wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 08:18 PM GMT

PDB ID	:	1JQB
Title	:	Alcohol Dehydrogenase from Clostridium Beijerinckii: Crystal Structure of Mutant with Enhanced Thermal Stability
Authors	:	Levin, I.; Frolow, F.; Bogin, O.; Peretz, M.; Hacham, Y.; Burstein, Y.
Deposited on	:	2001-08-05
Resolution	:	1.97 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 (i)) were used in the production of this report:

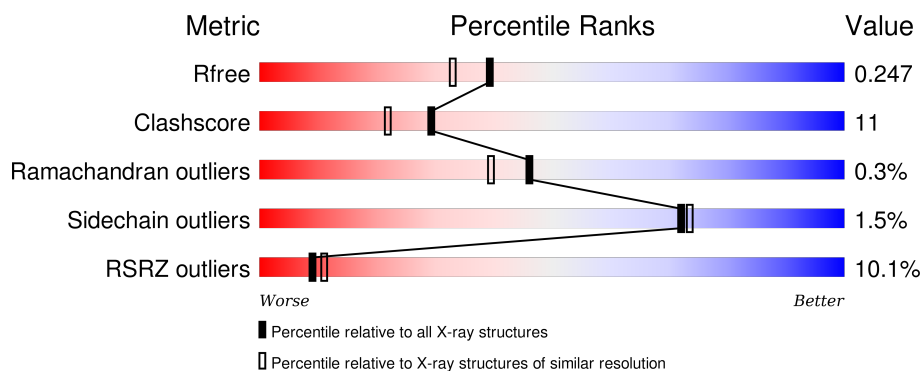|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.97 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	8664 (2.00-1.96)
Clashscore	102246	9905 (2.00-1.96)
Ramachandran outliers	100387	9792 (2.00-1.96)
Sidechain outliers	100360	9791 (2.00-1.96)
RSRZ outliers	91569	8679 (2.00-1.96)

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	351	7% / 79% / 19% / •
1	B	351	10% / 78% / 21% / •
1	C	351	17% / 78% / 21% / •
1	D	351	7% / 79% / 20% / •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11004 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 NADP-dependent Alcohol Dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	351	Total	C	N	O	S	0	0	0
			2647	1679	463	483	22			
1	B	351	Total	C	N	O	S	0	0	0
			2647	1679	463	483	22			
1	C	351	Total	C	N	O	S	0	0	0
			2647	1679	463	483	22			
1	D	351	Total	C	N	O	S	0	0	0
			2648	1679	463	484	22			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1154	THR	SER	CONFLICT	UNP P25984
A	1165	GLU	GLN	ENGINEERED	UNP P25984
A	1224	GLU	VAL	ENGINEERED	UNP P25984
A	1234	LYS	GLU	CONFLICT	UNP P25984
A	1254	LYS	SER	ENGINEERED	UNP P25984
A	1304	ARG	MET	ENGINEERED	UNP P25984
B	2154	THR	SER	CONFLICT	UNP P25984
B	2165	GLU	GLN	ENGINEERED	UNP P25984
B	2224	GLU	VAL	ENGINEERED	UNP P25984
B	2234	LYS	GLU	CONFLICT	UNP P25984
B	2254	LYS	SER	ENGINEERED	UNP P25984
B	2304	ARG	MET	ENGINEERED	UNP P25984
C	3154	THR	SER	CONFLICT	UNP P25984
C	3165	GLU	GLN	ENGINEERED	UNP P25984
C	3224	GLU	VAL	ENGINEERED	UNP P25984
C	3234	LYS	GLU	CONFLICT	UNP P25984
C	3254	LYS	SER	ENGINEERED	UNP P25984
C	3304	ARG	MET	ENGINEERED	UNP P25984
D	4154	THR	SER	CONFLICT	UNP P25984
D	4165	GLU	GLN	ENGINEERED	UNP P25984
D	4224	GLU	VAL	ENGINEERED	UNP P25984

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	4234	LYS	GLU	CONFLICT	UNP P25984
D	4254	LYS	SER	ENGINEERED	UNP P25984
D	4304	ARG	MET	ENGINEERED	UNP P25984

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0
2	D	1	Total Zn 1 1	0	0
2	C	1	Total Zn 1 1	0	0

- Molecule 3 is water.

