wwPDB X-ray Structure Validation Summary Report (i)

Aug 17, 2016 – 08:26 AM EDT

PDB ID	:	5KGH
Title	:	X-ray structure of a glucosamine N-Acetyltransferase from Clostridium acetobutylicum, mutant Y297F
Authors	:	Dopkins, B.J.; Thoden, J.B.; Tipton, P.A.; Holden, H.M.
Deposited on	:	2016-06-13
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 (i) symbol.

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

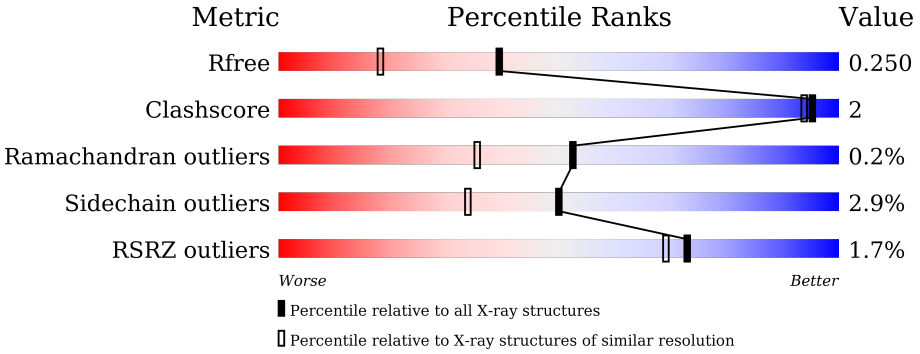
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 1.80 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 <=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	328	3% 89% 7% •
1	B	328	.% 90% • • 5%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	EDO	A	403	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5789 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.

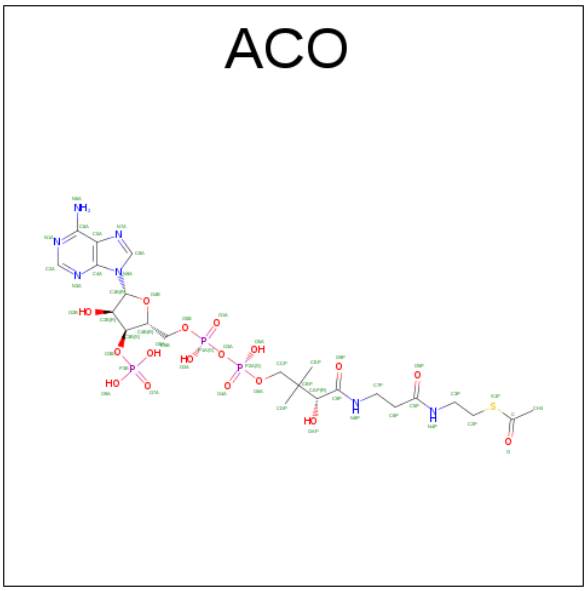
- Molecule 1 is a protein called Predicted acetyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	316	2610	1714	428	458	10	0	0	0
1	B	313	2595	1703	425	457	10	0	2	0

There are 18 discrepancies between the modelled and reference sequences:

