wwPDB X-ray Structure Validation Summary Report (i)

Jun 14, 2022 – 04:06 pm BST

PDB ID	:	7QZD
Title	:	Complex of rice blast (Magnaporthe oryzae) effector protein AVR-PikF with an engineered HMA domain of Pikp-1 (Pikp-SNK-EKE) from rice (Oryza sativa)
Authors	:	Maidment, J.H.R.; Franceschetti, M.; Longya, A.; Banfield, M.J.
Deposited on	:	2022-01-31
Resolution	:	2.20 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 (1)) were used in the production of this report:

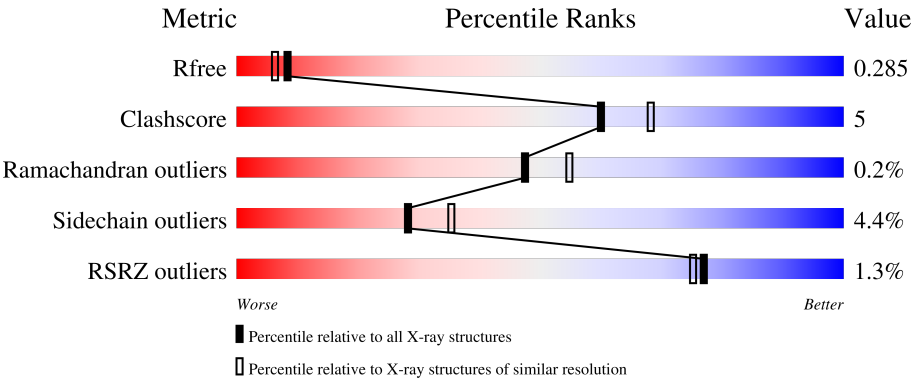
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	80	80% 10% • 9%
1	B	80	71% 24% 5%
1	E	80	79% 9% • 11%
1	F	80	80% 12% 8%
2	C	93	75% 13% 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	93	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3635 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 Resistance protein Pikp-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	73	Total	C	N	O	S	0	0	0
			541	341	95	102	3			
1	B	76	Total	C	N	O	S	0	0	0
			568	357	99	109	3			
1	E	71	Total	C	N	O	S	0	0	0
			529	333	93	100	3			
1	F	74	Total	C	N	O	S	0	0	0
			552	349	96	104	3			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	GLY	-	expression tag	UNP E9KPB5
A	185	PRO	-	expression tag	UNP E9KPB5
A	258	GLU	SER	engineered mutation	UNP E9KPB5
A	261	LYS	ASN	engineered mutation	UNP E9KPB5
A	262	GLU	LYS	engineered mutation	UNP E9KPB5
B	184	GLY	-	expression tag	UNP E9KPB5
B	185	PRO	-	expression tag	UNP E9KPB5
B	258	GLU	SER	engineered mutation	UNP E9KPB5
B	261	LYS	ASN	engineered mutation	UNP E9KPB5
B	262	GLU	LYS	engineered mutation	UNP E9KPB5
E	184	GLY	-	expression tag	UNP E9KPB5
E	185	PRO	-	expression tag	UNP E9KPB5
E	258	GLU	SER	engineered mutation	UNP E9KPB5
E	261	LYS	ASN	engineered mutation	UNP E9KPB5
E	262	GLU	LYS	engineered mutation	UNP E9KPB5
F	184	GLY	-	expression tag	UNP E9KPB5
F	185	PRO	-	expression tag	UNP E9KPB5
F	258	GLU	SER	engineered mutation	UNP E9KPB5
F	261	LYS	ASN	engineered mutation	UNP E9KPB5
F	262	GLU	LYS	engineered mutation	UNP E9KPB5

- Molecule 2 is a protein called Avr-Pik.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	82	Total	C	N	O	S	0	0	0
			668	428	115	120	5			
2	G	82	Total	C	N	O	S	0	0	0
			668	428	115	120	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	21	MET	-	initiating methionine	UNP A0A219T3Y8
G	21	MET	-	initiating methionine	UNP A0A219T3Y8

