



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 12:40 am BST

PDB ID : 2IVK  
Title : Crystal structure of the periplasmic endonuclease Vvn complexed with a 16-bp DNA  
Authors : Wang, Y.T.; Yang, W.J.; Li, C.L.; Doudeva, L.G.; Yuan, H.S.  
Deposited on : 2006-06-14  
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

<https://www.wwpdb.org/validation/2017/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
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

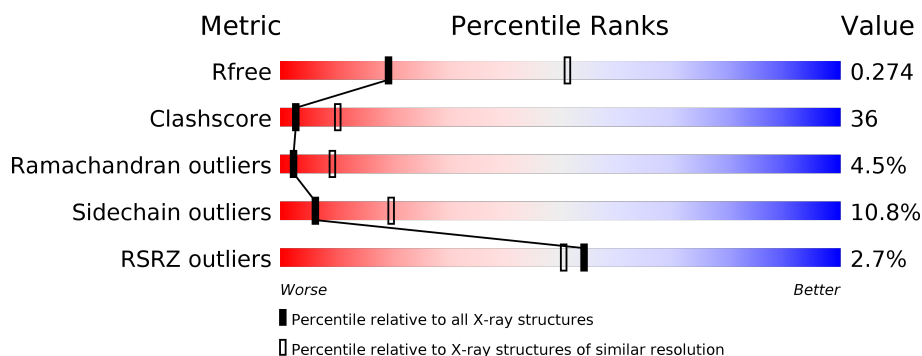
# 1 Overall quality at a glance

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.





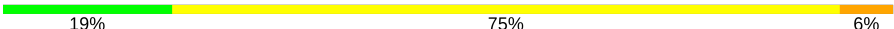
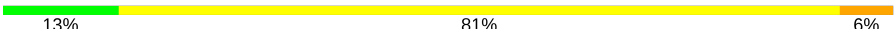
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 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  $\leq 5\%$ . 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	213	<div> <div>5%</div> <div>36%</div> <div>51%</div> <div>12%</div> <div>.</div> </div>
1	B	213	<div> <div>56%</div> <div>38%</div> <div>6%</div> </div>
1	C	213	<div> <div>54%</div> <div>39%</div> <div>7%</div> </div>
1	D	213	<div> <div>6%</div> <div>33%</div> <div>55%</div> <div>11%</div> <div>.</div> </div>
2	E	15	<div> <div>53%</div> <div>40%</div> <div>7%</div> </div>
2	G	15	<div> <div>80%</div> <div>20%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	15	 80% 20%
3	H	15	 7% 73% 20%
4	I	16	 19% 75% 6%
4	J	16	 13% 81% 6%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8881 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 ENDONUCLEASE I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	211	Total	C	N	O	S	0	0	1
			1713	1066	320	314	13			
1	B	213	Total	C	N	O	S	0	0	0
			1738	1080	324	321	13			
1	C	213	Total	C	N	O	S	0	0	0
			1738	1080	324	321	13			
1	D	211	Total	C	N	O	S	0	0	1
			1713	1066	320	314	13			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	80	ALA	HIS	engineered mutation	UNP Q7MHK3
B	80	ALA	HIS	engineered mutation	UNP Q7MHK3
C	80	ALA	HIS	engineered mutation	UNP Q7MHK3
D	80	ALA	HIS	engineered mutation	UNP Q7MHK3

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	15	Total	C	N	O	P	0	0	0
			306	148	56	88	14			
2	G	15	Total	C	N	O	P	0	0	0
			306	148	56	88	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	15	Total	C	N	O	P	0	0	0
			303	147	54	88	14			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	15	Total	C	N	O	P	0	0	0
			303	147	54	88	14			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	I	16	Total	C	N	O	P	0	0	0
			325	157	59	94	15			
4	J	16	Total	C	N	O	P	0	0	0
			325	157	59	94	15			