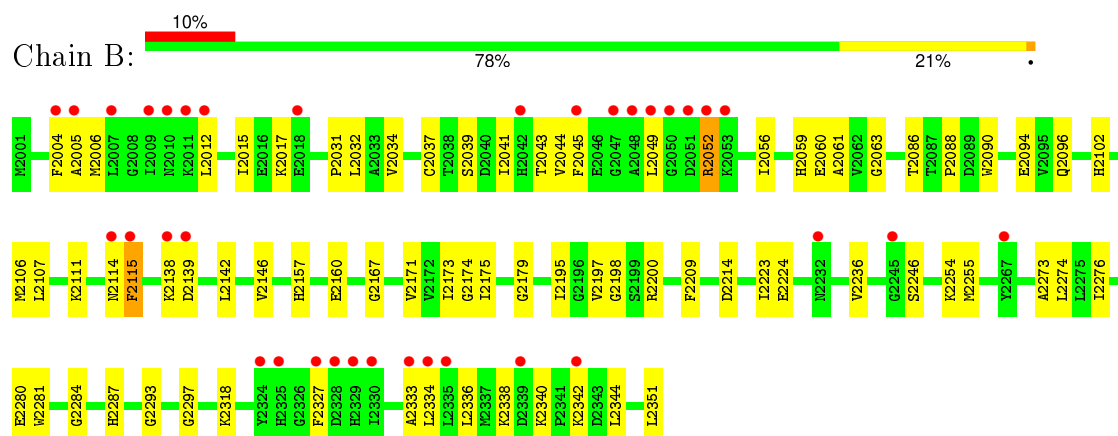| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	109	Total O 109 109	0	0
3	B	110	Total O 110 110	0	0
3	C	86	Total O 86 86	0	0
3	D	106	Total O 106 106	0	0

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.
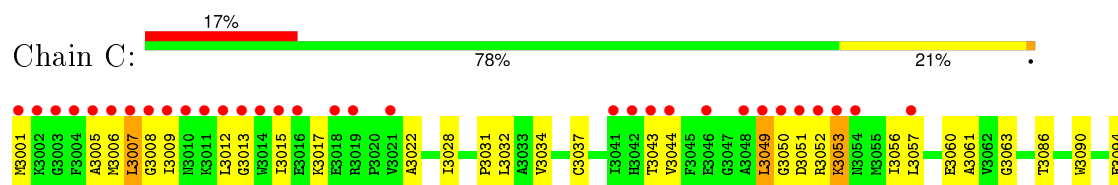
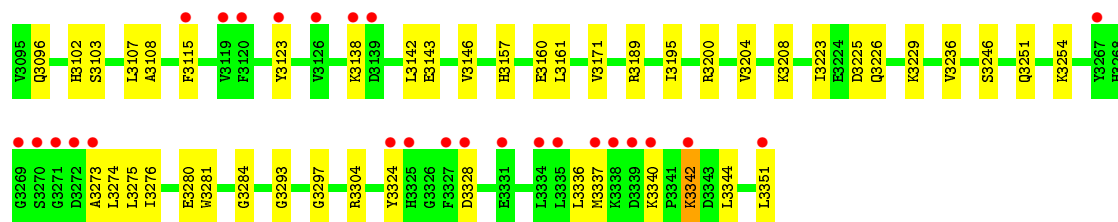- Molecule 1: NADP-dependent Alcohol Dehydrogenase

- Molecule 1: NADP-dependent Alcohol Dehydrogenase

- Molecule 1: NADP-dependent Alcohol Dehydrogenase

V3095 Q3096 H3102 S3103 L3107 A3108 F3115 V3119 F3120 Y3123 V3126 K3138 D3139 L3142 E3143 V3146 H3157 E3160 L3161 V3171 R3189 I3195 R3200 V3204 K3208 I3223 E3224 D3225 Q3226 K3229 V3236 S3246 Q3251 K3254 Y3267 H3268

G3269 S3270 G3271 D3272 A3273 L3274 L3275 I3276 E3280 W3281 G3284 G3293 G3297 R3304 Y3324 H3325 G3326 F3327 D3328 E3331 L3334 I3335 L3336 M3337 K3338 D3339 K3340 P3341 K3342 D3343 L3344 L3351

• **Molecule 1: NADP-dependent Alcohol Dehydrogenase**

7%

Chain D:
79%   20%

M4001 F4004 A4005 M4006 L4007 G4008 I4009 N4010 K4011 L4012 E4018 S4024 V4034 C4037 I4041 H4042 T4043 V4044 F4045 E4046 G4047 A4048 L4049 G4050 D4051 R4052 E4060 A4061 E4071 T4086 W4090 E4094 V4095 Q4096 Q4101 H4102 W4110 S4113 P4137 M4147 F4156 H4157

