



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:39 AM GMT

PDB ID : 2DUM  
Title : Crystal structure of hypothetical protein, PH0823  
Authors : Hosaka, T.; Kishishita, S.; Murayama, K.; Shirouzu, M.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2006-07-24  
Resolution : 2.75 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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

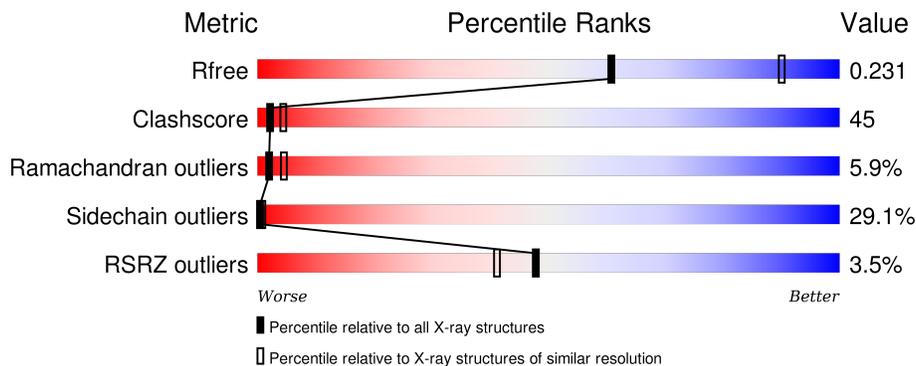
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	3340 (2.80-2.72)
Clashscore	102246	3829 (2.80-2.72)
Ramachandran outliers	100387	3767 (2.80-2.72)
Sidechain outliers	100360	3770 (2.80-2.72)
RSRZ outliers	91569	3352 (2.80-2.72)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	170	4% (Poor fit) / 22% (0 outliers) / 38% (1 outlier) / 20% (2 outliers) / 16% (3+ outliers)
1	B	170	2% (Poor fit) / 24% (0 outliers) / 37% (1 outlier) / 18% (2 outliers) / 18% (3+ outliers)
1	C	170	4% (Poor fit) / 19% (0 outliers) / 39% (1 outlier) / 22% (2 outliers) / 6% (3 outliers) / 14% (Not modelled)
1	D	170	2% (Poor fit) / 21% (0 outliers) / 40% (1 outlier) / 18% (2 outliers) / 17% (3+ outliers)

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4687 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 Hypothetical protein PH0823.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	143	1171	750	203	214	4	0	0	0
1	B	139	1143	733	197	209	4	0	0	0
1	C	146	1193	763	206	220	4	0	0	0
1	D	141	1152	738	200	210	4	0	0	0

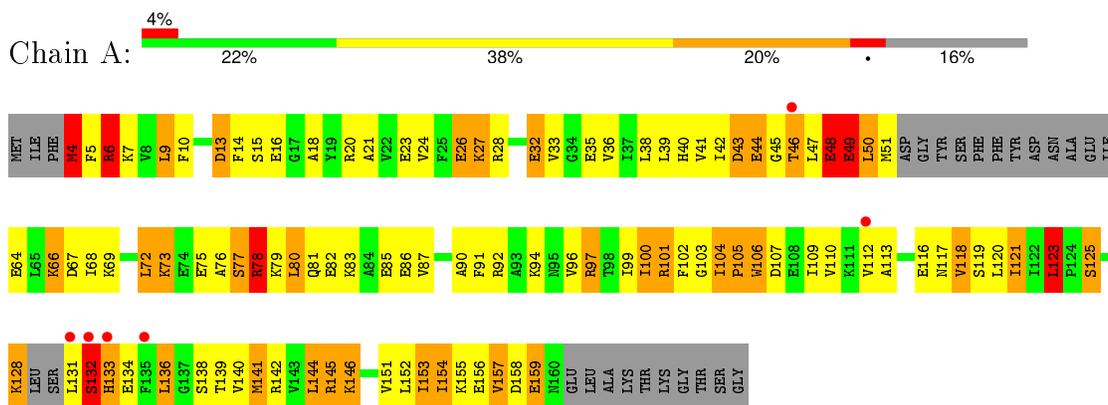
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	6	Total	O	0	0
			6	6		
2	B	8	Total	O	0	0
			8	8		
2	C	9	Total	O	0	0
			9	9		
2	D	5	Total	O	0	0
			5	5		

