



# Full wwPDB X-ray Structure Validation Report

Nov 26, 2014 – 01:36 PM EST

PDB ID : 4WSA  
Title : Crystal structure of Influenza B polymerase bound to the vRNA promoter (FluB1 form)  
Authors : Reich, S.; Guilligay, D.; Pflug, A.; Cusack, S.  
Deposited on : 2014-10-26  
Resolution : 3.40 Å(reported)

This is a full wwPDB validation 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>

---

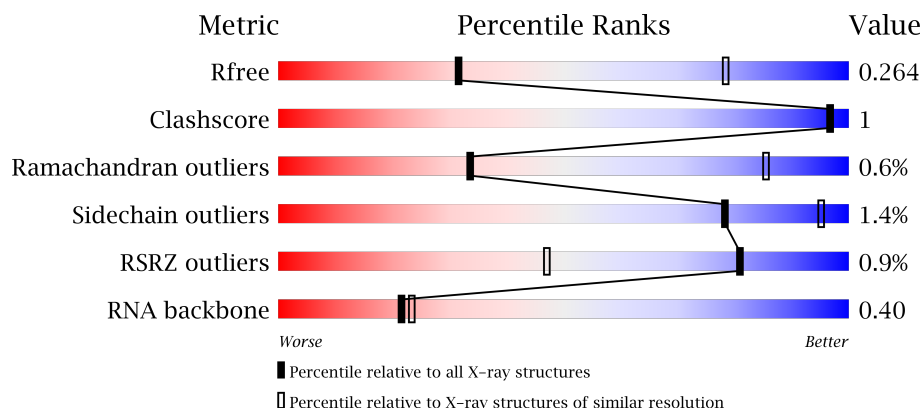
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24195  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24195

# 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)
RNA backbone	1838	1002 (4.02-2.76)

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	R	18	
2	V	14	
3	A	751	
4	B	772	
5	C	798	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 17940 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 RNA chain called Influenza B vRNA promoter 3' end.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	R	14	Total	C	N	O	P	0	0	0
			282	128	42	99	13			

- Molecule 2 is a RNA chain called Influenza B vRNA promoter 5' end.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	V	14	Total	C	N	O	P	0	0	0
			307	137	62	94	14			

- Molecule 3 is a protein called PA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	716	Total	C	N	O	S	0	0	0
			5747	3651	962	1094	40			

There are 25 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	GLY	-	expression tag	UNP Q5V8Z9
A	-12	SER	-	expression tag	UNP Q5V8Z9
A	-11	HIS	-	expression tag	UNP Q5V8Z9
A	-10	HIS	-	expression tag	UNP Q5V8Z9
A	-9	HIS	-	expression tag	UNP Q5V8Z9
A	-8	HIS	-	expression tag	UNP Q5V8Z9
A	-7	HIS	-	expression tag	UNP Q5V8Z9
A	-6	HIS	-	expression tag	UNP Q5V8Z9
A	-5	HIS	-	expression tag	UNP Q5V8Z9
A	-4	HIS	-	expression tag	UNP Q5V8Z9
A	-3	GLY	-	expression tag	UNP Q5V8Z9
A	-2	SER	-	expression tag	UNP Q5V8Z9
A	-1	GLY	-	expression tag	UNP Q5V8Z9
A	0	SER	-	expression tag	UNP Q5V8Z9

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	727	GLY	-	expression tag	UNP Q5V8Z9
A	728	SER	-	expression tag	UNP Q5V8Z9
A	729	GLY	-	expression tag	UNP Q5V8Z9
A	730	SER	-	expression tag	UNP Q5V8Z9
A	731	GLY	-	expression tag	UNP Q5V8Z9
A	732	GLU	-	expression tag	UNP Q5V8Z9
A	733	ASN	-	expression tag	UNP Q5V8Z9
A	734	LEU	-	expression tag	UNP Q5V8Z9
A	735	TYR	-	expression tag	UNP Q5V8Z9
A	736	PHE	-	expression tag	UNP Q5V8Z9
A	737	GLN	-	expression tag	UNP Q5V8Z9

