wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 08:32 am BST

PDB ID	:	2V9D
Title	:	Crystal Structure of YagE, a prophage protein belonging to the dihydrodipicolinic acid synthase family from E. coli K12
Authors	:	Manicka, S.; Peleg, Y.; Unger, T.; Albeck, S.; Dym, O.; Greenblatt, H.M.; Bourenkov, G.; Lamzin, V.; Krishnaswamy, S.; Sussman, J.L.
Deposited on	:	2007-08-23
Resolution	:	2.15 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:

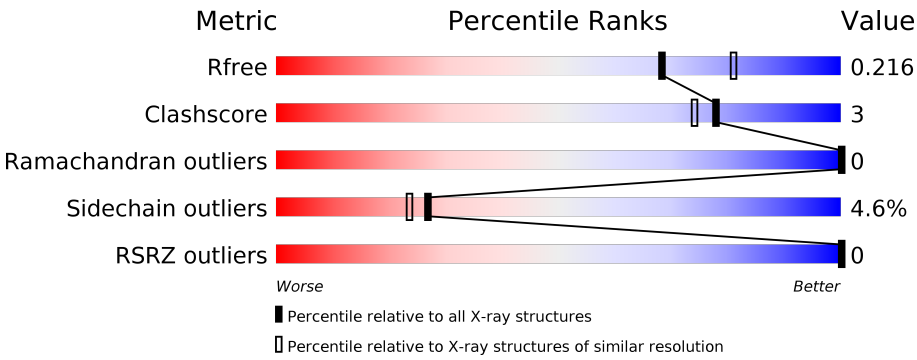
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	343	79% 8% • 13%
1	B	343	80% 6% • 13%
1	C	343	78% 9% • 13%
1	D	343	78% 8% • 13%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9660 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 YAGE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	298	2250	1443	381	420	3	3	0	0	0
1	B	298	2255	1445	385	419	3	3	0	0	0
1	C	298	2251	1444	382	419	3	3	0	0	0
1	D	298	2255	1445	385	419	3	3	0	0	0

- Molecule 2 is water.

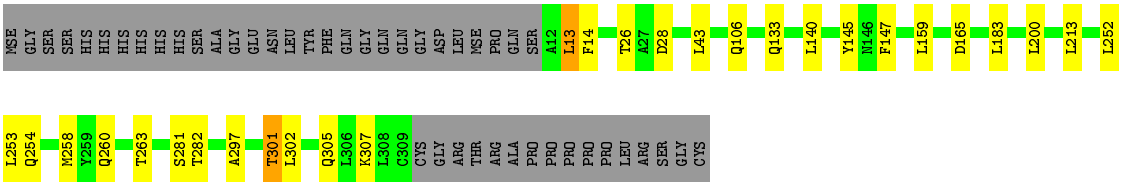
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	173	173	173	0	0
2	B	158	158	158	0	0
2	C	136	136	136	0	0
2	D	182	182	182	0	0

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.

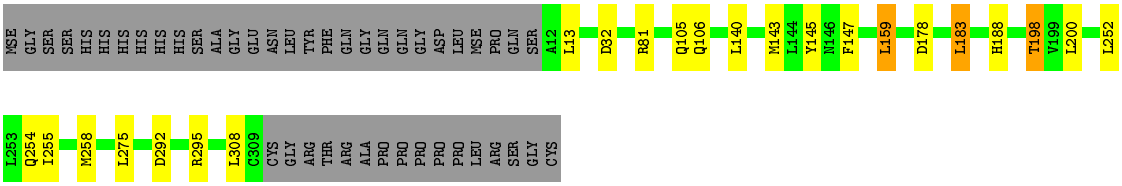
- Molecule 1: YAGE

Chain A:



- Molecule 1: YAGE

Chain B:

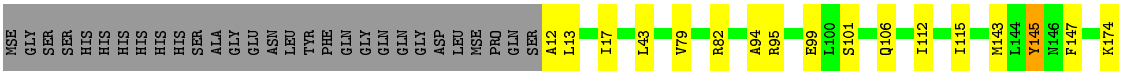


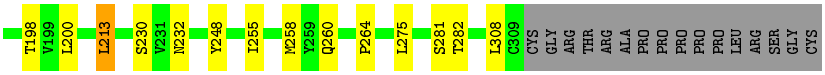
- Molecule 1: YAGE

Chain C:

