Full wwPDB X-ray Structure Validation Report (i)

Jun 23, 2024 – 02:45 PM EDT

PDB ID	:	5E3E
Title	:	Crystal structure of CdiA-CT/CdiI complex from Y. kristensenii 33638
Authors	:	Michalska, K.; Joachimiak, G.; Jedrzejczak, R.; Goulding, C.W.; Joachimiak, A.; Structure-Function Analysis of Polymorphic CDI Toxin-Immunity Protein Complexes (UC4CDI); Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2015-10-02
Resolution	:	1.70 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 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:

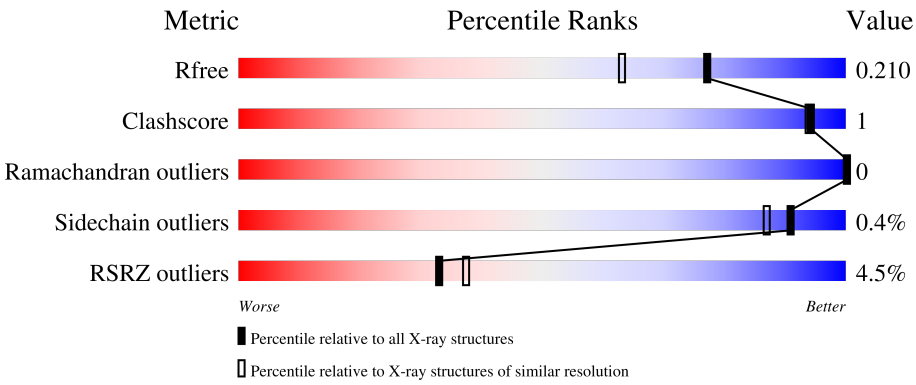
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	104	.% 95% 5%
1	C	104	2% 85% 6% 10%
1	E	104	.% 90% • • 5%
2	B	114	7% 95% 5%
2	D	114	4% 96% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	114	11% 96%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5591 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 CdiI immunity protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	99	847	546	131	165	2	3	0	1	0
1	C	94	824	530	128	161	2	3	0	3	0
1	E	99	889	571	137	175	2	4	0	6	0

