



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 10:23 pm BST

PDB ID : 3CYP  
Title : The crystal structure of the C-terminal domain of Helicobacter pylori MotB (residues 125-256).  
Authors : Roujeinikova, A.  
Deposited on : 2008-04-26  
Resolution : 1.60 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

<https://www.wwpdb.org/validation/2017/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
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

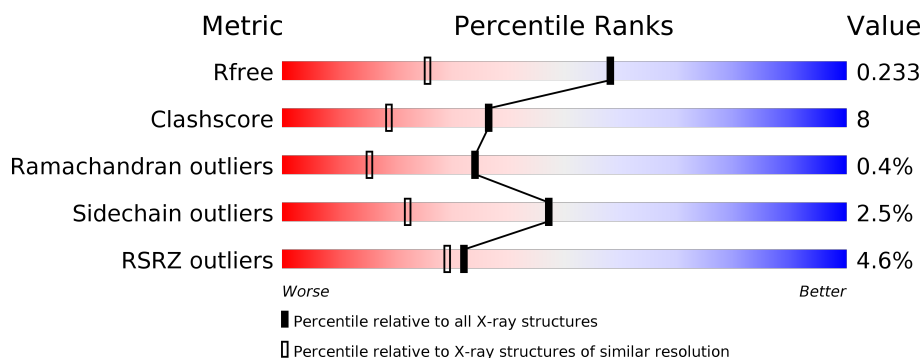
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.60 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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 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  $\leq 5\%$ . 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	B	138	<div> <div>5%</div> <div> <div></div> <div>82%</div> <div>10%</div> <div>7%</div> </div> </div>
1	C	138	<div> <div>2%</div> <div> <div></div> <div>87%</div> <div>8%</div> <div>...</div> </div> </div>
1	D	138	<div> <div>4%</div> <div> <div></div> <div>83%</div> <div>12%</div> <div>...</div> </div> </div>
1	E	138	<div> <div>7%</div> <div> <div></div> <div>81%</div> <div>13%</div> <div>...</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4968 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 Chemotaxis protein motB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	129	Total	C	N	O	S	0	3	0
			1062	670	184	203	5			
1	C	134	Total	C	N	O	S	0	7	0
			1120	707	197	212	4			
1	D	132	Total	C	N	O	S	0	2	0
			1077	678	189	206	4			
1	E	132	Total	C	N	O	S	0	4	0
			1098	691	194	209	4			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	119	GLY	-	EXPRESSION TAG	UNP P56427
B	120	ILE	-	EXPRESSION TAG	UNP P56427
B	121	ASP	-	EXPRESSION TAG	UNP P56427
B	122	PRO	-	EXPRESSION TAG	UNP P56427
B	123	PHE	-	EXPRESSION TAG	UNP P56427
B	124	THR	-	EXPRESSION TAG	UNP P56427
C	119	GLY	-	EXPRESSION TAG	UNP P56427
C	120	ILE	-	EXPRESSION TAG	UNP P56427
C	121	ASP	-	EXPRESSION TAG	UNP P56427
C	122	PRO	-	EXPRESSION TAG	UNP P56427
C	123	PHE	-	EXPRESSION TAG	UNP P56427
C	124	THR	-	EXPRESSION TAG	UNP P56427
D	119	GLY	-	EXPRESSION TAG	UNP P56427
D	120	ILE	-	EXPRESSION TAG	UNP P56427
D	121	ASP	-	EXPRESSION TAG	UNP P56427
D	122	PRO	-	EXPRESSION TAG	UNP P56427
D	123	PHE	-	EXPRESSION TAG	UNP P56427
D	124	THR	-	EXPRESSION TAG	UNP P56427
E	119	GLY	-	EXPRESSION TAG	UNP P56427
E	120	ILE	-	EXPRESSION TAG	UNP P56427
E	121	ASP	-	EXPRESSION TAG	UNP P56427

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	122	PRO	-	EXPRESSION TAG	UNP P56427
E	123	PHE	-	EXPRESSION TAG	UNP P56427
E	124	THR	-	EXPRESSION TAG	UNP P56427

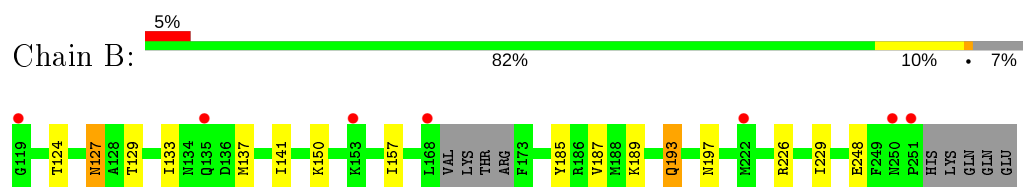
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	124	Total 124	O 124	0	0
2	C	183	Total 183	O 183	0	0
2	D	165	Total 165	O 165	0	0
2	E	139	Total 139	O 139	0	0

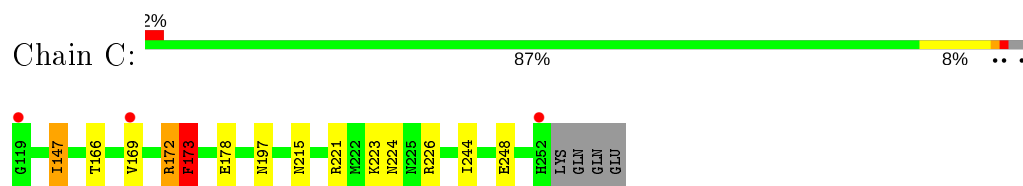
### 3 Residue-property plots [i](#)

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.

