



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 7, 2017 – 09:45 AM EST

PDB ID : 4G0H
Title : Crystal structure of the N-terminal domain of Helicobacter pylori CagA protein
Authors : Kaplan-Turkoz, B.; Remaut, H.; Dian, C.; Louche, A.; Terradot, L.
Deposited on : unknown
Resolution : 3.60 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

<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

PERCENTILES INFOmissingINFO

1 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3275 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 Cytotoxicity-associated immunodominant antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	450	Total	C	N	O	S	0	0	0
			3275	2040	572	658	5			

There are 32 discrepancies between the modelled and reference sequences:


Chain	Residue	Modelled	Actual	Comment	Reference
A	885	LYS	-	EXPRESSION TAG	UNP P55980
A	886	GLY	-	EXPRESSION TAG	UNP P55980
A	887	GLU	-	EXPRESSION TAG	UNP P55980
A	888	LEU	-	EXPRESSION TAG	UNP P55980
A	889	ASN	-	EXPRESSION TAG	UNP P55980
A	890	SER	-	EXPRESSION TAG	UNP P55980
A	891	LYS	-	EXPRESSION TAG	UNP P55980
A	892	LEU	-	EXPRESSION TAG	UNP P55980
A	893	GLU	-	EXPRESSION TAG	UNP P55980
A	894	GLY	-	EXPRESSION TAG	UNP P55980
A	895	LYS	-	EXPRESSION TAG	UNP P55980
A	896	PRO	-	EXPRESSION TAG	UNP P55980
A	897	ILE	-	EXPRESSION TAG	UNP P55980
A	898	PRO	-	EXPRESSION TAG	UNP P55980
A	899	ASN	-	EXPRESSION TAG	UNP P55980
A	900	PRO	-	EXPRESSION TAG	UNP P55980
A	901	LEU	-	EXPRESSION TAG	UNP P55980
A	902	LEU	-	EXPRESSION TAG	UNP P55980
A	903	GLY	-	EXPRESSION TAG	UNP P55980
A	904	LEU	-	EXPRESSION TAG	UNP P55980
A	905	ASP	-	EXPRESSION TAG	UNP P55980
A	906	SER	-	EXPRESSION TAG	UNP P55980
A	907	THR	-	EXPRESSION TAG	UNP P55980
A	908	ARG	-	EXPRESSION TAG	UNP P55980
A	909	THR	-	EXPRESSION TAG	UNP P55980
A	910	GLY	-	EXPRESSION TAG	UNP P55980
A	911	HIS	-	EXPRESSION TAG	UNP P55980

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	912	HIS	-	EXPRESSION TAG	UNP P55980
A	913	HIS	-	EXPRESSION TAG	UNP P55980
A	914	HIS	-	EXPRESSION TAG	UNP P55980
A	915	HIS	-	EXPRESSION TAG	UNP P55980
A	916	HIS	-	EXPRESSION TAG	UNP P55980

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 ($\text{RSRZ} > 2$). 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.

- Chain A: 