- Molecule 4 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	733	Total	C	N	O	S	0	0	0
			5761	3636	999	1074	52			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-8	GLY	-	expression tag	UNP Q5V8Y6
B	-7	SER	-	expression tag	UNP Q5V8Y6
B	-6	GLY	-	expression tag	UNP Q5V8Y6
B	-5	SER	-	expression tag	UNP Q5V8Y6
B	-4	GLY	-	expression tag	UNP Q5V8Y6
B	-3	SER	-	expression tag	UNP Q5V8Y6
B	-2	GLY	-	expression tag	UNP Q5V8Y6
B	-1	SER	-	expression tag	UNP Q5V8Y6
B	0	GLY	-	expression tag	UNP Q5V8Y6
B	753	GLY	-	expression tag	UNP Q5V8Y6
B	754	SER	-	expression tag	UNP Q5V8Y6
B	755	GLY	-	expression tag	UNP Q5V8Y6
B	756	SER	-	expression tag	UNP Q5V8Y6
B	757	GLY	-	expression tag	UNP Q5V8Y6
B	758	GLU	-	expression tag	UNP Q5V8Y6
B	759	ASN	-	expression tag	UNP Q5V8Y6
B	760	LEU	-	expression tag	UNP Q5V8Y6
B	761	TYR	-	expression tag	UNP Q5V8Y6
B	762	PHE	-	expression tag	UNP Q5V8Y6
B	763	GLN	-	expression tag	UNP Q5V8Y6

- Molecule 5 is a protein called PB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	C	730	Total	C	N	O	S	0	0	0
			5843	3715	1023	1065	40			

There are 28 discrepancies between the modelled and reference sequences:

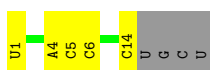
Chain	Residue	Modelled	Actual	Comment	Reference
C	-8	GLY	-	expression tag	UNP Q5V8X3
C	-7	SER	-	expression tag	UNP Q5V8X3
C	-6	GLY	-	expression tag	UNP Q5V8X3
C	-5	SER	-	expression tag	UNP Q5V8X3
C	-4	GLY	-	expression tag	UNP Q5V8X3
C	-3	SER	-	expression tag	UNP Q5V8X3
C	-2	GLY	-	expression tag	UNP Q5V8X3
C	-1	SER	-	expression tag	UNP Q5V8X3
C	0	GLY	-	expression tag	UNP Q5V8X3
C	771	GLY	-	expression tag	UNP Q5V8X3
C	772	TRP	-	expression tag	UNP Q5V8X3
C	773	SER	-	expression tag	UNP Q5V8X3
C	774	HIS	-	expression tag	UNP Q5V8X3
C	775	PRO	-	expression tag	UNP Q5V8X3
C	776	GLN	-	expression tag	UNP Q5V8X3
C	777	PHE	-	expression tag	UNP Q5V8X3
C	778	GLU	-	expression tag	UNP Q5V8X3
C	779	LYS	-	expression tag	UNP Q5V8X3
C	780	GLY	-	expression tag	UNP Q5V8X3
C	781	SER	-	expression tag	UNP Q5V8X3
C	782	GLY	-	expression tag	UNP Q5V8X3
C	783	SER	-	expression tag	UNP Q5V8X3
C	784	GLU	-	expression tag	UNP Q5V8X3
C	785	ASN	-	expression tag	UNP Q5V8X3
C	786	LEU	-	expression tag	UNP Q5V8X3
C	787	TYR	-	expression tag	UNP Q5V8X3
C	788	PHE	-	expression tag	UNP Q5V8X3
C	789	GLN	-	expression tag	UNP Q5V8X3

### 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: Influenza B vRNA promoter 3' end

Chain R: 



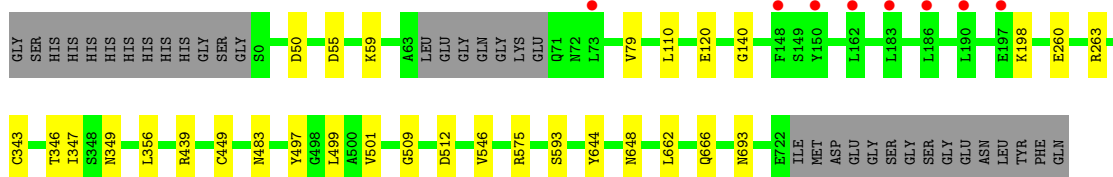
- Molecule 2: Influenza B vRNA promoter 5' end

Chain V: 



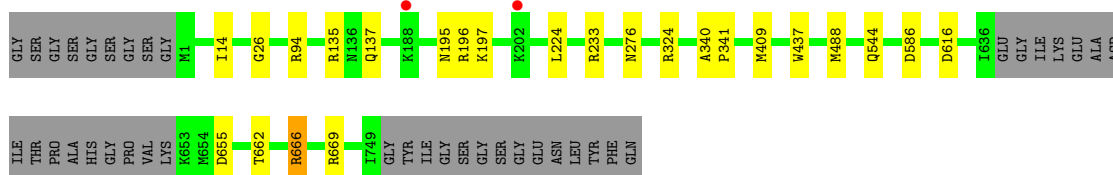
- Molecule 3: PA

Chain A: 



