.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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)

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	0.86	0/1176	0.91	0/1599
1	B	0.79	1/1275 (0.1%)	0.92	4/1735 (0.2%)
1	C	0.80	0/1159	0.87	0/1577
1	D	0.83	0/1274	0.87	3/1735 (0.2%)
All	All	0.82	1/4884 (0.0%)	0.89	7/6646 (0.1%)

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	A	0	1
1	B	0	1
1	D	0	1
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	32	GLU	CG-CD	5.13	1.59	1.51

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	79	GLU	N-CA-C	-5.44	96.30	111.00
1	B	74	LEU	CA-CB-CG	5.43	127.80	115.30
1	D	22	ASP	CB-CG-OD2	-5.19	113.63	118.30
1	D	154	SER	N-CA-C	5.14	124.88	111.00
1	B	142	ARG	NE-CZ-NH1	5.12	122.86	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	134	ALA	Peptide
1	B	78	SER	Peptide
1	D	160	HIS	Peptide

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	1155	0	1157	68	0
1	B	1247	0	1228	59	0
1	C	1137	0	1138	58	0
1	D	1246	0	1228	56	0
2	A	65	0	0	9	0
2	B	52	0	0	13	0
2	C	65	0	0	20	0
2	D	79	0	0	17	0
All	All	5046	0	4751	232	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 24.

The worst 5 of 232 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:D:160:HIS:CD2	1:D:161:HIS:H	1.51	1.29
1:A:50:CYS:H	1:A:119:GLN:NE2	1.44	1.13
1:B:50:CYS:H	1:B:119:GLN:NE2	1.47	1.10
1:C:112:TYR:HA	2:C:241:HOH:O	1.51	1.10
1:A:39:ARG:HH21	1:A:114:ARG:NH2	1.52	1.05

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	144/162 (89%)	135 (94%)	9 (6%)	0	100	100
1	B	156/162 (96%)	148 (95%)	6 (4%)	2 (1%)	15	37
1	C	141/162 (87%)	132 (94%)	7 (5%)	2 (1%)	14	35
1	D	156/162 (96%)	143 (92%)	11 (7%)	2 (1%)	15	37
All	All	597/648 (92%)	558 (94%)	33 (6%)	6 (1%)	19	45

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	121	GLY
1	D	120	ALA
1	B	124	GLY
1	D	135	GLU
1	C	124	GLY

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	126/140 (90%)	103 (82%)	23 (18%)	2	5
1	B	136/140 (97%)	109 (80%)	27 (20%)	1	4
1	C	124/140 (89%)	98 (79%)	26 (21%)	1	3
1	D	136/140 (97%)	109 (80%)	27 (20%)	1	4
All	All	522/560 (93%)	419 (80%)	103 (20%)	1	4

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	149	MET
1	C	46	MET
1	D	143	GLN
1	B	152	ARG
1	C	12	LEU

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	119	GLN
1	B	119	GLN
1	C	81	GLN
1	D	51	GLN
1	D	160	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 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	148/162 (91%)	0.19	16 (10%)	8	6	39, 54, 108, 122	0
1	B	158/162 (97%)	-0.05	14 (8%)	12	9	38, 56, 104, 118	0
1	C	145/162 (89%)	0.12	11 (7%)	17	15	42, 58, 114, 125	0
1	D	158/162 (97%)	0.01	12 (7%)	17	15	41, 55, 96, 115	0
All	All	609/648 (93%)	0.07	53 (8%)	13	10	38, 56, 105, 125	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	154	SER	8.1
1	A	155	ALA	6.7
1	D	5	GLU	6.6
1	D	159	HIS	6.1
1	C	4	PHE	5.8

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.