



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 11:46 PM GMT

PDB ID : 4G0J  
Title : Crystallographic Analysis of Rotavirus NSP2-RNA Complex Reveals Specific Recognition of 5'-GG Sequence for RTPase activity  
Authors : Hu, L.; Prasad, B.V.V.  
Deposited on : 2012-07-09  
Resolution : 3.40 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

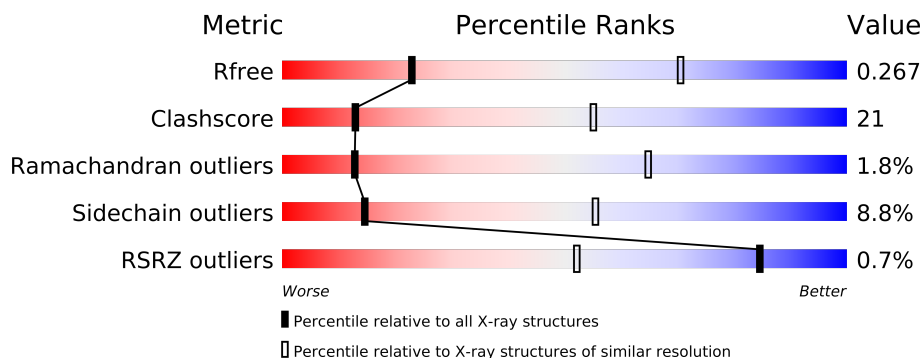
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.40 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))
$R_{free}$	66092	1017 (3.52-3.28)
Clashscore	79885	1214 (3.50-3.30)
Ramachandran outliers	78287	1177 (3.50-3.30)
Sidechain outliers	78261	1177 (3.50-3.30)
RSRZ outliers	66119	1017 (3.52-3.28)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	294	
1	B	294	
1	C	294	
1	D	294	
1	E	294	
1	F	294	
1	G	294	
1	H	294	
1	I	294	
1	J	294	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 23600 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Non-structural protein 2.

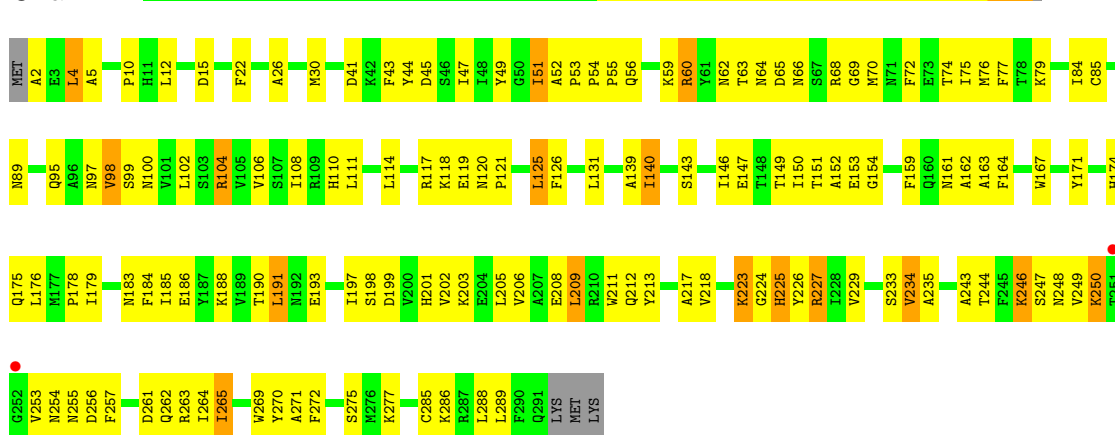
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	B	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	C	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	D	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	E	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	F	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	G	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	H	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	I	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			
1	J	290	Total	C	N	O	S	0	0	0
			2360	1518	400	430	12			

