



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 09:54 AM GMT

PDB ID : 3K54  
Title : Structures of human Bruton's tyrosine kinase in active and inactive conformations suggests a mechanism of activation for TEC family kinases.  
Authors : Marcotte, D.J.; Liu, Y.-T.; Arduini, R.M.; Hession, C.A.; Miatkowski, K.; Wildes, C.P.; Cullen, P.F.; Hopkins, B.T.; Mertsching, E.; Jenkins, T.J.; Romanowski, M.J.; Baker, D.P.; Silvian, L.F.  
Deposited on : 2009-10-06  
Resolution : 1.94 Å(reported)

This is a Full wwPDB X-ray Structure 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/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

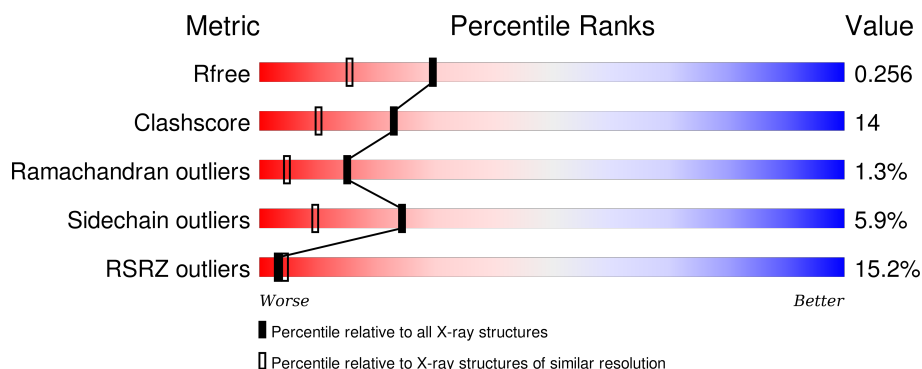
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.94 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	2910 (1.96-1.92)
Clashscore	102246	3095 (1.96-1.92)
Ramachandran outliers	100387	3062 (1.96-1.92)
Sidechain outliers	100360	3062 (1.96-1.92)
RSRZ outliers	91569	2915 (1.96-1.92)

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	283	<div> <div>13%</div> <div>65%</div> <div>16%</div> <div>• •</div> <div>14%</div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2212 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.

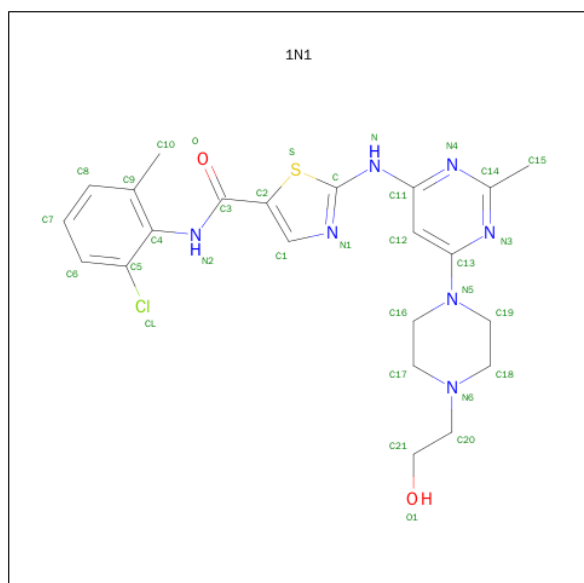
- Molecule 1 is a protein called Tyrosine-protein kinase BTK.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	243	2015	1301	334	362	18	0	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	377	GLY	-	EXPRESSION TAG	UNP Q06187
A	378	PRO	-	EXPRESSION TAG	UNP Q06187
A	379	LEU	-	EXPRESSION TAG	UNP Q06187
A	380	GLY	-	EXPRESSION TAG	UNP Q06187
A	381	SER	-	EXPRESSION TAG	UNP Q06187
A	551	GLU	TYR	ENGINEERED	UNP Q06187

