Full wwPDB X-ray Structure Validation Report (i)

May 15, 2020 – 11:30 pm BST

PDB ID	:	6R8M
Title	:	Complex of rice blast (Magnaporthe oryzae) effector protein AVR-PikE with an engineered HMA domain of Pikp-1 from rice (Oryza sativa)
Authors	:	De la Concepcion, J.C.; Franceschetti, M.; Banfield, M.J.
Deposited on	:	2019-04-02
Resolution	:	1.85 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 following versions of software and data (see references (1)) were used in the production of this report:

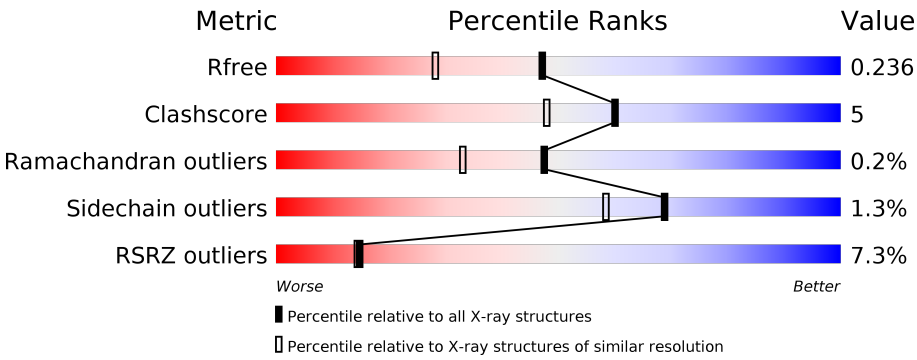
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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 on 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	11% 85% 6% • 8%
1	B	80	9% 89% 9% •
1	E	80	6% 84% 8% 9%
1	F	80	9% 86% 10% •
2	C	92	% 83% 7% 11%
2	G	92	5% 82% 9% 10%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3916 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 NBS-LRR class disease resistance protein Pikh-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	74	549	348	97	101	3	0	2	0
1	B	78	576	364	99	110	3	0	2	0
1	E	73	549	346	98	102	3	0	2	0
1	F	77	573	364	98	108	3	0	3	0

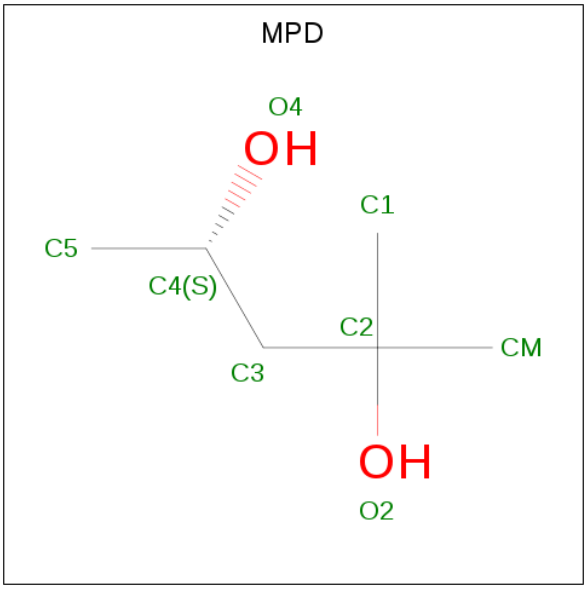
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	GLY	-	expression tag	UNP D5L9G5
A	185	PRO	-	expression tag	UNP D5L9G5
A	262	GLU	LYS	conflict	UNP D5L9G5
B	184	GLY	-	expression tag	UNP D5L9G5
B	185	PRO	-	expression tag	UNP D5L9G5
B	262	GLU	LYS	conflict	UNP D5L9G5
E	184	GLY	-	expression tag	UNP D5L9G5
E	185	PRO	-	expression tag	UNP D5L9G5
E	262	GLU	LYS	conflict	UNP D5L9G5
F	184	GLY	-	expression tag	UNP D5L9G5
F	185	PRO	-	expression tag	UNP D5L9G5
F	262	GLU	LYS	conflict	UNP D5L9G5

