wwPDB X-ray Structure Validation Summary Report (i)

Oct 2, 2021 – 04:18 PM EDT

PDB ID	:	3KAV
Title	:	Crystal Structure of THE MUTANT (L80M) PA2107 PROTEIN from Pseudomonas aeruginosa, Northeast Structural Genomics Consortium Target PaR198
Authors	:	Forouhar, F.; Neely, H.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccosanti, C.; Foote, E.L.; Zhao, L.; 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	:	2009-10-19
Resolution	:	2.50 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 (i)) were used in the production of this report:

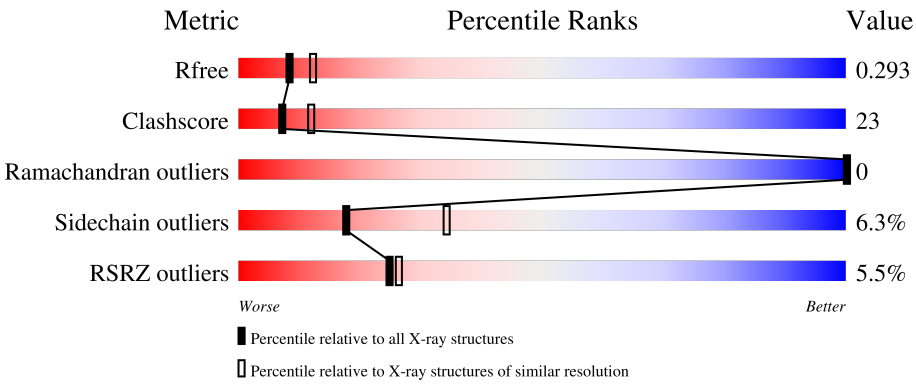
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	140	5% 46% 24% • 27%
1	B	140	6% 36% 31% • 30%
1	C	140	3% 50% 24% • 25%
1	D	140	% 37% 32% • 29%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3130 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 PA2107.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	102	778	467	154	144	12	1	0	0	0
1	B	98	748	447	149	139	12	1	0	0	0
1	C	105	794	477	157	147	12	1	0	0	0
1	D	100	768	461	152	142	12	1	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	80	MSE	LEU	engineered mutation	UNP Q9I208
A	133	LEU	-	expression tag	UNP Q9I208
A	134	GLU	-	expression tag	UNP Q9I208
A	135	HIS	-	expression tag	UNP Q9I208
A	136	HIS	-	expression tag	UNP Q9I208
A	137	HIS	-	expression tag	UNP Q9I208
A	138	HIS	-	expression tag	UNP Q9I208
A	139	HIS	-	expression tag	UNP Q9I208
A	140	HIS	-	expression tag	UNP Q9I208
B	80	MSE	LEU	engineered mutation	UNP Q9I208
B	133	LEU	-	expression tag	UNP Q9I208
B	134	GLU	-	expression tag	UNP Q9I208
B	135	HIS	-	expression tag	UNP Q9I208
B	136	HIS	-	expression tag	UNP Q9I208
B	137	HIS	-	expression tag	UNP Q9I208
B	138	HIS	-	expression tag	UNP Q9I208
B	139	HIS	-	expression tag	UNP Q9I208
B	140	HIS	-	expression tag	UNP Q9I208
C	80	MSE	LEU	engineered mutation	UNP Q9I208
C	133	LEU	-	expression tag	UNP Q9I208
C	134	GLU	-	expression tag	UNP Q9I208

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	135	HIS	-	expression tag	UNP Q9I208
C	136	HIS	-	expression tag	UNP Q9I208
C	137	HIS	-	expression tag	UNP Q9I208
C	138	HIS	-	expression tag	UNP Q9I208
C	139	HIS	-	expression tag	UNP Q9I208
C	140	HIS	-	expression tag	UNP Q9I208
D	80	MSE	LEU	engineered mutation	UNP Q9I208
D	133	LEU	-	expression tag	UNP Q9I208
D	134	GLU	-	expression tag	UNP Q9I208
D	135	HIS	-	expression tag	UNP Q9I208
D	136	HIS	-	expression tag	UNP Q9I208
D	137	HIS	-	expression tag	UNP Q9I208
D	138	HIS	-	expression tag	UNP Q9I208
D	139	HIS	-	expression tag	UNP Q9I208
D	140	HIS	-	expression tag	UNP Q9I208

- Molecule 2 is water.

