



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 26, 2017 – 08:57 PM EDT

PDB ID : 5KHS  
Title : Crystal structures of the Burkholderia multivorans hopanoid transporter HpnN  
Authors : Su, C.-C.; Yu, E.W.  
Deposited on : 2016-06-15  
Resolution : 3.76 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  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029077  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20029077

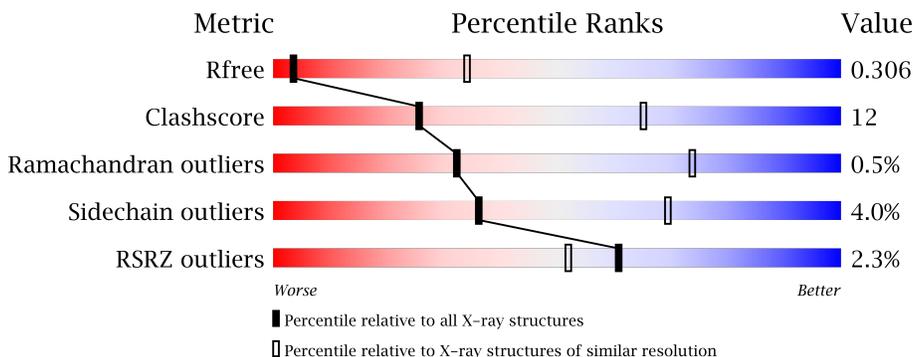
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.76 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))
$R_{free}$	100719	1423 (4.02-3.50)
Clashscore	112137	1087 (4.00-3.52)
Ramachandran outliers	110173	1047 (4.00-3.52)
Sidechain outliers	110143	1041 (4.00-3.52)
RSRZ outliers	101464	1011 (4.00-3.52)

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	883	
1	B	883	

## 2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 12679 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 Putative RND superfamily efflux pump membrane protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	853	6342	4094	1087	1143	18	0	0	0
1	A	852	6337	4091	1086	1142	18	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	360	GLY	ASP	engineered mutation	UNP A0A0H3KP92
B	362	ALA	ARG	engineered mutation	UNP A0A0H3KP92
B	365	ALA	HIS	engineered mutation	UNP A0A0H3KP92
B	878	HIS	-	expression tag	UNP A0A0H3KP92
B	879	HIS	-	expression tag	UNP A0A0H3KP92
B	880	HIS	-	expression tag	UNP A0A0H3KP92
B	881	HIS	-	expression tag	UNP A0A0H3KP92
B	882	HIS	-	expression tag	UNP A0A0H3KP92
B	883	HIS	-	expression tag	UNP A0A0H3KP92
A	360	GLY	ASP	engineered mutation	UNP A0A0H3KP92
A	362	ALA	ARG	engineered mutation	UNP A0A0H3KP92
A	365	ALA	HIS	engineered mutation	UNP A0A0H3KP92
A	878	HIS	-	expression tag	UNP A0A0H3KP92
A	879	HIS	-	expression tag	UNP A0A0H3KP92
A	880	HIS	-	expression tag	UNP A0A0H3KP92
A	881	HIS	-	expression tag	UNP A0A0H3KP92
A	882	HIS	-	expression tag	UNP A0A0H3KP92
A	883	HIS	-	expression tag	UNP A0A0H3KP92





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	112.16Å 143.37Å 112.66Å 90.00° 114.10° 90.00°	Depositor
Resolution (Å)	83.57 - 3.76 83.57 - 3.76	Depositor EDS
% Data completeness (in resolution range)	83.0 (83.57-3.76) 82.8 (83.57-3.76)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.80 (at 3.78Å)	Xtrriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.263 , 0.307 0.260 , 0.306	Depositor DCC
$R_{free}$ test set	1391 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	10.9	Xtrriage
Anisotropy	1.444	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , -7.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	0.387 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.82	EDS
Total number of atoms	12679	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.19% 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	A	0.25	0/6472	0.45	0/8852
1	B	0.26	1/6477 (0.0%)	0.44	1/8859 (0.0%)
All	All	0.25	1/12949 (0.0%)	0.45	1/17711 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	485	LYS	CE-NZ	-5.12	1.36	1.49

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	344	ASP	CB-CG-OD2	6.05	123.75	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	6337	0	6566	164	1
1	B	6342	0	6571	151	1
All	All	12679	0	13137	308	1

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 12.

The worst 5 of 308 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:A:269:GLU:HG2	1:A:485:LYS:HE3	1.48	0.95
1:B:127:PRO:HA	1:B:549:LYS:HE2	1.60	0.82
1:A:751:PRO:HG3	1:A:859:LEU:HD13	1.65	0.78
1:A:395:THR:HG21	1:A:719:ALA:HA	1.65	0.76
1:A:751:PRO:HB2	1:A:789:LYS:HE3	1.67	0.74

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:590:TYR:OH	1:A:451:ASP:OD2[1_554]	2.16	0.04

## 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	848/883 (96%)	787 (93%)	58 (7%)	3 (0%)	38 77
1	B	849/883 (96%)	791 (93%)	52 (6%)	6 (1%)	25 68
All	All	1697/1766 (96%)	1578 (93%)	110 (6%)	9 (0%)	32 74

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	297	ARG
1	B	811	ALA
1	B	859	LEU
1	B	771	ASN
1	A	214	PRO

### 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	662/690 (96%)	635 (96%)	27 (4%)	35 69
1	B	662/690 (96%)	636 (96%)	26 (4%)	37 70
All	All	1324/1380 (96%)	1271 (96%)	53 (4%)	36 69

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	742	PHE
1	A	120	ASN
1	A	657	LEU
1	B	747	ARG
1	A	52	GLU

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	222	GLN
1	A	150	ASN
1	A	480	ASN
1	A	683	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	A	852/883 (96%)	-0.04	21 (2%) 58 47	6, 46, 81, 112	0
1	B	853/883 (96%)	-0.01	18 (2%) 64 55	6, 44, 82, 112	0
All	All	1705/1766 (96%)	-0.02	39 (2%) 61 52	6, 45, 82, 112	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	207	ASP	5.8
1	A	170	THR	4.8
1	A	190	ASP	4.4
1	B	174	LYS	3.6
1	A	229	ALA	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 [i](#)

There are no such residues in this entry.