- Molecule 2 is a protein called AVR-Pik protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	82	676	435	116	119	6	0	2	0
2	G	83	681	436	119	120	6	0	1	0

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	C	1	8	6	2	0	0
			Total	C	O		
3	G	1	8	6	2	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
4	G	1	1	1	0	0
			Total	Cl		
4	C	1	1	1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	37	37	37	0	0
			Total	O		
5	B	47	47	47	0	0
			Total	O		
5	C	81	81	81	0	0
			Total	O		
5	E	35	35	35	0	0

Continued on next page...

Continued from previous page...

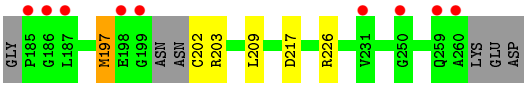
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	F	55	55	55	0	0
			Total	O		
5	G	39	39	39	0	0

3 Residue-property plots (i)

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 (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.

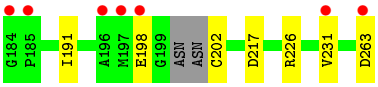
- Molecule 1: NBS-LRR class disease resistance protein Pikh-1

Chain A:



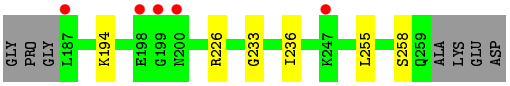
- Molecule 1: NBS-LRR class disease resistance protein Pikh-1

Chain B:



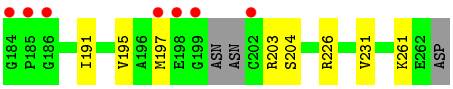
- Molecule 1: NBS-LRR class disease resistance protein Pikh-1

Chain E:



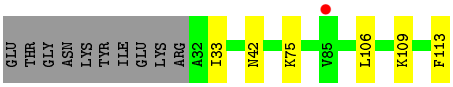
- Molecule 1: NBS-LRR class disease resistance protein Pikh-1

Chain F:



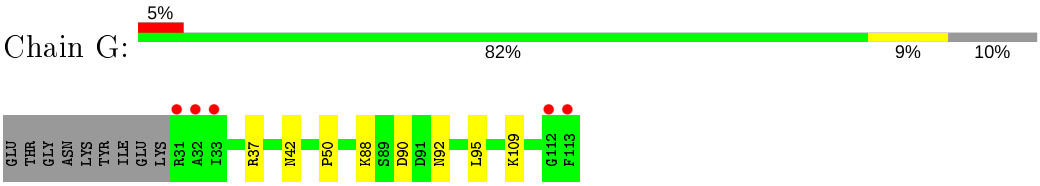
- Molecule 2: AVR-Pik protein

Chain C:



- Molecule 2: AVR-Pik protein

Chain G:


5%
82%
9%
10%
GLU
THR
GLY
ASN
LYS
TYR
ILE
GLU
LYS
R31
A32
I33
R37
N42
P50
K88
S89
D90
D91
N92
L95
K109
G112
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	66.46Å 80.70Å 105.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.50 – 1.85 29.50 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.50-1.85) 99.8 (29.50-1.85)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.92 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0232	Depositor
R, R_{free}	0.186 , 0.230 0.195 , 0.236	Depositor DCC
R_{free} test set	2518 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	29.5	Xtriage
Anisotropy	0.696	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3916	wwPDB-VP
Average B, all atoms (Å^2)	40.0	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 30.19 % 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 1.3638e-03. 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MPD, CL

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	0.77	0/556	0.87	0/742
1	B	0.77	0/583	0.89	0/778
1	E	0.75	0/556	0.92	0/744
1	F	0.83	0/583	0.89	0/779
2	C	0.71	0/706	0.83	0/955
2	G	0.67	0/708	0.78	0/958
All	All	0.75	0/3692	0.86	0/4956