### 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.

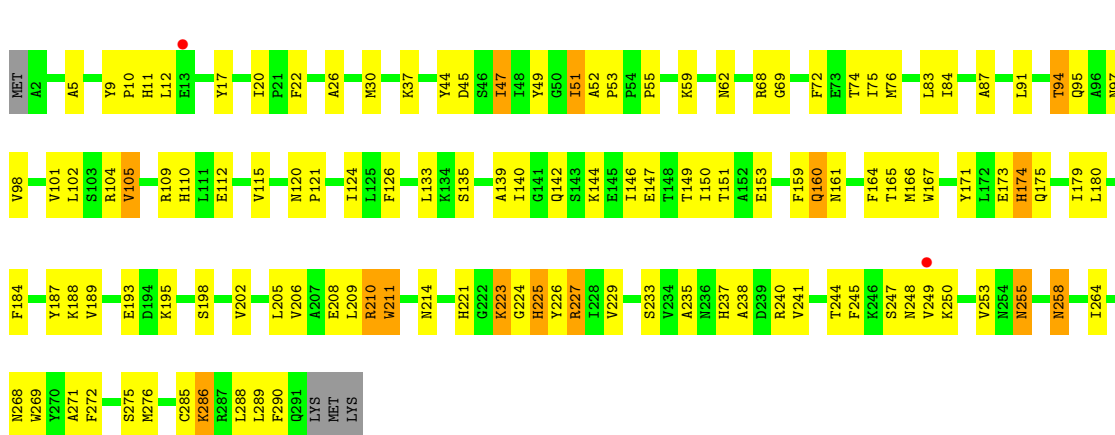
#### • Molecule 1: Non-structural protein 2

Chain A:



#### • Molecule 1: Non-structural protein 2

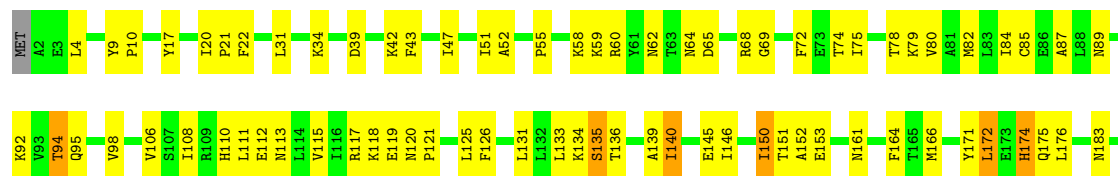
Chain B:

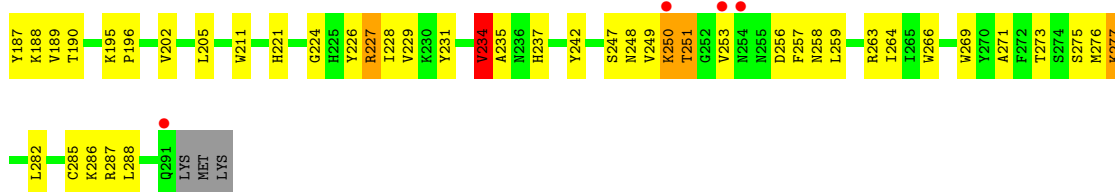


#### • Molecule 1: Non-structural protein 2

Chain C:

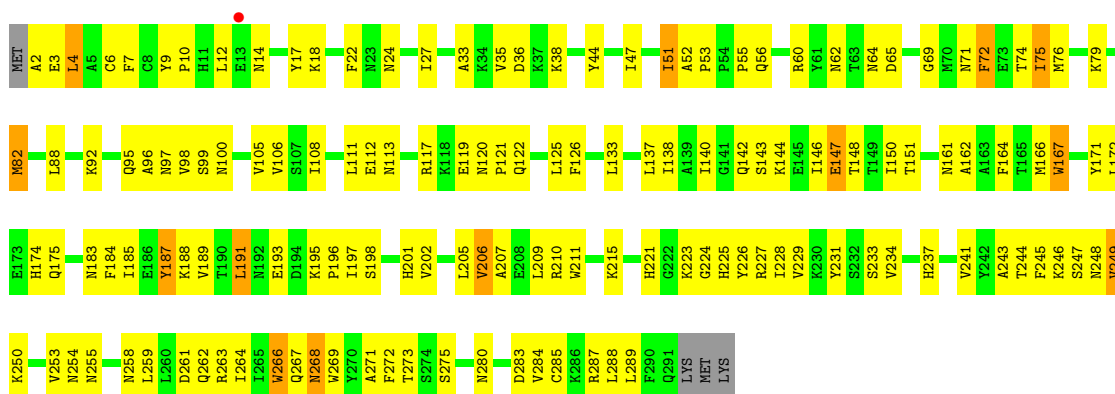






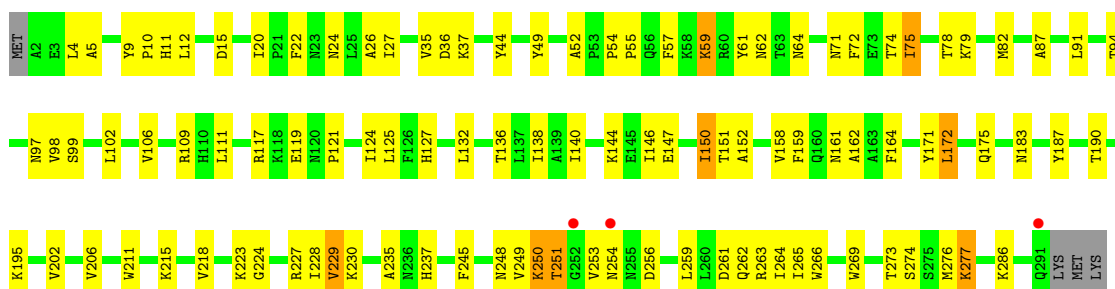
- Molecule 1: Non-structural protein 2

Chain G:



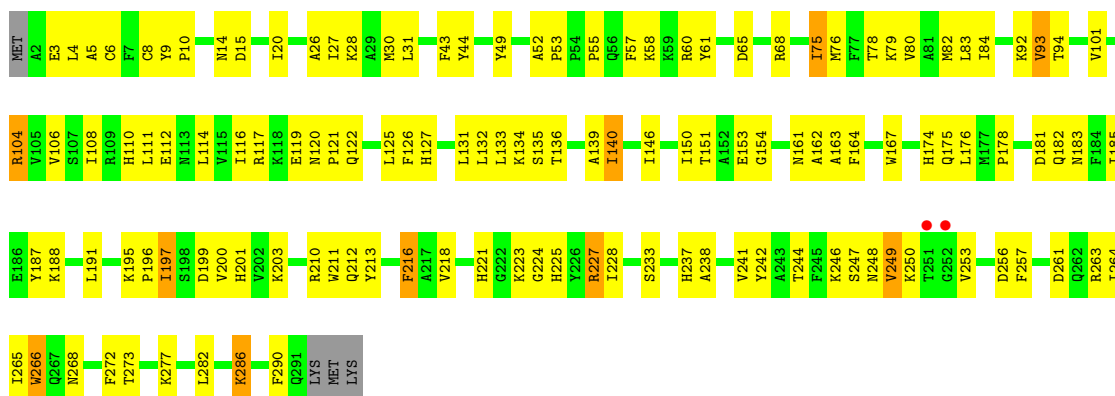
- Molecule 1: Non-structural protein 2

Chain H:



- Molecule 1: Non-structural protein 2

Chain I:



- Molecule 1: Non-structural protein 2

S274	N192	L102	MET
S275	E193	S103	A2
M276	D194	R104	E3
K277	K195	V105	L4
	L196	V106	A5
C285	T197		
K286	H201	L111	Y9
L287	V202	E112	P10
L288			N14
L289			
F290	L205	R117	Y17
G291	V206	K118	K18
LYS	L209	E119	
MET	R210	P120	A26
LYS	V211	P121	
	K215	I124	A29
		L125	M30
	V219	F126	L31
		H127	
			K34
	K223	D130	
G224	G224	L131	K37
H225	H225	L132	K38
V226	V226	L133	D39
R227	R227	K134	
L228	L228	S135	K42
V229	V229		F43
K230	K230	I140	Y44
V231	V231		
			I47
V234	V234	I146	
		E147	I51
H237	H237	T148	A52
A238	A238	T149	
		I150	K59
V248	V248	T151	R60
K249	K249	A152	G61
V250	V250	E153	N62
T251	T251	N161	T63
G252	G252		M64
V253	V253	F164	D65
	D256	W167	G69
V257	V257		
M258	M258	T170	T74
L259	L259	Y171	I75
L260	L260	L172	M76
D261	D261		F77
G262	G262	Q175	T78
R263	R263	L176	K79
L264	L264		V80
L265	L265	N183	A81
V266	V266	F184	
		I185	K92
V269	V269		V93
A271	A271	K188	T94
T272	T272	V189	Q95
F273	F273	L190	
T274	T274	L191	V98

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 4	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	122.48Å 122.48Å 301.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	301.31 – 3.40 19.96 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.3 (301.31-3.40) 99.9 (19.96-3.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.61 (at 3.44Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.237 , 0.294 0.214 , 0.267	Depositor DCC
$R_{free}$ test set	3109 reflections (5.43%)	DCC
Wilson B-factor (Å <sup>2</sup> )	67.0	Xtriage
Anisotropy	0.165	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 18.9	EDS
Estimated twinning fraction	0.299 for H, K, L 0.701 for -H, K, -L 0.236 for h,-k,-l	Xtriage
Reported twinning fraction	0.299 for H, K, L 0.701 for -H, K, -L	Depositor
L-test for twinning	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 60405 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	23600	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 33.80 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 7.5552e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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



## 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	0.65	2/2411 (0.1%)	0.78	1/3261 (0.0%)
1	B	0.64	2/2411 (0.1%)	0.78	0/3261
1	C	0.64	1/2411 (0.0%)	0.80	0/3261
1	D	0.64	3/2411 (0.1%)	0.77	0/3261
1	E	0.67	4/2411 (0.2%)	0.78	1/3261 (0.0%)
1	F	0.64	3/2411 (0.1%)	0.74	0/3261
1	G	0.65	4/2411 (0.2%)	0.79	1/3261 (0.0%)
1	H	0.65	3/2411 (0.1%)	0.78	0/3261
1	I	0.64	2/2411 (0.1%)	0.77	0/3261
1	J	0.63	3/2411 (0.1%)	0.74	0/3261
All	All	0.64	27/24110 (0.1%)	0.77	3/32610 (0.0%)

The worst 5 of 27 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	266	TRP	CD2-CE2	6.11	1.48	1.41
1	A	269	TRP	CD2-CE2	6.10	1.48	1.41
1	H	269	TRP	CD2-CE2	6.10	1.48	1.41
1	I	211	TRP	CD2-CE2	6.07	1.48	1.41
1	D	211	TRP	CD2-CE2	6.01	1.48	1.41

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	191	LEU	CA-CB-CG	6.96	131.30	115.30
1	E	259	LEU	CA-CB-CG	5.28	127.44	115.30
1	A	45	ASP	CB-CG-OD1	5.25	123.03	118.30

There are no chirality outliers.

There are no planarity outliers.

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2360	0	2374	126	0
1	B	2360	0	2374	94	0
1	C	2360	0	2374	92	1
1	D	2360	0	2374	94	0
1	E	2360	0	2374	101	0
1	F	2360	0	2374	89	1
1	G	2360	0	2374	136	0
1	H	2360	0	2374	81	0
1	I	2360	0	2374	112	0
1	J	2360	0	2374	108	0
All	All	23600	0	23740	982	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 21.

