W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2016 – 03:20 PM EST

PDB ID : 5KKU
Title : Crystal structure of an N-terminal dehydratase from difficidin assembly line
Authors : Keatinge-Clay, A.T.; Zeng, J.
Deposited on : 2016-06-22
Resolution : 2.22 Å(reported)

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

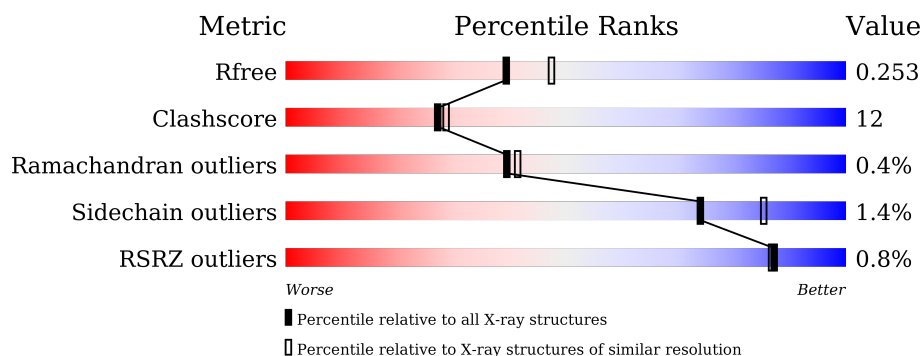|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 2.22 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	Percentile Ranks	Value
Rfree		0.253
Clashscore		12
Ramachandran outliers		0.4%
Sidechain outliers		1.4%
RSRZ outliers		0.8%

Worse   *Better*

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	4405 (2.24-2.20)
Clashscore	102246	5146 (2.24-2.20)
Ramachandran outliers	100387	5065 (2.24-2.20)
Sidechain outliers	100360	5066 (2.24-2.20)
RSRZ outliers	91569	4414 (2.24-2.20)

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	300	2% 66% 27% • 5%
1	B	300	% 74% 20% • •
1	C	300	77% 18% • •
1	D	300	76% 16% • 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9353 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 Polyketide synthase type I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	286	Total	C	N	O	S	0	0	0
			2333	1478	409	434	12			
1	B	289	Total	C	N	O	S	0	0	0
			2351	1488	412	439	12			
1	C	289	Total	C	N	O	S	0	0	0
			2351	1488	412	439	12			
1	D	284	Total	C	N	O	S	0	0	0
			2316	1468	407	429	12			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	294	GLU	-	expression tag	UNP Q1RS44
A	295	HIS	-	expression tag	UNP Q1RS44
A	296	HIS	-	expression tag	UNP Q1RS44
A	297	HIS	-	expression tag	UNP Q1RS44
A	298	HIS	-	expression tag	UNP Q1RS44
A	299	HIS	-	expression tag	UNP Q1RS44
A	300	HIS	-	expression tag	UNP Q1RS44
B	294	GLU	-	expression tag	UNP Q1RS44
B	295	HIS	-	expression tag	UNP Q1RS44
B	296	HIS	-	expression tag	UNP Q1RS44
B	297	HIS	-	expression tag	UNP Q1RS44
B	298	HIS	-	expression tag	UNP Q1RS44
B	299	HIS	-	expression tag	UNP Q1RS44
B	300	HIS	-	expression tag	UNP Q1RS44
C	294	GLU	-	expression tag	UNP Q1RS44
C	295	HIS	-	expression tag	UNP Q1RS44
C	296	HIS	-	expression tag	UNP Q1RS44
C	297	HIS	-	expression tag	UNP Q1RS44
C	298	HIS	-	expression tag	UNP Q1RS44
C	299	HIS	-	expression tag	UNP Q1RS44
C	300	HIS	-	expression tag	UNP Q1RS44

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	294	GLU	-	expression tag	UNP Q1RS44
D	295	HIS	-	expression tag	UNP Q1RS44
D	296	HIS	-	expression tag	UNP Q1RS44
D	297	HIS	-	expression tag	UNP Q1RS44
D	298	HIS	-	expression tag	UNP Q1RS44
D	299	HIS	-	expression tag	UNP Q1RS44
D	300	HIS	-	expression tag	UNP Q1RS44