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	E	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	258	SER	Peptide

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	549	0	608	5	0
1	B	576	0	629	5	0
1	E	549	0	605	5	0
1	F	573	0	631	13	0
2	C	676	0	642	3	0
2	G	681	0	642	9	0
3	C	8	0	14	1	0
3	G	8	0	14	1	0
4	C	1	0	0	0	0
4	G	1	0	0	0	0
5	A	37	0	0	0	0
5	B	47	0	0	0	0
5	C	81	0	0	0	0
5	E	35	0	0	1	0
5	F	55	0	0	0	0
5	G	39	0	0	2	0
All	All	3916	0	3785	34	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.

All (34) 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:F:195:VAL:CG2	1:F:197:MET:SD	2.63	0.87
1:F:195:VAL:HG23	1:F:197:MET:SD	2.19	0.82
5:E:326:HOH:O	1:F:226:ARG:HD3	1.79	0.82
1:F:226:ARG:NH1	2:G:42[B]:ASN:HD21	1.84	0.75
1:F:226:ARG:CZ	2:G:42[B]:ASN:HD21	2.02	0.71
1:F:195:VAL:HG21	1:F:197:MET:SD	2.32	0.69
1:E:226[A]:ARG:HG2	1:E:226[A]:ARG:HH21	1.64	0.63
1:F:197:MET:SD	1:F:203:ARG:HA	2.39	0.62
1:B:198:GLU:HB3	1:B:202:CYS:SG	2.40	0.61
1:F:226:ARG:HG2	2:G:42[B]:ASN:ND2	2.15	0.60
1:F:195:VAL:HG23	1:F:197:MET:HG3	1.84	0.60
1:F:261:LYS:O	2:G:50:PRO:HG2	2.03	0.58
1:F:191:ILE:HB	1:F:231:VAL:HG13	1.87	0.57
1:F:195:VAL:HG23	1:F:197:MET:CG	2.36	0.56
3:G:201:MPD:O2	3:G:201:MPD:O4	2.19	0.56
2:G:37:ARG:CD	5:G:302:HOH:O	2.58	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:226[A]:ARG:HG2	1:E:226[A]:ARG:NH2	2.24	0.51
2:C:109:LYS:HE2	2:C:113:PHE:OXT	2.11	0.51
1:A:226[B]:ARG:HG3	1:A:226[B]:ARG:HH21	1.77	0.50
1:B:226:ARG:HG2	2:C:42[A]:ASN:ND2	2.27	0.49
1:E:226[B]:ARG:HH21	1:E:226[B]:ARG:HG3	1.78	0.47
1:A:217:ASP:OD1	1:B:226:ARG:NH2	2.48	0.46
1:E:194:LYS:HB2	1:E:255:LEU:HD11	1.98	0.45
1:B:191:ILE:HB	1:B:231[A]:VAL:CG1	2.47	0.45
1:A:197:MET:HA	1:A:202:CYS:HB3	1.99	0.44
2:G:37:ARG:HD2	5:G:302:HOH:O	2.16	0.43
1:A:226[A]:ARG:NH1	1:B:217:ASP:OD1	2.48	0.43
2:C:33:ILE:HG21	2:C:106:LEU:CD1	2.48	0.43
1:E:233:GLY:HA3	1:E:236:ILE:HD11	2.01	0.43
1:A:209:LEU:C	1:A:209:LEU:HD23	2.40	0.41
1:F:226:ARG:NH1	2:G:42[B]:ASN:ND2	2.62	0.41
2:G:90:ASP:OD1	2:G:92:ASN:N	2.49	0.41
3:C:201:MPD:H12	3:C:201:MPD:O4	2.20	0.41
2:G:88:LYS:O	2:G:95:LEU:HA	2.21	0.41