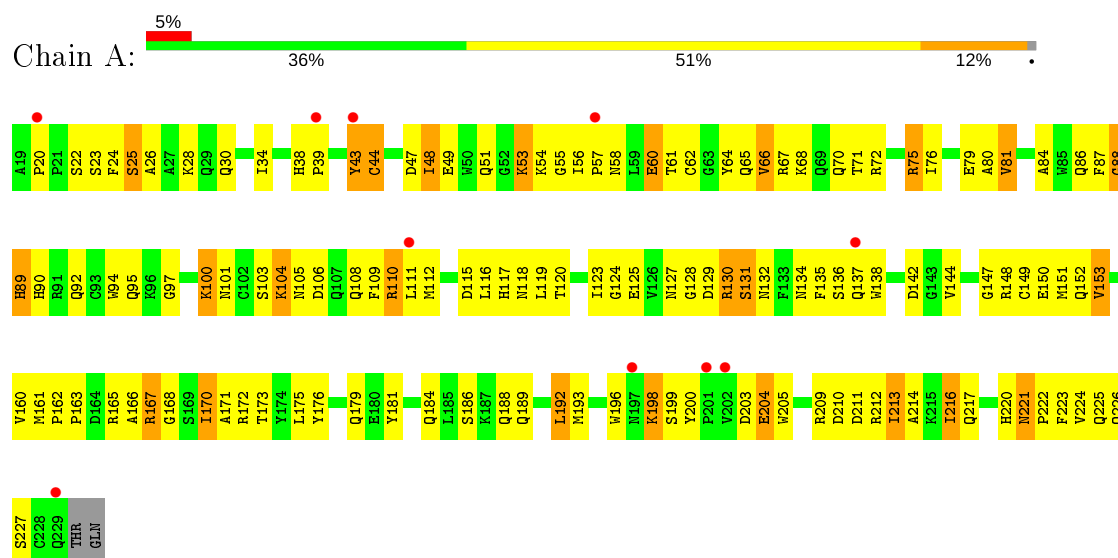
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	17	Total	O	0	0
			17	17		
5	B	20	Total	O	0	0
			20	20		
5	C	19	Total	O	0	0
			19	19		
5	D	30	Total	O	0	0
			30	30		
5	E	7	Total	O	0	0
			7	7		
5	G	5	Total	O	0	0
			5	5		
5	H	2	Total	O	0	0
			2	2		
5	I	6	Total	O	0	0
			6	6		
5	J	5	Total	O	0	0
			5	5		

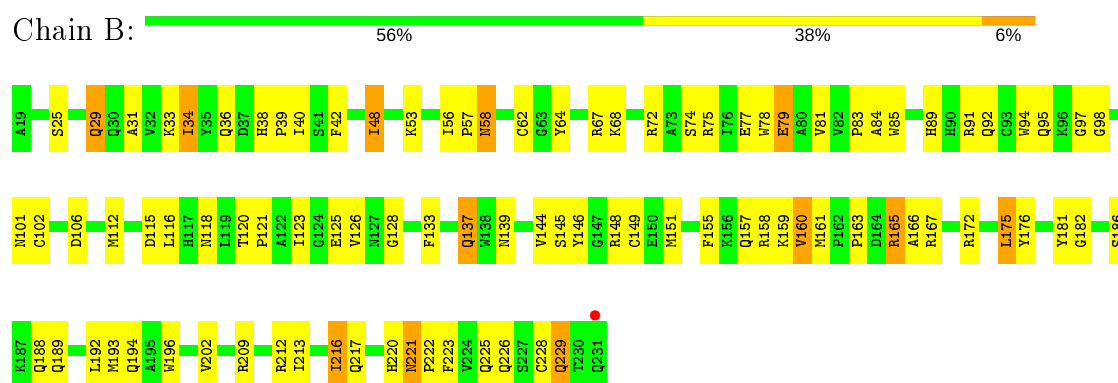
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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 ( $\text{RSRZ} > 2$ ). 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.

