

# Full wwPDB X-ray Structure Validation Report (i)

#### Sep 7, 2020 – 02:17 PM BST

PDB ID : 5H2Y

Title : Crystal structure of reduced DapF from Corynebacterium glutamicum

Authors : Sagong, H.-Y.; Kim, K.-J.

Deposited on : 2016-10-18

Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 Xtriage (Phenix) : 1.13

EDS : 2.14.2

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)

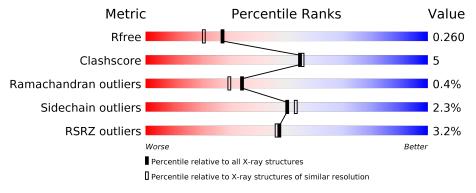
Validation Pipeline (wwPDB-VP) : 2.14.2

# 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 2.00 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar \; resolution} \\ (\#{\rm Entries, \; resolution \; range(\AA)}) \end{array}$
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	283	83%	13%	
1	В	283	85%	11%	••



# 2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4273 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 Diaminopimelate epimerase.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	278	Total 2053	C 1276	N 355	O 407	S 15	0	0	0
1	В	276	Total 2035	C 1265		O 405	S 14	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	278	HIS	-	expression tag	UNP Q8NP73
A	279	HIS	-	expression tag	UNP Q8NP73
A	280	HIS	_	expression tag	UNP Q8NP73
A	281	HIS	-	expression tag	UNP Q8NP73
A	282	HIS	_	expression tag	UNP Q8NP73
A	283	HIS	_	expression tag	UNP Q8NP73
В	278	HIS	-	expression tag	UNP Q8NP73
В	279	HIS	_	expression tag	UNP Q8NP73
В	280	HIS	-	expression tag	UNP Q8NP73
В	281	HIS	-	expression tag	UNP Q8NP73
В	282	HIS	-	expression tag	UNP Q8NP73
В	283	HIS	_	expression tag	UNP Q8NP73

• Molecule 2 is water.

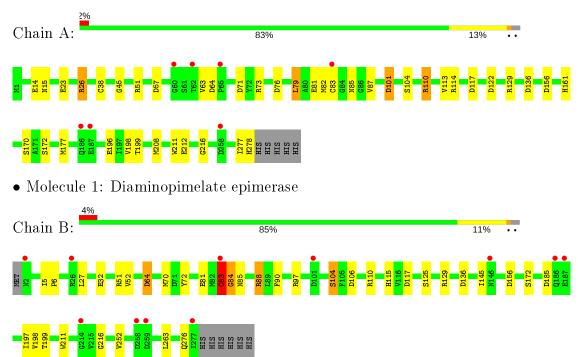
Mol	Chain	Residues	${f Atoms}$	ZeroOcc	AltConf
2	A	119	Total O 119 119	0	0
2	В	66	Total O 66 66	0	0



# 3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Diaminopimelate epimerase





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants	101.76Å 118.94Å 153.33Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 - 2.00	Depositor
Resolution (A)	33.12 - 2.00	EDS
% Data completeness	98.8 (50.00-2.00)	Depositor
(in resolution range)	98.8 (33.12-2.00)	EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	3.98 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
P. P.	0.207 , 0.252	Depositor
$R, R_{free}$	0.221 , $0.260$	DCC
$R_{free}$ test set	3114  reflections  (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.5	Xtriage
Anisotropy	0.297	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.37, 45.4	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.45, < L^2>=0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4273	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.99% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



 $<sup>^{1}</sup>$ Intensities estimated from amplitudes.

# 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		Bo	nd lengths	Bond angles		
WIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z >5	
1	A	1.15	3/2088 (0.1%)	1.18	$18/2842 \ (0.6\%)$	
1	В	1.05	3/2069 (0.1%)	1.13	$15/2817 \; (0.5\%)$	
All	All	1.10	6/4157 (0.1%)	1.16	$33/5659 \ (0.6\%)$	

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	${ m Observed}({ m \AA})$	$Ideal(\AA)$
1	В	125	SER	CB-OG	-6.63	1.33	1.42
1	В	32	GLU	CD-OE1	6.39	1.32	1.25
1	A	212	GLU	CD-OE2	6.32	1.32	1.25
1	В	32	GLU	CG-CD	6.19	1.61	1.51
1	A	170	SER	CB-OG	5.38	1.49	1.42
1	A	14	GLU	CD-OE1	5.26	1.31	1.25

