



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 27, 2017 – 09:31 AM EDT

PDB ID : 1VHY  
Title : Crystal structure of Haemophilus influenzae protein HI0303, Pfam DUF558  
Authors : Structural GenomiX; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : unknown  
Resolution : 1.90 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-20030345
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-20030345

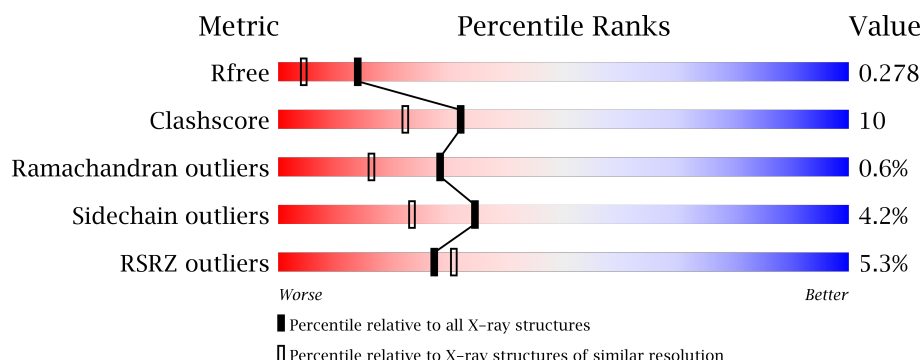
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

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	257	 10% 77% 14% • 6%
1	B	257	 10% 70% 19% •• 8%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 4004 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 HI0303.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	241	Total	C	N	O	S	Se	0	3	0
			1852	1169	323	349	6	5			
1	B	237	Total	C	N	O	S	Se	0	3	0
			1814	1147	317	339	7	4			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MSE	-	cloning artifact	UNP P44627
A	0	SER	-	cloning artifact	UNP P44627
A	1	LEU	-	cloning artifact	UNP P44627
A	34	MSE	MET	modified residue	UNP P44627
A	92	MSE	MET	modified residue	UNP P44627
A	124	MSE	MET	modified residue	UNP P44627
A	154	MSE	MET	modified residue	UNP P44627
A	246	GLU	-	cloning artifact	UNP P44627
A	247	GLY	-	cloning artifact	UNP P44627
A	248	GLY	-	cloning artifact	UNP P44627
A	249	SER	-	cloning artifact	UNP P44627
A	250	HIS	-	cloning artifact	UNP P44627
A	251	HIS	-	cloning artifact	UNP P44627
A	252	HIS	-	cloning artifact	UNP P44627
A	253	HIS	-	cloning artifact	UNP P44627
A	254	HIS	-	cloning artifact	UNP P44627
A	255	HIS	-	cloning artifact	UNP P44627
B	-1	MSE	-	cloning artifact	UNP P44627
B	0	SER	-	cloning artifact	UNP P44627
B	1	LEU	-	cloning artifact	UNP P44627
B	34	MSE	MET	modified residue	UNP P44627
B	92	MSE	MET	modified residue	UNP P44627
B	124	MSE	MET	modified residue	UNP P44627
B	154	MSE	MET	modified residue	UNP P44627
B	246	GLU	-	cloning artifact	UNP P44627

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	247	GLY	-	cloning artifact	UNP P44627
B	248	GLY	-	cloning artifact	UNP P44627
B	249	SER	-	cloning artifact	UNP P44627
B	250	HIS	-	cloning artifact	UNP P44627
B	251	HIS	-	cloning artifact	UNP P44627
B	252	HIS	-	cloning artifact	UNP P44627
B	253	HIS	-	cloning artifact	UNP P44627
B	254	HIS	-	cloning artifact	UNP P44627
B	255	HIS	-	cloning artifact	UNP P44627

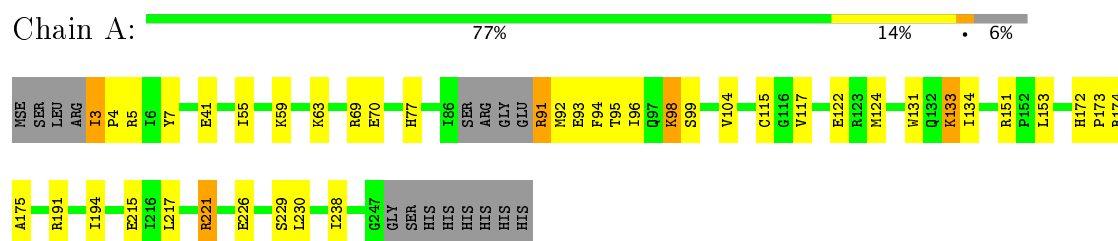
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	205	Total O 205 205	0	0
2	B	133	Total O 133 133	0	0

