Full wwPDB X-ray Structure Validation Report (i)

Feb 19, 2016 – 09:19 PM GMT

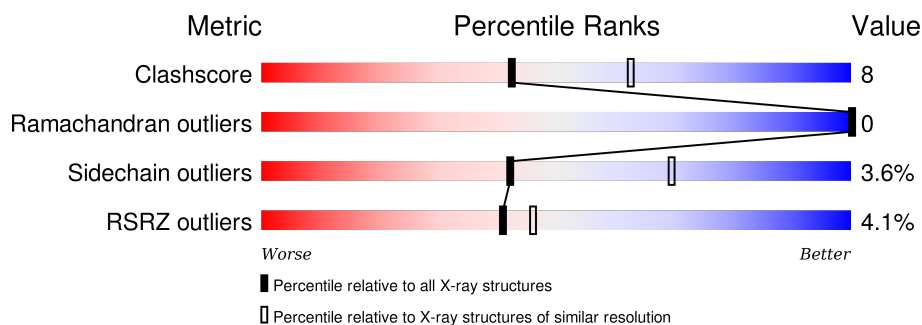|		
PDB ID	:	5A77
Title	:	Crystal structure of the homing endonuclease I-CvuI in complex with I- CreI target (C1221) in the presence of 2 mM Mg revealing DNA cleaved
Authors	:	Molina, R.; Redondo, P.; LopezMendez, B.; Villate, M.; Merino, N.; Blanco, F.J.; Valton, J.; Grizot, S.; Duchateau, P.; Prieto, J.; Montoya, G.
Deposited on	:	2015-07-03
Resolution	:	2.50 Å(reported)

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	Percentile Ranks	Value
Clashscore		8
Ramachandran outliers		0
Sidechain outliers		3.6%
RSRZ outliers		4.1%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	172	4% 74% 17% • 9%
1	B	172	5% 73% 17% • 8%
2	C	14	57% 43%
2	D	14	86% 14%
3	E	10	70% 20% 10%
3	F	10	40% 50% 10%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3471 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 DNA ENDONUCLEASE I-CVUI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	157	Total	C	N	O	S	0	0	0
			1233	791	215	221	6			
1	B	158	Total	C	N	O	S	0	0	0
			1239	794	216	223	6			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	ALA	-	EXPRESSION TAG	UNP P56347
A	54	GLN	ARG	CONFLICT	UNP P56347
A	163	SER	-	EXPRESSION TAG	UNP P56347
A	164	ALA	-	EXPRESSION TAG	UNP P56347
A	165	ALA	-	EXPRESSION TAG	UNP P56347
A	166	LEU	-	EXPRESSION TAG	UNP P56347
A	167	GLU	-	EXPRESSION TAG	UNP P56347
A	168	HIS	-	EXPRESSION TAG	UNP P56347
A	169	HIS	-	EXPRESSION TAG	UNP P56347
A	170	HIS	-	EXPRESSION TAG	UNP P56347
A	171	HIS	-	EXPRESSION TAG	UNP P56347
A	172	HIS	-	EXPRESSION TAG	UNP P56347
A	173	HIS	-	EXPRESSION TAG	UNP P56347
B	2	ALA	-	EXPRESSION TAG	UNP P56347
B	54	GLN	ARG	CONFLICT	UNP P56347
B	163	SER	-	EXPRESSION TAG	UNP P56347
B	164	ALA	-	EXPRESSION TAG	UNP P56347
B	165	ALA	-	EXPRESSION TAG	UNP P56347
B	166	LEU	-	EXPRESSION TAG	UNP P56347
B	167	GLU	-	EXPRESSION TAG	UNP P56347
B	168	HIS	-	EXPRESSION TAG	UNP P56347
B	169	HIS	-	EXPRESSION TAG	UNP P56347
B	170	HIS	-	EXPRESSION TAG	UNP P56347
B	171	HIS	-	EXPRESSION TAG	UNP P56347
B	172	HIS	-	EXPRESSION TAG	UNP P56347

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	173	HIS	-	EXPRESSION TAG	UNP P56347

- Molecule 2 is a DNA chain called 14MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP* TP*CP*GP*TP*AP*DCP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	14	Total	C	N	O	P	0	0	0
			282	136	53	80	13			
2	D	14	Total	C	N	O	P	0	0	0
			282	136	53	80	13			