All (33) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}(^{o})$
1	A	51	ARG	NE-CZ-NH1	9.12	124.86	120.30
1	A	73	ARG	NE-CZ-NH2	-8.97	115.81	120.30
1	В	97	ARG	NE-CZ-NH1	8.72	124.66	120.30
1	В	51	ARG	NE-CZ-NH2	-8.42	116.09	120.30
1	A	76	ASP	CB-CG-OD1	7.53	125.07	118.30
1	A	51	ARG	NE-CZ-NH2	-7.38	116.61	120.30
1	A	136	ASP	CB-CG-OD1	7.12	124.70	118.30
1	В	110	ARG	NE-CZ-NH2	-7.00	116.80	120.30
1	A	76	ASP	CB-CG-OD2	-6.69	112.28	118.30
1	A	110	ARG	NE-CZ-NH2	6.62	123.61	120.30
1	В	84	GLY	N-CA-C	6.48	129.31	113.10
1	В	97	ARG	NE-CZ-NH2	-6.46	117.07	120.30
1	В	64	ASP	CB-CG-OD2	-6.39	112.55	118.30
1	В	51	ARG	NE-CZ-NH1	6.37	123.49	120.30
1	A	122	ASP	CB-CG-OD1	6.37	124.03	118.30
1	В	106	ASP	CB-CG-OD1	6.29	123.96	118.30

Continued on next page...



Continued from previous page...

Mol	Chain	${f Res}$	$\mathbf{Type}$	Atoms	$\mathbf{Z}$	$\mathbf{Observed}(^o)$	$  \ \mathbf{Ideal}(^o) \  $
1	В	129	ARG	NE-CZ-NH2	-6.26	117.17	120.30
1	В	185	ASP	CB-CG-OD2	-6.20	112.72	118.30
1	A	26	ARG	NE-CZ-NH1	6.17	123.38	120.30
1	A	208	MET	CA-CB-CG	5.88	123.30	113.30
1	В	110	ARG	NE-CZ-NH1	5.88	123.24	120.30
1	A	64	ASP	CB-CG-OD1	5.85	123.56	118.30
1	A	114	ARG	NE-CZ-NH1	5.83	123.22	120.30
1	В	83	CYS	N-CA-C	5.79	126.64	111.00
1	A	110	ARG	NE-CZ-NH1	-5.71	117.44	120.30
1	A	156	ASP	CB-CG-OD1	5.68	123.42	118.30
1	В	72	TYR	CB-CG-CD1	5.55	124.33	121.00
1	В	88	ARG	NE-CZ-NH2	-5.22	117.69	120.30
1	A	71	ASP	CB-CG-OD2	-5.16	113.66	118.30
1	В	129	ARG	NE-CZ-NH1	5.11	122.85	120.30
1	A	177	MET	CA-CB-CG	5.07	121.91	113.30
1	A	129	ARG	NE-CZ-NH1	5.03	122.82	120.30
1	A	57	ASP	CB-CG-OD2	5.02	122.82	118.30

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	2053	0	1993	23	0
1	В	2035	0	1974	19	0
2	A	119	0	0	0	0
2	В	66	0	0	3	0
All	All	4273	0	3967	42	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 5.