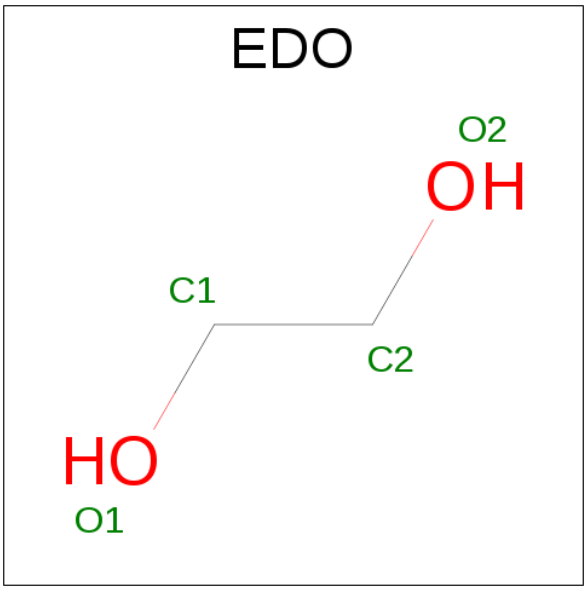
Chain	Residue	Modelled	Actual	Comment	Reference
A	297	PHE	TYR	engineered mutation	UNP Q97ML2
A	321	GLY	-	expression tag	UNP Q97ML2
A	322	GLY	-	expression tag	UNP Q97ML2
A	323	HIS	-	expression tag	UNP Q97ML2
A	324	HIS	-	expression tag	UNP Q97ML2
A	325	HIS	-	expression tag	UNP Q97ML2
A	326	HIS	-	expression tag	UNP Q97ML2
A	327	HIS	-	expression tag	UNP Q97ML2
A	328	HIS	-	expression tag	UNP Q97ML2
B	297	PHE	TYR	engineered mutation	UNP Q97ML2
B	321	GLY	-	expression tag	UNP Q97ML2
B	322	GLY	-	expression tag	UNP Q97ML2
B	323	HIS	-	expression tag	UNP Q97ML2
B	324	HIS	-	expression tag	UNP Q97ML2
B	325	HIS	-	expression tag	UNP Q97ML2
B	326	HIS	-	expression tag	UNP Q97ML2
B	327	HIS	-	expression tag	UNP Q97ML2
B	328	HIS	-	expression tag	UNP Q97ML2

- Molecule 2 is ACETYL COENZYME *A (three-letter code: ACO) (formula: $C_{23}H_{38}N_7O_{17}P_3S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	P	S		
2	A	1	51	23	7	17	3	1	0	0
			Total	C	N	O	P	S		
2	A	1	29	13	2	11	2	1	0	0
			Total	C	N	O	P	S		
2	B	1	51	23	7	17	3	1	0	0
			Total	C	N	O	P	S		
2	B	1	29	13	2	11	2	1	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	4	2	2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
4	B	1	1	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	175	175	175	0	0
			Total	O		
5	B	244	244	244	0	0

3 Residue-property plots (i)

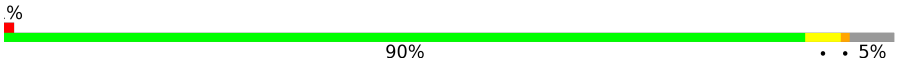
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Predicted acetyltransferase

Chain A:

- Molecule 1: Predicted acetyltransferase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	64.29Å 65.94Å 90.45Å 90.00° 106.57° 90.00°	Depositor
Resolution (Å)	50.00 – 1.80 29.61 – 1.80	Depositor EDS
% Data completeness (in resolution range)	98.1 (50.00-1.80) 98.1 (29.61-1.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	1.87 (at 1.80Å)	Xtriage
Refinement program	REFMAC 5.8.0124	Depositor
R, R_{free}	0.205 , 0.249 0.211 , 0.250	Depositor DCC
R_{free} test set	3346 reflections (5.34%)	DCC
Wilson B-factor (Å^2)	18.3	Xtriage
Anisotropy	0.131	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5789	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: ACO, EDO, CL

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	0.42	0/2678	0.80	0/3599
1	B	0.50	0/2667	0.88	3/3583 (0.1%)
All	All	0.46	0/5345	0.84	3/7182 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	178	ARG	NE-CZ-NH2	-5.75	117.42	120.30
1	B	153	ASP	CB-CG-OD1	5.73	123.45	118.30
1	B	299	ARG	NE-CZ-NH1	5.02	122.81	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	2610	0	2629	10	0
1	B	2595	0	2604	7	0
2	A	80	0	54	0	0
2	B	80	0	57	0	0
3	A	4	0	6	0	0
4	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	175	0	0	1	0
5	B	244	0	0	3	1
All	All	5789	0	5350	17	1