- Molecule 2 is N-(2-CHLORO-6-METHYLPHENYL)-2-({6-[4-(2-HYDROXYETHYL)PIPERAZIN-1-YL]-2-METHYLPYRIMIDIN-4-YL}AMINO)-1,3-THIAZOLE-5-CARBOXAMIDE (three-letter code: 1N1) (formula: C<sub>22</sub>H<sub>26</sub>ClN<sub>7</sub>O<sub>2</sub>S).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	S	0	0
			33	22	1	7	2	1		

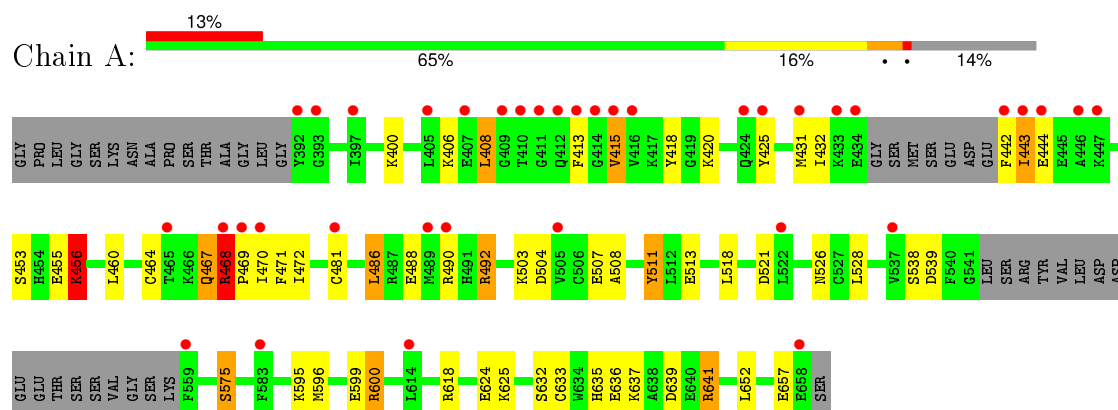
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	164	Total	O	0	0
			164	164		

### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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: Tyrosine-protein kinase BTK



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.44Å 104.62Å 38.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.40 – 1.94 28.40 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.2 (28.40-1.94) 99.2 (28.40-1.94)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.33 (at 1.93Å)	Xtriage
Refinement program	REFMAC 5.5.0088	Depositor
R, $R_{free}$	0.199 , 0.258 0.199 , 0.256	Depositor DCC
$R_{free}$ test set	1125 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	28.6	Xtriage
Anisotropy	0.276	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 56.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 22266 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2212	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.16% 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: 1N1

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	1.13	7/2067 (0.3%)	0.98	5/2783 (0.2%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	636	GLU	CD-OE1	8.14	1.34	1.25
1	A	636	GLU	CG-CD	7.61	1.63	1.51
1	A	575	SER	CA-CB	5.63	1.61	1.52
1	A	596	MET	C-O	5.57	1.33	1.23
1	A	511	TYR	CG-CD1	5.26	1.46	1.39
1	A	456	LYS	CB-CG	-5.21	1.38	1.52
1	A	513	GLU	CD-OE2	5.12	1.31	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	486	LEU	CB-CG-CD1	7.95	124.51	111.00
1	A	521	ASP	CB-CG-OD1	6.30	123.97	118.30
1	A	641	ARG	NE-CZ-NH2	-5.80	117.40	120.30
1	A	641	ARG	NE-CZ-NH1	5.79	123.20	120.30
1	A	636	GLU	CG-CD-OE1	5.19	128.68	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

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	2015	0	1991	54	0
2	A	33	0	26	1	0
3	A	164	0	0	17	0
All	All	2212	0	2017	55	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 14.

