Full wwPDB X-ray Structure Validation Report (i)

Jun 16, 2024 – 11:40 PM EDT

PDB ID	:	5GNG
Title	:	Crystal Structure of BioG from Haemophilus influenzae at 1.26 Angstroms resolution
Authors	:	Shi, J.; Guo, Z.
Deposited on	:	2016-07-20
Resolution	:	1.26 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 (i) 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 (1)) were used in the production of this report:

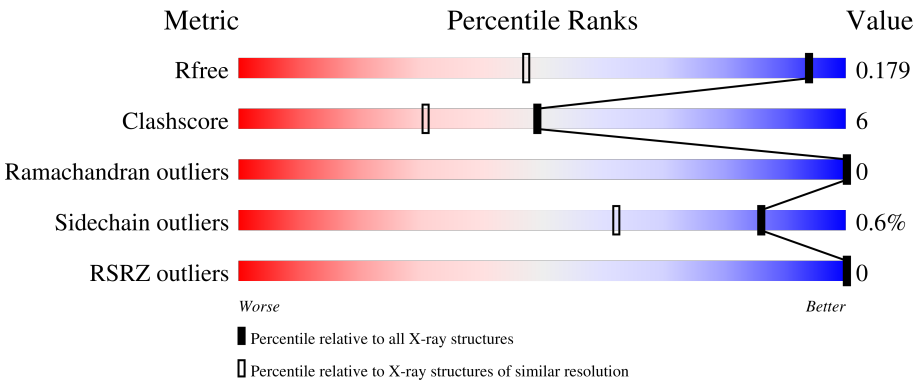
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.26 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	1023 (1.28-1.24)
Clashscore	141614	1060 (1.28-1.24)
Ramachandran outliers	138981	1029 (1.28-1.24)
Sidechain outliers	138945	1028 (1.28-1.24)
RSRZ outliers	127900	1004 (1.28-1.24)

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	223	86% 9% ·
1	B	223	84% 12% ·

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4272 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 Uncharacterized protein HI_1552.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	214	1772	1149	301	315	7	0	2	0
1	B	214	1779	1158	302	312	7	0	3	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	216	LEU	-	expression tag	UNP P44251
A	217	GLU	-	expression tag	UNP P44251
A	218	HIS	-	expression tag	UNP P44251
A	219	HIS	-	expression tag	UNP P44251
A	220	HIS	-	expression tag	UNP P44251
A	221	HIS	-	expression tag	UNP P44251
A	222	HIS	-	expression tag	UNP P44251
A	223	HIS	-	expression tag	UNP P44251
B	216	LEU	-	expression tag	UNP P44251
B	217	GLU	-	expression tag	UNP P44251
B	218	HIS	-	expression tag	UNP P44251
B	219	HIS	-	expression tag	UNP P44251
B	220	HIS	-	expression tag	UNP P44251
B	221	HIS	-	expression tag	UNP P44251
B	222	HIS	-	expression tag	UNP P44251
B	223	HIS	-	expression tag	UNP P44251

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	363	363	363	0	0
2	B	358	358	358	0	0

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: Uncharacterized protein HI_1552

Chain A:

- Molecule 1: Uncharacterized protein HI_1552

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	45.95Å 67.83Å 67.99Å 90.00° 90.01° 90.00°	Depositor
Resolution (Å)	30.39 – 1.26 30.39 – 1.26	Depositor EDS
% Data completeness (in resolution range)	93.9 (30.39-1.26) 93.9 (30.39-1.26)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.57 (at 1.26Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.145 , 0.181 0.146 , 0.179	Depositor DCC
R_{free} test set	1986 reflections (1.88%)	wwPDB-VP
Wilson B-factor (Å^2)	12.1	Xtriage
Anisotropy	0.342	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.058 for -h,l,k 0.057 for -h,-l,-k 0.450 for h,-k,-l	Xtriage
Reported twinning fraction	0.460 for -h,k,l	Depositor
Outliers	0 of 105604 reflections	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	4272	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.52% 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 (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	0.31	0/1833	0.45	0/2494
1	B	0.32	0/1843	0.47	0/2507
All	All	0.31	0/3676	0.46	0/5001

There are no bond length outliers.

There are no bond angle outliers.