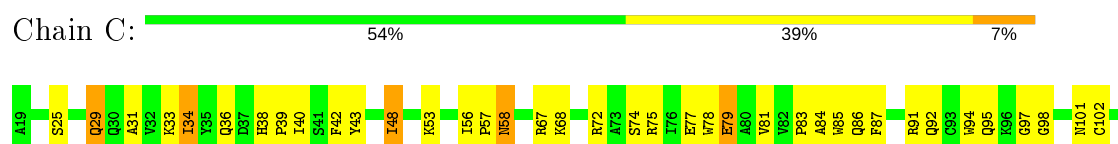
#### • Molecule 1: ENDONUCLEASE I

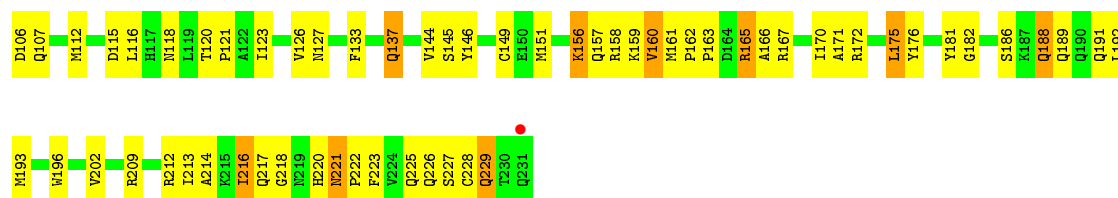


#### • Molecule 1: ENDONUCLEASE I

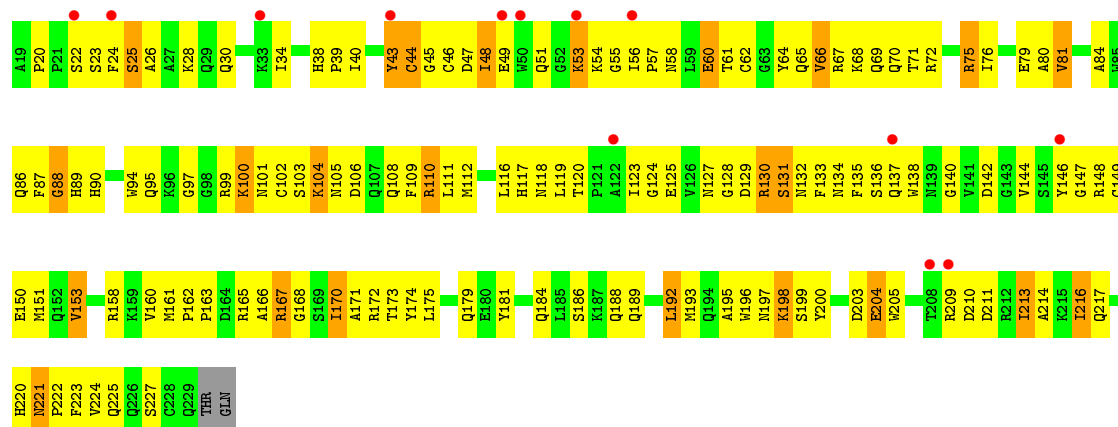


#### • Molecule 1: ENDONUCLEASE I

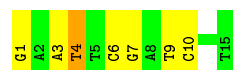




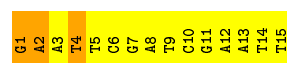
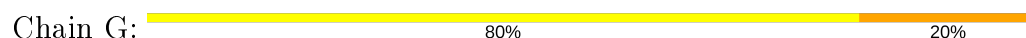
• Molecule 1: ENDONUCLEASE I



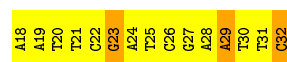
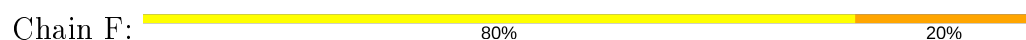
• Molecule 2: 5'-D(\*GP\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP \*GP\*AP\*AP\*TP\*T)-3'



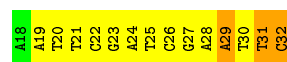
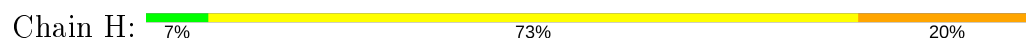
• Molecule 2: 5'-D(\*GP\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP \*GP\*AP\*AP\*TP\*T)-3'



• Molecule 3: 5'-D(\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP\*GP \*AP\*AP\*TP\*TP\*C)-3'

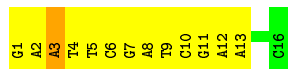


• Molecule 3: 5'-D(\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP\*GP \*AP\*AP\*TP\*TP\*C)-3'

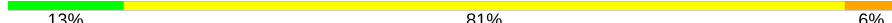


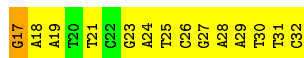
- Molecule 4: 5'-D(\*GP\*AP\*AP\*TP\*TP\*CP\*GP\*AP\*TP\*CP \*GP\*AP\*AP\*TP\*TP\*C)-3',

Chain I:  19% 75% 6%



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

Chain J:  13% 81% 6%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.64Å 64.70Å 79.16Å 74.45° 73.56° 75.56°	Depositor
Resolution (Å)	37.09 – 2.90 37.09 – 2.86	Depositor EDS
% Data completeness (in resolution range)	95.4 (37.09-2.90) 90.8 (37.09-2.86)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.14 (at 2.85Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.237 , 0.284 0.228 , 0.274	Depositor DCC
$R_{free}$ test set	1990 reflections (7.58%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.6	Xtriage
Anisotropy	0.356	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 40.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.429 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	8881	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.75% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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	0.47	0/1758	0.69	0/2372
1	B	0.51	0/1783	0.72	0/2404
1	C	0.52	0/1783	0.72	0/2404
1	D	0.47	0/1758	0.69	0/2372
2	E	0.51	0/343	0.84	0/528
2	G	0.57	0/343	0.86	0/528
3	F	0.54	0/339	0.99	1/521 (0.2%)
3	H	0.55	0/339	1.00	1/521 (0.2%)
4	I	0.43	0/364	0.87	0/560
4	J	0.46	0/364	0.89	0/560
All	All	0.50	0/9174	0.76	2/12770 (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	C	0	1
2	E	0	2
2	G	0	3
3	F	0	2
3	H	0	2
4	I	0	1
4	J	0	2
All	All	0	13

