



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 01:15 PM GMT

PDB ID : 4B3Q  
Title : Structures of HIV-1 RT and RNA-DNA Complex Reveal a Unique RT Con-  
formation and Substrate Interface  
Authors : Lapkouski, M.; Tian, L.; Miller, J.T.; Le Grice, S.F.J.; Yang, W.  
Deposited on : 2012-07-25  
Resolution : 5.00 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>

---

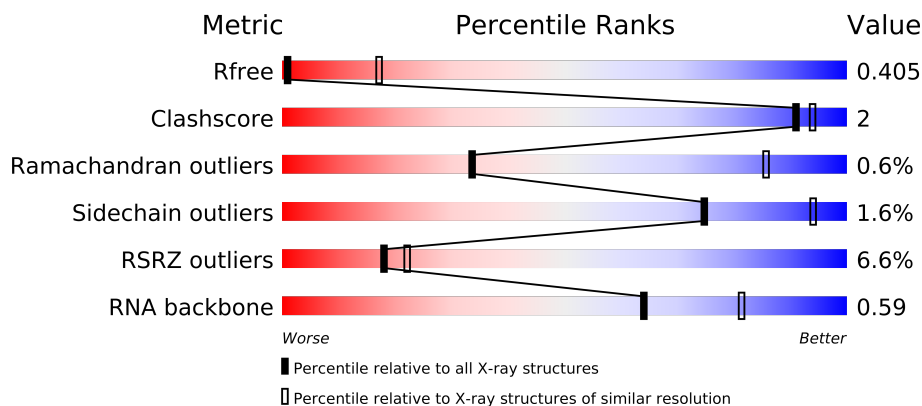
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 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 5.00 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	1052 (6.50-3.50)
Clashscore	79885	1327 (6.50-3.50)
Ramachandran outliers	78287	1242 (6.50-3.50)
Sidechain outliers	78261	1221 (6.50-3.50)
RSRZ outliers	66119	1051 (6.50-3.50)
RNA backbone	1838	1037 (7.00-2.80)

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	560	
2	B	454	
3	D	25	
4	R	34	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8433 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 REVERSE TRANSCRIPTASE/RIBONUCLEASEH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	543	Total	C	N	O	S	0	0	0
			4271	2758	702	803	8			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	68	GLY	SER	ENGINEERED MUTATION	UNP P04585
A	83	LYS	ARG	ENGINEERED MUTATION	UNP P04585
A	411	VAL	ILE	ENGINEERED MUTATION	UNP P04585
A	447	SER	ASN	ENGINEERED MUTATION	UNP P04585
A	461	LYS	ARG	ENGINEERED MUTATION	UNP P04585
A	483	HIS	TYR	ENGINEERED MUTATION	UNP P04585
A	559	ILE	VAL	ENGINEERED MUTATION	UNP P04585

- Molecule 2 is a protein called P51 RT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	410	Total	C	N	O	S	0	0	0
			3264	2122	528	607	7			

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	MET	-	EXPRESSION TAG	UNP P04585
B	-12	ARG	-	EXPRESSION TAG	UNP P04585
B	-11	GLY	-	EXPRESSION TAG	UNP P04585
B	-10	SER	-	EXPRESSION TAG	UNP P04585
B	-9	HIS	-	EXPRESSION TAG	UNP P04585
B	-8	HIS	-	EXPRESSION TAG	UNP P04585
B	-7	HIS	-	EXPRESSION TAG	UNP P04585
B	-6	HIS	-	EXPRESSION TAG	UNP P04585
B	-5	HIS	-	EXPRESSION TAG	UNP P04585

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-4	HIS	-	EXPRESSION TAG	UNP P04585
B	-3	GLY	-	EXPRESSION TAG	UNP P04585
B	-2	SER	-	EXPRESSION TAG	UNP P04585
B	-1	GLN	-	EXPRESSION TAG	UNP P04585
B	0	LEU	-	EXPRESSION TAG	UNP P04585
B	68	GLY	SER	ENGINEERED MUTATION	UNP P04585
B	83	LYS	ARG	ENGINEERED MUTATION	UNP P04585
B	411	VAL	ILE	ENGINEERED MUTATION	UNP P04585

