WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2017 – 04:22 AM EDT

PDB ID : 2FZV

Title : Crystal Structure of an apo form of a Flavin-binding Protein from Shigella flexneri

Authors : Vorontsov, I.I.; Minasov, G.; Brunzelle, J.S.; Shuvalova, L.; Collart, F.R.; Joachimiak, A.; Anderson, W.F.; Midwest Center for Structural Genomics (MCSG)

Deposited on : unknown

Resolution : 1.70 Å(reported)

The following versions of software and data (see references (1)) were used in the production of this report:

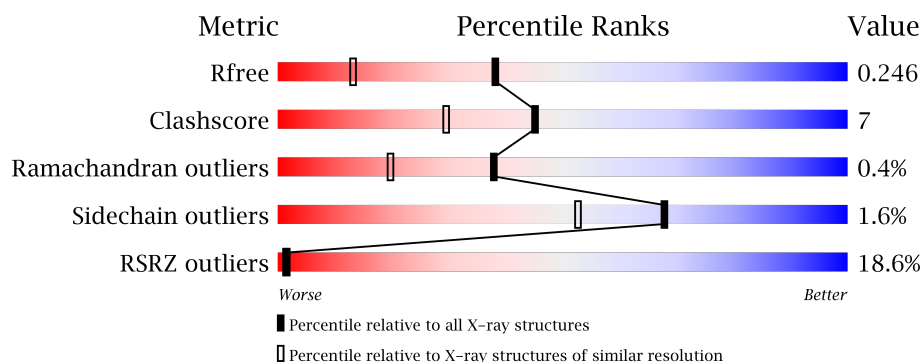|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

1 Overall quality at a glance ⓘ

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	Percentile Ranks	Value
Rfree		0.246
Clashscore		7
Ramachandran outliers		0.4%
Sidechain outliers		1.6%
RSRZ outliers		18.6%

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))
R_{free}	100719	3453 (1.70-1.70)
Clashscore	112137	3876 (1.70-1.70)
Ramachandran outliers	110173	3815 (1.70-1.70)
Sidechain outliers	110143	3815 (1.70-1.70)
RSRZ outliers	101464	3491 (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 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	279	9% 74% 9% · 16%
1	B	279	17% 77% 11% · 11%
1	C	279	13% 76% 8% · 15%
1	D	279	24% 71% 12% · 15%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 9117 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 putative arsenical resistance protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	235	Total	C	N	O	S	0	19	0
			2009	1255	375	370	9			
1	B	247	Total	C	N	O	S	0	19	0
			2090	1307	383	391	9			
1	C	236	Total	C	N	O	S	0	12	0
			1977	1242	368	359	8			
1	D	236	Total	C	N	O	S	0	6	0
			1909	1197	351	353	8			

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-23	MET	-	INITIATING METHIONINE	GB 30042143
A	-22	HIS	-	EXPRESSION TAG	GB 30042143
A	-21	HIS	-	EXPRESSION TAG	GB 30042143
A	-20	HIS	-	EXPRESSION TAG	GB 30042143
A	-19	HIS	-	EXPRESSION TAG	GB 30042143
A	-18	HIS	-	EXPRESSION TAG	GB 30042143
A	-17	HIS	-	EXPRESSION TAG	GB 30042143
A	-16	SER	-	CLONING ARTIFACT	GB 30042143
A	-15	SER	-	CLONING ARTIFACT	GB 30042143
A	-14	GLY	-	CLONING ARTIFACT	GB 30042143
A	-13	VAL	-	CLONING ARTIFACT	GB 30042143
A	-12	ASP	-	CLONING ARTIFACT	GB 30042143
A	-11	LEU	-	CLONING ARTIFACT	GB 30042143
A	-10	GLY	-	CLONING ARTIFACT	GB 30042143
A	-9	THR	-	CLONING ARTIFACT	GB 30042143
A	-8	GLU	-	CLONING ARTIFACT	GB 30042143
A	-7	ASN	-	CLONING ARTIFACT	GB 30042143
A	-6	LEU	-	CLONING ARTIFACT	GB 30042143
A	-5	TYR	-	CLONING ARTIFACT	GB 30042143
A	-4	PHE	-	CLONING ARTIFACT	GB 30042143
A	-3	GLN	-	CLONING ARTIFACT	GB 30042143