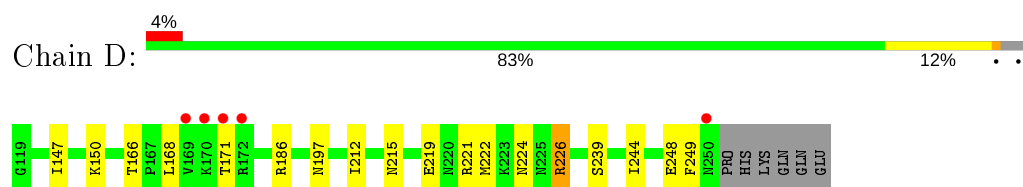
- Molecule 1: Chemotaxis protein motB



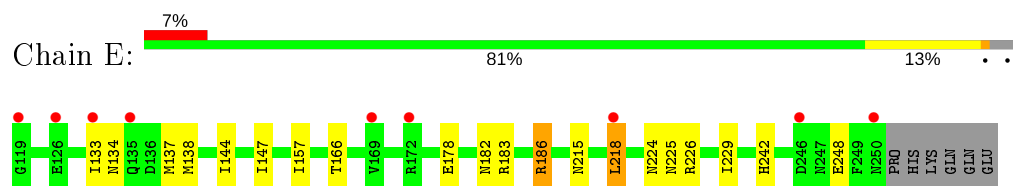
- Molecule 1: Chemotaxis protein motB



- Molecule 1: Chemotaxis protein motB



- Molecule 1: Chemotaxis protein motB



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.82Å 89.48Å 66.32Å 90.00° 112.55° 90.00°	Depositor
Resolution (Å)	50.57 – 1.60 14.99 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (50.57-1.60) 99.6 (14.99-1.60)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.21 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.185 , 0.227 0.191 , 0.233	Depositor DCC
$R_{free}$ test set	3630 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.3	Xtriage
Anisotropy	0.195	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 49.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.094 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4968	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<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, 0.333 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	B	0.62	0/1083	0.69	0/1464
1	C	0.67	0/1152	0.82	5/1560 (0.3%)
1	D	0.69	0/1098	0.75	0/1487
1	E	0.66	0/1119	0.74	0/1513
All	All	0.66	0/4452	0.75	5/6024 (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	C	0	1

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	C	173	PHE	N-CA-CB	8.09	125.17	110.60
1	C	172	ARG	C-N-CA	7.75	141.07	121.70
1	C	172	ARG	CA-C-N	6.13	130.68	117.20
1	C	172	ARG	N-CA-C	5.78	126.61	111.00
1	C	173	PHE	N-CA-C	-5.11	97.20	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	172	ARG	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	B	1062	0	1038	17	0
1	C	1120	0	1103	14	0
1	D	1077	0	1056	22	0
1	E	1098	0	1076	19	0
2	B	124	0	0	2	0
2	C	183	0	0	3	0
2	D	165	0	0	11	0
2	E	139	0	0	3	0
All	All	4968	0	4273	68	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 8.

The worst 5 of 68 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:D:222:MET:CE	2:D:524:HOH:O	2.01	1.07
1:E:166:THR:HG21	2:E:497:HOH:O	1.68	0.93
1:E:183:ARG:HH12	1:E:225:ASN:HD22	1.18	0.92
1:C:147[A]:ILE:HD11	1:C:248:GLU:HB2	1.53	0.90
1:D:222:MET:HE3	2:D:524:HOH:O	1.64	0.85

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	B	128/138 (93%)	126 (98%)	2 (2%)	0	100	100
1	C	138/138 (100%)	135 (98%)	2 (1%)	1 (1%)	22	7
1	D	132/138 (96%)	131 (99%)	1 (1%)	0	100	100
1	E	134/138 (97%)	133 (99%)	0	1 (1%)	22	7
All	All	532/552 (96%)	525 (99%)	5 (1%)	2 (0%)	34	15

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	173	PHE
1	E	134	ASN

### 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	B	120/126 (95%)	117 (98%)	3 (2%)	47	22
1	C	127/126 (101%)	123 (97%)	4 (3%)	40	15
1	D	121/126 (96%)	118 (98%)	3 (2%)	47	22
1	E	123/126 (98%)	119 (97%)	4 (3%)	38	14
All	All	491/504 (97%)	477 (97%)	14 (3%)	47	18

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

Mol	Chain	Res	Type
1	C	226	ARG
1	D	219	GLU
1	E	186[B]	ARG
1	C	173	PHE
1	E	186[A]	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	220	ASN
1	C	247	ASN
1	E	200	GLN
1	C	215	ASN
1	D	215	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 [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<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	B	129/138 (93%)	0.15	7 (5%) 25 23	13, 24, 37, 44	0
1	C	134/138 (97%)	-0.08	3 (2%) 62 60	12, 18, 35, 42	0
1	D	132/138 (95%)	0.08	5 (3%) 40 37	11, 21, 35, 44	0
1	E	132/138 (95%)	0.18	9 (6%) 17 16	13, 22, 38, 48	0
All	All	527/552 (95%)	0.08	24 (4%) 32 29	11, 21, 37, 48	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	169	VAL	10.3
1	B	251	PRO	7.2
1	C	169	VAL	4.4
1	D	172	ARG	4.0
1	B	119	GLY	3.8

### 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.