



# Full wwPDB X-ray Structure Validation Report ⓘ

May 17, 2020 – 08:41 pm BST

PDB ID : 4I61  
Title : Crystal structure of a trimeric bacterial microcompartment shell protein PduB  
Authors : Pang, A.H.; Pickersgill, R.W.  
Deposited on : 2012-11-29  
Resolution : 1.64 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](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

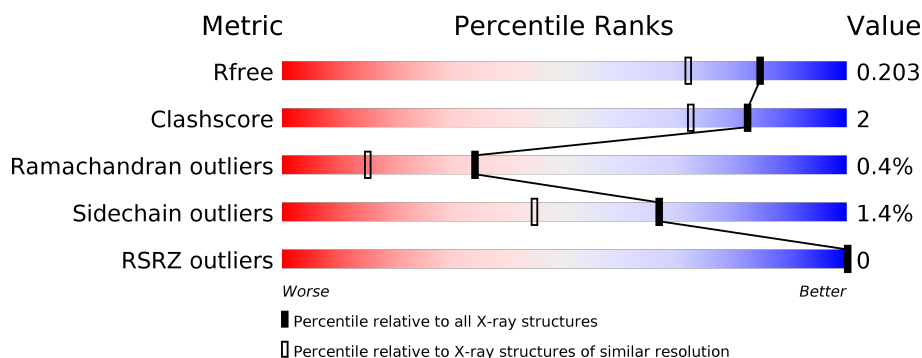
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.64 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	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

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	A	258	<div> <div>81%</div> <div>7% • 12%</div> </div>
1	B	258	<div> <div>81%</div> <div>7% 12%</div> </div>
1	C	258	<div> <div>82%</div> <div>5% • 12%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5879 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 Propanediol utilization protein PduB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	228	Total	C	N	O	S	0	4	0
			1698	1059	298	330	11			
1	B	227	Total	C	N	O	S	0	6	0
			1711	1067	304	329	11			
1	C	227	Total	C	N	O	S	0	4	0
			1690	1056	295	328	11			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP F8DQ39
A	-18	GLY	-	EXPRESSION TAG	UNP F8DQ39
A	-17	SER	-	EXPRESSION TAG	UNP F8DQ39
A	-16	SER	-	EXPRESSION TAG	UNP F8DQ39
A	-15	HIS	-	EXPRESSION TAG	UNP F8DQ39
A	-14	HIS	-	EXPRESSION TAG	UNP F8DQ39
A	-13	HIS	-	EXPRESSION TAG	UNP F8DQ39
A	-12	PRO	-	EXPRESSION TAG	UNP F8DQ39
A	-11	PRO	-	EXPRESSION TAG	UNP F8DQ39
A	-10	LYS	-	EXPRESSION TAG	UNP F8DQ39
A	-9	SER	-	EXPRESSION TAG	UNP F8DQ39
A	-8	SER	-	EXPRESSION TAG	UNP F8DQ39
A	-7	GLY	-	EXPRESSION TAG	UNP F8DQ39
A	-6	LEU	-	EXPRESSION TAG	UNP F8DQ39
A	-5	VAL	-	EXPRESSION TAG	UNP F8DQ39
A	-4	PRO	-	EXPRESSION TAG	UNP F8DQ39
A	-3	ARG	-	EXPRESSION TAG	UNP F8DQ39
A	-2	GLY	-	EXPRESSION TAG	UNP F8DQ39
A	-1	SER	-	EXPRESSION TAG	UNP F8DQ39
A	0	HIS	-	EXPRESSION TAG	UNP F8DQ39
B	-19	MET	-	EXPRESSION TAG	UNP F8DQ39
B	-18	GLY	-	EXPRESSION TAG	UNP F8DQ39
B	-17	SER	-	EXPRESSION TAG	UNP F8DQ39