Residue	Position	Conservation Score (bits)
MET	1	0.1
THR	2	0.1
ASN	3	0.1
PHE	4	0.1
GLU	5	0.1
GLY	6	0.1
THR	7	0.1
ILE	8	0.1
ASP	9	0.1
GLN	10	0.1
THR	11	0.1
ARG	12	0.1
GLU	13	0.1
THR	14	0.1
PRO	15	0.1
ASP	16	0.1
GLN	17	0.1
THR	18	0.1
ASN	19	0.1
GLN	20	0.1
LYS	21	0.1
ALA	22	0.1
SER	23	0.1
GLN	24	0.1
THR	25	0.1
ALA	26	0.1
ALA	27	0.1
PHE	28	0.1
PRO	29	0.1
ASP	30	0.1
THR	31	0.1
LYS	32	0.1
PRO	33	0.1
GLN	34	0.1
GLN	35	0.1
ASN	36	0.1
GLN	37	0.1
PHE	38	0.1
ILE	39	0.1
ILE	40	0.1
THR	41	0.1
GLY	42	0.1
PHE	43	0.1
ASN	44	0.1
LEU	45	0.1
GLN	46	0.1
VAL	47	0.1
ALA	48	0.1
ASP	49	0.1
SER	50	0.1
PHE	51	0.1
ILE	52	0.1
ASP	53	0.1
ASP	54	0.1
LEU	55	0.1
ILE	56	0.1
ASN	57	0.1
VAL	58	0.1
VAL	59	0.1
LYS	60	0.1
ALA	61	0.1
ASP	62	0.1
ASN	63	0.1
VAL	64	0.1
GLU	65	0.1
GLU	66	0.1
SER	67	0.1
VAL	68	0.1
ALA	69	0.1
GLY	70	0.1
PHE	71	0.1
LYS	72	0.1
THR	73	0.1
ILE	74	0.1
GLN	75	0.1
ILE	76	0.1
GLY	77	0.1
GLY	78	0.1
VAL	79	0.1
PRO	80	0.1
ASN	81	0.1
GLN	82	0.1
THR	83	0.1
GLY	84	0.1
GLY	85	0.1
THR	86	0.1
GLY	87	0.1
GLY	88	0.1
THR	89	0.1
ASP	90	0.1
ILE	91	0.1
LYS	92	0.1
ASN	93	0.1
SER	94	0.1
VAL	95	0.1
ASN	96	0.1
GLY	97	0.1
LYS	98	0.1
THR	99	0.1
ASN	100	0.1
THR	101	0.1
ASN	102	0.1
THR	103	0.1
VAL	104	0.1
LYS	105	0.1
ASP	106	0.1
PHE	107	0.1
VAL	108	0.1
VAL	109	0.1
GLY	110	0.1
GLY	111	0.1
THR	112	0.1
LYS	113	0.1
ASN	114	0.1
THR	115	0.1
ASN	116	0.1
THR	117	0.1
ASN	118	0.1
THR	119	0.1
ASN	120	0.1
THR	121	0.1
ASN	122	0.1
THR	123	0.1
ASN	124	0.1
THR	125	0.1
ASN	126	0.1
THR	127	0.1
ASN	128	0.1
THR	129	0.1
ASN	130	0.1
THR	131	0.1
ASN	132	0.1
THR	133	0.1
ASN	134	0.1
THR	135	0.1
ASN	136	0.1
THR	137	0.1
ASN	138	0.1
THR	139	0.1
ASN	140	0.1
THR	141	0.1
ASN	142	0.1
THR	143	0.1
ASN	144	0.1
THR	145	0.1
ASN	146	0.1
THR	147	0.1
ASN	148	0.1

3 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	97.83Å 97.83Å 245.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	62.75 – 3.60 62.75 – 3.60	Depositor EDS
% Data completeness (in resolution range)	91.8 (62.75-3.60) 91.8 (62.75-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.57 (at 3.57Å)	Xtriage
Refinement program	PHENIX 1.8_1056	Depositor
R, R_{free}	0.329 , 0.341 0.324 , 0.337	Depositor DCC
R_{free} test set	669 reflections (5.01%)	DCC
Wilson B-factor (Å ²)	159.9	Xtriage
Anisotropy	0.064	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 160.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	3275	wwPDB-VP
Average B, all atoms (Å ²)	158.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.45% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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.

4 Model quality [i](#)

4.1 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	0.22	0/3304	0.37	0/4458

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

4.2 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	3275	0	3025	28	0
All	All	3275	0	3025	28	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 4.

The worst 5 of 28 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:334:HIS:NE2	1:A:337:ASN:OD1	2.27	0.65
1:A:507:ASP:N	1:A:507:ASP:OD2	2.30	0.64
1:A:676:ALA:O	1:A:680:ASN:ND2	2.30	0.63
1:A:319:LYS:HB3	1:A:342:TYR:HB2	1.82	0.61
1:A:410:GLN:HG3	1:A:471:LEU:HD21	1.84	0.59

There are no symmetry-related clashes.

4.3 Torsion angles [i](#)

4.3.1 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	432/916 (47%)	414 (96%)	18 (4%)	0	100	100

There are no Ramachandran outliers to report.

4.3.2 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	316/801 (40%)	303 (96%)	13 (4%)	35	72

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	474	THR
1	A	507	ASP
1	A	687	GLU
1	A	399	ARG
1	A	684	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	467	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	810	GLN
1	A	468	ASN
1	A	446	ASN
1	A	669	ASN

4.3.3 RNA [i](#)

There are no RNA molecules in this entry.

4.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

4.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

4.6 Ligand geometry [i](#)

There are no ligands in this entry.

4.7 Other polymers [i](#)

There are no such residues in this entry.

4.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

5 Fit of model and data [i](#)

5.1 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	450/916 (49%)	-0.04	9 (2%) 65 51	92, 156, 214, 255	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	569	GLY	3.4
1	A	633	VAL	2.9
1	A	630	GLU	2.7
1	A	565	LEU	2.6
1	A	812	LEU	2.6

5.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.4 Ligands [i](#)

There are no ligands in this entry.

5.5 Other polymers [i](#)

There are no such residues in this entry.