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total O 1 1	0	0
2	C	1	Total O 1 1	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: Polyketide synthase type I
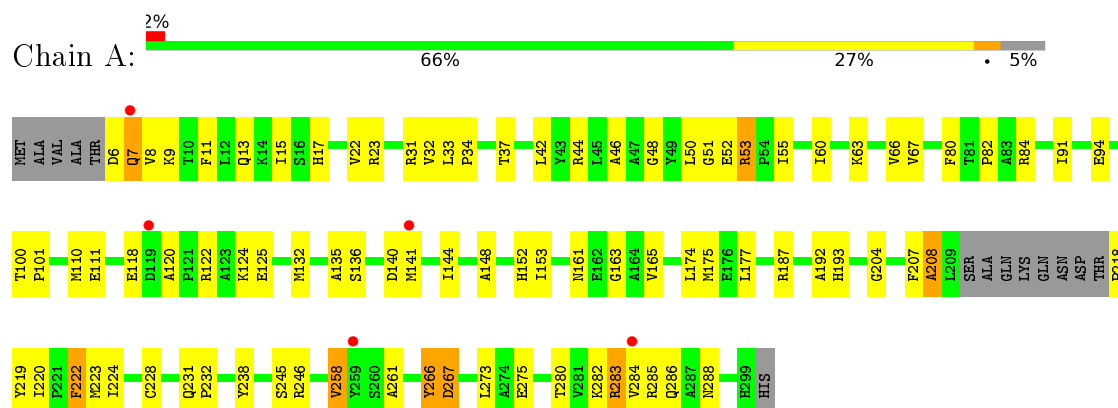
- Molecule 1: Polyketide synthase type I
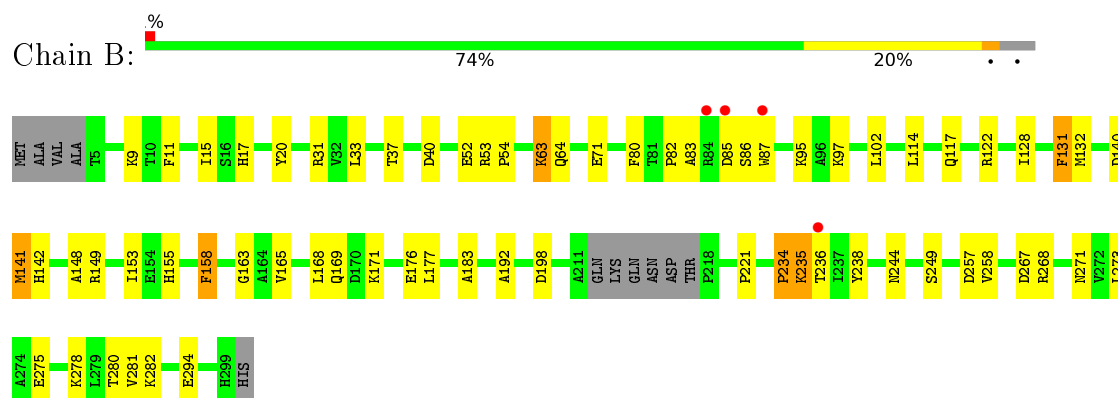
- Molecule 1: Polyketide synthase type I
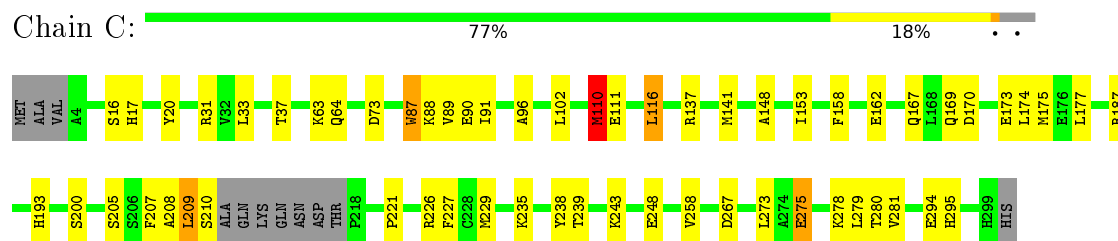
- Molecule 1: Polyketide synthase type I

Chain D:

76%　　　　　　16%　　•　5%

MET ALA VAL ALA T5 D6 Q7 H17 Y20 H28 Q29 I30 R31 V32 G51 E52 I60 K63 V67 K74 W87 K88 E94 K95 A96 L102 M110 E111 C112 Q113 Q117 GLU ASP A120 P121 K133 S136 I144 F158 V165 L168

Q169 D170 L174 L179 A183 F207 A208 L209 SER ALA GLN LYS GLN ASN ASP THR PRO Y219 I220 P221 I224 Y230 Y238 S241 V258 Y259 S260 I265 Y266 D267 R268 L273 A274 E275 K278 L279 T280 V281 K282 R283 V284 R285 Q286 H299 HIS