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 17 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:B:292[B]:ASP:OD1	5:B:501:HOH:O	1.99	0.79
1:A:164:ILE:HD12	1:A:267:LEU:HD21	1.75	0.68
1:B:292[B]:ASP:OD1	5:B:502:HOH:O	2.16	0.60
1:A:299:ARG:HB3	1:A:299:ARG:HH11	1.70	0.56
1:A:165:LYS:NZ	1:A:262:GLN:OE1	2.40	0.54

All (1) 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)
5:B:504:HOH:O	5:B:716:HOH:O[1_455]	2.08	0.12

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	314/328 (96%)	306 (98%)	7 (2%)	1 (0%)	46 29
1	B	311/328 (95%)	304 (98%)	7 (2%)	0	100 100
All	All	625/656 (95%)	610 (98%)	14 (2%)	1 (0%)	52 35

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	216	LYS

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	281/291 (97%)	272 (97%)	9 (3%)	46	29
1	B	279/291 (96%)	272 (98%)	7 (2%)	55	39
All	All	560/582 (96%)	544 (97%)	16 (3%)	50	34

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	257	LYS
1	A	299	ARG
1	B	145	ASP
1	A	169	LYS
1	B	178	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	49	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)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 for Mogul analysis.

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	ACO	A	401	-	44,53,53	0.88	1 (2%)	54,79,79	1.73	9 (16%)
2	ACO	A	402	-	23,28,53	1.76	2 (8%)	29,40,79	1.31	4 (13%)
3	EDO	A	403	-	3,3,3	0.64	0	2,2,2	0.28	0
2	ACO	B	401	-	44,53,53	0.93	2 (4%)	54,79,79	1.95	11 (20%)
2	ACO	B	402	-	23,28,53	0.76	0	29,40,79	1.37	5 (17%)

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	ACO	A	401	-	-	0/47/67/67	0/3/3/3
2	ACO	A	402	-	-	0/36/36/67	0/0/0/3
3	EDO	A	403	-	-	0/1/1/1	0/0/0/0
2	ACO	B	401	-	-	0/47/67/67	0/3/3/3
2	ACO	B	402	-	-	0/36/36/67	0/0/0/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	402	ACO	CH3-C	-6.01	1.23	1.50
2	B	401	ACO	O4B-C1B	2.85	1.45	1.41
2	B	401	ACO	C5A-C4A	2.90	1.47	1.40
2	A	401	ACO	C5A-C4A	3.49	1.48	1.40
2	A	402	ACO	O-C	5.18	1.49	1.21

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401	ACO	N3A-C2A-N1A	-8.53	122.17	128.87
2	B	401	ACO	N3A-C2A-N1A	-7.54	122.95	128.87
2	B	401	ACO	C1B-N9A-C4A	-6.69	119.34	126.81
2	B	401	ACO	O5P-C5P-C6P	-2.98	116.80	121.97
2	B	401	ACO	C4B-O4B-C1B	-2.81	106.66	109.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	316/328 (96%)	-0.10	9 (2%)	56	51	13, 28, 53, 79	0
1	B	313/328 (95%)	-0.44	2 (0%)	90	88	9, 19, 39, 60	0
All	All	629/656 (95%)	-0.27	11 (1%)	73	69	9, 24, 48, 79	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	163	SER	3.5
1	A	164	ILE	3.5
1	B	159	PHE	2.8
1	A	161	LYS	2.7
1	B	245	HIS	2.7

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
3	EDO	A	403	4/4	0.76	0.16	9.34	37,38,38,45	0
2	ACO	A	401	51/51	0.94	0.11	-0.24	15,25,77,79	0
2	ACO	B	401	51/51	0.97	0.07	-0.73	9,15,19,21	0
2	ACO	A	402	29/51	0.96	0.08	-0.84	19,27,39,41	0
2	ACO	B	402	29/51	0.98	0.07	-0.84	14,19,26,34	0
4	CL	B	403	1/1	0.97	0.05	-1.00	33,33,33,33	0

6.5 Other polymers (i)

There are no such residues in this entry.