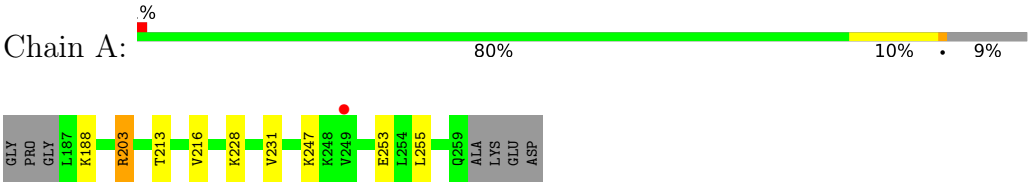
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	8	Total	O	0	0
			8	8		
3	B	16	Total	O	0	0
			16	16		
3	C	28	Total	O	0	0
			28	28		
3	E	10	Total	O	0	0
			10	10		
3	F	21	Total	O	0	0
			21	21		
3	G	26	Total	O	0	0
			26	26		

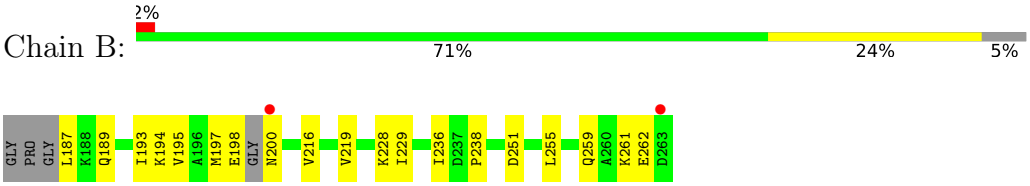
3 Residue-property plots (i)

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.

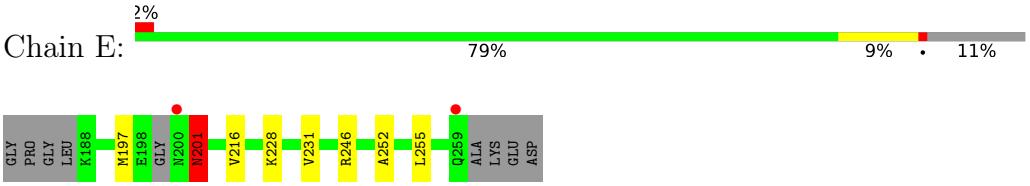
• Molecule 1: Resistance protein Pikp-1



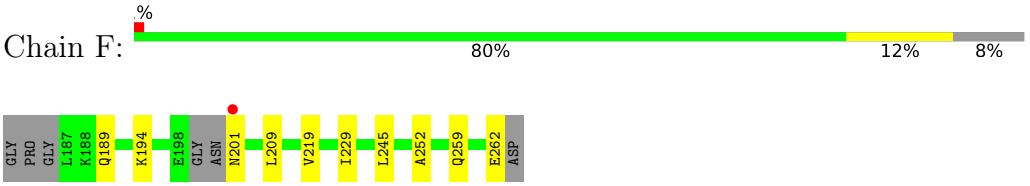
• Molecule 1: Resistance protein Pikp-1



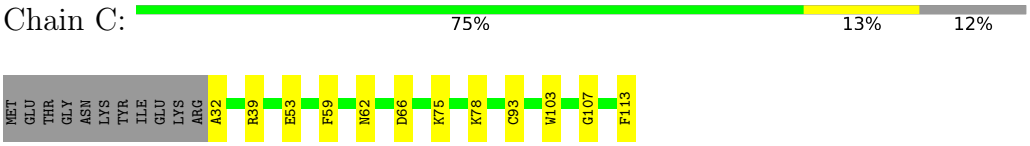
• Molecule 1: Resistance protein Pikp-1



• Molecule 1: Resistance protein Pikp-1

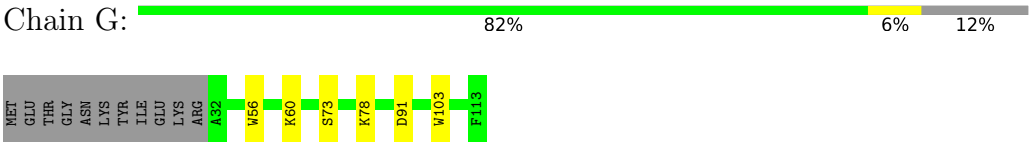


• Molecule 2: Avr-Pik



• Molecule 2: Avr-Pik

Chain G:


82%
6%
12%
MET
GLU
THR
GLY
ASN
LYS
TYR
ILE
GLU
LYS
ARG
A32
W56
K60
S73
K78
D91
W103
F113

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	67.38Å 80.78Å 103.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	63.83 – 2.20 63.75 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.8 (63.83-2.20) 96.7 (63.75-2.20)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.49 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.231 , 0.285 0.234 , 0.285	Depositor DCC
R_{free} test set	1437 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	34.5	Xtriage
Anisotropy	0.617	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3635	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.68 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.5717e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

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	0.74	0/542	0.92	0/726
1	B	0.73	0/568	0.91	0/759
1	E	0.69	0/529	0.91	0/707
1	F	0.74	0/552	0.91	0/737
2	C	0.75	0/691	0.88	0/935
2	G	0.68	0/691	0.82	0/935
All	All	0.72	0/3573	0.89	0/4799

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

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	541	0	588	5	0
1	B	568	0	612	11	0
1	E	529	0	573	6	0
1	F	552	0	602	6	0
2	C	668	0	626	6	0
2	G	668	0	626	3	0
3	A	8	0	0	0	0
3	B	16	0	0	0	0
3	C	28	0	0	0	0
3	E	10	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	F	21	0	0	0	0
3	G	26	0	0	0	0
All	All	3635	0	3627	33	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 5.

The worst 5 of 33 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:189:GLN:HG2	1:B:259:GLN:HE22	1.49	0.77
1:F:189:GLN:HG2	1:F:259:GLN:HE22	1.59	0.66
1:E:201:ASN:HD22	1:E:201:ASN:C	2.02	0.62
1:B:189:GLN:HG2	1:B:259:GLN:NE2	2.15	0.61
1:E:246:ARG:NH1	1:E:252:ALA:O	2.32	0.61

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	71/80 (89%)	70 (99%)	1 (1%)	0	100 100
1	B	72/80 (90%)	68 (94%)	4 (6%)	0	100 100
1	E	67/80 (84%)	64 (96%)	2 (3%)	1 (2%)	10 8
1	F	70/80 (88%)	69 (99%)	1 (1%)	0	100 100
2	C	80/93 (86%)	79 (99%)	1 (1%)	0	100 100
2	G	80/93 (86%)	77 (96%)	3 (4%)	0	100 100
All	All	440/506 (87%)	427 (97%)	12 (3%)	1 (0%)	47 55

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	201	ASN

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	59/63 (94%)	55 (93%)	4 (7%)	16	17
1	B	62/63 (98%)	57 (92%)	5 (8%)	11	12
1	E	58/63 (92%)	57 (98%)	1 (2%)	60	74
1	F	60/63 (95%)	58 (97%)	2 (3%)	38	49
2	C	73/83 (88%)	69 (94%)	4 (6%)	21	26
2	G	73/83 (88%)	72 (99%)	1 (1%)	67	80
All	All	385/418 (92%)	368 (96%)	17 (4%)	28	35

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	201	ASN
2	G	73	SER
1	B	251	ASP
1	B	262	GLU
2	C	39	ARG

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

Mol	Chain	Res	Type
1	E	201	ASN
1	F	189	GLN
1	F	259	GLN
1	B	259	GLN
1	B	189	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

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	73/80 (91%)	0.10	1 (1%)	75	73	29, 49, 81, 87	0
1	B	76/80 (95%)	0.09	2 (2%)	56	53	27, 41, 69, 85	0
1	E	71/80 (88%)	0.16	2 (2%)	53	51	32, 48, 80, 87	0
1	F	74/80 (92%)	-0.11	1 (1%)	75	73	28, 37, 68, 96	0
2	C	82/93 (88%)	-0.31	0	100	100	24, 34, 48, 60	0
2	G	82/93 (88%)	-0.30	0	100	100	28, 38, 59, 81	0
All	All	458/506 (90%)	-0.07	6 (1%)	77	75	24, 40, 73, 96	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	249	VAL	5.7
1	B	263	ASP	3.7
1	E	200	ASN	3.4
1	F	201	ASN	2.9
1	B	200	ASN	2.3

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.