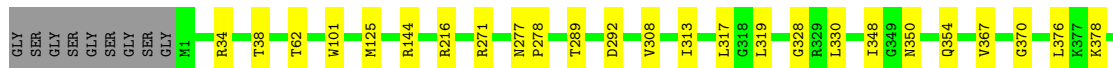
- Molecule 4: RNA-directed RNA polymerase catalytic subunit

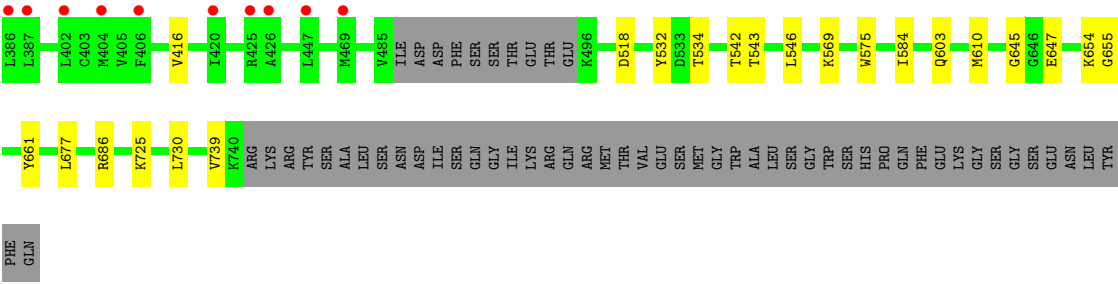
Chain B: 



- Molecule 5: PB2

Chain C: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	199.70Å 199.70Å 252.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.98 – 3.40 48.98 – 3.40	Depositor EDS
% Data completeness (in resolution range)	98.5 (48.98-3.40) 98.6 (48.98-3.40)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.71 (at 3.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1760)	Depositor
R, $R_{free}$	0.229 , 0.265 0.229 , 0.264	Depositor DCC
$R_{free}$ test set	3819 reflections (4.82%)	DCC
Wilson B-factor (Å <sup>2</sup> )	129.1	Xtriage
Anisotropy	0.639	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 94.3	EDS
Estimated twinning fraction	0.047 for -h,-k,l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	1 of 79266 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	17940	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	158.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% of the height of the origin peak. No significant pseudotranslation is detected.*

<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	R	0.26	0/312	0.81	0/482
2	V	0.62	1/345 (0.3%)	0.70	0/535
3	A	0.23	0/5862	0.38	0/7904
4	B	0.23	0/5873	0.39	0/7915
5	C	0.23	0/5943	0.40	0/7987
All	All	0.24	1/18335 (0.0%)	0.41	0/24823

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	V	1	A	OP3-P	-10.62	1.48	1.61

There are no bond angle outliers.

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	R	282	0	146	0	0
2	V	307	0	153	0	0
3	A	5747	0	5718	13	0
4	B	5761	0	5783	14	0
5	C	5843	0	6015	19	0
All	All	17940	0	17815	37	0

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 1.

All (37) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:B:195:ASN:O	4:B:197:LYS:N	2.32	0.62
3:A:648:ASN:OD1	3:A:693:ASN:ND2	2.40	0.55
3:A:346:THR:O	3:A:349:ASN:ND2	2.40	0.54
3:A:509:GLY:N	3:A:512:ASP:OD2	2.42	0.52
4:B:669:ARG:NH1	5:C:38:THR:OG1	2.44	0.52
5:C:330:LEU:HD11	5:C:367:VAL:CG1	2.42	0.50
3:A:666:GLN:HB2	4:B:14:ILE:HD12	1.92	0.50
3:A:439:ARG:NH1	4:B:544:GLN:OE1	2.45	0.49
4:B:276:ASN:OD1	5:C:144:ARG:NH1	2.45	0.49
5:C:725:LYS:HG2	5:C:739:VAL:HG22	1.94	0.49
4:B:340:ALA:HB3	4:B:341:PRO:HD3	1.94	0.48
5:C:661:TYR:CE1	5:C:677:LEU:HD12	2.48	0.48
3:A:662:LEU:HB3	4:B:14:ILE:HD13	1.97	0.47
3:A:55:ASP:OD1	3:A:59:LYS:N	2.47	0.47
5:C:546:LEU:HB2	5:C:584:ILE:HD11	1.98	0.46
3:A:644:TYR:HA	4:B:26:GLY:HA2	1.99	0.45
4:B:655:ASP:OD2	5:C:216:ARG:NH2	2.50	0.45
5:C:367:VAL:HG21	5:C:376:LEU:HB2	1.99	0.44
5:C:543:THR:HG23	5:C:584:ILE:HG21	1.99	0.44
3:A:79:VAL:HA	3:A:110:LEU:HD23	2.00	0.43
4:B:586:ASP:N	4:B:586:ASP:OD1	2.50	0.43
3:A:347:ILE:HD12	3:A:499:LEU:HD13	2.00	0.43
5:C:532:TYR:O	5:C:534:THR:N	2.48	0.42
4:B:224:LEU:HD13	4:B:409:MET:HG3	2.02	0.42
3:A:483:ASN:HB3	3:A:497:TYR:HE1	1.85	0.42
5:C:350:ASN:HB3	5:C:416:VAL:HG12	2.02	0.42
5:C:313:ILE:HG22	5:C:317:LEU:HD12	2.01	0.41
5:C:348:ILE:HD12	5:C:354:GLN:HG3	2.02	0.41
4:B:666:ARG:NH2	5:C:62:THR:OG1	2.53	0.41
4:B:135:ARG:NH1	4:B:137:GLN:OE1	2.52	0.41
5:C:277:ASN:N	5:C:278:PRO:HD3	2.35	0.41
5:C:289:THR:HG22	5:C:532:TYR:CD1	2.55	0.41
3:A:347:ILE:CD1	3:A:499:LEU:HD22	2.51	0.41
5:C:654:LYS:HG2	5:C:655:GLY:N	2.35	0.41
5:C:546:LEU:CB	5:C:584:ILE:HD11	2.51	0.41
3:A:343:CYS:SG	3:A:501:VAL:HG22	2.61	0.40
4:B:662:THR:HG21	5:C:101:TRP:CD1	2.56	0.40