- Molecule 3 is a DNA chain called 10MER DNA, 5'-D(*GP*AP*CP*GP*TP*TP*TP*TP* GP*AP*DGP*AP*CP*GP*TP*TP*TP*TP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	10	Total	C	N	O	P	0	0	0
			208	99	36	63	10			
3	F	10	Total	C	N	O	P	0	0	0
			208	99	36	63	10			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	D	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	12	Total	O	0	0
			12	12		
5	B	5	Total	O	0	0
			5	5		

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: DNA ENDONUCLEASE I-CVUI
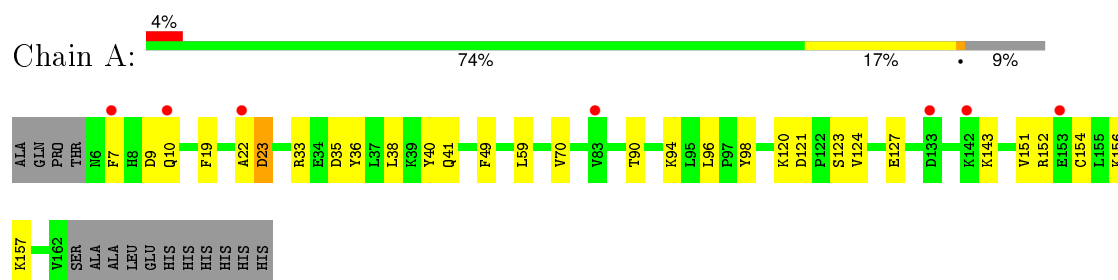
• Molecule 1: DNA ENDONUCLEASE I-CVUI
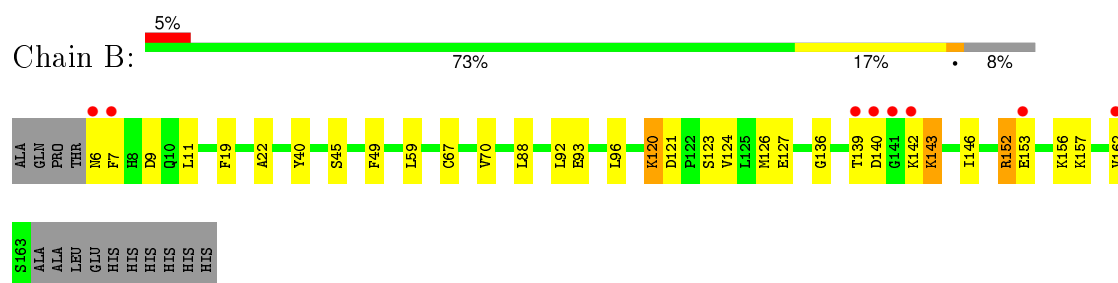
• Molecule 2: 14MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP*TP*CP*GP*TP*AP*DC P)-3'

• Molecule 2: 14MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP*TP*CP*GP*TP*AP*DC P)-3'
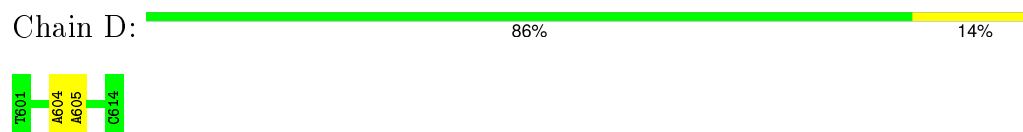
• Molecule 3: 10MER DNA, 5'-D(*GP*AP*CP*GP*TP*TP*TP*TP* GP*AP*DGP*AP*CP*G P*TP*TP*TP*TP*GP*A)-3'

Chain E:

70%   20%   10%

G515 A516 C517 G523 A524
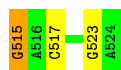
• Molecule 3: 10MER DNA, 5'-D(*GP*AP*CP*GP*TP*TP*TP*TP* GP*AP*DGP*AP*CP*G
P*TP*TP*TP*TP*GP*A)-3'

Chain F:

40%   50%   10%

G615 A616 C617 G618 T619 T620 T621 T622 G623 A624

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.57Å 95.08Å 52.15Å 90.00° 103.71° 90.00°	Depositor
Resolution (Å)	47.54 – 2.50 47.54 – 2.50	Depositor EDS
% Data completeness (in resolution range)	97.9 (47.54-2.50) 98.7 (47.54-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.86 (at 2.51Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.196 , 0.265 0.214 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	36.9	Xtriage
Anisotropy	0.970	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 35.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$<|L|> = 0.50, <L^2> = 0.34$	Xtriage
Outliers	7 of 15795 reflections (0.044%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3471	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.52% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: MG

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	0.49	0/1249	0.62	0/1683
1	B	0.44	0/1255	0.64	0/1691
2	C	0.79	0/316	1.08	0/485
2	D	0.83	0/316	1.01	0/485
3	E	1.00	1/232 (0.4%)	1.11	1/355 (0.3%)
3	F	1.05	1/232 (0.4%)	1.18	1/355 (0.3%)
All	All	0.63	2/3600 (0.1%)	0.81	2/5054 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	515	DG	OP3-P	-10.11	1.49	1.61
3	F	615	DG	OP3-P	-10.01	1.49	1.61

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	523	DG	O4'-C1'-N9	6.46	112.52	108.00
3	F	623	DG	O4'-C1'-N9	5.38	111.77	108.00

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	1233	0	1307	21	1
1	B	1239	0	1312	32	0
2	C	282	0	159	5	0
2	D	282	0	159	2	0
3	E	208	0	115	2	0
3	F	208	0	115	7	1
4	B	1	0	0	0	0
4	D	1	0	0	0	0
5	A	12	0	0	2	0
5	B	5	0	0	1	0
All	All	3471	0	3167	55	1

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 8.

All (55) 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:7:PHE:HA	1:A:10:GLN:HG2	1.57	0.84
1:B:136:GLY:O	1:B:142:LYS:NZ	2.11	0.83
1:B:139:THR:O	5:B:2002:HOH:O	2.10	0.69
1:B:40:TYR:HE1	1:B:146:ILE:HD11	1.58	0.68
1:B:152:ARG:HH12	1:B:156:LYS:NZ	1.91	0.67
1:B:143:LYS:HZ3	3:F:616:DA:H1'	1.61	0.65
2:C:507:DC:H2"	2:C:508:DG:H5'	1.82	0.61
1:A:9:ASP:OD1	5:A:2001:HOH:O	2.17	0.60
1:B:40:TYR:CE1	1:B:146:ILE:HD11	2.38	0.56
1:B:49:PHE:CE2	3:F:615:DG:H2'	2.44	0.53
1:B:126:MET:HG2	1:B:162:VAL:HG11	1.91	0.52
1:B:146:ILE:HG22	3:F:618:DG:OP1	2.09	0.51
1:A:22:ALA:HB2	1:B:19:PHE:CD1	2.45	0.51
1:B:6:ASN:CG	1:B:7:PHE:H	2.14	0.51
2:D:604:DA:H2"	2:D:605:DA:C8	2.46	0.51
1:A:59:LEU:HD23	1:A:70:VAL:HG21	1.91	0.51
1:B:153:GLU:O	1:B:157:LYS:HG3	2.11	0.50
1:B:59:LEU:HD23	1:B:70:VAL:HG21	1.93	0.50
1:B:140:ASP:HA	1:B:142:LYS:NZ	2.28	0.49
1:B:120:LYS:O	1:B:120:LYS:HE3	2.13	0.49
1:B:123:SER:O	1:B:127:GLU:HG3	2.13	0.48
1:A:143:LYS:HD2	3:E:517:DC:H1'	1.95	0.48
1:A:123:SER:O	1:A:127:GLU:HG3	2.14	0.48
1:B:152:ARG:HH12	1:B:156:LYS:HZ2	1.62	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:143:LYS:NZ	3:F:616:DA:H1'	2.31	0.46
1:B:120:LYS:HD3	2:C:502:DC:OP1	2.16	0.46
1:A:7:PHE:HD1	1:A:10:GLN:HB2	1.80	0.46
1:A:90:THR:O	1:A:94:LYS:HG2	2.16	0.45
1:A:35:ASP:OD1	1:A:35:ASP:N	2.49	0.45
1:B:45:SER:HA	1:B:88:LEU:HD22	1.99	0.45
2:D:604:DA:H2"	2:D:605:DA:H8	1.83	0.44
3:F:620:DT:H1'	3:F:621:DT:H5'	1.99	0.44
1:A:38:LEU:HD11	1:A:120:LYS:O	2.18	0.44
1:B:143:LYS:HZ3	3:F:616:DA:C1'	2.29	0.44
1:B:146:ILE:HG23	1:B:146:ILE:O	2.18	0.44
1:A:36:TYR:CE2	1:A:41:GLN:HB2	2.53	0.43
1:A:154:CYS:SG	5:A:2012:HOH:O	2.43	0.43
1:A:96:LEU:HD12	1:A:96:LEU:HA	1.88	0.43
2:C:506:DA:C5	2:C:507:DC:C5	3.08	0.42
1:A:40:TYR:CD2	1:A:151:VAL:HG22	2.55	0.42
2:C:504:DA:H2"	2:C:505:DA:C8	2.54	0.42
1:A:7:PHE:CE1	1:B:11:LEU:HG	2.55	0.41
2:C:506:DA:C4	2:C:507:DC:C6	3.09	0.41
1:A:23:ASP:HB2	1:B:22:ALA:HB1	2.01	0.41
1:A:19:PHE:CD1	1:B:22:ALA:HB2	2.56	0.41
1:B:92:LEU:HA	1:B:92:LEU:HD23	1.80	0.41
1:A:121:ASP:OD2	1:A:124:VAL:HG23	2.21	0.41
1:B:7:PHE:N	1:B:7:PHE:CD1	2.88	0.41
1:B:92:LEU:O	1:B:96:LEU:HB2	2.21	0.41
1:A:7:PHE:HE1	1:B:11:LEU:HG	1.85	0.41
1:B:143:LYS:HZ1	3:F:616:DA:H4'	1.85	0.41
1:B:152:ARG:HH12	1:B:156:LYS:HZ3	1.65	0.40
1:B:121:ASP:OD2	1:B:124:VAL:HG23	2.21	0.40
1:A:10:GLN:HB3	1:A:98:TYR:OH	2.21	0.40
1:A:49:PHE:CE2	3:E:515:DG:H2'	2.57	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:157:LYS:NZ	3:F:620:DT:OP1[2_445]	2.18	0.02