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	32	DC	C4'-C3'-O3'	6.87	126.89	109.70
3	H	32	DC	C4'-C3'-O3'	5.60	123.69	109.70

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	43	TYR	Sidechain
2	E	1	DG	Sidechain
2	E	4	DT	Sidechain
3	F	23	DG	Sidechain
3	F	29	DA	Sidechain

## 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	1713	0	1617	133	0
1	B	1738	0	1640	83	0
1	C	1738	0	1640	83	1
1	D	1713	0	1617	141	0
2	E	306	0	172	11	0
2	G	306	0	172	23	0
3	F	303	0	172	33	0
3	H	303	0	172	30	0
4	I	325	0	183	32	0
4	J	325	0	183	30	1
5	A	17	0	0	8	0
5	B	20	0	0	8	0
5	C	19	0	0	6	0
5	D	30	0	0	14	0
5	E	7	0	0	0	0
5	G	5	0	0	2	0
5	H	2	0	0	3	0
5	I	6	0	0	4	0
5	J	5	0	0	3	0
All	All	8881	0	7568	580	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 36.

The worst 5 of 580 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)
3:F:23:DG:H2''	3:F:24:DA:H5'	1.41	0.98
4:J:29:DA:H2''	4:J:30:DT:H5''	1.48	0.94
1:B:67:ARG:HD2	1:B:144:VAL:HG21	1.50	0.94
4:I:12:DA:H1'	4:I:13:DA:O5'	1.68	0.94
1:D:40:ILE:HG23	5:D:2004:HOH:O	1.72	0.90

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:C:191:GLN:OE1	4:J:21:DT:OP1[1_565]	2.17	0.03

## 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	209/213 (98%)	155 (74%)	37 (18%)	17 (8%)	1	2
1	B	211/213 (99%)	176 (83%)	32 (15%)	3 (1%)	11	36
1	C	211/213 (99%)	175 (83%)	33 (16%)	3 (1%)	11	36
1	D	209/213 (98%)	156 (75%)	38 (18%)	15 (7%)	1	3
All	All	840/852 (99%)	662 (79%)	140 (17%)	38 (4%)	2	9

5 of 38 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	167	ARG
1	A	199	SER
1	D	199	SER
1	A	22	SER
1	A	66	VAL

### 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	182/185 (98%)	159 (87%)	23 (13%)	4	13
1	B	185/185 (100%)	169 (91%)	16 (9%)	10	30
1	C	185/185 (100%)	168 (91%)	17 (9%)	9	27
1	D	182/185 (98%)	159 (87%)	23 (13%)	4	13
All	All	734/740 (99%)	655 (89%)	79 (11%)	6	20

5 of 79 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	192	LEU
1	C	79	GLU
1	D	188	GLN
1	B	216	ILE
1	C	34	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	191	GLN
1	C	101	ASN
1	D	137	GLN
1	C	29	GLN
1	C	107	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](#)

There are no ligands in this entry.

## 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	211/213 (99%)	0.27	10 (4%) 31 28	43, 69, 88, 95	0
1	B	213/213 (100%)	-0.49	1 (0%) 91 91	17, 44, 69, 89	0
1	C	213/213 (100%)	-0.41	1 (0%) 91 91	13, 44, 67, 90	0
1	D	211/213 (99%)	0.29	13 (6%) 20 16	45, 70, 88, 94	0
2	E	15/15 (100%)	-0.44	0 100 100	43, 66, 92, 99	0
2	G	15/15 (100%)	-0.51	0 100 100	47, 66, 97, 99	0
3	F	15/15 (100%)	-0.57	0 100 100	40, 82, 93, 94	0
3	H	15/15 (100%)	-0.47	0 100 100	41, 85, 93, 97	0
4	I	16/16 (100%)	-0.71	0 100 100	52, 83, 92, 94	0
4	J	16/16 (100%)	-0.57	0 100 100	55, 87, 92, 92	0
All	All	940/944 (99%)	-0.13	25 (2%) 54 50	13, 59, 89, 99	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	43	TYR	4.0
1	D	208	THR	3.5
1	A	229	GLN	3.2
1	B	231	GLN	3.2
1	D	22	SER	3.0

### 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 [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.