There are 90 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	see remark 999	UNP C4TSS4
A	2	ILE	-	see remark 999	UNP C4TSS4
A	3	ILE	-	see remark 999	UNP C4TSS4
A	4	ASN	-	see remark 999	UNP C4TSS4
A	5	GLU	-	see remark 999	UNP C4TSS4
A	6	LYS	-	see remark 999	UNP C4TSS4
A	7	TYR	-	see remark 999	UNP C4TSS4
A	8	PRO	-	see remark 999	UNP C4TSS4
A	9	TYR	-	see remark 999	UNP C4TSS4
A	10	LEU	-	see remark 999	UNP C4TSS4
A	11	SER	-	see remark 999	UNP C4TSS4
A	12	TYR	-	see remark 999	UNP C4TSS4
A	13	LEU	-	see remark 999	UNP C4TSS4
A	14	LEU	-	see remark 999	UNP C4TSS4
A	15	ARG	-	see remark 999	UNP C4TSS4
A	16	CYS	-	see remark 999	UNP C4TSS4
A	17	TYR	-	see remark 999	UNP C4TSS4
A	18	PHE	-	see remark 999	UNP C4TSS4
A	19	ASN	-	see remark 999	UNP C4TSS4
A	20	GLN	-	see remark 999	UNP C4TSS4
A	21	ASP	-	see remark 999	UNP C4TSS4
A	22	PHE	-	see remark 999	UNP C4TSS4
A	23	GLU	-	see remark 999	UNP C4TSS4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	VAL	-	see remark 999	UNP C4TSS4
A	99	HIS	-	expression tag	UNP C4TSS4
A	100	HIS	-	expression tag	UNP C4TSS4
A	101	HIS	-	expression tag	UNP C4TSS4
A	102	HIS	-	expression tag	UNP C4TSS4
A	103	HIS	-	expression tag	UNP C4TSS4
A	104	HIS	-	expression tag	UNP C4TSS4
C	1	MSE	-	see remark 999	UNP C4TSS4
C	2	ILE	-	see remark 999	UNP C4TSS4
C	3	ILE	-	see remark 999	UNP C4TSS4
C	4	ASN	-	see remark 999	UNP C4TSS4
C	5	GLU	-	see remark 999	UNP C4TSS4
C	6	LYS	-	see remark 999	UNP C4TSS4
C	7	TYR	-	see remark 999	UNP C4TSS4
C	8	PRO	-	see remark 999	UNP C4TSS4
C	9	TYR	-	see remark 999	UNP C4TSS4
C	10	LEU	-	see remark 999	UNP C4TSS4
C	11	SER	-	see remark 999	UNP C4TSS4
C	12	TYR	-	see remark 999	UNP C4TSS4
C	13	LEU	-	see remark 999	UNP C4TSS4
C	14	LEU	-	see remark 999	UNP C4TSS4
C	15	ARG	-	see remark 999	UNP C4TSS4
C	16	CYS	-	see remark 999	UNP C4TSS4
C	17	TYR	-	see remark 999	UNP C4TSS4
C	18	PHE	-	see remark 999	UNP C4TSS4
C	19	ASN	-	see remark 999	UNP C4TSS4
C	20	GLN	-	see remark 999	UNP C4TSS4
C	21	ASP	-	see remark 999	UNP C4TSS4
C	22	PHE	-	see remark 999	UNP C4TSS4
C	23	GLU	-	see remark 999	UNP C4TSS4
C	24	VAL	-	see remark 999	UNP C4TSS4
C	99	HIS	-	expression tag	UNP C4TSS4
C	100	HIS	-	expression tag	UNP C4TSS4
C	101	HIS	-	expression tag	UNP C4TSS4
C	102	HIS	-	expression tag	UNP C4TSS4
C	103	HIS	-	expression tag	UNP C4TSS4
C	104	HIS	-	expression tag	UNP C4TSS4
E	1	MSE	-	see remark 999	UNP C4TSS4
E	2	ILE	-	see remark 999	UNP C4TSS4
E	3	ILE	-	see remark 999	UNP C4TSS4
E	4	ASN	-	see remark 999	UNP C4TSS4
E	5	GLU	-	see remark 999	UNP C4TSS4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	6	LYS	-	see remark 999	UNP C4TSS4
E	7	TYR	-	see remark 999	UNP C4TSS4
E	8	PRO	-	see remark 999	UNP C4TSS4
E	9	TYR	-	see remark 999	UNP C4TSS4
E	10	LEU	-	see remark 999	UNP C4TSS4
E	11	SER	-	see remark 999	UNP C4TSS4
E	12	TYR	-	see remark 999	UNP C4TSS4
E	13	LEU	-	see remark 999	UNP C4TSS4
E	14	LEU	-	see remark 999	UNP C4TSS4
E	15	ARG	-	see remark 999	UNP C4TSS4
E	16	CYS	-	see remark 999	UNP C4TSS4
E	17	TYR	-	see remark 999	UNP C4TSS4
E	18	PHE	-	see remark 999	UNP C4TSS4
E	19	ASN	-	see remark 999	UNP C4TSS4
E	20	GLN	-	see remark 999	UNP C4TSS4
E	21	ASP	-	see remark 999	UNP C4TSS4
E	22	PHE	-	see remark 999	UNP C4TSS4
E	23	GLU	-	see remark 999	UNP C4TSS4
E	24	VAL	-	see remark 999	UNP C4TSS4
E	99	HIS	-	expression tag	UNP C4TSS4
E	100	HIS	-	expression tag	UNP C4TSS4
E	101	HIS	-	expression tag	UNP C4TSS4
E	102	HIS	-	expression tag	UNP C4TSS4
E	103	HIS	-	expression tag	UNP C4TSS4
E	104	HIS	-	expression tag	UNP C4TSS4

- Molecule 2 is a protein called Large exoprotein involved in heme utilization or adhesion.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
2	B	114	871	535	157	178	1	0	2	0
			Total	C	N	O	Se			
2	D	113	847	521	150	175	1	0	0	0
			Total	C	N	O	Se			
2	F	114	875	536	156	182	1	0	3	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
3	B	1	1	1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
3	E	1	1	1	0	0

- Molecule 4 is water.

