Full wwPDB X-ray Structure Validation Report ⓘ

Nov 13, 2024 – 12:22 pm GMT

PDB ID	:	8R4P
Title	:	Structure of BabA from Helicobacter pylori strain 17875
Authors	:	Persson, K.; Boren, T.; Bugaytsova, J.
Deposited on	:	2023-11-14
Resolution	:	2.00 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

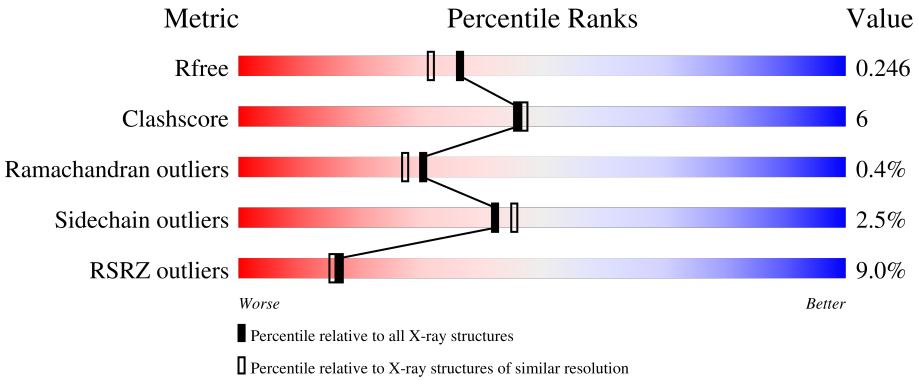
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.00 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}	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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 of 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	571	8% 71% 12% • 16%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3843 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 Adhesin binding fucosylated histo-blood group antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	477	3618	2232	623	751	12	0	2	0

There are 47 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	initiating methionine	UNP O52269
A	-20	LYS	-	expression tag	UNP O52269
A	-19	TYR	-	expression tag	UNP O52269
A	-18	LEU	-	expression tag	UNP O52269
A	-17	LEU	-	expression tag	UNP O52269
A	-16	PRO	-	expression tag	UNP O52269
A	-15	THR	-	expression tag	UNP O52269
A	-14	ALA	-	expression tag	UNP O52269
A	-13	ALA	-	expression tag	UNP O52269
A	-12	ALA	-	expression tag	UNP O52269
A	-11	GLY	-	expression tag	UNP O52269
A	-10	LEU	-	expression tag	UNP O52269
A	-9	LEU	-	expression tag	UNP O52269
A	-8	LEU	-	expression tag	UNP O52269
A	-7	LEU	-	expression tag	UNP O52269
A	-6	ALA	-	expression tag	UNP O52269
A	-5	ALA	-	expression tag	UNP O52269
A	-4	GLN	-	expression tag	UNP O52269
A	-3	PRO	-	expression tag	UNP O52269
A	-2	ALA	-	expression tag	UNP O52269
A	-1	MET	-	expression tag	UNP O52269
A	0	ALA	-	expression tag	UNP O52269
A	525	LYS	-	expression tag	UNP O52269
A	526	LYS	-	expression tag	UNP O52269
A	527	LYS	-	expression tag	UNP O52269
A	528	LYS	-	expression tag	UNP O52269
A	529	LYS	-	expression tag	UNP O52269

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	530	LYS	-	expression tag	UNP O52269
A	531	GLY	-	expression tag	UNP O52269
A	532	SER	-	expression tag	UNP O52269
A	533	GLU	-	expression tag	UNP O52269
A	534	GLN	-	expression tag	UNP O52269
A	535	LYS	-	expression tag	UNP O52269
A	536	LEU	-	expression tag	UNP O52269
A	537	ILE	-	expression tag	UNP O52269
A	538	SER	-	expression tag	UNP O52269
A	539	GLU	-	expression tag	UNP O52269
A	540	GLU	-	expression tag	UNP O52269
A	541	ASP	-	expression tag	UNP O52269
A	542	LEU	-	expression tag	UNP O52269
A	543	SER	-	expression tag	UNP O52269
A	544	HIS	-	expression tag	UNP O52269
A	545	HIS	-	expression tag	UNP O52269
A	546	HIS	-	expression tag	UNP O52269
A	547	HIS	-	expression tag	UNP O52269
A	548	HIS	-	expression tag	UNP O52269
A	549	HIS	-	expression tag	UNP O52269

- Molecule 2 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	10	Total	Cd	0	0
			10	10		

