



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 10:20 PM GMT

PDB ID : 4XI2  
Title : Crystal Structure of an auto-inhibited form of Bruton's Tryrosine Kinase  
Authors : Vogan, E.M.; Harrison, S.C.  
Deposited on : 2015-01-06  
Resolution : 2.60 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

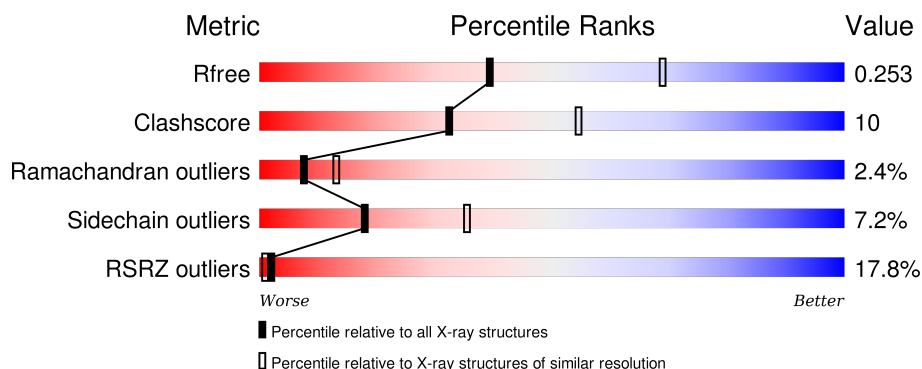
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	446	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	AU	A	702	-	-	-	X

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6835 atoms, of which 3375 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
1	A	421	Total	C	H	N	O	S	54	0	0
			6833	2215	3375	577	642	24			

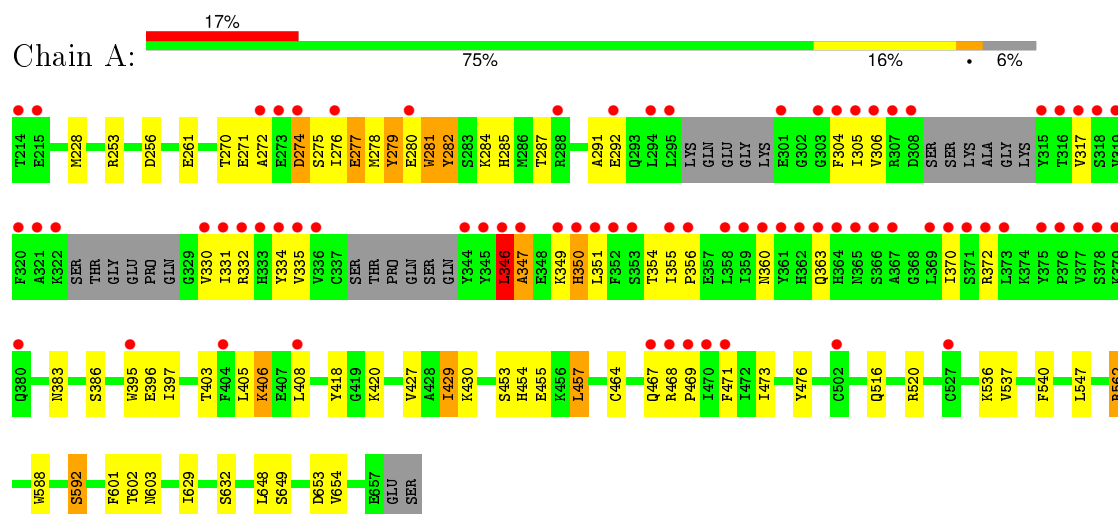
- Molecule 2 is GOLD ION (three-letter code: AU) (formula: Au).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Au	0	0
			2	2		

### 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: Tyrosine-protein kinase BTK



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	132.20 Å   132.20 Å   107.63 Å 90.00°   90.00°   120.00°	Depositor
Resolution (Å)	43.27 – 2.60 43.27 – 2.60	Depositor EDS
% Data completeness (in resolution range)	89.2 (43.27-2.60) 91.8 (43.27-2.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.46 (at 2.61 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1839)	Depositor
R, $R_{free}$	0.230 , 0.248 0.227 , 0.253	Depositor DCC
$R_{free}$ test set	1524 reflections (4.93%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.2	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 73.8	EDS
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 30937 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6835	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	122.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.02% 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: AU

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	0.35	0/3540	0.54	0/4774

There are no bond length outliers.

There are no bond angle outliers.

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	3458	3375	3375	71	6
2	A	2	0	0	0	0
All	All	3460	3375	3375	71	6

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 10.