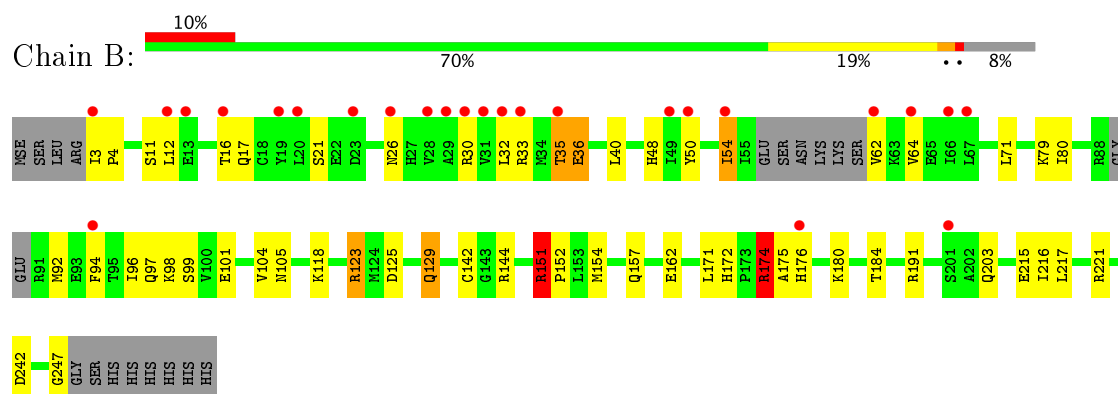
### 3 Residue-property plots [i](#)

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: Hypothetical protein HI0303



#### • Molecule 1: Hypothetical protein HI0303



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.90Å 76.41Å 62.11Å 90.00° 108.43° 90.00°	Depositor
Resolution (Å)	38.07 – 1.90 38.21 – 1.90	Depositor EDS
% Data completeness (in resolution range)	(Not available) (38.07-1.90) 99.7 (38.21-1.90)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.87 (at 1.89Å)	Xtriage
Refinement program	REFMAC 4.0	Depositor
R, $R_{free}$	0.229 , 0.300 0.217 , 0.278	Depositor DCC
$R_{free}$ test set	2026 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.4	Xtriage
Anisotropy	0.688	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 61.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4004	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.05% 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	0.73	0/1888	1.19	10/2548 (0.4%)
1	B	0.65	0/1849	1.13	9/2496 (0.4%)
All	All	0.69	0/3737	1.16	19/5044 (0.4%)

There are no bond length outliers.

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	221	ARG	NE-CZ-NH1	-15.98	112.31	120.30
1	A	191	ARG	NE-CZ-NH2	13.40	127.00	120.30
1	B	151	ARG	CD-NE-CZ	12.54	141.16	123.60
1	B	191	ARG	NE-CZ-NH2	11.88	126.24	120.30
1	B	221	ARG	NE-CZ-NH1	10.11	125.36	120.30

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	1852	0	1852	35	0
1	B	1814	0	1813	44	0
2	A	205	0	0	5	0
2	B	133	0	0	4	0
All	All	4004	0	3665	74	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 10.

The worst 5 of 74 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:B:157:GLN:HE22	1:B:203:GLN:HG3	1.30	0.93
1:B:80:ILE:H	1:B:105:ASN:HD22	1.23	0.85
1:A:124:MSE:HG3	1:A:153:LEU:HD22	1.65	0.78
1:B:157:GLN:NE2	1:B:203:GLN:HG3	1.99	0.76
1:B:174:ARG:HG2	1:B:174:ARG:HH11	1.54	0.73

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	240/257 (93%)	236 (98%)	4 (2%)	0	100	100
1	B	234/257 (91%)	222 (95%)	9 (4%)	3 (1%)	14	4
All	All	474/514 (92%)	458 (97%)	13 (3%)	3 (1%)	28	16

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	36	GLU
1	B	54	ILE
1	B	21	SER

### 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	197/213 (92%)	191 (97%)	6 (3%)	46	37
1	B	192/213 (90%)	182 (95%)	10 (5%)	27	16
All	All	389/426 (91%)	373 (96%)	16 (4%)	34	24

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	35	THR
1	B	40	LEU
1	B	142	CYS
1	B	16	THR
1	B	151	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	105	ASN
1	B	211	GLN
1	B	157	GLN
1	A	27	HIS
1	B	129	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 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	237/257 (92%)	-0.02	0 <span>100</span> <span>100</span>	15, 26, 43, 57	0
1	B	233/257 (90%)	0.61	25 (10%) <span>7</span> <span>7</span>	16, 33, 66, 78	0
All	All	470/514 (91%)	0.30	25 (5%) <span>27</span> <span>31</span>	15, 29, 55, 78	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	54	ILE	5.4
1	B	29	ALA	5.3
1	B	62	VAL	5.0
1	B	31	VAL	4.9
1	B	33	ARG	4.1

### 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.