The worst 5 of 982 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:H:74:THR:HB	1:H:175:GLN:HE21	1.12	1.14
1:I:65:ASP:HA	1:J:183:ASN:HD21	1.04	1.11
1:H:202:VAL:O	1:H:206:VAL:HG23	1.52	1.09
1:E:74:THR:HB	1:E:175:GLN:HE21	1.10	1.08
1:E:202:VAL:HG13	1:E:264:ILE:HD11	1.35	1.07

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	Distance(Å)	Clash(Å)
1:C:239:ASP:OD2	1:F:251:THR:OG1[2_654]	2.14	0.06

## 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	288/294 (98%)	246 (85%)	35 (12%)	7 (2%)	9	58
1	B	288/294 (98%)	246 (85%)	36 (12%)	6 (2%)	11	61
1	C	288/294 (98%)	243 (84%)	41 (14%)	4 (1%)	16	69
1	D	288/294 (98%)	256 (89%)	29 (10%)	3 (1%)	22	78
1	E	288/294 (98%)	260 (90%)	22 (8%)	6 (2%)	11	61
1	F	288/294 (98%)	252 (88%)	27 (9%)	9 (3%)	7	52
1	G	288/294 (98%)	252 (88%)	32 (11%)	4 (1%)	16	69
1	H	288/294 (98%)	254 (88%)	28 (10%)	6 (2%)	11	61
1	I	288/294 (98%)	257 (89%)	28 (10%)	3 (1%)	22	78
1	J	288/294 (98%)	256 (89%)	28 (10%)	4 (1%)	16	69
All	All	2880/2940 (98%)	2522 (88%)	306 (11%)	52 (2%)	13	65

5 of 52 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	249	VAL
1	E	249	VAL
1	F	235	ALA
1	F	250	LYS
1	G	249	VAL

### 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	262/266 (98%)	237 (90%)	25 (10%)	12	50

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	262/266 (98%)	232 (88%)	30 (12%)	8	38
1	C	262/266 (98%)	233 (89%)	29 (11%)	9	40
1	D	262/266 (98%)	237 (90%)	25 (10%)	12	50
1	E	262/266 (98%)	246 (94%)	16 (6%)	26	73
1	F	262/266 (98%)	241 (92%)	21 (8%)	17	61
1	G	262/266 (98%)	241 (92%)	21 (8%)	17	61
1	H	262/266 (98%)	242 (92%)	20 (8%)	19	64
1	I	262/266 (98%)	242 (92%)	20 (8%)	19	64
1	J	262/266 (98%)	239 (91%)	23 (9%)	14	55
All	All	2620/2660 (98%)	2390 (91%)	230 (9%)	14	55

5 of 230 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	223	LYS
1	F	68	ARG
1	J	104	ARG
1	D	267	GLN
1	E	174	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 61 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	160	GLN
1	F	161	ASN
1	I	267	GLN
1	E	175	GLN
1	E	267	GLN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	290/294 (98%)	-0.15	2 (0%)	84 52	37, 73, 144, 188	0
1	B	290/294 (98%)	-0.17	2 (0%)	84 52	43, 72, 141, 183	0
1	C	290/294 (98%)	-0.18	2 (0%)	84 52	40, 70, 143, 180	0
1	D	290/294 (98%)	-0.21	1 (0%)	91 73	41, 71, 123, 182	0
1	E	290/294 (98%)	-0.23	2 (0%)	84 52	37, 66, 136, 212	0
1	F	290/294 (98%)	-0.17	4 (1%)	72 35	43, 73, 124, 206	0
1	G	290/294 (98%)	-0.20	1 (0%)	91 73	41, 71, 137, 196	0
1	H	290/294 (98%)	-0.24	3 (1%)	79 44	38, 70, 122, 205	0
1	I	290/294 (98%)	-0.29	2 (0%)	84 52	39, 67, 131, 224	0
1	J	290/294 (98%)	-0.17	2 (0%)	84 52	41, 71, 140, 240	0
All	All	2900/2940 (98%)	-0.20	21 (0%)	84 52	37, 71, 138, 240	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	J	252	GLY	5.1
1	F	254	ASN	3.5
1	E	252	GLY	3.2
1	F	250	LYS	3.1
1	B	13	GLU	3.1

### 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 ⓘ

There are no ligands in this entry.

### 6.5 Other polymers ⓘ

There are no such residues in this entry.