- Molecule 1: YAGE

Chain D:




T198
V199
L200
L213
S230
V231
N232
Y248
I255
M258
Y259
Q260
P264
L275
S281
T282
L308
C309
CYS
GLY
ARG
THR
ARG
ALA
PRO
PRO
PRO
PRO
PRO
LEU
ARG
SER
GLY
CYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	56.28Å 149.81Å 78.97Å 90.00° 108.42° 90.00°	Depositor
Resolution (Å)	36.35 – 2.15 36.34 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.7 (36.35-2.15) 99.7 (36.34-2.15)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	9.60 (at 2.14Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.166 , 0.218 0.166 , 0.216	Depositor DCC
R_{free} test set	3466 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	17.0	Xtriage
Anisotropy	0.135	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 32.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.021 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9660	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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	0.86	0/2295	0.79	2/3124 (0.1%)
1	B	0.86	0/2300	0.81	3/3130 (0.1%)
1	C	0.90	0/2296	0.83	4/3125 (0.1%)
1	D	0.89	0/2300	0.81	3/3130 (0.1%)
All	All	0.88	0/9191	0.81	12/12509 (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	C	0	1
1	D	0	1
All	All	0	3

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	82	ARG	NE-CZ-NH2	-10.79	114.90	120.30
1	C	82	ARG	NE-CZ-NH1	8.83	124.71	120.30
1	A	282	THR	N-CA-C	7.44	131.10	111.00
1	C	282	THR	N-CA-C	-6.61	93.14	111.00
1	B	183	LEU	CA-CB-CG	6.23	129.63	115.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	281	SER	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	C	281	SER	Peptide
1	D	281	SER	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	2250	0	2268	11	0
1	B	2255	0	2279	12	0
1	C	2251	0	2272	17	0
1	D	2255	0	2279	18	0
2	A	173	0	0	4	0
2	B	158	0	0	5	0
2	C	136	0	0	4	0
2	D	182	0	0	5	0
All	All	9660	0	9098	53	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 3.

The worst 5 of 53 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:188:HIS:HD2	2:B:2099:HOH:O	1.35	1.09
1:B:13:LEU:O	1:B:198:THR:HG21	1.68	0.93
1:C:79:VAL:O	1:C:82:ARG:HD3	1.73	0.88
1:B:188:HIS:CD2	2:B:2099:HOH:O	2.18	0.82
1:D:12:ALA:N	2:D:2001:HOH:O	2.23	0.69

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	296/343 (86%)	294 (99%)	2 (1%)	0	100	100
1	B	296/343 (86%)	291 (98%)	5 (2%)	0	100	100
1	C	296/343 (86%)	292 (99%)	4 (1%)	0	100	100
1	D	296/343 (86%)	288 (97%)	8 (3%)	0	100	100
All	All	1184/1372 (86%)	1165 (98%)	19 (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	240/275 (87%)	227 (95%)	13 (5%)	22	18
1	B	241/275 (88%)	232 (96%)	9 (4%)	34	32
1	C	240/275 (87%)	227 (95%)	13 (5%)	22	18
1	D	241/275 (88%)	232 (96%)	9 (4%)	34	32
All	All	962/1100 (88%)	918 (95%)	44 (5%)	27	23

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	252	LEU
1	C	106	GLN
1	D	200	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	275	LEU
1	C	13	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	102	GLN
1	B	188	HIS
1	D	62	GLN
1	B	62	GLN
1	C	188	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	295/343 (86%)	-0.43	0	100	100	8, 14, 24, 33	0
1	B	295/343 (86%)	-0.44	0	100	100	8, 15, 23, 29	0
1	C	295/343 (86%)	-0.47	0	100	100	8, 14, 23, 28	0
1	D	295/343 (86%)	-0.51	0	100	100	7, 14, 24, 29	0
All	All	1180/1372 (86%)	-0.46	0	100	100	7, 14, 24, 33	0

There are no RSRZ outliers to report.

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.