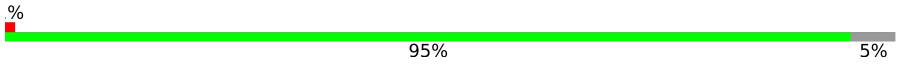
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	92	92	92	0	3
			Total	O		
4	B	85	85	85	0	2
			Total	O		
4	C	81	81	81	0	2
			Total	O		
4	D	56	56	56	0	0
			Total	O		
4	E	69	69	69	0	1
			Total	O		
4	F	53	53	53	0	0

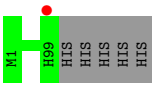
3 Residue-property plots (i)

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: CdiI immunity protein

Chain A:



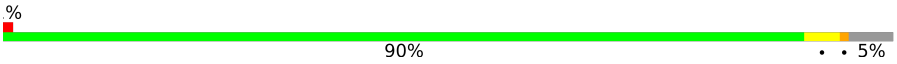


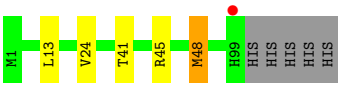
• Molecule 1: CdiI immunity protein

Chain C:

• Molecule 1: CdiI immunity protein

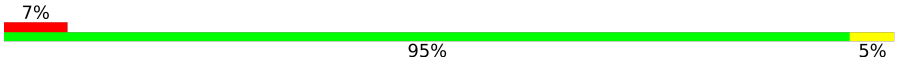
Chain E:

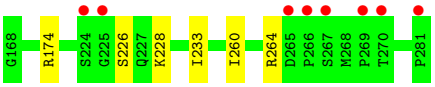




• Molecule 2: Large exoprotein involved in heme utilization or adhesion

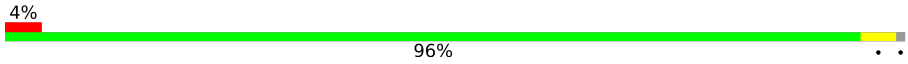
Chain B:

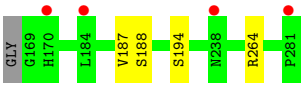




• Molecule 2: Large exoprotein involved in heme utilization or adhesion

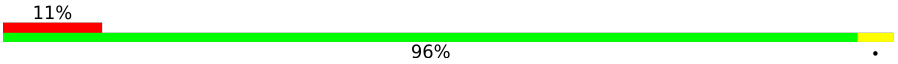
Chain D:

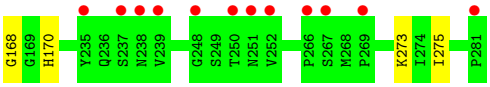




• Molecule 2: Large exoprotein involved in heme utilization or adhesion

Chain F:


11%
96%


G168
G169
H170
Y235
Q236
S237
N238
V239
G248
S249
T250
N251
V252
P266
S267
M268
P269
K273
I274
I275
P281

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	48.89Å 57.15Å 64.30Å 78.25° 90.21° 79.11°	Depositor
Resolution (Å)	30.00 – 1.70 27.44 – 1.70	Depositor EDS
% Data completeness (in resolution range)	96.8 (30.00-1.70) 96.8 (27.44-1.70)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.8.0073, PHENIX	Depositor
R, R_{free}	0.174 , 0.202 0.183 , 0.210	Depositor DCC
R_{free} test set	1423 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å^2)	18.0	Xtriage
Anisotropy	0.174	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 45.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5591	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.76% 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)

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

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	0.64	0/866	0.73	0/1170
1	C	0.65	0/842	0.66	0/1136
1	E	0.62	0/908	0.72	2/1226 (0.2%)
2	B	0.60	0/882	0.69	0/1191
2	D	0.53	0/858	0.69	1/1161 (0.1%)
2	F	0.54	0/886	0.70	0/1199
All	All	0.60	0/5242	0.70	3/7083 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	264	ARG	NE-CZ-NH1	-5.50	117.55	120.30
1	E	48[A]	MSE	CA-CB-CG	-5.04	104.73	113.30
1	E	48[B]	MSE	CA-CB-CG	-5.04	104.73	113.30

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	847	0	807	0	0
1	C	824	0	780	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	889	0	840	4	0
2	B	871	0	873	4	0
2	D	847	0	847	1	0
2	F	875	0	868	2	0
3	B	1	0	0	0	0
3	E	1	0	0	0	0
4	A	92	0	0	0	0
4	B	85	0	0	0	0
4	C	81	0	0	0	0
4	D	56	0	0	0	0
4	E	69	0	0	0	0
4	F	53	0	0	1	0
All	All	5591	0	5015	13	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 1.