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	CLONING ARTIFACT	GB 30042143
A	-1	ASN	-	CLONING ARTIFACT	GB 30042143
A	0	ALA	-	CLONING ARTIFACT	GB 30042143
B	-23	MET	-	INITIATING METHIONINE	GB 30042143
B	-22	HIS	-	EXPRESSION TAG	GB 30042143
B	-21	HIS	-	EXPRESSION TAG	GB 30042143
B	-20	HIS	-	EXPRESSION TAG	GB 30042143
B	-19	HIS	-	EXPRESSION TAG	GB 30042143
B	-18	HIS	-	EXPRESSION TAG	GB 30042143
B	-17	HIS	-	EXPRESSION TAG	GB 30042143
B	-16	SER	-	CLONING ARTIFACT	GB 30042143
B	-15	SER	-	CLONING ARTIFACT	GB 30042143
B	-14	GLY	-	CLONING ARTIFACT	GB 30042143
B	-13	VAL	-	CLONING ARTIFACT	GB 30042143
B	-12	ASP	-	CLONING ARTIFACT	GB 30042143
B	-11	LEU	-	CLONING ARTIFACT	GB 30042143
B	-10	GLY	-	CLONING ARTIFACT	GB 30042143
B	-9	THR	-	CLONING ARTIFACT	GB 30042143
B	-8	GLU	-	CLONING ARTIFACT	GB 30042143
B	-7	ASN	-	CLONING ARTIFACT	GB 30042143
B	-6	LEU	-	CLONING ARTIFACT	GB 30042143
B	-5	TYR	-	CLONING ARTIFACT	GB 30042143
B	-4	PHE	-	CLONING ARTIFACT	GB 30042143
B	-3	GLN	-	CLONING ARTIFACT	GB 30042143
B	-2	SER	-	CLONING ARTIFACT	GB 30042143
B	-1	ASN	-	CLONING ARTIFACT	GB 30042143
B	0	ALA	-	CLONING ARTIFACT	GB 30042143
C	-23	MET	-	INITIATING METHIONINE	GB 30042143
C	-22	HIS	-	EXPRESSION TAG	GB 30042143
C	-21	HIS	-	EXPRESSION TAG	GB 30042143
C	-20	HIS	-	EXPRESSION TAG	GB 30042143
C	-19	HIS	-	EXPRESSION TAG	GB 30042143
C	-18	HIS	-	EXPRESSION TAG	GB 30042143
C	-17	HIS	-	EXPRESSION TAG	GB 30042143
C	-16	SER	-	CLONING ARTIFACT	GB 30042143
C	-15	SER	-	CLONING ARTIFACT	GB 30042143
C	-14	GLY	-	CLONING ARTIFACT	GB 30042143
C	-13	VAL	-	CLONING ARTIFACT	GB 30042143
C	-12	ASP	-	CLONING ARTIFACT	GB 30042143
C	-11	LEU	-	CLONING ARTIFACT	GB 30042143
C	-10	GLY	-	CLONING ARTIFACT	GB 30042143
C	-9	THR	-	CLONING ARTIFACT	GB 30042143

