wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 07:39 PM EDT

PDB ID	:	3U07
Title	:	Crystal Structure of the VPA0106 protein from Vibrio parahaemolyticus. Northeast Structural Genomics Consortium Target VpR106.
Authors	:	Vorobiev, S.; Neely, H.; Seetharaman, J.; Patel, P.; Xiao, R.; Ciccosanti, C.; Wang, H.; Everett, J.K.; Nair, R.; Acton, T.B.; Rost, B.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2011-09-28
Resolution	:	2.08 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

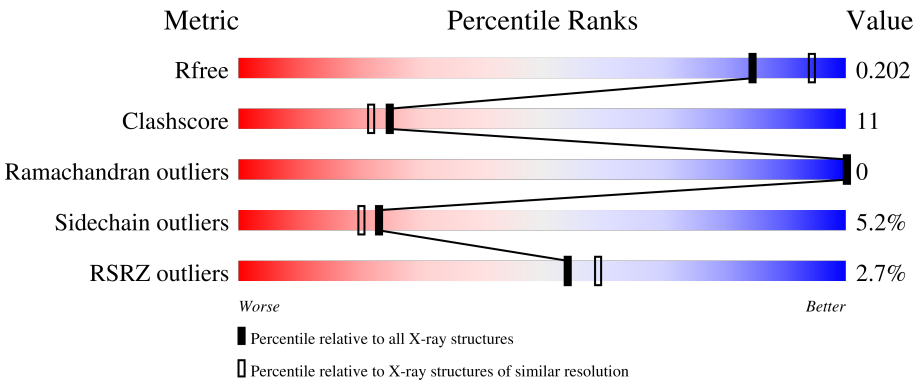
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.08 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	6189 (2.10-2.06)
Clashscore	141614	6738 (2.10-2.06)
Ramachandran outliers	138981	6663 (2.10-2.06)
Sidechain outliers	138945	6664 (2.10-2.06)
RSRZ outliers	127900	6057 (2.10-2.06)

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	443	2% 71% 16% • 10%
1	B	443	2% 71% 17% • 10%
1	C	443	3% 73% 16% • 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 10329 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 VPA0106.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	398	3180	2013	537	611	2	17	0	0	0
1	B	398	3182	2014	537	612	2	17	0	0	0
1	C	399	3187	2017	538	613	2	17	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	60	GLU	ARG	engineered mutation	UNP Q87JZ2
A	64	ARG	ILE	engineered mutation	UNP Q87JZ2
A	436	LEU	-	EXPRESSION TAG	UNP Q87JZ2
A	437	GLU	-	EXPRESSION TAG	UNP Q87JZ2
A	438	HIS	-	EXPRESSION TAG	UNP Q87JZ2
A	439	HIS	-	EXPRESSION TAG	UNP Q87JZ2
A	440	HIS	-	EXPRESSION TAG	UNP Q87JZ2
A	441	HIS	-	EXPRESSION TAG	UNP Q87JZ2
A	442	HIS	-	EXPRESSION TAG	UNP Q87JZ2
A	443	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	60	GLU	ARG	engineered mutation	UNP Q87JZ2
B	64	ARG	ILE	engineered mutation	UNP Q87JZ2
B	436	LEU	-	EXPRESSION TAG	UNP Q87JZ2
B	437	GLU	-	EXPRESSION TAG	UNP Q87JZ2
B	438	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	439	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	440	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	441	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	442	HIS	-	EXPRESSION TAG	UNP Q87JZ2
B	443	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	60	GLU	ARG	engineered mutation	UNP Q87JZ2
C	64	ARG	ILE	engineered mutation	UNP Q87JZ2
C	436	LEU	-	EXPRESSION TAG	UNP Q87JZ2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	437	GLU	-	EXPRESSION TAG	UNP Q87JZ2
C	438	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	439	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	440	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	441	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	442	HIS	-	EXPRESSION TAG	UNP Q87JZ2
C	443	HIS	-	EXPRESSION TAG	UNP Q87JZ2

- Molecule 2 is water.