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	SER	-	EXPRESSION TAG	UNP F8DQ39
B	-15	HIS	-	EXPRESSION TAG	UNP F8DQ39
B	-14	HIS	-	EXPRESSION TAG	UNP F8DQ39
B	-13	HIS	-	EXPRESSION TAG	UNP F8DQ39
B	-12	PRO	-	EXPRESSION TAG	UNP F8DQ39
B	-11	PRO	-	EXPRESSION TAG	UNP F8DQ39
B	-10	LYS	-	EXPRESSION TAG	UNP F8DQ39
B	-9	SER	-	EXPRESSION TAG	UNP F8DQ39
B	-8	SER	-	EXPRESSION TAG	UNP F8DQ39
B	-7	GLY	-	EXPRESSION TAG	UNP F8DQ39
B	-6	LEU	-	EXPRESSION TAG	UNP F8DQ39
B	-5	VAL	-	EXPRESSION TAG	UNP F8DQ39
B	-4	PRO	-	EXPRESSION TAG	UNP F8DQ39
B	-3	ARG	-	EXPRESSION TAG	UNP F8DQ39
B	-2	GLY	-	EXPRESSION TAG	UNP F8DQ39
B	-1	SER	-	EXPRESSION TAG	UNP F8DQ39
B	0	HIS	-	EXPRESSION TAG	UNP F8DQ39
C	-19	MET	-	EXPRESSION TAG	UNP F8DQ39
C	-18	GLY	-	EXPRESSION TAG	UNP F8DQ39
C	-17	SER	-	EXPRESSION TAG	UNP F8DQ39
C	-16	SER	-	EXPRESSION TAG	UNP F8DQ39
C	-15	HIS	-	EXPRESSION TAG	UNP F8DQ39
C	-14	HIS	-	EXPRESSION TAG	UNP F8DQ39
C	-13	HIS	-	EXPRESSION TAG	UNP F8DQ39
C	-12	PRO	-	EXPRESSION TAG	UNP F8DQ39
C	-11	PRO	-	EXPRESSION TAG	UNP F8DQ39
C	-10	LYS	-	EXPRESSION TAG	UNP F8DQ39
C	-9	SER	-	EXPRESSION TAG	UNP F8DQ39
C	-8	SER	-	EXPRESSION TAG	UNP F8DQ39
C	-7	GLY	-	EXPRESSION TAG	UNP F8DQ39
C	-6	LEU	-	EXPRESSION TAG	UNP F8DQ39
C	-5	VAL	-	EXPRESSION TAG	UNP F8DQ39
C	-4	PRO	-	EXPRESSION TAG	UNP F8DQ39
C	-3	ARG	-	EXPRESSION TAG	UNP F8DQ39
C	-2	GLY	-	EXPRESSION TAG	UNP F8DQ39
C	-1	SER	-	EXPRESSION TAG	UNP F8DQ39
C	0	HIS	-	EXPRESSION TAG	UNP F8DQ39

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	246	Total O 246 246	0	0

*Continued on next page...*


*Continued from previous page...*

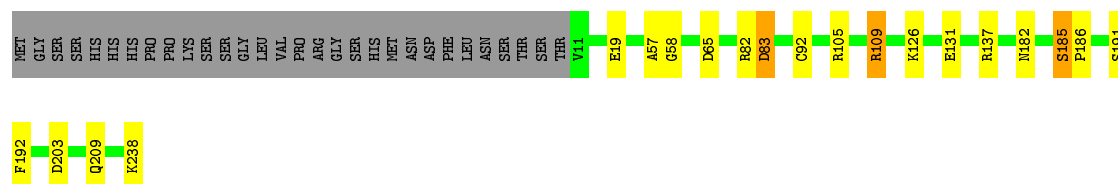
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	270	Total 270	O 270	0	0
2	C	264	Total 264	O 264	0	0

### 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.

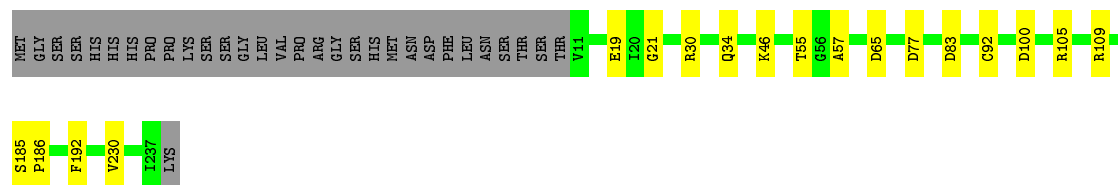
- Molecule 1: Propanediol utilization protein PduB

Chain A: 




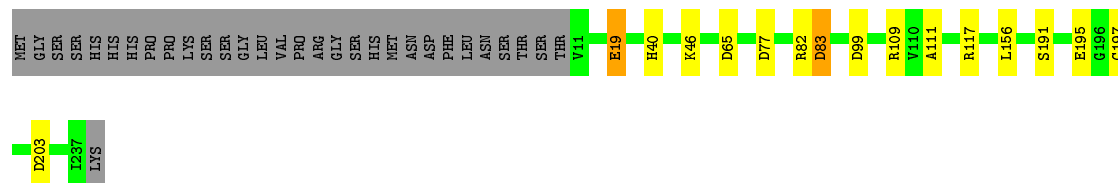
- Molecule 1: Propanediol utilization protein PduB

Chain B: 



- Molecule 1: Propanediol utilization protein PduB

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.01Å 119.55Å 147.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	59.77 – 1.64 59.77 – 1.64	Depositor EDS
% Data completeness (in resolution range)	99.8 (59.77-1.64) 99.8 (59.77-1.64)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.57 (at 1.64Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.153 , 0.200 0.157 , 0.203	Depositor DCC
$R_{free}$ test set	3749 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	9.1	Xtriage
Anisotropy	0.162	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 44.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.149 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.186 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5879	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	11.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.33% 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

### 5.1 Standard geometry

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	1.11	2/1726 (0.1%)	0.98	5/2333 (0.2%)
1	B	1.18	0/1738	1.03	4/2347 (0.2%)
1	C	1.18	1/1717 (0.1%)	1.00	5/2320 (0.2%)
All	All	1.16	3/5181 (0.1%)	1.00	14/7000 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	185	SER	CB-OG	6.33	1.50	1.42
1	A	209	GLN	CG-CD	6.23	1.65	1.51
1	C	111	ALA	CA-CB	5.29	1.63	1.52