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	GLU	-	CLONING ARTIFACT	GB 30042143
C	-7	ASN	-	CLONING ARTIFACT	GB 30042143
C	-6	LEU	-	CLONING ARTIFACT	GB 30042143
C	-5	TYR	-	CLONING ARTIFACT	GB 30042143
C	-4	PHE	-	CLONING ARTIFACT	GB 30042143
C	-3	GLN	-	CLONING ARTIFACT	GB 30042143
C	-2	SER	-	CLONING ARTIFACT	GB 30042143
C	-1	ASN	-	CLONING ARTIFACT	GB 30042143
C	0	ALA	-	CLONING ARTIFACT	GB 30042143
D	-23	MET	-	INITIATING METHIONINE	GB 30042143
D	-22	HIS	-	EXPRESSION TAG	GB 30042143
D	-21	HIS	-	EXPRESSION TAG	GB 30042143
D	-20	HIS	-	EXPRESSION TAG	GB 30042143
D	-19	HIS	-	EXPRESSION TAG	GB 30042143
D	-18	HIS	-	EXPRESSION TAG	GB 30042143
D	-17	HIS	-	EXPRESSION TAG	GB 30042143
D	-16	SER	-	CLONING ARTIFACT	GB 30042143
D	-15	SER	-	CLONING ARTIFACT	GB 30042143
D	-14	GLY	-	CLONING ARTIFACT	GB 30042143
D	-13	VAL	-	CLONING ARTIFACT	GB 30042143
D	-12	ASP	-	CLONING ARTIFACT	GB 30042143
D	-11	LEU	-	CLONING ARTIFACT	GB 30042143
D	-10	GLY	-	CLONING ARTIFACT	GB 30042143
D	-9	THR	-	CLONING ARTIFACT	GB 30042143
D	-8	GLU	-	CLONING ARTIFACT	GB 30042143
D	-7	ASN	-	CLONING ARTIFACT	GB 30042143
D	-6	LEU	-	CLONING ARTIFACT	GB 30042143
D	-5	TYR	-	CLONING ARTIFACT	GB 30042143
D	-4	PHE	-	CLONING ARTIFACT	GB 30042143
D	-3	GLN	-	CLONING ARTIFACT	GB 30042143
D	-2	SER	-	CLONING ARTIFACT	GB 30042143
D	-1	ASN	-	CLONING ARTIFACT	GB 30042143
D	0	ALA	-	CLONING ARTIFACT	GB 30042143

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total Ca 1 1		0	1

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total 2	Cl 2	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	357	Total 366	O 366	0	16
4	B	285	Total 288	O 288	0	6
4	C	285	Total 286	O 286	0	5
4	D	184	Total 189	O 189	0	6