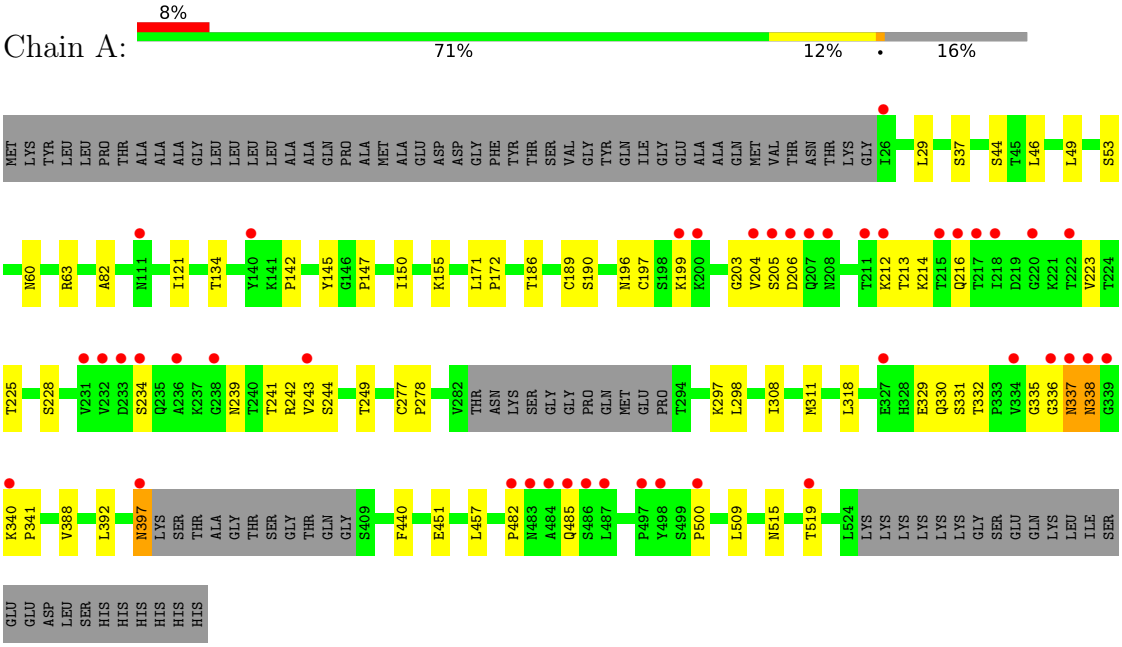
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	205	Total	O	0	10
			215	215		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Adhesin binding fucosylated histo-blood group antigen



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	55.61Å 96.95Å 97.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.64 – 2.00 36.64 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.5 (36.64-2.00) 91.8 (36.64-2.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 2.00Å)	Xtriage
Refinement program	PHENIX (1.17.1_3660: ???)	Depositor
R, R_{free}	0.211 , 0.247 0.211 , 0.246	Depositor DCC
R_{free} test set	34013 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	30.8	Xtriage
Anisotropy	0.151	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.008 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3843	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.19% 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

Bond lengths and bond angles in the following residue types are not validated in this section: CD

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	0.46	1/3671 (0.0%)	0.56	3/4992 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	451	GLU	CD-OE1	-6.37	1.18	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	337	ASN	CB-CA-C	6.85	124.10	110.40
1	A	397	ASN	CB-CA-C	6.57	123.53	110.40
1	A	485	GLN	CB-CA-C	5.15	120.70	110.40

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	3618	0	3519	41	0
2	A	10	0	0	0	0
3	A	215	0	0	1	0
All	All	3843	0	3519	41	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 6.