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	
3	A	710/751 (94%)	666 (94%)	41 (6%)	3 (0%)	43	90
4	B	729/772 (94%)	688 (94%)	40 (6%)	1 (0%)	59	96
5	C	726/798 (91%)	676 (93%)	42 (6%)	8 (1%)	21	76
All	All	2165/2321 (93%)	2030 (94%)	123 (6%)	12 (1%)	33	85

All (12) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	B	196	ARG
5	C	292	ASP
5	C	569	LYS
5	C	730	LEU
3	A	50	ASP
5	C	308	VAL
3	A	140	GLY
3	A	546	VAL
5	C	686	ARG
5	C	645	GLY
5	C	328	GLY
5	C	370	GLY

### 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	
3	A	637/664 (96%)	629 (99%)	8 (1%)	80	95
4	B	631/657 (96%)	624 (99%)	7 (1%)	84	96

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	C	638/694 (92%)	627 (98%)	11 (2%)	73	94
All	All	1906/2015 (95%)	1880 (99%)	26 (1%)	78	95

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	120	GLU
3	A	198	LYS
3	A	260	GLU
3	A	263	ARG
3	A	356	LEU
3	A	449	CYS
3	A	575	ARG
3	A	593	SER
4	B	94	ARG
4	B	233	ARG
4	B	324	ARG
4	B	437	TRP
4	B	488	MET
4	B	616	ASP
4	B	666	ARG
5	C	34	ARG
5	C	125	MET
5	C	271	ARG
5	C	319	LEU
5	C	378	LYS
5	C	518	ASP
5	C	542	THR
5	C	575	TRP
5	C	603	GLN
5	C	610	MET
5	C	647	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	666	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	13/18 (72%)	4 (30%)	1 (7%)
2	V	13/14 (92%)	2 (15%)	0
All	All	26/32 (81%)	6 (23%)	1 (3%)

All (6) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	4	A
1	R	5	C
1	R	6	C
1	R	14	C
2	V	7	A
2	V	11	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	R	1	U

## 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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	164:GLU	C	165:GLU	N	3.00

## 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	R	14/18 (77%)	-0.46	0 100 100	110, 129, 172, 209	0
2	V	14/14 (100%)	-0.43	0 100 100	99, 113, 127, 127	0
3	A	716/751 (95%)	0.05	8 (1%) 77 40	96, 151, 238, 271	0
4	B	733/772 (94%)	-0.01	2 (0%) 91 73	101, 148, 178, 209	0
5	C	730/798 (91%)	0.15	10 (1%) 72 35	101, 157, 234, 262	0
All	All	2207/2353 (93%)	0.06	20 (0%) 81 47	96, 151, 230, 271	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	C	425	ARG	4.1
3	A	73	LEU	3.4
5	C	386	LEU	3.3
5	C	387	LEU	3.1
3	A	150	TYR	2.9
3	A	148	PHE	2.9
3	A	190	LEU	2.8
5	C	402	LEU	2.8
3	A	162	LEU	2.7
3	A	186	LEU	2.5
5	C	406	PHE	2.5
5	C	447	LEU	2.5
3	A	197	GLU	2.5
5	C	420	ILE	2.4
5	C	469	MET	2.4
5	C	426	ALA	2.4
5	C	404	MET	2.3
4	B	188	LYS	2.2
4	B	202	LYS	2.1
3	A	183	LEU	2.1

## 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.