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 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.
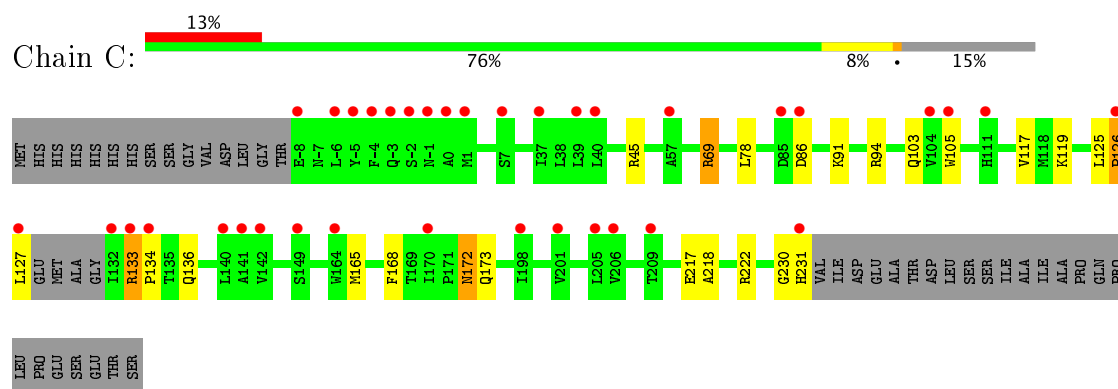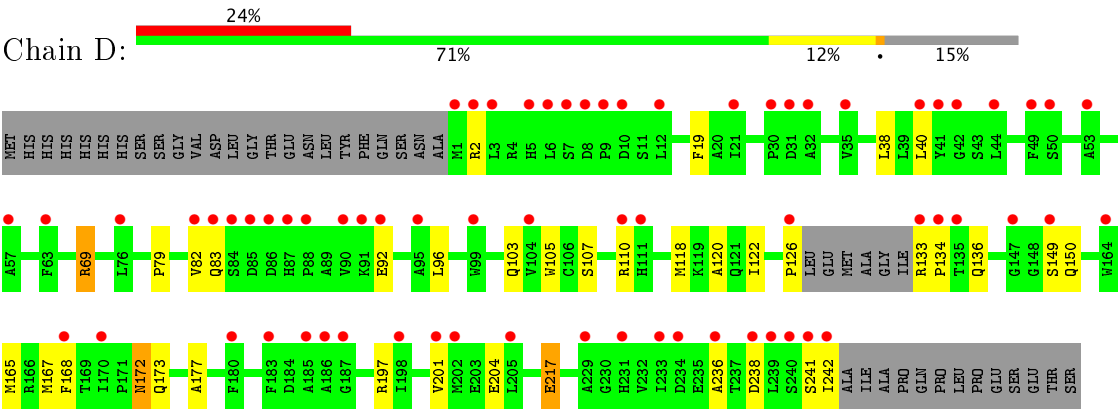
- Molecule 1: putative arsenical resistance protein
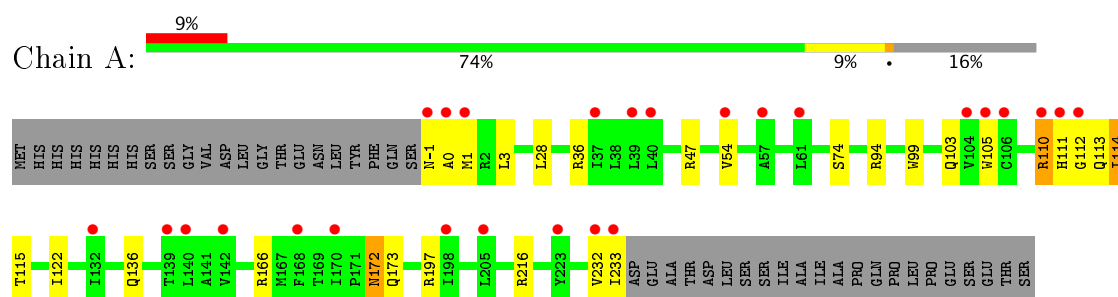
- Molecule 1: putative arsenical resistance protein

- Molecule 1: putative arsenical resistance protein

- Molecule 1: putative arsenical resistance protein

Chain D:
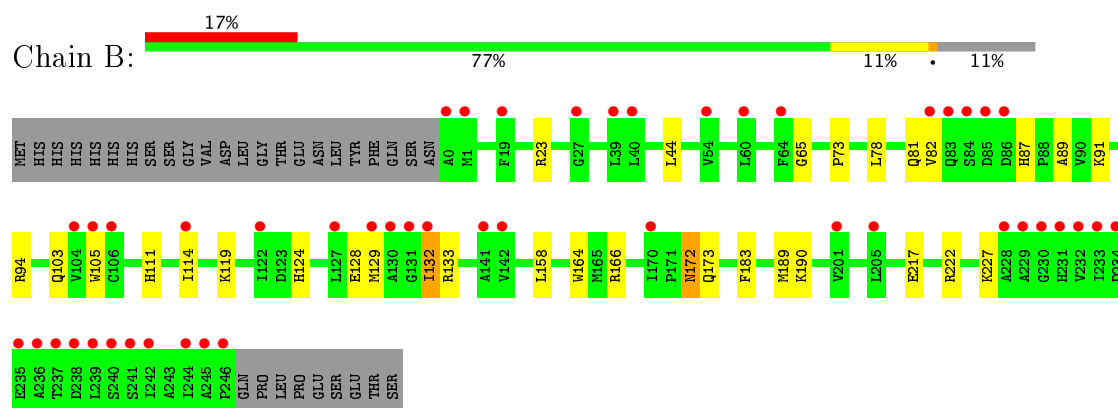
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	117.82Å 117.82Å 154.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.81 – 1.70 29.81 – 1.70	Depositor EDS
% Data completeness (in resolution range)	98.3 (29.81-1.70) 98.3 (29.81-1.70)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.29 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.199 , 0.246 0.200 , 0.246	Depositor DCC
R_{free} test set	5843 reflections (5.27%)	DCC
Wilson B-factor (Å^2)	20.7	Xtriage
Anisotropy	0.191	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 35.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9117	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.51% 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: CA, 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	0.49	0/2052	0.63	0/2774
1	B	0.42	0/2133	0.63	0/2888
1	C	0.44	0/2023	0.62	0/2738
1	D	0.58	2/1950 (0.1%)	0.61	1/2640 (0.0%)
All	All	0.48	2/8158 (0.0%)	0.62	1/11040 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	238	ASP	CG-OD1	15.56	1.61	1.25
1	D	238	ASP	CG-OD2	11.03	1.50	1.25