All (41) 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:239:ASN:OD1	1:A:241:THR:HG22	1.77	0.84
1:A:338:ASN:OD1	1:A:338:ASN:N	2.11	0.76
1:A:239:ASN:OD1	1:A:241:THR:CG2	2.33	0.75
1:A:60:ASN:OD1	1:A:63[B]:ARG:NH1	2.25	0.70
1:A:199:LYS:O	1:A:203:GLY:N	2.25	0.69
1:A:63[B]:ARG:NH2	1:A:332:THR:O	2.28	0.67
1:A:340:LYS:HB3	1:A:341:PRO:HD2	1.77	0.66
1:A:341:PRO:HG3	1:A:500:PRO:HG2	1.80	0.63
1:A:243:VAL:O	1:A:243:VAL:HG12	2.01	0.59
1:A:186:THR:HG22	1:A:249:THR:HG22	1.86	0.58
1:A:29:LEU:HD11	1:A:482:PRO:HG3	1.86	0.58
1:A:278:PRO:HD2	1:A:298:LEU:HB2	1.86	0.58
1:A:82:ALA:HB1	1:A:318:LEU:HD21	1.87	0.55
1:A:234:SER:HB3	1:A:244:SER:H	1.73	0.53
1:A:29:LEU:HD11	1:A:482:PRO:CD	2.40	0.52
1:A:189:CYS:SG	1:A:197:CYS:N	2.83	0.52
1:A:142:PRO:HA	1:A:147:PRO:HG2	1.92	0.51
1:A:277:CYS:SG	1:A:297:LYS:HB2	2.50	0.51
1:A:63[A]:ARG:NH1	1:A:330:GLN:OE1	2.45	0.50
1:A:212:LYS:HD2	1:A:213:THR:N	2.28	0.49
1:A:44:SER:HB2	1:A:509:LEU:HD22	1.94	0.48
1:A:29:LEU:CD1	1:A:482:PRO:CD	2.92	0.48
1:A:29:LEU:HD11	1:A:482:PRO:CG	2.46	0.46
1:A:308:ILE:HA	1:A:311:MET:HE3	1.97	0.46
1:A:121:ILE:HG12	1:A:134:THR:HG23	1.97	0.45
1:A:216:GLN:O	1:A:223:VAL:HB	2.16	0.45
1:A:216:GLN:OE1	1:A:225:THR:OG1	2.34	0.44
1:A:53:SER:HB3	1:A:457:LEU:HG	2.00	0.43
1:A:171:LEU:HD12	1:A:172:PRO:HD2	2.01	0.43
1:A:388:VAL:HA	1:A:392:LEU:HB3	2.01	0.43
1:A:515:ASN:O	1:A:519:THR:HG23	2.19	0.43
1:A:155:LYS:NZ	3:A:711:HOH:O	2.52	0.43
1:A:335:GLY:HA3	1:A:336:GLY:HA3	1.48	0.42
1:A:298:LEU:HD23	1:A:298:LEU:HA	1.90	0.42
1:A:46:LEU:HD12	1:A:46:LEU:HA	1.86	0.41
1:A:239:ASN:OD1	1:A:241:THR:HG23	2.15	0.41
1:A:150:ILE:HG13	1:A:196:ASN:ND2	2.35	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:49:LEU:HD23	1:A:49:LEU:HA	1.95	0.41
1:A:329:GLU:OE1	1:A:331:SER:OG	2.28	0.41
1:A:336:GLY:N	1:A:338:ASN:OD1	2.53	0.41
1:A:228:SER:HB2	1:A:249:THR:OG1	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	473/571 (83%)	454 (96%)	17 (4%)	2 (0%)	30 27

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	205	SER
1	A	204	VAL

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	404/475 (85%)	394 (98%)	10 (2%)	42 45

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	37	SER
1	A	145	TYR
1	A	190	SER
1	A	206	ASP
1	A	214	LYS
1	A	242	ARG
1	A	337	ASN
1	A	338	ASN
1	A	397	ASN
1	A	440	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	369	HIS
1	A	373	GLN
1	A	439	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 10 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	477/571 (83%)	0.47	43 (9%)	17	15	18, 38, 86, 108	2 (0%)

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	484	ALA	6.2
1	A	337	ASN	5.8
1	A	500	PRO	5.2
1	A	243	VAL	5.1
1	A	336	GLY	4.9
1	A	338	ASN	4.9
1	A	482	PRO	3.8
1	A	26	ILE	3.8
1	A	217	THR	3.7
1	A	498	TYR	3.6
1	A	216	GLN	3.6
1	A	218	ILE	3.5
1	A	215	THR	3.3
1	A	397	ASN	3.3
1	A	339	GLY	3.2
1	A	206	ASP	3.1
1	A	208	ASN	3.0
1	A	236	ALA	2.7
1	A	140	TYR	2.7
1	A	486	SER	2.7
1	A	231	VAL	2.7
1	A	207	GLN	2.6
1	A	232	VAL	2.6
1	A	340	LYS	2.5
1	A	234	SER	2.5
1	A	327	GLU	2.5
1	A	200	LYS	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	205	SER	2.5
1	A	222	THR	2.5
1	A	483	ASN	2.5
1	A	519	THR	2.4
1	A	497	PRO	2.4
1	A	111	ASN	2.4
1	A	487	LEU	2.3
1	A	199	LYS	2.3
1	A	485	GLN	2.3
1	A	204	VAL	2.2
1	A	220	GLY	2.2
1	A	211	THR	2.1
1	A	233	ASP	2.1
1	A	334	VAL	2.1
1	A	212	LYS	2.1
1	A	238	GLY	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	CD	A	610	1/1	0.61	0.16	71,71,71,71	1
2	CD	A	607	1/1	0.85	0.12	103,103,103,103	0
2	CD	A	606	1/1	0.89	0.10	77,77,77,77	1
2	CD	A	605	1/1	0.92	0.10	62,62,62,62	1
2	CD	A	604	1/1	0.93	0.09	50,50,50,50	1
2	CD	A	609	1/1	0.95	0.06	47,47,47,47	1
2	CD	A	601	1/1	0.95	0.07	70,70,70,70	1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	CD	A	608	1/1	0.98	0.04	42,42,42,42	1
2	CD	A	603	1/1	0.99	0.03	35,35,35,35	0
2	CD	A	602	1/1	0.99	0.03	35,35,35,35	0

6.5 Other polymers (i)

There are no such residues in this entry.