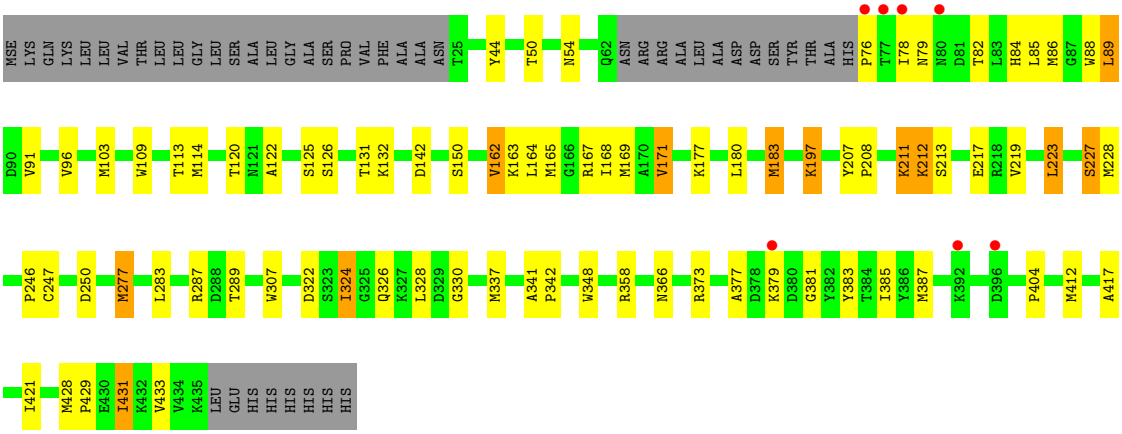
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	245	245	245	0	0
			Total	O		
2	B	265	265	265	0	0
			Total	O		
2	C	270	270	270	0	0

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.

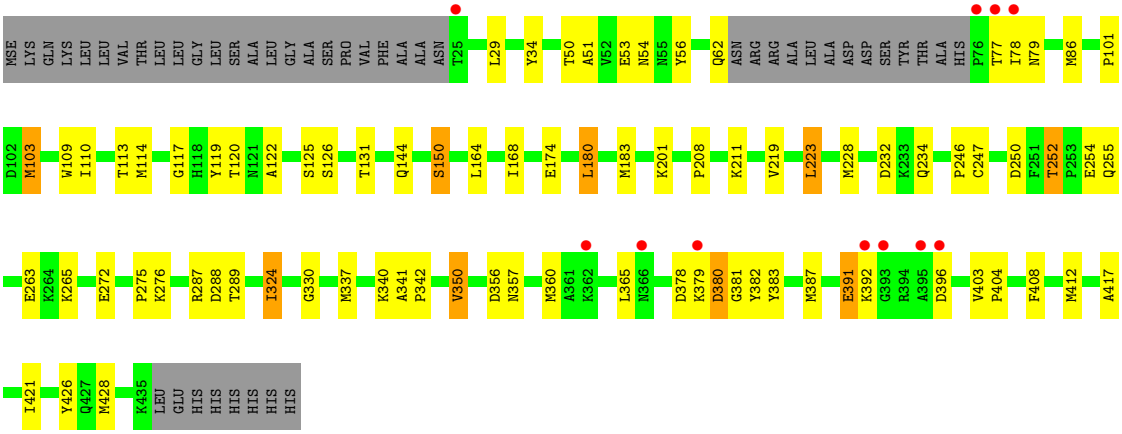
• Molecule 1: uncharacterized protein VPA0106

Chain A:



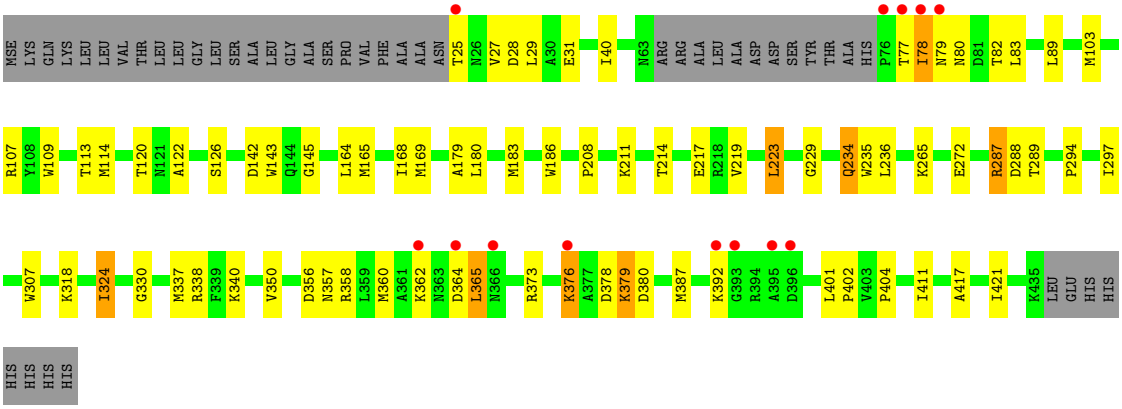
• Molecule 1: uncharacterized protein VPA0106