All (55) 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:468:ARG:HB2	1:A:469:PRO:CD	1.78	1.12
1:A:599:GLU:OE2	3:A:194:HOH:O	1.74	1.05
1:A:468:ARG:HB2	1:A:469:PRO:HD2	1.51	0.93
1:A:492:ARG:HH11	1:A:492:ARG:HG3	1.32	0.92
1:A:420:LYS:HB3	3:A:109:HOH:O	1.72	0.90
1:A:507:GLU:O	3:A:143:HOH:O	1.93	0.86
1:A:468:ARG:CB	1:A:469:PRO:CD	2.54	0.85
1:A:455:GLU:HG3	3:A:133:HOH:O	1.77	0.84
1:A:624:GLU:OE1	3:A:63:HOH:O	1.98	0.81
1:A:618:ARG:NE	3:A:155:HOH:O	2.21	0.73
1:A:468:ARG:CB	1:A:469:PRO:HD3	2.18	0.72
1:A:464:CYS:HB2	1:A:471:PHE:HB2	1.74	0.70
1:A:490:ARG:O	1:A:492:ARG:NH1	2.26	0.69
1:A:600:ARG:HG2	1:A:600:ARG:HH11	1.59	0.68
1:A:503:LYS:HD2	1:A:652:LEU:HD21	1.75	0.67
1:A:639:ASP:OD2	3:A:156:HOH:O	2.11	0.67
1:A:468:ARG:HB2	1:A:469:PRO:HD3	1.71	0.64
1:A:618:ARG:NH2	3:A:119:HOH:O	2.31	0.62
2:A:1:1N1:S	2:A:1:1N1:N4	2.73	0.61
1:A:635[A]:HIS:HD2	1:A:641:ARG:HA	1.68	0.59
1:A:467:GLN:HG2	1:A:468:ARG:HD3	1.83	0.58
1:A:481:CYS:HA	1:A:528:LEU:HD23	1.85	0.57
1:A:456:LYS:HB3	1:A:508:ALA:HB2	1.87	0.56
1:A:432:ILE:HD13	1:A:472:ILE:HD11	1.88	0.55
1:A:595:LYS:HG3	1:A:599:GLU:OE2	2.06	0.55
1:A:633:CYS:O	1:A:641:ARG:HD3	2.07	0.55
1:A:635[B]:HIS:HD2	1:A:637:LYS:H	1.56	0.54
1:A:632:SER:O	1:A:635[A]:HIS:CD2	2.61	0.54

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:492:ARG:NH1	1:A:492:ARG:HG3	2.11	0.54
1:A:456:LYS:HG2	1:A:504:ASP:O	2.08	0.53
1:A:600:ARG:CD	3:A:125:HOH:O	2.56	0.53
1:A:468:ARG:NH1	3:A:86:HOH:O	2.42	0.53
1:A:470:ILE:HD12	1:A:470:ILE:H	1.75	0.51
1:A:600:ARG:HD3	3:A:125:HOH:O	2.09	0.51
1:A:406:LYS:NZ	3:A:93:HOH:O	2.43	0.50
1:A:467:GLN:CG	1:A:468:ARG:HD3	2.41	0.49
1:A:468:ARG:HB3	1:A:469:PRO:HD3	1.93	0.48
1:A:503:LYS:NZ	3:A:131:HOH:O	2.41	0.48
1:A:467:GLN:HG2	1:A:468:ARG:N	2.29	0.48
1:A:443:ILE:HG23	1:A:444:GLU:HG2	1.97	0.47
1:A:511:TYR:HB2	3:A:143:HOH:O	2.14	0.47
1:A:625:LYS:CE	1:A:657:GLU:OE1	2.63	0.47
1:A:492:ARG:HH11	1:A:492:ARG:CG	2.14	0.46
1:A:618:ARG:CZ	3:A:155:HOH:O	2.63	0.45
1:A:625:LYS:NZ	1:A:657:GLU:OE1	2.45	0.45
1:A:408:LEU:HD22	1:A:418:TYR:N	2.32	0.44
1:A:600:ARG:CG	1:A:600:ARG:HH11	2.30	0.44
1:A:600:ARG:HD2	3:A:125:HOH:O	2.18	0.43
1:A:442:PHE:O	1:A:444:GLU:N	2.52	0.42
1:A:431:MET:HG2	1:A:471:PHE:CE1	2.54	0.42
1:A:518:LEU:HD21	1:A:575:SER:O	2.20	0.42
1:A:488:GLU:HG3	1:A:490:ARG:HG2	2.01	0.42
1:A:432:ILE:HD12	1:A:442:PHE:HB2	2.02	0.41
1:A:425:TYR:N	3:A:109:HOH:O	2.54	0.41
1:A:526:ASN:O	1:A:538:SER:HB3	2.22	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	238/283 (84%)	229 (96%)	6 (2%)	3 (1%)	15 5

