



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:20 PM GMT

PDB ID : 4RI8  
Title : FAN1 Nuclease bound to 5' phosphorylated p(dG)/3'(dT-dT-dT-dT) double flap DNA  
Authors : Pavletich, N.P.; Wang, R.  
Deposited on : 2014-10-05  
Resolution : 2.90 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

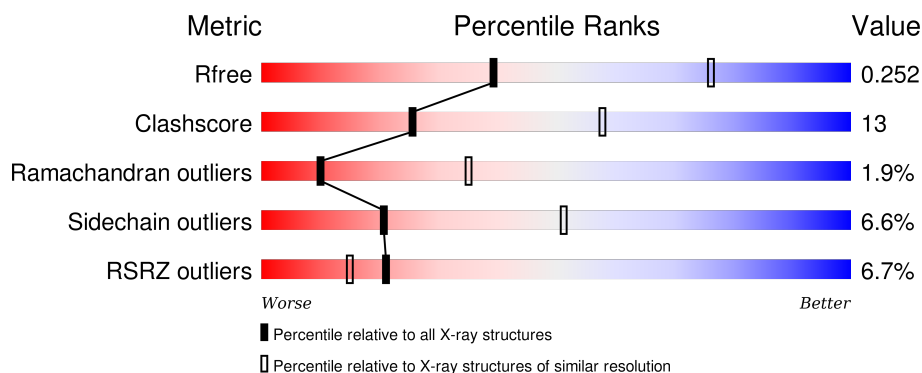
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	91344	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	651	<div> <div>7%</div> <div>67%</div> <div>24%</div> <div>6%</div> </div>
1	B	651	<div> <div>7%</div> <div>67%</div> <div>25%</div> <div>6%</div> </div>
2	E	14	<div> <div>57%</div> <div>43%</div> </div>
2	H	14	<div> <div>57%</div> <div>43%</div> </div>
3	F	8	<div> <div>50%</div> <div>38%</div> <div>13%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	I	8	 50% 38% 13%
4	G	19	 68% 32%
4	J	19	 63% 37%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11558 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 Fanconi-associated nuclease 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	615	Total	C	N	O	S	0	0	0
			4942	3145	880	890	27			
1	B	615	Total	C	N	O	S	0	0	0
			4942	3145	880	890	27			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	358	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
A	359	ALA	-	EXPRESSION TAG	UNP Q9Y2M0
A	360	HIS	-	EXPRESSION TAG	UNP Q9Y2M0
A	361	MET	-	EXPRESSION TAG	UNP Q9Y2M0
A	362	THR	-	EXPRESSION TAG	UNP Q9Y2M0
A	363	ARG	-	EXPRESSION TAG	UNP Q9Y2M0
A	364	ASN	-	EXPRESSION TAG	UNP Q9Y2M0
A	365	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
A	366	PRO	-	EXPRESSION TAG	UNP Q9Y2M0
A	367	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
A	368	GLN	-	EXPRESSION TAG	UNP Q9Y2M0
A	369	THR	-	EXPRESSION TAG	UNP Q9Y2M0
A	487	ALA	VAL	ENGINEERED MUTATION	UNP Q9Y2M0
A	?	-	CYS	DELETION	UNP Q9Y2M0
A	?	-	THR	DELETION	UNP Q9Y2M0
A	?	-	TRP	DELETION	UNP Q9Y2M0
A	?	-	GLY	DELETION	UNP Q9Y2M0
A	?	-	LYS	DELETION	UNP Q9Y2M0
A	?	-	ASN	DELETION	UNP Q9Y2M0
A	?	-	LYS	DELETION	UNP Q9Y2M0
A	?	-	PRO	DELETION	UNP Q9Y2M0
A	?	-	GLY	DELETION	UNP Q9Y2M0
B	358	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
B	359	ALA	-	EXPRESSION TAG	UNP Q9Y2M0
B	360	HIS	-	EXPRESSION TAG	UNP Q9Y2M0

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	361	MET	-	EXPRESSION TAG	UNP Q9Y2M0
B	362	THR	-	EXPRESSION TAG	UNP Q9Y2M0
B	363	ARG	-	EXPRESSION TAG	UNP Q9Y2M0
B	364	ASN	-	EXPRESSION TAG	UNP Q9Y2M0
B	365	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
B	366	PRO	-	EXPRESSION TAG	UNP Q9Y2M0
B	367	GLY	-	EXPRESSION TAG	UNP Q9Y2M0
B	368	GLN	-	EXPRESSION TAG	UNP Q9Y2M0
B	369	THR	-	EXPRESSION TAG	UNP Q9Y2M0
B	487	ALA	VAL	ENGINEERED MUTATION	UNP Q9Y2M0
B	?	-	CYS	DELETION	UNP Q9Y2M0
B	?	-	THR	DELETION	UNP Q9Y2M0
B	?	-	TRP	DELETION	UNP Q9Y2M0
B	?	-	GLY	DELETION	UNP Q9Y2M0
B	?	-	LYS	DELETION	UNP Q9Y2M0
B	?	-	ASN	DELETION	UNP Q9Y2M0
B	?	-	LYS	DELETION	UNP Q9Y2M0
B	?	-	PRO	DELETION	UNP Q9Y2M0
B	?	-	GLY	DELETION	UNP Q9Y2M0