All (71) 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:292:GLU:HG2	1:A:331:ILE:HD12	1.21	1.18
1:A:292:GLU:CG	1:A:331:ILE:HD12	1.99	0.91

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:292:GLU:HA	1:A:331:ILE:CD1	2.01	0.90
1:A:349:LYS:NZ	1:A:350:HIS:HB2	1.86	0.90
1:A:292:GLU:HA	1:A:331:ILE:HD13	1.61	0.82
1:A:292:GLU:HG2	1:A:331:ILE:CD1	2.06	0.81
1:A:349:LYS:HZ2	1:A:350:HIS:HB2	1.52	0.73
1:A:429:ILE:HD12	1:A:430:LYS:N	2.06	0.71
1:A:349:LYS:HZ3	1:A:350:HIS:HB2	1.54	0.70
1:A:429:ILE:HD12	1:A:429:ILE:C	2.14	0.68
1:A:354:THR:C	1:A:356:PRO:HD2	2.13	0.68
1:A:429:ILE:HD11	1:A:471:PHE:CD2	2.30	0.67
1:A:355:ILE:N	1:A:356:PRO:HD2	2.10	0.67
1:A:304:PHE:O	1:A:305:ILE:CG2	2.43	0.67
1:A:349:LYS:HG3	1:A:350:HIS:N	2.10	0.66
1:A:306:VAL:HG22	1:A:317:VAL:HG22	1.79	0.64
1:A:292:GLU:CG	1:A:331:ILE:CD1	2.73	0.62
1:A:304:PHE:O	1:A:305:ILE:HG23	2.00	0.62
1:A:354:THR:CB	1:A:356:PRO:HD2	2.30	0.62
1:A:562:ARG:NH1	1:A:603:ASN:OD1	2.33	0.62
1:A:276:ILE:CG2	1:A:282:TYR:CD1	2.85	0.59
1:A:355:ILE:N	1:A:356:PRO:CD	2.65	0.59
1:A:334:TYR:CD2	1:A:346:LEU:HD21	2.38	0.58
1:A:629:ILE:O	1:A:632:SER:OG	2.19	0.58
1:A:253:ARG:NH2	1:A:261:GLU:OE1	2.37	0.58
1:A:349:LYS:HG3	1:A:350:HIS:H	1.69	0.58
1:A:292:GLU:CA	1:A:331:ILE:CD1	2.79	0.57
1:A:349:LYS:CG	1:A:350:HIS:H	2.17	0.57
1:A:304:PHE:C	1:A:305:ILE:HG23	2.26	0.56
1:A:335:VAL:O	1:A:346:LEU:O	2.24	0.56
1:A:349:LYS:CG	1:A:350:HIS:N	2.68	0.56
1:A:354:THR:HB	1:A:356:PRO:HD2	1.90	0.54
1:A:276:ILE:HG21	1:A:282:TYR:CD1	2.42	0.54
1:A:562:ARG:NH2	1:A:601:PHE:O	2.41	0.54
1:A:291:ALA:HB1	1:A:305:ILE:HD11	1.91	0.53
1:A:588:TRP:O	1:A:592:SER:OG	2.26	0.53
1:A:356:PRO:O	1:A:360:ASN:OD1	2.26	0.53
1:A:408:LEU:HD11	1:A:476:TYR:HE1	1.73	0.53
1:A:334:TYR:HD2	1:A:346:LEU:HD21	1.73	0.52
1:A:464:CYS:SG	1:A:471:PHE:HB2	2.52	0.50
1:A:363:GLN:O	1:A:372:ARG:HD2	2.12	0.50
1:A:330:VAL:HG22	1:A:331:ILE:N	2.27	0.49
1:A:279:TYR:N	1:A:279:TYR:CD1	2.82	0.47

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:457:LEU:HD22	1:A:540:PHE:CE1	2.50	0.46
1:A:349:LYS:CD	1:A:350:HIS:H	2.29	0.46
1:A:370:ILE:O	1:A:370:ILE:HG22	2.16	0.45
1:A:405:LEU:HG	1:A:406:LYS:HD2	1.99	0.45
1:A:334:TYR:HB3	1:A:346:LEU:HG	1.99	0.45
1:A:304:PHE:C	1:A:305:ILE:CG2	2.85	0.45
1:A:277:GLU:H	1:A:277:GLU:HG2	1.39	0.45
1:A:429:ILE:CD1	1:A:429:ILE:C	2.85	0.44
1:A:397:ILE:HG13	1:A:473:ILE:HD13	2.00	0.44
1:A:276:ILE:HG22	1:A:282:TYR:CG	2.52	0.44
1:A:454:HIS:HB3	1:A:457:LEU:HB2	2.00	0.44
1:A:279:TYR:H	1:A:279:TYR:HD1	1.63	0.44
1:A:351:LEU:HA	1:A:351:LEU:HD12	1.71	0.44
1:A:304:PHE:O	1:A:305:ILE:HG22	2.16	0.43
1:A:455:GLU:O	1:A:536:LYS:HD3	2.19	0.43
1:A:468:ARG:HA	1:A:469:PRO:C	2.40	0.42
1:A:274:ASP:HB3	1:A:279:TYR:OH	2.20	0.41
1:A:405:LEU:HD22	1:A:420:LYS:HE3	2.02	0.41
1:A:429:ILE:HD11	1:A:471:PHE:CG	2.55	0.41
1:A:346:LEU:HB3	1:A:347:ALA:H	1.63	0.41
1:A:408:LEU:HD11	1:A:418:TYR:HB2	2.02	0.41
1:A:349:LYS:HD2	1:A:350:HIS:H	1.84	0.41
1:A:284:LYS:HG2	1:A:285:HIS:H	1.85	0.41
1:A:270:THR:OG1	1:A:271:GLU:N	2.53	0.41
1:A:292:GLU:CB	1:A:331:ILE:HD12	2.47	0.41
1:A:282:TYR:CE2	1:A:284:LYS:N	2.84	0.41
1:A:395:TRP:CG	1:A:396:GLU:N	2.88	0.40
1:A:284:LYS:HG2	1:A:285:HIS:N	2.37	0.40