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	41.35Å 73.98Å 94.18Å 90.20° 90.08° 90.66°	Depositor
Resolution (Å)	39.79 – 2.22 41.35 – 2.22	Depositor EDS
% Data completeness (in resolution range)	89.0 (39.79-2.22) 87.3 (41.35-2.22)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.71 (at 2.22Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.217 , 0.248 0.225 , 0.253	Depositor DCC
R_{free} test set	2513 reflections (5.12%)	DCC
Wilson B-factor (Å^2)	25.6	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 9.5	EDS
L-test for twinning [2]	$<\|L\|> = 0.50$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	0.269 for h,-k,-l 0.012 for -h,k,-l 0.014 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9353	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.00% 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	1.45	5/2388 (0.2%)	1.10	25/3222 (0.8%)
1	B	1.44	5/2406 (0.2%)	0.92	5/3247 (0.2%)
1	C	1.45	6/2406 (0.2%)	1.03	11/3247 (0.3%)
1	D	1.40	7/2369 (0.3%)	0.93	8/3195 (0.3%)
All	All	1.44	23/9569 (0.2%)	1.00	49/12911 (0.4%)

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	C	0	1
1	D	0	3
All	All	0	6

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	294	GLU	CD-OE1	-7.61	1.17	1.25
1	A	267	ASP	N-CA	-7.30	1.31	1.46
1	B	275	GLU	CD-OE1	-6.77	1.18	1.25
1	B	275	GLU	CD-OE2	-6.66	1.18	1.25
1	D	136	SER	CB-OG	-6.53	1.33	1.42

The worst 5 of 49 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	110	MET	C-N-CA	16.91	163.98	121.70
1	C	110	MET	O-C-N	-16.63	96.09	122.70
1	C	64	GLN	N-CA-CB	12.72	133.50	110.60

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	63	LYS	CB-CA-C	-12.26	85.88	110.40
1	C	110	MET	CA-C-N	11.27	142.00	117.20

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	266	TYR	Peptide
1	B	131	PHE	Peptide
1	C	110	MET	Mainchain
1	D	52	GLU	Mainchain
1	D	63	LYS	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	2333	0	2277	90	0
1	B	2351	0	2293	51	0
1	C	2351	0	2295	46	0
1	D	2316	0	2266	36	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
All	All	9353	0	9131	222	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 12.

The worst 5 of 222 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:7:GLN:HA	1:A:82:PRO:CG	1.23	1.44
1:A:7:GLN:CA	1:A:82:PRO:HG2	1.52	1.30
1:A:6:ASP:O	1:A:7:GLN:HG2	1.12	1.23
1:C:88:LYS:NZ	1:C:90:GLU:OE1	1.70	1.22
1:A:7:GLN:CA	1:A:82:PRO:CG	2.04	1.18

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	282/300 (94%)	266 (94%)	14 (5%)	2 (1%)	26	25
1	B	285/300 (95%)	272 (95%)	11 (4%)	2 (1%)	26	25
1	C	285/300 (95%)	277 (97%)	8 (3%)	0	100	100
1	D	278/300 (93%)	268 (96%)	9 (3%)	1 (0%)	39	41
All	All	1130/1200 (94%)	1083 (96%)	42 (4%)	5 (0%)	39	41

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	7	GLN
1	B	141	MET
1	B	234	PRO
1	A	125	GLU
1	D	112	CYS

5.3.2 Protein sidechains ⓘ

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	252/263 (96%)	247 (98%)	5 (2%)	63	75
1	B	254/263 (97%)	249 (98%)	5 (2%)	63	75
1	C	254/263 (97%)	254 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	250/263 (95%)	246 (98%)	4 (2%)	70	82
All	All	1010/1052 (96%)	996 (99%)	14 (1%)	74	85

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	85	ASP
1	B	122	ARG
1	D	6	ASP
1	B	63	LYS
1	D	5	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	169	GLN
1	D	296	HIS
1	B	286	GLN
1	A	167	GLN
1	C	286	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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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(Å²)	Q<0.9
1	A	286/300 (95%)	-0.26	5 (1%)	73	72	20, 28, 48, 61	0
1	B	289/300 (96%)	-0.40	4 (1%)	78	77	17, 27, 40, 59	0
1	C	289/300 (96%)	-0.47	0	100	100	16, 26, 43, 55	0
1	D	284/300 (94%)	-0.44	0	100	100	19, 27, 43, 52	0
All	All	1148/1200 (95%)	-0.39	9 (0%)	87	87	16, 27, 44, 61	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	141	MET	8.8
1	B	85	ASP	5.2
1	A	7	GLN	4.0
1	A	284	VAL	3.5
1	B	84	ARG	2.6

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.