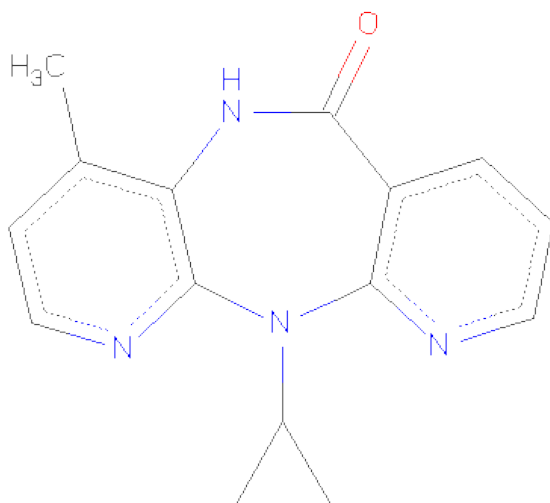
- Molecule 3 is a DNA chain called PRIMER DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	21	Total	C	N	O	P	0	0	0
			430	206	73	130	21			

- Molecule 4 is a RNA chain called TEMPLATE RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	21	Total	C	N	O	P	0	0	0
			448	201	83	143	21			

- Molecule 5 is 11-CYCLOPROPYL-5,11-DIHYDRO-4-METHYL-6H-DIPYRIDO[3,2-B:2',3'-E][1,4]DIAZEPIN-6-ONE (three-letter code: NVP) (formula: C<sub>15</sub>H<sub>14</sub>N<sub>4</sub>O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			20	15	4	1		



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	272.70Å 272.70Å 233.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.99 – 5.00 29.99 – 5.00	Depositor EDS
% Data completeness (in resolution range)	89.8 (29.99-5.00) 89.9 (29.99-5.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 5.06Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.361 , 0.407 0.361 , 0.405	Depositor DCC
$R_{free}$ test set	661 reflections (5.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	170.4	Xtriage
Anisotropy	0.486	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 37.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 13022 reflections	Xtriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	8433	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	149.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.76% 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

Bond lengths and bond angles in the following residue types are not validated in this section: NVP

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	0.50	6/4380 (0.1%)	0.47	0/5978
2	B	0.53	2/3353 (0.1%)	0.46	0/4574
3	D	0.28	0/480	0.72	0/739
4	R	0.22	0/501	0.65	0/778
All	All	0.49	8/8714 (0.1%)	0.50	0/12069

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	426	TRP	CD2-CE2	5.06	1.47	1.41
1	A	410	TRP	CD2-CE2	5.05	1.47	1.41
1	A	266	TRP	CD2-CE2	5.02	1.47	1.41
2	B	426	TRP	CD2-CE2	5.01	1.47	1.41
1	A	88	TRP	CD2-CE2	5.01	1.47	1.41
2	B	153	TRP	CD2-CE2	5.01	1.47	1.41
1	A	212	TRP	CD2-CE2	5.00	1.47	1.41
1	A	337	TRP	CD2-CE2	5.00	1.47	1.41

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	A	4271	0	4186	13	0
2	B	3264	0	3199	10	0
3	D	430	0	240	7	0
4	R	448	0	228	4	0
5	A	20	0	12	0	0
All	All	8433	0	7865	33	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 2.