- Molecule 2 is a DNA chain called DNA (5'-D(\*TP\*TP\*TP\*GP\*AP\*GP\*GP\*AP\*GP\*TP\*CP\*TP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	14	Total	C	N	O	P	0	0	0
			286	139	47	87	13			
2	H	14	Total	C	N	O	P	0	0	0
			286	139	47	87	13			

- Molecule 3 is a DNA chain called DNA (5'-D(P\*GP\*AP\*GP\*GP\*CP\*GP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	8	Total	C	N	O	P	0	0	0
			171	79	35	49	8			
3	I	8	Total	C	N	O	P	0	0	0
			171	79	35	49	8			

- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*AP\*CP\*AP\*CP\*GP\*CP\*CP\*TP\*AP\*GP\*AP\*CP\*TP\*CP\*CP\*TP\*CP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	19	Total	C	N	O	P	0	0	0
			379	182	70	109	18			
4	J	19	Total	C	N	O	P	0	0	0
			379	182	70	109	18			

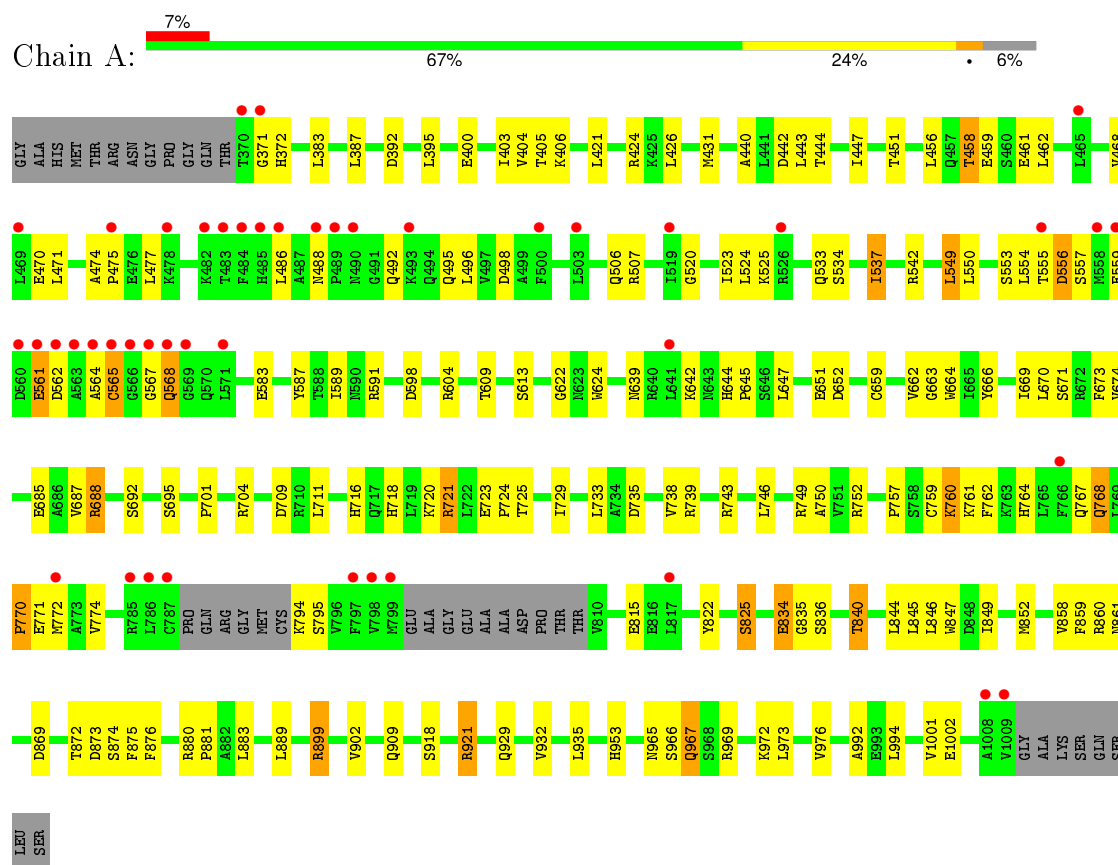
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Ca	0	0
			1	1		
5	A	1	Total	Ca	0	0
			1	1		