All (13) 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:13:LEU:HD22	1:C:48[B]:MSE:SE	2.47	0.65
1:C:37:LYS:NZ	1:C:97:MSE:O	2.43	0.51
1:E:13:LEU:HD22	1:E:48[A]:MSE:SE	2.60	0.50
2:F:273:LYS:HE3	2:F:275:ILE:HD13	1.95	0.49
1:E:41:THR:O	1:E:45[A]:ARG:HG3	2.15	0.47
2:B:226:SER:O	2:B:264[A]:ARG:NH1	2.49	0.46
2:B:228[A]:LYS:HE2	2:B:228[A]:LYS:HB3	1.80	0.43
1:C:11:SER:OG	1:C:15[A]:ARG:HD2	2.17	0.43
1:E:48[A]:MSE:HE2	1:E:48[A]:MSE:HB3	1.80	0.41
2:B:174:ARG:HD2	1:E:24:VAL:O	2.21	0.41
2:D:187:VAL:HG11	2:D:194:SER:HA	2.03	0.40
2:B:233:ILE:HG13	2:B:260:ILE:HB	2.02	0.40
2:F:168:GLY:HA3	4:F:304:HOH:O	2.21	0.40

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	98/104 (94%)	94 (96%)	4 (4%)	0	100	100
1	C	95/104 (91%)	93 (98%)	2 (2%)	0	100	100
1	E	103/104 (99%)	99 (96%)	4 (4%)	0	100	100
2	B	114/114 (100%)	113 (99%)	1 (1%)	0	100	100
2	D	111/114 (97%)	110 (99%)	1 (1%)	0	100	100
2	F	115/114 (101%)	113 (98%)	2 (2%)	0	100	100
All	All	636/654 (97%)	622 (98%)	14 (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	92/93 (99%)	92 (100%)	0	100	100
1	C	89/93 (96%)	89 (100%)	0	100	100
1	E	97/93 (104%)	97 (100%)	0	100	100
2	B	98/95 (103%)	98 (100%)	0	100	100
2	D	96/95 (101%)	95 (99%)	1 (1%)	76	67
2	F	99/95 (104%)	98 (99%)	1 (1%)	76	67
All	All	571/564 (101%)	569 (100%)	2 (0%)	91	87

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	188	SER
2	F	170	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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	96/104 (92%)	-0.12	1 (1%)	82	85	11, 14, 22, 53	0
1	C	92/104 (88%)	-0.08	2 (2%)	62	66	12, 18, 26, 53	0
1	E	96/104 (92%)	-0.20	1 (1%)	82	85	13, 18, 27, 64	0
2	B	113/114 (99%)	0.03	8 (7%)	16	18	12, 20, 35, 63	0
2	D	112/114 (98%)	0.24	4 (3%)	42	47	13, 29, 47, 56	0
2	F	113/114 (99%)	0.31	12 (10%)	6	7	17, 25, 44, 55	0
All	All	622/654 (95%)	0.04	28 (4%)	33	37	11, 20, 41, 64	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	99	HIS	5.4
2	F	238	ASN	5.2
1	E	99	HIS	5.0
2	B	266	PRO	5.0
2	F	250	THR	4.7
2	D	238	ASN	4.2
2	F	266	PRO	3.9
2	F	248	GLY	3.6
1	C	5	GLU	3.6
2	B	225	GLY	3.5
2	D	170	HIS	3.5
2	B	267	SER	3.1
2	F	252	VAL	2.8
2	F	269	PRO	2.7
2	B	224	SER	2.7
2	F	235	TYR	2.6
2	B	265	ASP	2.6
2	B	270	THR	2.6
2	B	281	PRO	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	269	PRO	2.4
2	D	281	PRO	2.4
2	F	239	VAL	2.3
2	F	267	SER	2.3
1	C	98	ASN	2.3
2	F	237	SER	2.2
2	F	281	PRO	2.2
2	F	251	ASN	2.1
2	D	184	LEU	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 monosaccharides 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	NA	B	301	1/1	0.94	0.24	32,32,32,32	0
3	NA	E	201	1/1	0.98	0.17	24,24,24,24	0

6.5 Other polymers (i)

There are no such residues in this entry.