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	109	ARG	NE-CZ-NH2	-6.94	116.83	120.30
1	B	109	ARG	NE-CZ-NH2	-6.89	116.86	120.30
1	C	83	ASP	CB-CG-OD2	6.77	124.39	118.30
1	C	82	ARG	NE-CZ-NH2	-6.28	117.16	120.30
1	B	105	ARG	NE-CZ-NH1	-5.68	117.46	120.30
1	A	109	ARG	NE-CZ-NH2	-5.37	117.62	120.30
1	C	203	ASP	CB-CG-OD2	-5.36	113.48	118.30
1	A	137	ARG	NE-CZ-NH1	-5.33	117.64	120.30
1	C	77	ASP	CB-CG-OD2	-5.31	113.52	118.30
1	A	83	ASP	CB-CG-OD2	5.20	122.98	118.30
1	B	109	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	B	100	ASP	CB-CG-OD1	5.16	122.94	118.30
1	A	203	ASP	CB-CG-OD1	5.14	122.92	118.30
1	A	82	ARG	NE-CZ-NH2	-5.12	117.74	120.30

There are no chirality outliers.

There are no planarity outliers.



## 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	1698	0	1679	9	0
1	B	1711	0	1698	8	0
1	C	1690	0	1674	10	0
2	A	246	0	0	2	0
2	B	270	0	0	3	0
2	C	264	0	0	5	0
All	All	5879	0	5051	25	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 2.

All (25) 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:156[B]:LEU:HD11	1:C:197:CYS:SG	2.04	0.97
1:C:156[B]:LEU:CD1	1:C:197:CYS:SG	2.60	0.90
1:B:46:LYS:HG2	2:B:540:HOH:O	1.87	0.75
1:C:40:HIS:HD2	2:C:309:HOH:O	1.73	0.70
1:C:156[B]:LEU:HD12	1:C:197:CYS:SG	2.39	0.62
1:B:77:ASP:HA	2:B:565:HOH:O	2.02	0.59
1:B:30[A]:ARG:NH1	1:B:46:LYS:HD2	2.18	0.57
1:C:19:GLU:HG2	2:C:536:HOH:O	2.09	0.51
1:B:185:SER:HB2	1:B:186:PRO:CD	2.42	0.49
1:C:46:LYS:HE3	1:C:99:ASP:O	2.14	0.48
1:C:46:LYS:HG2	2:C:404:HOH:O	2.14	0.47
1:A:57:ALA:HA	1:A:92:CYS:HB2	1.98	0.46
1:A:185:SER:HB2	1:A:186:PRO:CD	2.47	0.44
1:A:192:PHE:CG	1:C:191:SER:HB3	2.54	0.43
1:A:182[B]:ASN:ND2	2:A:359:HOH:O	2.51	0.43
1:C:117:ARG:HD2	2:C:450:HOH:O	2.18	0.43
1:A:105:ARG:O	1:A:109:ARG:HG3	2.18	0.43
1:A:191:SER:HB3	1:B:192:PHE:CG	2.53	0.43
1:A:58:GLY:HA3	1:A:131:GLU:OE2	2.20	0.42
1:B:57:ALA:HA	1:B:92:CYS:HB2	2.02	0.42
1:B:34:GLN:NE2	2:B:423:HOH:O	2.41	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:19:GLU:HG2	2:A:442:HOH:O	2.20	0.41
1:B:21:GLY:HA2	1:B:55:THR:HB	2.02	0.41
1:A:126:LYS:HB2	1:A:126:LYS:HE3	1.84	0.40
1:C:195:GLU:HB2	2:C:554:HOH:O	2.22	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	Percentiles	
1	A	230/258 (89%)	223 (97%)	6 (3%)	1 (0%)	34	15
1	B	230/258 (89%)	221 (96%)	8 (4%)	1 (0%)	34	15
1	C	228/258 (88%)	220 (96%)	7 (3%)	1 (0%)	34	15
All	All	688/774 (89%)	664 (96%)	21 (3%)	3 (0%)	34	15

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	83	ASP
1	A	83	ASP
1	C	83	ASP

### 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	177/200 (88%)	175 (99%)	2 (1%)	73	55
1	B	178/200 (89%)	174 (98%)	4 (2%)	52	25
1	C	176/200 (88%)	174 (99%)	2 (1%)	73	55
All	All	531/600 (88%)	523 (98%)	8 (2%)	67	42

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	65	ASP
1	A	238	LYS
1	B	19	GLU
1	B	65	ASP
1	B	230[A]	VAL
1	B	230[B]	VAL
1	C	19	GLU
1	C	65	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	34	GLN
1	C	34	GLN
1	C	40	HIS
1	C	124	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

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, 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	228/258 (88%)	-0.73	0 100 100	3, 9, 19, 31	0
1	B	227/258 (87%)	-0.77	0 100 100	4, 8, 16, 24	0
1	C	227/258 (87%)	-0.76	0 100 100	4, 8, 17, 24	0
All	All	682/774 (88%)	-0.75	0 100 100	3, 8, 17, 31	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.