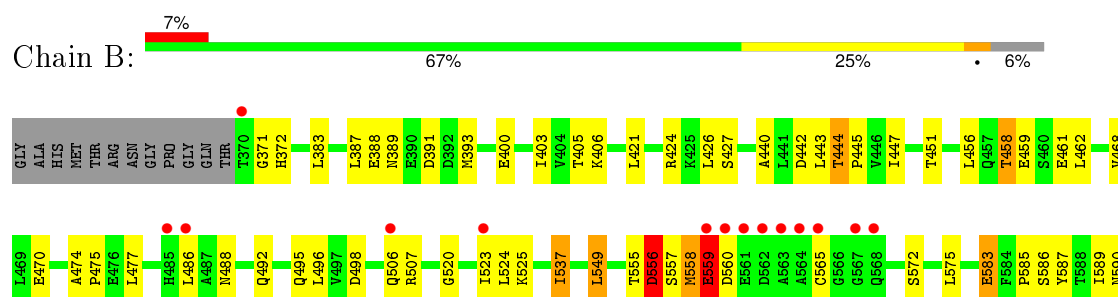
### 3 Residue-property plots

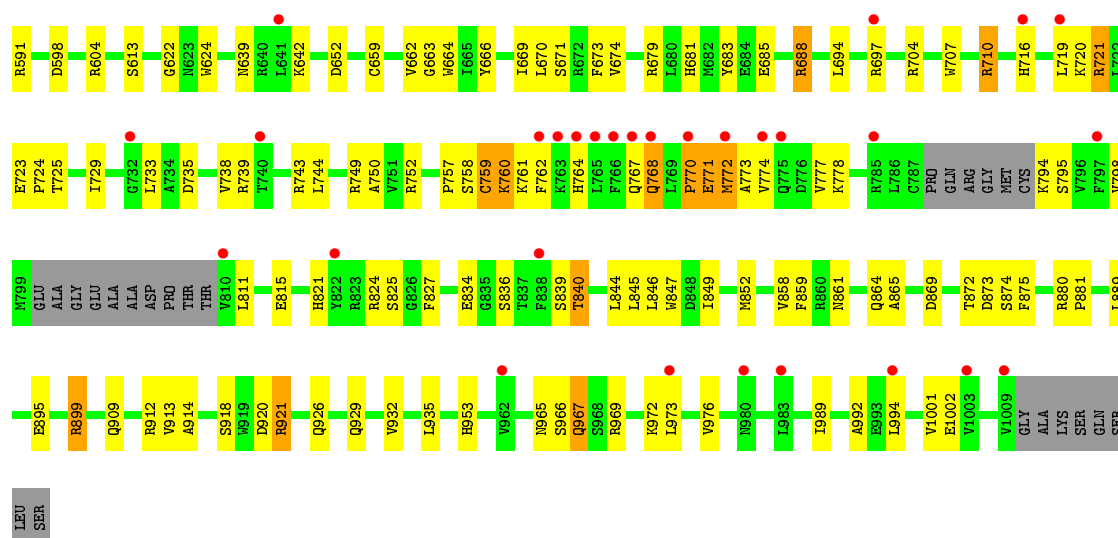
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Fanconi-associated nuclease 1

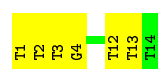


- Molecule 1: Fanconi-associated nuclease 1

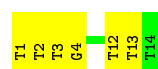




- Molecule 2: DNA (5'-D(\*TP\*TP\*TP\*GP\*AP\*GP\*GP\*AP\*GP\*TP\*CP\*TP\*TP\*T)-3')



- Molecule 2: DNA (5'-D(\*TP\*TP\*TP\*GP\*AP\*GP\*GP\*AP\*GP\*TP\*CP\*TP\*TP\*T)-3')



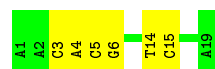
- Molecule 3: DNA (5'-D(P\*GP\*AP\*GP\*GP\*CP\*GP\*TP\*G)-3')



- Molecule 3: DNA (5'-D(P\*GP\*AP\*GP\*GP\*CP\*GP\*TP\*G)-3')

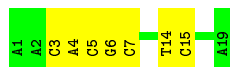


- Molecule 4: DNA (5'-D(\*AP\*AP\*CP\*AP\*CP\*GP\*CP\*CP\*TP\*AP\*GP\*AP\*CP\*TP\*CP\*CP\*TP\*CP\*A)-3')