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	72/80 (90%)	67 (93%)	4 (6%)	1 (1%)	11 3
1	B	76/80 (95%)	73 (96%)	3 (4%)	0	100 100
1	E	73/80 (91%)	71 (97%)	2 (3%)	0	100 100
1	F	76/80 (95%)	76 (100%)	0	0	100 100
2	C	82/92 (89%)	81 (99%)	1 (1%)	0	100 100
2	G	82/92 (89%)	81 (99%)	1 (1%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	461/504 (92%)	449 (97%)	11 (2%)	1 (0%)	47 33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	197	MET

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	60/63 (95%)	59 (98%)	1 (2%)	60 47
1	B	63/63 (100%)	62 (98%)	1 (2%)	62 49
1	E	61/63 (97%)	61 (100%)	0	100 100
1	F	63/63 (100%)	61 (97%)	2 (3%)	39 22
2	C	75/82 (92%)	73 (97%)	2 (3%)	44 29
2	G	75/82 (92%)	74 (99%)	1 (1%)	69 58
All	All	397/416 (95%)	390 (98%)	7 (2%)	69 45

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	203	ARG
1	B	263	ASP
2	C	75[A]	LYS
2	C	75[B]	LYS
1	F	204[A]	SER
1	F	204[B]	SER
2	G	109	LYS

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

Mol	Chain	Res	Type
1	E	256	GLN
1	E	259	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	MPD	C	201	-	7,7,7	0.07	0	9,10,10	0.51	0
3	MPD	G	201	-	7,7,7	0.20	0	9,10,10	0.63	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MPD	C	201	-	-	1/5/5/5	-
3	MPD	G	201	-	-	2/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	201	MPD	C1-C2-C3-C4
3	G	201	MPD	O2-C2-C3-C4
3	G	201	MPD	C2-C3-C4-O4

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	201	MPD	1	0
3	G	201	MPD	1	0

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	74/80 (92%)	0.45	9 (12%)	4	4	26, 38, 85, 116	0
1	B	78/80 (97%)	0.47	7 (8%)	9	9	24, 35, 72, 114	0
1	E	73/80 (91%)	0.20	5 (6%)	17	16	28, 39, 67, 81	0
1	F	77/80 (96%)	0.32	7 (9%)	9	8	26, 33, 72, 112	0
2	C	82/92 (89%)	-0.18	1 (1%)	79	79	22, 29, 43, 54	0
2	G	83/92 (90%)	0.09	5 (6%)	21	21	30, 41, 64, 101	0
All	All	467/504 (92%)	0.22	34 (7%)	15	14	22, 36, 70, 116	0

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	185	PRO	8.0
1	B	263	ASP	7.0
1	E	200	ASN	5.5
1	F	198	GLU	5.2
1	B	198	GLU	4.9
1	F	185	PRO	4.8
2	G	31	ARG	4.7
1	A	187	LEU	4.6
1	B	184	GLY	4.4
1	A	250	GLY	4.3
2	G	32	ALA	4.1
1	A	186	GLY	4.0
1	F	197	MET	3.9
1	F	199	GLY	3.9
1	B	197	MET	3.7
1	B	185	PRO	3.7
1	F	186	GLY	3.2
1	F	184	GLY	3.2
1	A	198	GLU	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	260	ALA	3.0
1	A	199	GLY	2.9
2	G	33	ILE	2.7
1	E	187	LEU	2.5
1	B	196	ALA	2.5
1	E	198	GLU	2.4
1	B	231[A]	VAL	2.4
1	E	199	GLY	2.2
2	G	112	GLY	2.2
1	E	247	LYS	2.2
1	A	259	GLN	2.2
1	F	202	CYS	2.1
1	A	231	VAL	2.1
2	C	85	VAL	2.1
2	G	113	PHE	2.0

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 carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	MPD	C	201	8/8	0.92	0.14	50,53,61,64	0
3	MPD	G	201	8/8	0.96	0.12	52,56,63,64	0
4	CL	G	202	1/1	0.99	0.08	39,39,39,39	0
4	CL	C	202	1/1	1.00	0.07	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.