All (33) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:D:17:DT:H2'	3:D:18:DG:H8	1.65	0.62
3:D:16:DT:H2'	3:D:17:DT:C6	2.43	0.54
3:D:17:DT:H2'	3:D:18:DG:C8	2.44	0.53
2:B:264:LEU:HD12	2:B:274:ILE:HG21	1.91	0.53
1:A:317:VAL:HG21	1:A:347:LYS:HB3	1.91	0.52
2:B:263:LYS:HE2	2:B:425:LEU:HB3	1.91	0.52
1:A:343:GLN:HG3	1:A:349:LEU:HD11	1.91	0.51
2:B:255:ASN:HB2	2:B:289:LEU:HB3	1.93	0.50
1:A:253:THR:HA	1:A:292:VAL:HG22	1.95	0.49
1:A:444:GLY:HA2	1:A:552:VAL:HG11	1.95	0.48
1:A:54:ASN:HB3	1:A:143:ARG:HH21	1.78	0.48
4:R:6:G:H2'	4:R:7:G:C8	2.49	0.48
4:R:9:C:H2'	4:R:10:A:C8	2.49	0.47
2:B:325:LEU:HD23	2:B:387:PRO:HB3	1.97	0.47
1:A:120:LEU:HD21	1:A:124:PHE:HD2	1.80	0.47
1:A:376:THR:O	1:A:380:ILE:HG12	2.16	0.46
2:B:362:THR:HB	2:B:367:GLN:HE21	1.81	0.45
1:A:363:ASN:HD22	1:A:509:GLN:HB2	1.81	0.45
2:B:254:VAL:HB	2:B:289:LEU:HA	1.98	0.45
1:A:480:GLN:HE21	1:A:517:LEU:HD11	1.81	0.45
1:A:408:ALA:HB2	2:B:331:LYS:HD2	1.98	0.45
3:D:18:DG:H2'	3:D:19:DT:C6	2.53	0.44
4:R:9:C:H2'	4:R:10:A:H8	1.82	0.43
3:D:8:DT:H2'	3:D:9:DA:C8	2.53	0.43
1:A:420:PRO:HB2	1:A:421:PRO:HD3	2.01	0.43
2:B:275:LYS:HD3	2:B:277:ARG:HH12	1.84	0.43
3:D:12:DG:H2'	3:D:13:DT:C6	2.53	0.43
1:A:312:GLU:HA	1:A:313:PRO:HD3	1.88	0.42

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:101:LYS:HB3	2:B:382:ILE:HA	2.02	0.42
2:B:149:LEU:HA	2:B:150:PRO:HD3	1.84	0.41
1:A:169:GLU:HB3	1:A:170:PRO:HD3	2.02	0.41
4:R:14:U:H2'	4:R:15:A:C8	2.56	0.40
3:D:8:DT:H2'	3:D:9:DA:H8	1.86	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	
1	A	539/560 (96%)	505 (94%)	29 (5%)	5 (1%)	25	81
2	B	406/454 (89%)	392 (97%)	13 (3%)	1 (0%)	56	93
All	All	945/1014 (93%)	897 (95%)	42 (4%)	6 (1%)	33	86

All (6) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	6	GLU
1	A	184	MET
1	A	253	THR
1	A	363	ASN
1	A	292	VAL
2	B	184	MET

### 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	453/499 (91%)	443 (98%)	10 (2%)	64	91
2	B	347/411 (84%)	344 (99%)	3 (1%)	87	96
All	All	800/910 (88%)	787 (98%)	13 (2%)	75	94

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2	ILE
1	A	17	ASP
1	A	185	ASP
1	A	192	ASP
1	A	246	LEU
1	A	295	LEU
1	A	323	LYS
1	A	330	GLN
1	A	353	LYS
1	A	488	ASP
2	B	86	ASP
2	B	113	ASP
2	B	214	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	222	GLN
1	A	265	ASN
1	A	330	GLN
1	A	340	GLN
1	A	367	GLN
1	A	494	ASN
2	B	23	GLN
2	B	394	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	R	20/34 (58%)	5 (25%)	0

All (5) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	R	12	A
4	R	15	A
4	R	16	A
4	R	18	U
4	R	19	A

There are no RNA pucker outliers to report.