- Molecule 4: DNA (5'-D(\*AP\*AP\*CP\*AP\*CP\*GP\*CP\*CP\*TP\*AP\*GP\*AP\*CP\*TP\*CP\*CP\*TP\*CP\*A)-3')

Chain J:  63% 37%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	62.26 Å   210.88 Å   83.22 Å 90.00°   107.61°   90.00°	Depositor
Resolution (Å)	70.00 – 2.90 74.24 – 2.90	Depositor EDS
% Data completeness (in resolution range)	90.1 (70.00-2.90) 90.1 (74.24-2.90)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.95 (at 2.91 Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.221   ,   0.255 0.219   ,   0.252	Depositor DCC
$R_{free}$ test set	1716 reflections (4.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	66.2	Xtriage
Anisotropy	0.636	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 49.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 42242 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11558	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	97.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.12% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: CA

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	0.51	0/5044	0.70	1/6818 (0.0%)
1	B	0.46	0/5044	0.67	1/6818 (0.0%)
2	E	0.42	0/319	0.70	0/492
2	H	0.38	0/319	0.70	0/492
3	F	0.85	1/192 (0.5%)	0.68	0/294
3	I	0.82	1/192 (0.5%)	0.68	0/294
4	G	0.41	0/424	0.74	0/650
4	J	0.37	0/424	0.73	0/650
All	All	0.49	2/11958 (0.0%)	0.69	2/16508 (0.0%)

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	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	0	DG	OP3-P	-10.74	1.48	1.61
3	I	0	DG	OP3-P	-10.53	1.48	1.61

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	549	LEU	CA-CB-CG	-5.34	103.02	115.30
1	B	549	LEU	CA-CB-CG	-5.05	103.69	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	567	GLY	Peptide

## 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	4942	0	4978	112	0
1	B	4942	0	4978	121	0
2	E	286	0	163	7	0
2	H	286	0	163	9	0
3	F	171	0	90	6	0
3	I	171	0	90	5	0
4	G	379	0	214	15	0
4	J	379	0	214	16	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	11558	0	10890	278	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 13.

The worst 5 of 278 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:752:ARG:NH1	4:G:3:DC:H5''	1.60	1.16
4:G:3:DC:H2''	4:G:4:DA:H5''	1.24	1.14
4:J:3:DC:H2''	4:J:4:DA:H5''	1.29	1.09
4:J:3:DC:H2''	4:J:4:DA:C5'	1.93	0.97
1:A:752:ARG:HH12	4:G:3:DC:H5''	1.22	0.96

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	609/651 (94%)	541 (89%)	59 (10%)	9 (2%)	13	42
1	B	609/651 (94%)	550 (90%)	45 (7%)	14 (2%)	8	30
All	All	1218/1302 (94%)	1091 (90%)	104 (8%)	23 (2%)	10	35

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	371	GLY
1	A	556	ASP
1	B	371	GLY
1	B	556	ASP
1	B	559	GLU

### 5.3.2 Protein sidechains

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	538/563 (96%)	503 (94%)	35 (6%)	21	52
1	B	538/563 (96%)	502 (93%)	36 (7%)	20	50
All	All	1076/1126 (96%)	1005 (93%)	71 (7%)	21	51

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	966	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	444	THR
1	B	899	ARG
1	A	967	GLN
1	B	405	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	980	ASN
1	B	389	ASN
1	B	764	HIS
1	A	953	HIS
1	A	967	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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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 ⓘ

### 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<sup>th</sup> 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 <sup>2</sup> )	Q < 0.9
1	A	615/651 (94%)	0.46	45 (7%) 18 12	42, 84, 169, 197	0
1	B	615/651 (94%)	0.38	43 (6%) 19 13	56, 90, 163, 220	0
2	E	14/14 (100%)	-0.44	0 100 100	85, 102, 145, 169	0
2	H	14/14 (100%)	-0.47	0 100 100	79, 101, 139, 180	0
3	F	8/8 (100%)	-0.32	0 100 100	97, 108, 118, 119	0
3	I	8/8 (100%)	-0.35	0 100 100	111, 118, 137, 141	0
4	G	19/19 (100%)	-0.65	0 100 100	82, 103, 118, 128	0
4	J	19/19 (100%)	-0.62	0 100 100	85, 98, 128, 129	0
All	All	1312/1384 (94%)	0.36	88 (6%) 21 15	42, 89, 168, 220	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	563	ALA	12.5
1	A	562	ASP	10.9
1	A	567	GLY	10.0
1	B	772	MET	8.7
1	B	770	PRO	8.0

### 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 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
5	CA	B	1101	1/1	0.94	0.09	-1.62	77,77,77,77	0
5	CA	A	1101	1/1	0.96	0.11	-1.81	59,59,59,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.