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	155/172 (90%)	149 (96%)	6 (4%)	0	100	100
1	B	156/172 (91%)	149 (96%)	7 (4%)	0	100	100
All	All	311/344 (90%)	298 (96%)	13 (4%)	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	137/149 (92%)	133 (97%)	4 (3%)	50	77
1	B	138/149 (93%)	132 (96%)	6 (4%)	35	61
All	All	275/298 (92%)	265 (96%)	10 (4%)	42	69

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	23	ASP
1	A	33	ARG
1	A	152	ARG
1	A	156	LYS
1	B	9	ASP
1	B	67	CYS
1	B	93	GLU
1	B	120	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	143	LYS
1	B	152	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	10	GLN
1	A	29	GLN
1	B	111	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 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	157/172 (91%)	0.47	7 (4%)	37	42	34, 43, 62, 77	0
1	B	158/172 (91%)	0.60	8 (5%)	32	36	33, 45, 63, 76	0
2	C	14/14 (100%)	-0.25	0	100	100	36, 44, 58, 59	0
2	D	14/14 (100%)	-0.27	0	100	100	37, 43, 52, 53	0
3	E	10/10 (100%)	-0.31	0	100	100	45, 46, 48, 49	0
3	F	10/10 (100%)	-0.08	0	100	100	43, 46, 46, 47	0
All	All	363/392 (92%)	0.43	15 (4%)	41	46	33, 44, 62, 77	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	7	PHE	4.7
1	B	141	GLY	4.2
1	A	153	GLU	3.4
1	B	140	ASP	3.1
1	B	153	GLU	3.0
1	B	6	ASN	2.9
1	B	139	THR	2.9
1	B	7	PHE	2.8
1	A	142	LYS	2.7
1	A	22	ALA	2.6
1	A	83	VAL	2.3
1	A	10	GLN	2.2
1	B	142	LYS	2.1
1	B	162	VAL	2.1
1	A	133	ASP	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
4	MG	D	1615	1/1	0.82	0.10	-2.49	52,52,52,52	0
4	MG	B	1164	1/1	0.83	0.08	-5.01	55,55,55,55	0

6.5 Other polymers (i)

There are no such residues in this entry.