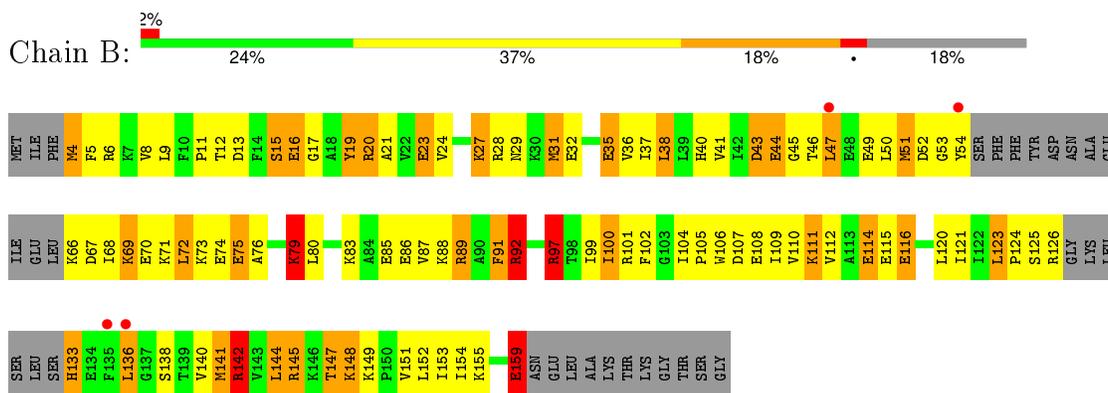
### 3 Residue-property plots [i](#)

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.

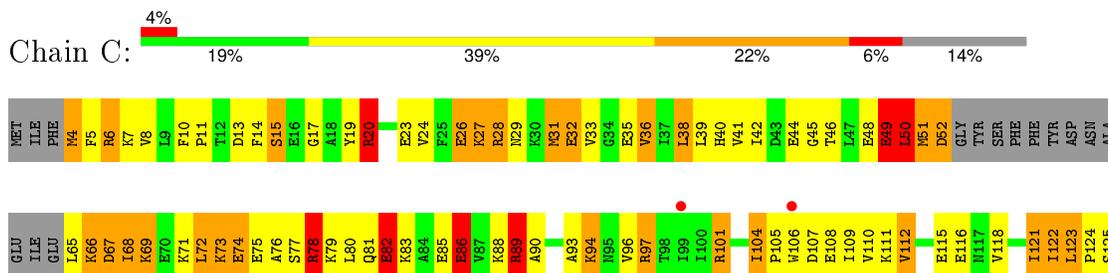
- Molecule 1: Hypothetical protein PH0823



- Molecule 1: Hypothetical protein PH0823

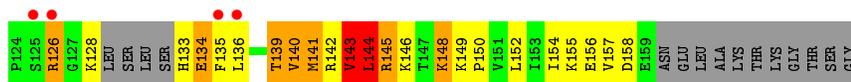
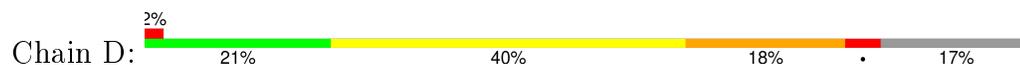


- Molecule 1: Hypothetical protein PH0823





• Molecule 1: Hypothetical protein PH0823



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	120.79Å 120.79Å 127.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.75 49.71 – 2.75	Depositor EDS
% Data completeness (in resolution range)	87.2 (40.00-2.75) 87.2 (49.71-2.75)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.53 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.216 , 0.279 0.222 , 0.231	Depositor DCC
$R_{free}$ test set	1025 reflections (4.93%)	DCC
Wilson B-factor (Å <sup>2</sup> )	55.5	Xtriage
Anisotropy	0.191	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 65.2	EDS
Estimated twinning fraction	0.000 for -h,l,k 0.000 for -l,-k,-h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Outliers	6 of 38113 reflections (0.016%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4687	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 61.28 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.3409e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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](#)