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	238	ASP	CB-CG-OD2	-5.71	113.17	118.30

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	2009	0	1993	38	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2090	0	2077	32	0
1	C	1977	0	1959	26	0
1	D	1909	0	1898	31	0
2	A	1	0	0	0	0
3	A	2	0	0	0	0
4	A	366	0	0	7	0
4	B	288	0	0	6	0
4	C	286	0	0	2	0
4	D	189	0	0	3	0
All	All	9117	0	7927	119	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 7.

The worst 5 of 119 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:129:MET:O	1:B:132[A]:ILE:HG22	1.65	0.96
1:B:23[B]:ARG:HH21	1:B:65:GLY:HA3	1.31	0.94
1:D:133:ARG:O	1:D:136:GLN:HG3	1.74	0.88
1:C:69:ARG:HH21	1:C:69:ARG:CG	1.88	0.86
1:B:23[B]:ARG:NH2	1:B:65:GLY:HA3	1.88	0.85

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	252/279 (90%)	244 (97%)	4 (2%)	4 (2%)	11	2
1	B	264/279 (95%)	254 (96%)	10 (4%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	244/279 (88%)	237 (97%)	6 (2%)	1 (0%)	38	20
1	D	237/279 (85%)	229 (97%)	7 (3%)	1 (0%)	38	20
All	All	997/1116 (89%)	964 (97%)	27 (3%)	6 (1%)	38	12

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	114[A]	ILE
1	A	114[B]	ILE
1	C	126	PRO
1	A	110[A]	ARG
1	A	110[B]	ARG

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	215/236 (91%)	213 (99%)	2 (1%)	82	74
1	B	225/236 (95%)	222 (99%)	3 (1%)	73	60
1	C	212/236 (90%)	207 (98%)	5 (2%)	54	35
1	D	205/236 (87%)	200 (98%)	5 (2%)	54	35
All	All	857/944 (91%)	842 (98%)	15 (2%)	68	47

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	133	ARG
1	C	168	PHE
1	D	172	ASN
1	C	69	ARG
1	D	168	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	62	GLN
1	C	103	GLN
1	D	172	ASN
1	C	83	GLN
1	C	113	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 3 ligands modelled in this entry, 3 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(Å²)	Q<0.9
1	A	235/279 (84%)	0.87	26 (11%)	6	7	14, 20, 33, 49	0
1	B	247/279 (88%)	1.24	47 (19%)	1	1	15, 20, 41, 49	0
1	C	236/279 (84%)	1.01	36 (15%)	2	3	10, 20, 36, 55	0
1	D	236/279 (84%)	1.65	68 (28%)	1	0	11, 21, 35, 57	0
All	All	954/1116 (85%)	1.19	177 (18%)	1	1	10, 20, 36, 57	0

The worst 5 of 177 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	242	ILE	15.4
1	C	-1	ASN	12.7
1	A	233	ILE	9.2
1	D	86	ASP	8.4
1	C	127	LEU	7.9

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
3	CL	A	257	1/1	0.99	0.17	1.76	13,13,13,13	0
2	CA	A	256[B]	1/1	0.99	0.10	-	8,8,8,8	1
3	CL	A	258	1/1	0.99	0.17	-	11,11,11,11	0

6.5 Other polymers (i)

There are no such residues in this entry.