E4160 V4171 I4175 G4198 S4199 R4200 P4201 I4202 F4209 L4216 N4217 Y4218 K4219 N4220 G4221 H4222 I4223 E4224 D4225 Q4226 V4227 M4228 K4229 L4230 T4231 N4232 G4233 K4234 G4235 V4236 S4246 Y4267 D4272 A4273 L4274 L4275 I4276 W4281 G4284 H4287 L4294 G4297 G4298 R4299 L4300 R4301 A4302

E4303 R4304 L4305 R4306 L4316 S4317 K4318 F4327 I4330 E4331 L4336 D4339 K4340 P4341 K4342 D4343 L4344 L4351

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	86.70Å 147.80Å 125.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.01 – 1.97 20.01 – 1.97	Depositor EDS
% Data completeness (in resolution range)	97.9 (20.01-1.97) 98.1 (20.01-1.97)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.92 (at 1.97Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.219 , 0.247 0.218 , 0.247	Depositor DCC
R_{free} test set	11159 reflections (10.00%)	DCC
Wilson B-factor (Å^2)	24.8	Xtriage
Anisotropy	0.306	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 47.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Outliers	1 of 111587 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11004	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.94% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ZN

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	0.31	0/2696	0.62	0/3635
1	B	0.32	0/2696	0.62	0/3635
1	C	0.30	0/2696	0.60	0/3635
1	D	0.31	0/2697	0.62	0/3635
All	All	0.31	0/10785	0.62	0/14540

There are no bond length outliers.

There are no bond angle outliers.

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	2647	0	2685	65	0
1	B	2647	0	2685	68	0
1	C	2647	0	2685	72	0
1	D	2648	0	2685	56	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	109	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	110	0	0	0	0
3	C	86	0	0	5	0
3	D	106	0	0	2	0
All	All	11004	0	10740	242	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 11.

The worst 5 of 242 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:2102:HIS:HD2	1:B:2107:LEU:H	1.08	1.02
1:A:1102:HIS:HD2	1:A:1107:LEU:H	1.03	0.94
1:B:2200:ARG:HD2	1:B:2342:LYS:HD3	1.49	0.92
1:B:2336:LEU:HD13	1:B:2344:LEU:HD22	1.54	0.89
1:C:3102:HIS:HD2	1:C:3107:LEU:H	1.19	0.87

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	349/351 (99%)	330 (95%)	18 (5%)	1 (0%)	46	39
1	B	349/351 (99%)	333 (95%)	14 (4%)	2 (1%)	30	21
1	C	349/351 (99%)	323 (93%)	25 (7%)	1 (0%)	46	39
1	D	349/351 (99%)	329 (94%)	20 (6%)	0	100	100
All	All	1396/1404 (99%)	1315 (94%)	77 (6%)	4 (0%)	46	39

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	2115	PHE
1	A	1219	LYS
1	C	3044	VAL
1	B	2340	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	279/279 (100%)	276 (99%)	3 (1%)	80	82
1	B	279/279 (100%)	278 (100%)	1 (0%)	93	95
1	C	279/279 (100%)	272 (98%)	7 (2%)	55	53
1	D	279/279 (100%)	273 (98%)	6 (2%)	60	59
All	All	1116/1116 (100%)	1099 (98%)	17 (2%)	72	73

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	3143	GLU
1	C	3161	LEU
1	D	4218	TYR
1	C	3053	LYS
1	D	4267	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	2232	ASN
1	B	2312	ASN
1	C	3232	ASN
1	B	2102	HIS
1	C	3220	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 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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	351/351 (100%)	0.33	24 (6%)	20	24	15, 25, 51, 59	0
1	B	351/351 (100%)	0.42	35 (9%)	9	11	12, 23, 58, 70	0
1	C	351/351 (100%)	0.75	58 (16%)	2	3	14, 29, 64, 70	0
1	D	351/351 (100%)	0.39	25 (7%)	19	22	16, 26, 51, 61	0
All	All	1404/1404 (100%)	0.47	142 (10%)	9	11	12, 26, 58, 70	0

The worst 5 of 142 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	3013	GLY	7.4
1	B	2050	GLY	6.9
1	D	4218	TYR	6.8
1	D	4008	GLY	6.6
1	C	3049	LEU	6.5

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)

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
2	ZN	D	4353	1/1	0.98	0.03	-2.61	36,36,36,36	0
2	ZN	A	1353	1/1	0.97	0.04	-2.81	35,35,35,35	0
2	ZN	C	3353	1/1	0.97	0.05	-2.95	41,41,41,41	0
2	ZN	B	2353	1/1	0.90	0.06	-4.85	41,41,41,41	0

6.5 Other polymers (i)

There are no such residues in this entry.