## 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 ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
5	NVP	A	999	-	18,23,23	7.85	7 (38%)	19,34,34	21.30	10 (52%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	NVP	A	999	-	-	0/0/6/6	0/0/4/4

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	999	NVP	CA-N1	32.25	2.57	1.50
5	A	999	NVP	OE-C9	-3.87	1.16	1.24
5	A	999	NVP	C10-C15	3.59	1.47	1.41
5	A	999	NVP	C6-C7	3.14	1.45	1.42
5	A	999	NVP	C12-C11	2.59	1.42	1.36
5	A	999	NVP	C13-N14	2.56	1.37	1.32
5	A	999	NVP	C12-C13	2.21	1.44	1.37

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	999	NVP	CB-CA-N1	-81.77	31.28	118.41
5	A	999	NVP	CC-CA-N1	-42.90	72.71	118.41
5	A	999	NVP	C4-N3-C2	4.06	122.27	116.70
5	A	999	NVP	C13-N14-C15	4.06	122.27	116.70
5	A	999	NVP	C5-C6-C7	3.64	119.18	116.82
5	A	999	NVP	C6-C7-C2	-3.51	117.88	122.58
5	A	999	NVP	C5-C4-N3	-2.89	119.83	124.48
5	A	999	NVP	C12-C13-N14	-2.80	119.65	123.89
5	A	999	NVP	C10-C15-N14	-2.53	120.32	123.55
5	A	999	NVP	C7-C2-N3	-2.28	120.70	122.81

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	543/560 (96%)	0.62	39 (7%) 15 19	132, 150, 159, 170	0
2	B	410/454 (90%)	0.55	20 (4%) 28 28	133, 149, 157, 163	0
3	D	21/25 (84%)	0.69	3 (14%) 3 7	137, 148, 156, 160	0
4	R	21/34 (61%)	0.79	3 (14%) 3 7	140, 149, 159, 160	0
All	All	995/1073 (92%)	0.59	65 (6%) 18 22	132, 149, 158, 170	0

All (65) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	133	PRO	5.0
1	A	457	TYR	4.4
3	D	4	DT	3.8
3	D	5	DG	3.7
1	A	427	TYR	3.7
1	A	19	PRO	3.5
1	A	359	GLY	3.5
1	A	219	LYS	3.4
1	A	29	GLU	3.4
1	A	426	TRP	3.4
1	A	25	PRO	3.3
4	R	26	U	3.2
2	B	243	PRO	3.1
1	A	456	GLY	3.1
3	D	3	DA	3.1
1	A	16	MET	3.0
4	R	25	A	3.0
1	A	290	THR	2.9
2	B	360	ALA	2.9
1	A	1	PRO	2.8
2	B	359	GLY	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
4	R	24	C	2.8
1	A	291	GLU	2.8
1	A	462	GLY	2.8
1	A	27	THR	2.7
2	B	357	MET	2.7
1	A	91	GLN	2.7
2	B	62	ALA	2.6
1	A	173	LYS	2.5
1	A	249	LYS	2.5
2	B	179	VAL	2.5
1	A	248	GLU	2.5
2	B	85	GLN	2.4
1	A	28	GLU	2.4
1	A	282	LEU	2.4
1	A	288	ALA	2.4
2	B	144	TYR	2.4
1	A	507	GLN	2.3
2	B	358	ARG	2.3
1	A	26	LEU	2.3
1	A	132	ILE	2.3
1	A	488	ASP	2.3
1	A	504	GLY	2.3
2	B	294	PRO	2.2
1	A	531	VAL	2.2
2	B	183	TYR	2.2
1	A	18	GLY	2.2
1	A	20	LYS	2.2
1	A	540	LYS	2.2
2	B	240	THR	2.2
1	A	144	TYR	2.2
2	B	241	VAL	2.2
1	A	30	LYS	2.2
1	A	464	GLN	2.2
2	B	71	TRP	2.1
1	A	446	ALA	2.1
2	B	367	GLN	2.1
2	B	334	GLN	2.1
1	A	343	GLN	2.1
2	B	68	GLY	2.1
2	B	409	THR	2.0
2	B	346	PHE	2.0
1	A	556	ILE	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	242	GLN	2.0
1	A	289	LEU	2.0

## 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 ⓘ

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	NVP	A	999	20/20	0.48	1.24	131,148,159,161	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.