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	1.60	11/1184 (0.9%)	1.56	20/1580 (1.3%)
1	B	1.71	17/1157 (1.5%)	1.53	14/1545 (0.9%)
1	C	1.71	16/1207 (1.3%)	1.48	13/1613 (0.8%)
1	D	1.65	14/1165 (1.2%)	1.47	13/1554 (0.8%)
All	All	1.67	58/4713 (1.2%)	1.51	60/6292 (1.0%)

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	2
All	All	0	3

The worst 5 of 58 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	85	GLU	CG-CD	9.43	1.66	1.51
1	B	85	GLU	CG-CD	9.39	1.66	1.51
1	C	49	GLU	CG-CD	8.52	1.64	1.51
1	D	85	GLU	CG-CD	8.31	1.64	1.51
1	C	32	GLU	CG-CD	7.88	1.63	1.51

The worst 5 of 60 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	6	ARG	NE-CZ-NH2	-14.96	112.82	120.30
1	A	6	ARG	NE-CZ-NH1	11.39	126.00	120.30
1	C	126	ARG	NE-CZ-NH1	9.09	124.84	120.30
1	D	141	MET	CG-SD-CE	9.08	114.73	100.20
1	B	144	LEU	CB-CG-CD2	-9.07	95.58	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	48	GLU	Peptide
1	C	49	GLU	Peptide
1	C	50	LEU	Peptide

## 5.2 Too-close contacts

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	1171	0	1233	113	0
1	B	1143	0	1194	103	0
1	C	1193	0	1254	131	0
1	D	1152	0	1212	107	0
2	A	6	0	0	1	0
2	B	8	0	0	2	0
2	C	9	0	0	10	0
2	D	5	0	0	7	0
All	All	4687	0	4893	427	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 45.

The worst 5 of 427 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:C:13:ASP:OD1	1:C:15:SER:HB3	1.31	1.29
1:B:51:MET:HA	1:B:51:MET:CE	1.79	1.13
1:B:31:MET:HE2	1:C:31:MET:HE3	1.13	1.10
1:B:51:MET:HA	1:B:51:MET:HE2	1.31	1.08
1:A:48:GLU:HA	1:A:48:GLU:OE2	1.56	1.04

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	137/170 (81%)	108 (79%)	20 (15%)	9 (7%)	1	3
1	B	133/170 (78%)	108 (81%)	22 (16%)	3 (2%)	8	23
1	C	142/170 (84%)	110 (78%)	23 (16%)	9 (6%)	2	4
1	D	135/170 (79%)	104 (77%)	20 (15%)	11 (8%)	1	2
All	All	547/680 (80%)	430 (79%)	85 (16%)	32 (6%)	2	4

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	PHE
1	A	49	GLU
1	B	116	GLU
1	C	68	ILE
1	C	128	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	130/152 (86%)	86 (66%)	44 (34%)	0	0
1	B	126/152 (83%)	92 (73%)	34 (27%)	0	1
1	C	133/152 (88%)	93 (70%)	40 (30%)	0	1
1	D	127/152 (84%)	95 (75%)	32 (25%)	1	1
All	All	516/608 (85%)	366 (71%)	150 (29%)	0	1

5 of 150 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	114	GLU
1	C	28	ARG
1	D	101	ARG
1	B	126	ARG
1	B	151	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	40	HIS
1	D	133	HIS
1	C	160	ASN
1	B	40	HIS
1	C	117	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](#)

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

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	143/170 (84%)	-0.01	6 (4%) 40 33	28, 57, 101, 120	0
1	B	139/170 (81%)	-0.10	4 (2%) 55 48	27, 52, 98, 106	0
1	C	146/170 (85%)	-0.12	6 (4%) 41 34	30, 52, 96, 113	0
1	D	141/170 (82%)	-0.23	4 (2%) 56 50	29, 53, 88, 105	0
All	All	569/680 (83%)	-0.11	20 (3%) 48 41	27, 54, 98, 120	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	132	SER	5.9
1	B	135	PHE	4.8
1	A	135	PHE	4.7
1	B	54	TYR	4.2
1	A	133	HIS	3.5

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.