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	1772	0	1691	20	0
1	B	1779	0	1721	21	0
2	A	363	0	0	11	5
2	B	358	0	0	11	4
All	All	4272	0	3412	41	6

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:B:57:ARG:NH1	2:B:301:HOH:O	1.88	1.03

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:139:ARG:NH2	2:B:304:HOH:O	2.18	0.77
1:A:194:GLN:NE2	2:A:302:HOH:O	2.16	0.76
1:B:185:GLN:CD	2:B:307:HOH:O	2.26	0.74
1:A:143[B]:GLU:OE1	2:A:301:HOH:O	2.07	0.73
1:B:57:ARG:NH2	2:B:306:HOH:O	2.24	0.71
1:B:143:GLU:OE1	2:B:302:HOH:O	2.09	0.70
1:B:65:SER:OG	2:B:303:HOH:O	2.10	0.69
1:A:57:ARG:HG2	2:A:558:HOH:O	2.00	0.62
1:A:168:ASN:HD22	1:A:194:GLN:HE22	1.45	0.61
1:A:97:PHE:HE1	1:A:185:GLN:HG2	1.69	0.57
1:B:116:ARG:NH2	1:B:139:ARG:O	2.27	0.56
1:A:140:PRO:HB2	1:A:143[B]:GLU:HG2	1.88	0.56
1:B:28:ASN:HB3	2:B:521:HOH:O	2.05	0.55
1:B:12:HIS:ND1	1:B:58:HIS:HB2	2.25	0.52
1:A:195:GLU:OE1	2:A:303:HOH:O	2.19	0.52
1:A:204:SER:HB2	2:A:460:HOH:O	2.08	0.52
1:A:97:PHE:CE1	1:A:185:GLN:HG2	2.45	0.52
1:B:25:ASP:OD1	1:B:28:ASN:ND2	2.45	0.50
1:B:161:ILE:HD12	1:B:190:ARG:NH2	2.27	0.49
1:A:52:ASP:OD2	2:A:305:HOH:O	2.20	0.49
1:B:32[B]:LEU:HD21	1:B:38:LEU:HB2	1.94	0.49
1:A:156:GLN:OE1	2:A:304:HOH:O	2.19	0.48
1:A:197:GLU:OE2	2:A:306:HOH:O	2.20	0.46
1:B:15[B]:LEU:HD23	1:B:75:LEU:HD11	1.97	0.46
1:B:185:GLN:CG	2:B:307:HOH:O	2.62	0.46
1:A:132:ARG:HD3	2:A:315:HOH:O	2.17	0.45
1:B:35:ASN:HA	2:B:325:HOH:O	2.15	0.45
1:B:159:LYS:HD2	2:B:639:HOH:O	2.15	0.44
1:B:3:THR:HB	1:B:39:LEU:HD11	1.99	0.44
1:A:5:PHE:CE2	1:A:7:ASP:HB2	2.52	0.44
1:B:153:ALA:O	1:B:157:GLN:HG3	2.20	0.42
1:A:76:GLN:NE2	1:A:163:LEU:HD22	2.35	0.42
1:B:87:ASN:HA	1:B:183:GLN:OE1	2.20	0.41
1:B:1:MET:N	1:B:46:ASP:OD1	2.54	0.41
1:A:104:PHE:CD2	1:A:152:PHE:CD1	3.08	0.41
1:A:128:ALA:HB1	2:A:469:HOH:O	2.19	0.41
1:A:168:ASN:ND2	1:A:194:GLN:HE22	2.16	0.41
1:A:185:GLN:OE1	2:A:307:HOH:O	2.22	0.41
1:A:61:LEU:HD22	1:A:75:LEU:HD12	2.02	0.40
1:B:139:ARG:HG3	2:B:391:HOH:O	2.21	0.40

All (6) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:637:HOH:O	2:A:655:HOH:O[2_554]	2.02	0.18
2:A:608:HOH:O	2:B:628:HOH:O[1_554]	2.05	0.15
2:A:512:HOH:O	2:B:482:HOH:O[1_554]	2.10	0.10
2:A:602:HOH:O	2:A:619:HOH:O[2_544]	2.10	0.10
2:B:538:HOH:O	2:B:566:HOH:O[2_555]	2.11	0.09
2:A:598:HOH:O	2:B:553:HOH:O[1_554]	2.18	0.02

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	214/223 (96%)	208 (97%)	6 (3%)	0	100	100
1	B	215/223 (96%)	208 (97%)	7 (3%)	0	100	100
All	All	429/446 (96%)	416 (97%)	13 (3%)	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	183/194 (94%)	182 (100%)	1 (0%)	88	66
1	B	185/194 (95%)	184 (100%)	1 (0%)	88	66
All	All	368/388 (95%)	366 (100%)	2 (0%)	86	66

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	104	PHE
1	B	104	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	156	GLN
1	A	168	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 monosaccharides 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, 95th 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	214/223 (95%)	-0.61	0	100	100	7, 12, 19, 25	0
1	B	214/223 (95%)	-0.60	0	100	100	7, 11, 19, 32	0
All	All	428/446 (95%)	-0.60	0	100	100	7, 11, 19, 32	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 monosaccharides 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.