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	415	VAL
1	A	443	ILE
1	A	468	ARG

### 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	220/251 (88%)	207 (94%)	13 (6%)	24 9

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

Mol	Chain	Res	Type
1	A	400	LYS
1	A	408	LEU
1	A	413	PHE
1	A	415	VAL
1	A	453	SER
1	A	456	LYS
1	A	460	LEU
1	A	467	GLN
1	A	468	ARG
1	A	486	LEU
1	A	492	ARG
1	A	539	ASP
1	A	600	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	467	GLN
1	A	484	ASN
1	A	609	HIS

### 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](#)

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	$\# Z  > 2$
2	1N1	A	1	-	32,36,36	0.90	2 (6%)	39,50,50	2.33	15 (38%)

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
2	1N1	A	1	-	-	0/13/29/29	0/4/4/4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1	1N1	C4-N2	-2.80	1.37	1.43
2	A	1	1N1	C5-CL	2.35	1.79	1.73

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1	1N1	C17-C16-N5	-3.48	103.73	110.63
2	A	1	1N1	N4-C14-N3	-3.11	119.46	125.58
2	A	1	1N1	O-C3-C2	-2.12	116.73	121.23
2	A	1	1N1	C12-C11-N4	-2.11	118.57	123.43
2	A	1	1N1	C21-C20-N6	-2.05	105.78	113.41
2	A	1	1N1	C-N-C11	-2.01	122.42	128.93
2	A	1	1N1	C20-N6-C17	2.08	116.60	111.27
2	A	1	1N1	C8-C9-C4	2.13	121.09	117.78
2	A	1	1N1	C15-C14-N4	2.44	121.36	117.20
2	A	1	1N1	C9-C4-N2	3.27	123.85	118.99
2	A	1	1N1	C20-N6-C18	3.54	120.36	111.27
2	A	1	1N1	C2-C3-N2	3.67	119.09	113.75
2	A	1	1N1	C14-N3-C13	3.84	118.72	115.64
2	A	1	1N1	C19-N5-C16	6.18	124.58	111.59
2	A	1	1N1	C18-N6-C17	6.53	123.03	108.90

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1	1N1	1	0

## 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	243/283 (85%)	0.87	37 (15%) <b>3</b> <b>4</b>	9, 19, 49, 64	4 (1%)

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	415	VAL	8.4
1	A	442	PHE	7.1
1	A	443	ILE	7.0
1	A	470	ILE	6.8
1	A	413	PHE	5.9
1	A	469	PRO	5.4
1	A	431	MET	5.3
1	A	559	PHE	5.2
1	A	446	ALA	4.9
1	A	481	CYS	4.7
1	A	414	GLY	4.4
1	A	411	GLY	4.3
1	A	425	TYR	4.2
1	A	410	THR	4.0
1	A	412	GLN	3.9
1	A	447	LYS	3.8
1	A	392	TYR	3.7
1	A	468	ARG	3.6
1	A	433	LYS	3.6
1	A	397	ILE	3.5
1	A	407	GLU	3.3
1	A	393	GLY	3.3
1	A	434	GLU	3.2
1	A	490	ARG	3.2
1	A	465	THR	3.1
1	A	522	LEU	2.9
1	A	444	GLU	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	489	MET	2.6
1	A	505	VAL	2.6
1	A	614	LEU	2.5
1	A	583	PHE	2.3
1	A	405	LEU	2.2
1	A	658	GLU	2.2
1	A	409	GLY	2.2
1	A	424	GLN	2.2
1	A	537	VAL	2.2
1	A	416	VAL	2.1

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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](#)

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
2	1N1	A	1	33/33	0.91	0.18	0.60	29,41,52,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.