All (6) 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:A:332:ARG:HG3	1:A:370:ILE:CD1[4_556]	0.79	0.81
1:A:332:ARG:HG3	1:A:370:ILE:HD11[4_556]	0.93	0.67
1:A:332:ARG:HG3	1:A:370:ILE:HD13[4_556]	0.94	0.66
1:A:332:ARG:CG	1:A:370:ILE:CD1[4_556]	1.71	0.49
1:A:332:ARG:HG3	1:A:370:ILE:HD12[4_556]	1.24	0.36
1:A:332:ARG:CG	1:A:370:ILE:HD11[4_556]	1.51	0.09



## 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	411/446 (92%)	374 (91%)	27 (7%)	10 (2%)	<b>7</b> <b>13</b>

All (10) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	272	ALA
1	A	274	ASP
1	A	287	THR
1	A	275	SER
1	A	346	LEU
1	A	347	ALA
1	A	280	GLU
1	A	282	TYR
1	A	383	ASN
1	A	281	TRP

### 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	376/397 (95%)	349 (93%)	27 (7%)	<b>18</b> <b>35</b>

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	228	MET

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	256	ASP
1	A	277	GLU
1	A	278	MET
1	A	279	TYR
1	A	281	TRP
1	A	346	LEU
1	A	350	HIS
1	A	386	SER
1	A	403	THR
1	A	406	LYS
1	A	427	VAL
1	A	429	ILE
1	A	453	SER
1	A	457	LEU
1	A	467	GLN
1	A	516	GLN
1	A	520	ARG
1	A	537	VAL
1	A	547	LEU
1	A	562	ARG
1	A	592	SER
1	A	602	THR
1	A	648	LEU
1	A	649	SER
1	A	653	ASP
1	A	654	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules 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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	421/446 (94%)	1.30	75 (17%) 2 1	53, 94, 214, 219	6 (1%)

All (75) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	358	LEU	12.0
1	A	344	TYR	11.7
1	A	318	SER	10.7
1	A	378	SER	10.3
1	A	316	THR	10.1
1	A	331	ILE	10.1
1	A	321	ALA	8.9
1	A	319	VAL	8.1
1	A	372	ARG	7.8
1	A	288	ARG	7.6
1	A	332	ARG	7.4
1	A	359	ILE	7.3
1	A	305	ILE	7.1
1	A	307	ARG	7.0
1	A	320	PHE	7.0
1	A	380	GLN	7.0
1	A	304	PHE	6.8
1	A	295	LEU	6.8
1	A	335	VAL	6.7
1	A	362	HIS	6.4
1	A	346	LEU	6.0
1	A	355	ILE	6.0
1	A	274	ASP	5.8
1	A	373	LEU	5.8
1	A	375	TYR	5.7
1	A	330	VAL	5.6
1	A	370	ILE	5.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	334	TYR	5.4
1	A	349	LYS	5.3
1	A	306	VAL	5.2
1	A	336	VAL	5.0
1	A	352	PHE	4.9
1	A	303	GLY	4.8
1	A	361	TYR	4.8
1	A	369	LEU	4.8
1	A	366	SER	4.7
1	A	333	HIS	4.6
1	A	364	HIS	4.5
1	A	360	ASN	4.5
1	A	317	VAL	4.3
1	A	356	PRO	4.3
1	A	272	ALA	4.1
1	A	353	SER	4.1
1	A	371	SER	4.0
1	A	350	HIS	3.9
1	A	395	TRP	3.9
1	A	214	THR	3.7
1	A	292	GLU	3.7
1	A	315	TYR	3.5
1	A	351	LEU	3.4
1	A	467	GLN	3.2
1	A	301	GLU	3.1
1	A	377	VAL	3.0
1	A	408	LEU	2.9
1	A	308	ASP	2.8
1	A	322	LYS	2.7
1	A	404	PHE	2.7
1	A	471	PHE	2.7
1	A	365	ASN	2.7
1	A	379	LYS	2.7
1	A	273	GLU	2.6
1	A	468	ARG	2.5
1	A	527	CYS	2.5
1	A	280	GLU	2.5
1	A	469	PRO	2.4
1	A	376	PRO	2.3
1	A	294	LEU	2.2
1	A	363	GLN	2.2
1	A	470	ILE	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	276	ILE	2.1
1	A	215	GLU	2.1
1	A	345	TYR	2.0
1	A	347	ALA	2.0
1	A	367	ALA	2.0
1	A	502	CYS	2.0

## 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	AU	A	702	1/1	0.89	0.35	18.81	218,218,218,218	0
2	AU	A	701	1/1	0.98	0.16	-1.23	191,191,191,191	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.