wwPDB X-ray Structure Validation Summary Report (i)

May 16, 2020 – 07:58 am BST

PDB ID	:	4ZH0
Title	:	Structure of Helicobacter pylori adhesin BabA determined by SeMet SAD
Authors	:	Howard, T.D.; Hage, N.; Phillips, C.; Brassington, C.A.; Debreczeni, J.; Overman, R.; Gellert, P.; Stolnik, S.; Winkler, G.S.; Falcone, F.H.
Deposited on	:	2015-04-24
Resolution	:	1.91 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

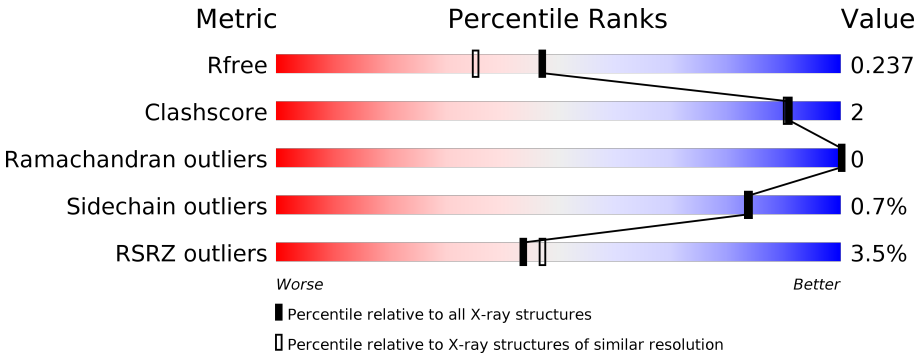
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.91 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	7937 (1.94-1.90)
Clashscore	141614	8644 (1.94-1.90)
Ramachandran outliers	138981	8530 (1.94-1.90)
Sidechain outliers	138945	8530 (1.94-1.90)
RSRZ outliers	127900	7793 (1.94-1.90)

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 >=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 <=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	543	3% 84% 5% 10%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3795 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 Outer membrane protein-adhesin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	486	3662	2258	638	754	8	4	0	0	0

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	528	LYS	-	expression tag	UNP Q9ZKV2
A	529	LYS	-	expression tag	UNP Q9ZKV2
A	530	LYS	-	expression tag	UNP Q9ZKV2
A	531	LYS	-	expression tag	UNP Q9ZKV2
A	532	LYS	-	expression tag	UNP Q9ZKV2
A	533	LYS	-	expression tag	UNP Q9ZKV2
A	534	GLY	-	expression tag	UNP Q9ZKV2
A	535	SER	-	expression tag	UNP Q9ZKV2
A	536	GLU	-	expression tag	UNP Q9ZKV2
A	537	GLN	-	expression tag	UNP Q9ZKV2
A	538	LYS	-	expression tag	UNP Q9ZKV2
A	539	LEU	-	expression tag	UNP Q9ZKV2
A	540	ILE	-	expression tag	UNP Q9ZKV2
A	541	SER	-	expression tag	UNP Q9ZKV2
A	542	GLU	-	expression tag	UNP Q9ZKV2
A	543	GLU	-	expression tag	UNP Q9ZKV2
A	544	ASP	-	expression tag	UNP Q9ZKV2
A	545	LEU	-	expression tag	UNP Q9ZKV2
A	546	SER	-	expression tag	UNP Q9ZKV2
A	547	HIS	-	expression tag	UNP Q9ZKV2
A	548	HIS	-	expression tag	UNP Q9ZKV2
A	549	HIS	-	expression tag	UNP Q9ZKV2
A	550	HIS	-	expression tag	UNP Q9ZKV2
A	551	HIS	-	expression tag	UNP Q9ZKV2
A	552	HIS	-	expression tag	UNP Q9ZKV2

- Molecule 2 is water.

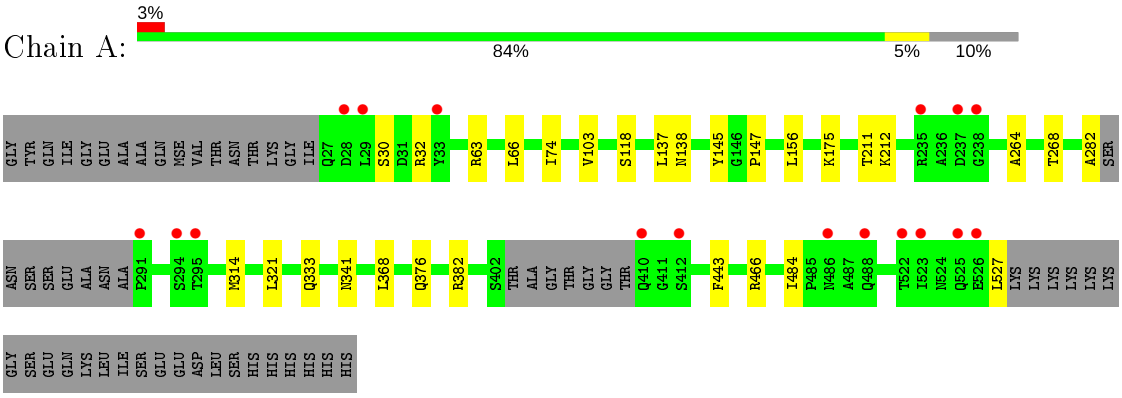
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	133	Total	O	0	0
			133	133		

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: Outer membrane protein-adhesin

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	60.83Å 93.04Å 96.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.12 – 1.91 48.46 – 1.91	Depositor EDS
% Data completeness (in resolution range)	96.7 (67.12-1.91) 96.7 (48.46-1.91)	Depositor EDS
R_{merge}	0.24	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.188 , 0.230 0.200 , 0.237	Depositor DCC
R_{free} test set	2071 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.226	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 41.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.017 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3795	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.17% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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	0.89	0/3713	0.88	7/5038 (0.1%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	466	ARG	NE-CZ-NH1	8.80	124.70	120.30
1	A	63	ARG	NE-CZ-NH1	8.60	124.60	120.30
1	A	341	ASN	CB-CA-C	6.26	122.93	110.40
1	A	466	ARG	NE-CZ-NH2	-6.02	117.29	120.30
1	A	63	ARG	NE-CZ-NH2	-5.60	117.50	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	3662	0	3561	12	0
2	A	133	0	0	0	0
All	All	3795	0	3561	12	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.

The worst 5 of 12 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:74:ILE:HG22	1:A:175:LYS:HB3	1.55	0.86
1:A:103:VAL:HG11	1:A:282:ALA:HB2	1.88	0.54
1:A:376:GLN:HG2	1:A:382:ARG:CZ	2.40	0.52
1:A:118:SER:H	1:A:138:ASN:HD21	1.58	0.49
1:A:264:ALA:O	1:A:268:THR:HG23	2.13	0.48

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	480/543 (88%)	462 (96%)	18 (4%)	0	100 100

There are no Ramachandran outliers to report.

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	407/447 (91%)	404 (99%)	3 (1%)	84 83

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	30	SER
1	A	145	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	527	LEU

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	A	138	ASN
1	A	491	GLN
1	A	369	ASN
1	A	123	ASN
1	A	178	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

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

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	482/543 (88%)	0.10	17 (3%) 44 47	19, 30, 59, 85	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	235	ARG	5.0
1	A	237	ASP	4.1
1	A	29	LEU	3.9
1	A	410	GLN	3.9
1	A	525	GLN	3.6

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.