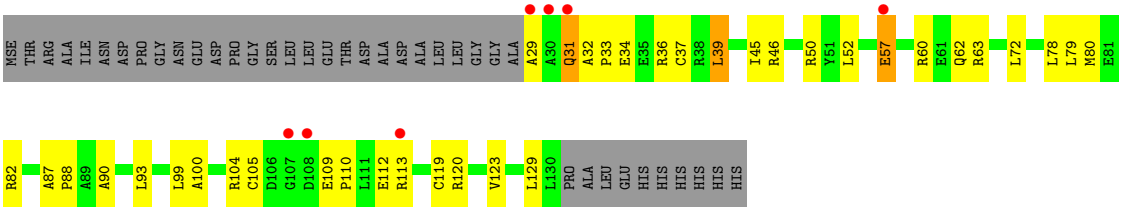
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	13	13	13	0	0
			Total	O		
2	B	6	6	6	0	0
			Total	O		
2	C	10	10	10	0	0
			Total	O		
2	D	13	13	13	0	0

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.

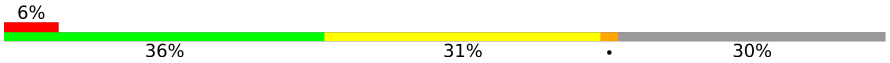
• Molecule 1: uncharacterized protein PA2107

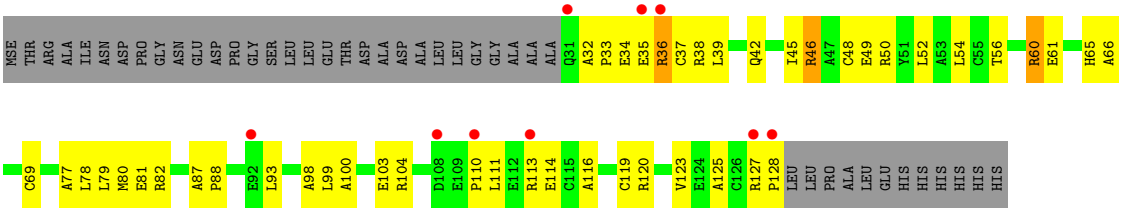
Chain A:



• Molecule 1: uncharacterized protein PA2107

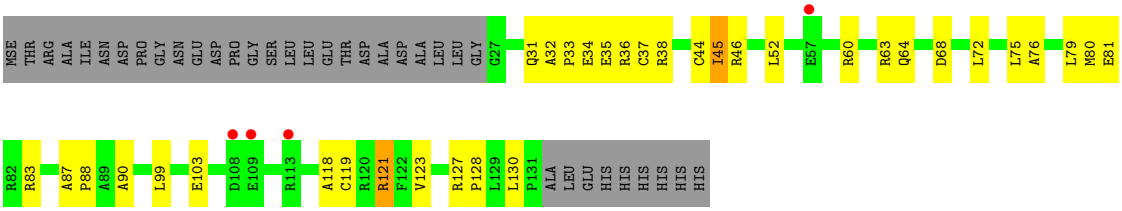
Chain B:





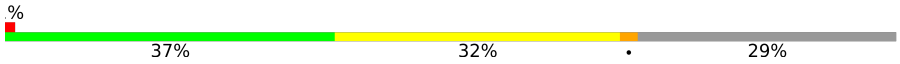
• Molecule 1: uncharacterized protein PA2107

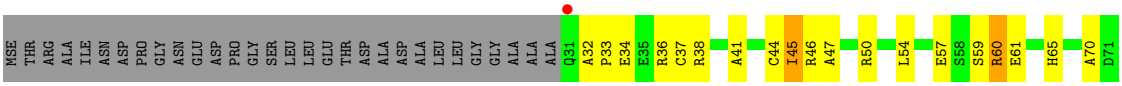
Chain C:

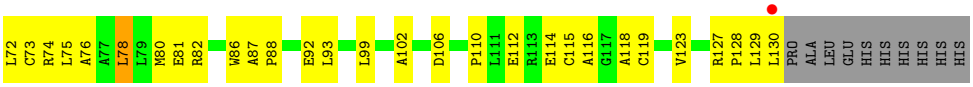


• Molecule 1: uncharacterized protein PA2107

Chain D:




MSE
THR
ARG
ALA
ILE
ASN
ASP
PRO
GLY
ASN
GLU
ASP
PRO
GLY
SER
LEU
LEU
GLU
THR
ASP
ALA
ASP
ALA
LEU
LEU
GLY
GLY
ALA
ALA
ALA
Q31
A32
P33
E34
E35
R36
C37
R38
A41
C44
I45
R46
A47
R50
L54
E57
S58
S59
R60
E61
H65
A70
D71


L72
C73
R74
L75
A76
A77
L78
L79
M80
E81
R82
W86
A87
P88
E92
L93
L99
A102
D106
P110
L111
E112
R113
E114
C115
A116
G117
A118
C119
V123
R127
P128
L129
L130
PRO
ALA
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	148.10Å 46.55Å 87.53Å 90.00° 122.16° 90.00°	Depositor
Resolution (Å)	19.97 – 2.50 28.79 – 2.50	Depositor EDS
% Data completeness (in resolution range)	79.5 (19.97-2.50) 96.9 (28.79-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.51Å)	Xtriage
Refinement program	CNS 1.2, REFMAC	Depositor
R, R_{free}	0.205 , 0.266 0.227 , 0.293	Depositor DCC
R_{free} test set	1690 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	32.7	Xtriage
Anisotropy	0.474	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 68.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	3130	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 12.48% 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	0.37	0/787	0.53	0/1059
1	B	0.34	0/757	0.49	0/1018
1	C	0.36	0/804	0.52	0/1083
1	D	0.39	0/777	0.54	0/1045
All	All	0.36	0/3125	0.52	0/4205

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	778	0	760	33	0
1	B	748	0	722	39	0
1	C	794	0	775	29	0
1	D	768	0	750	43	0
2	A	13	0	0	0	0
2	B	6	0	0	0	0
2	C	10	0	0	1	0
2	D	13	0	0	2	0
All	All	3130	0	3007	138	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 23.

The worst 5 of 138 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:C:34:GLU:HG3	1:C:80:MSE:HE2	1.37	1.06
1:A:34:GLU:HG3	1:A:80:MSE:HE2	1.35	1.03
1:B:37:CYS:HB3	1:B:80:MSE:HE1	1.52	0.92
1:B:34:GLU:HG2	1:B:80:MSE:HE2	1.58	0.83
1:A:37:CYS:HB3	1:A:80:MSE:HE1	1.63	0.81

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	100/140 (71%)	97 (97%)	3 (3%)	0	100	100
1	B	96/140 (69%)	95 (99%)	1 (1%)	0	100	100
1	C	103/140 (74%)	102 (99%)	1 (1%)	0	100	100
1	D	98/140 (70%)	96 (98%)	2 (2%)	0	100	100
All	All	397/560 (71%)	390 (98%)	7 (2%)	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	76/103 (74%)	71 (93%)	5 (7%)	16	32
1	B	73/103 (71%)	69 (94%)	4 (6%)	21	41
1	C	77/103 (75%)	73 (95%)	4 (5%)	23	44
1	D	76/103 (74%)	70 (92%)	6 (8%)	12	24
All	All	302/412 (73%)	283 (94%)	19 (6%)	18	34

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	45	ILE
1	D	65	HIS
1	D	78	LEU
1	D	60	ARG
1	B	93	LEU

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	31	GLN
1	C	42	GLN
1	D	42	GLN

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	101/140 (72%)	0.27	7 (6%)	16	17	17, 36, 64, 76	0
1	B	97/140 (69%)	0.61	9 (9%)	8	8	33, 53, 78, 84	0
1	C	104/140 (74%)	0.28	4 (3%)	40	43	23, 39, 63, 69	0
1	D	99/140 (70%)	0.12	2 (2%)	65	68	25, 36, 59, 78	0
All	All	401/560 (71%)	0.32	22 (5%)	25	26	17, 40, 67, 84	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	110	PRO	4.5
1	A	30	ALA	4.5
1	C	108	ASP	4.1
1	B	127	ARG	3.4
1	A	31	GLN	3.3

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.