Chain B:



• Molecule 1: uncharacterized protein VPA0106

Chain C:


MSE LYS GLN LYS LEU LEU VAL THR LEU LEU GLY LEU SER ALA LEU GLY ALA SER PRO VAL PHE ALA ALA ASN T25 N26 V27 D28 L29 A30 E31 I40 N63 ARG ARG ALA LEU ALA ASP ASP SER TYR THR ALA HIS P76 T77 I78 N79 N80 D81 T82 L83 L89 M103
R107 Y108 W109 T113 M114 T120 N121 A122 S126 D142 W143 Q144 G145 L164 M165 I168 M169 A179 L180 M183 W186 P208 K211 T214 E217 R218 V219 L223 G229 Q234 W235 L236 K265 E272 R287 D288 T289 P294 I297
W307 K318 I324 G330 M337 R338 F339 K340 V350 D356 N357 R358 L359 M360 A361 K362 N363 D364 L365 N366 R373 K376 A377 D378 K379 D380 M387 K392 G393 R394 A395 D396 L401 P402 V403 P404 I411 A417 I421 K435 LEU GLU HIS HIS
HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	51.45Å 96.32Å 100.54Å 60.61° 80.87° 86.55°	Depositor
Resolution (Å)	46.96 – 2.08 46.96 – 2.08	Depositor EDS
% Data completeness (in resolution range)	94.8 (46.96-2.08) 96.1 (46.96-2.08)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.27 (at 2.08Å)	Xtriage
Refinement program	PHENIX 1.7.1_743	Depositor
R, R_{free}	0.194 , 0.216 0.179 , 0.202	Depositor DCC
R_{free} test set	4763 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	25.2	Xtriage
Anisotropy	0.356	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 55.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.007 for -h,k-l,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10329	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.95% 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	0.35	0/3242	0.64	0/4363
1	B	0.33	0/3244	0.63	0/4366
1	C	0.34	0/3249	0.64	0/4373
All	All	0.34	0/9735	0.64	0/13102

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	3180	0	3044	76	0
1	B	3182	0	3049	69	0
1	C	3187	0	3051	61	0
2	A	245	0	0	10	0
2	B	265	0	0	10	0
2	C	270	0	0	16	0
All	All	10329	0	9144	202	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 11.

The worst 5 of 202 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:168:ILE:CD1	1:B:183:MSE:HE3	1.75	1.14
1:B:228:MSE:HE2	1:B:287:ARG:HH12	1.21	1.04
1:B:168:ILE:HD13	1:B:183:MSE:HE3	1.36	1.02
1:A:324:ILE:HG13	1:C:289:THR:O	1.60	1.01
1:A:289:THR:O	1:B:324:ILE:HG13	1.62	0.98

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	394/443 (89%)	384 (98%)	10 (2%)	0	100	100
1	B	394/443 (89%)	382 (97%)	12 (3%)	0	100	100
1	C	395/443 (89%)	386 (98%)	9 (2%)	0	100	100
All	All	1183/1329 (89%)	1152 (97%)	31 (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	338/358 (94%)	323 (96%)	15 (4%)	28	27
1	B	339/358 (95%)	319 (94%)	20 (6%)	19	16
1	C	339/358 (95%)	321 (95%)	18 (5%)	22	20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1016/1074 (95%)	963 (95%)	53 (5%)	23 20

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	272	GLU
1	C	29	LEU
1	C	365	LEU
1	B	288	ASP
1	B	380	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	49	HIS
1	A	84	HIS
1	A	144	GLN
1	B	49	HIS
1	C	368	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 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	381/443 (86%)	-0.36	7 (1%)	68	71	14, 28, 62, 86	0
1	B	381/443 (86%)	-0.31	11 (2%)	51	56	15, 27, 67, 97	0
1	C	382/443 (86%)	-0.35	13 (3%)	45	50	12, 25, 61, 100	0
All	All	1144/1329 (86%)	-0.34	31 (2%)	54	59	12, 27, 63, 100	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	76	PRO	10.7
1	B	78	ILE	6.7
1	B	76	PRO	5.8
1	A	76	PRO	5.6
1	A	78	ILE	5.3

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.