All (42) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



A 1 4	A 4 9	Interatomic	Clash
Atom-1	Atom-2	${ m distance} \; ({ m \AA})$	overlap (Å)
1:A:82:MET:CE	1:A:87:VAL:HG21	1.93	0.98
1:A:82:MET:SD	1:A:87:VAL:HG21	2.17	0.85
1:A:82:MET:HE1	1:A:87:VAL:HG21	1.63	0.80
1:A:82:MET:SD	1:A:87:VAL:CG2	2.76	0.74
1:A:161:HIS:HD2	1:A:196:GLU:OE2	1.75	0.70
1:A:15:ASN:HD22	1:A:85:ASN:HD22	1.40	0.70
1:A:79:LEU:HD13	1:A:110:ARG:CZ	2.22	0.69
1:B:70:MET:CE	1:B:83:CYS:HB3	2.29	0.62
1:B:85:ASN:HD22	1:B:88:ARG:HH12	1.50	0.59
1:A:211:TRP:CE2	1:A:216:GLY:HA2	2.37	0.59
1:B:198:VAL:HG12	1:B:199:THR:O	2.05	0.57
1:A:81:GLU:HA	1:A:110:ARG:HB2	1.87	0.56
1:A:81:GLU:N	1:A:81:GLU:OE1	2.39	0.56
1:A:104:SER:HB2	1:A:117:ASP:OD1	2.08	0.54
1:B:70:MET:HE3	1:B:83:CYS:HB3	1.90	0.54
1:B:211:TRP:CE2	1:B:216:GLY:HA2	2.43	0.53
1:B:83:CYS:HB2	2:B:346:HOH:O	2.08	0.53
1:B:70:MET:HE1	1:B:83:CYS:HB3	1.92	0.52
1:A:83:CYS:O	1:A:87:VAL:HG23	2.11	0.51
1:A:82:MET:CE	1:A:87:VAL:CG2	2.81	0.50
1:B:85:ASN:HD22	1:B:88:ARG:NH1	2.09	0.49
1:B:104:SER:HB2	1:B:117:ASP:OD1	2.13	0.48
1:A:79:LEU:HD13	1:A:110:ARG:NE	2.29	0.47
1:B:145:ILE:CD1	1:B:197:ILE:HD11	2.44	0.46
1:A:101:ASP:OD1	1:A:101:ASP:N	2.48	0.46
1:B:145:ILE:O	2:B:302:HOH:O	2.20	0.46
1:B:52:VAL:HG22	1:B:90:PHE:CE1	2.50	0.46
1:A:277:ILE:HG22	1:A:278:HIS:CD2	2.50	0.46
1:A:161:HIS:CD2	1:A:196:GLU:OE2	2.63	0.46
1:A:79:LEU:CD1	1:A:110:ARG:CZ	2.94	0.45
1:B:115:HIS:HB2	2:B:354:HOH:O	2.16	0.45
1:A:79:LEU:CD1	1:A:110:ARG:NE	2.80	0.44
1:B:145:ILE:HD11	1:B:197:ILE:HD11	2.01	0.43
1:B:85:ASN:ND2	1:B:88:ARG:HH12	2.15	0.43
1:A:81:GLU:N	1:A:81:GLU:CD	2.72	0.43
1:A:198:VAL:HG12	1:A:199:THR:O	2.19	0.42
1:A:63:VAL:HG12	1:A:113:VAL:HG23	2.02	0.42
1:B:252:VAL:HG21	1:B:263:LEU:HD11	2.01	0.42
1:A:38:CYS:O	1:A:45:GLY:HA2	2.20	0.42
1:B:64:ASP:C	1:B:64:ASP:OD1	2.59	0.41
1:B:136:ASP:HB3	1:B:156:ASP:HB3	2.02	0.40
1:B:5:ILE:HA	1:B:6:PRO:HD3	1.80	0.40



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	Perce	$\mathbf{ntiles}$
1	A	$276/283 \ (98\%)$	269 (98%)	7 (2%)	0	100	100
1	В	274/283 (97%)	258 (94%)	14 (5%)	2 (1%)	22	16
All	All	550/566 (97%)	527 (96%)	21 (4%)	2 (0%)	34	30

#### All (2) Ramachandran outliers are listed below:

Mol	Chain	${f Res}$	$\mathbf{Type}$
1	В	83	CYS
1	В	84	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	218/223 (98%)	213 (98%)	5 (2%)	50 53
1	В	216/223 (97%)	211 (98%)	5 (2%)	50 53
All	All	434/446 (97%)	424 (98%)	10 (2%)	50 53

All (10) residues with a non-rotameric sidechain are listed below:

$\mathbf{Mol}$	Chain	${f Res}$	$\mathbf{Type}$
1 A		23	GLU

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	A	26	ARG
1	A	79	LEU
1	A	101	ASP
1	A	172	SER
1	В	27	LEU
1	В	81	GLU
1	В	104	SER
1	В	172	SER
1	В	276	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	15	ASN
1	A	92	HIS
1	A	161	HIS
1	A	276	GLN
1	A	278	HIS
1	В	15	ASN
1	В	74	ASN
1	В	85	ASN
1	В	124	HIS
1	В	127	GLN
1	В	186	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 monosaccharides 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,  $95^{th}$  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	$\mathbf{OWAB}(\mathbf{\mathring{A}}^2)$	Q < 0.9
1	A	278/283 (98%)	0.09	7 (2%) 57	56	22, 36, 66, 84	0
1	В	276/283 (97%)	0.20	11 (3%) 38	37	23, 44, 74, 108	0
All	All	554/566 (97%)	0.14	18 (3%) 47	46	22, 40, 70, 108	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	83	CYS	9.4
1	В	2	ASN	4.6
1	В	146	ASN	3.6
1	A	62	THR	3.5
1	В	258	ASP	3.0
1	В	277	ILE	2.9
1	В	186	GLN	2.8
1	В	187	GLU	2.7
1	A	258	ASP	2.7
1	A	187	GLU	2.6
1	В	214	GLY	2.6
1	A	60	GLY	2.6
1	В	259	ASP	2.6
1	A	65	PRO	2.2
1	В	26	ARG	2.2
1	A	186	GLN	2.1
1	A	83	CYS	2.0
1	В	101	ASP	2.0

## 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 monosaccharides 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.

