wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 08:34 pm BST

PDB ID	:	2XH5
Title	:	Structure of 4-(4-tert-Butylbenzyl)-1-(7H-pyrrolo(2,3-d)pyrimidin-4-yl)piperidin-4-amine bound to PKB
Authors	:	Davies, T.G.; McHardy, T.; Caldwell, J.J.; Cheung, K.M.; Hunter, L.J.; Taylor, K.; Rowlands, M.; Ruddle, R.; Henley, A.; Brandon, A.D.; Valenti, M.; Fazal, L.; Seavers, L.; Raynaud, F.I.; Eccles, S.A.; Aherne, G.W.; Garrett, M.D.; Collins, I.
Deposited on	:	2010-06-09
Resolution	:	2.72 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

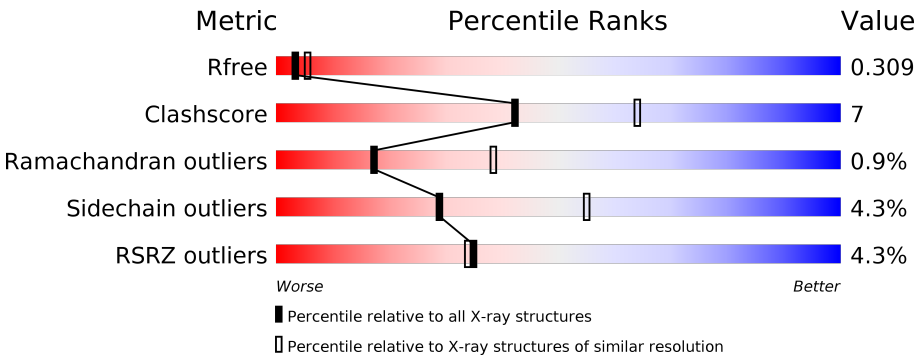
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.72 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}	130704	3359 (2.74-2.70)
Clashscore	141614	3686 (2.74-2.70)
Ramachandran outliers	138981	3622 (2.74-2.70)
Sidechain outliers	138945	3623 (2.74-2.70)
RSRZ outliers	127900	3276 (2.74-2.70)

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 >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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 <=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	342	4% 75% 17% 7%
2	C	10	10% 90% 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2771 atoms, of which 30 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 RAC-BETA SERINE/THREONINE-PROTEIN KINASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
1	A	317	2591	1665	434	476	1	15	0	0	0

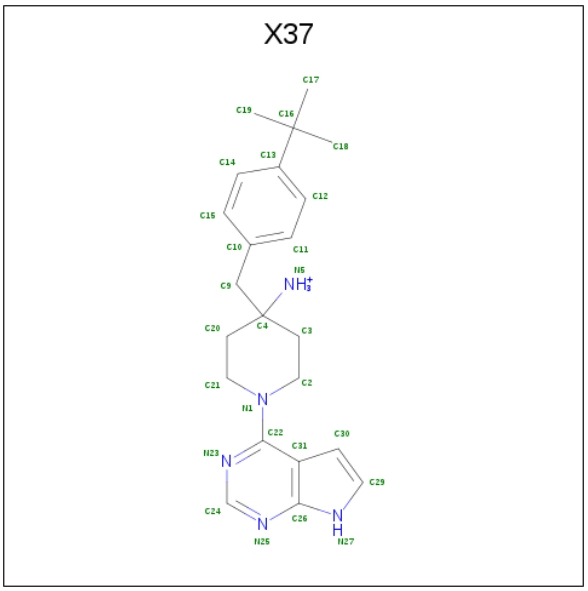
There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	141	GLY	-	expression tag	UNP P31751
A	142	ALA	-	expression tag	UNP P31751
A	143	MET	-	expression tag	UNP P31751
A	144	ASP	-	expression tag	UNP P31751
A	145	PRO	-	expression tag	UNP P31751
A	464C	ARG	-	insertion	UNP P31751
A	465	GLU	-	insertion	UNP P31751
A	466	GLU	-	insertion	UNP P31751
A	467	GLN	ARG	engineered mutation	UNP P31751
A	468	GLU	THR	engineered mutation	UNP P31751
A	469	MET	HIS	engineered mutation	UNP P31751
A	471	GLU	PRO	engineered mutation	UNP P31751
A	472	ASP	GLN	engineered mutation	UNP P31751
A	474	ASP	SER	engineered mutation	UNP P31751
A	476	ILE	SER	engineered mutation	UNP P31751
A	478	ASP	SER	engineered mutation	UNP P31751
A	479	TRP	ILE	engineered mutation	UNP P31751

- Molecule 2 is a protein called GLYCOGEN SYNTHASE KINASE-3 BETA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	10	79	47	16	16	0	0	0

- Molecule 3 is 4-(4-tert-butylbenzyl)-1-(7H-pyrrolo[2,3-d]pyrimidin-4-yl)piperidin-4-aminium (three-letter code: X37) (formula: $C_{22}H_{30}N_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	N		
3	A	1	57	22	30	5	0	0

- Molecule 4 is water.

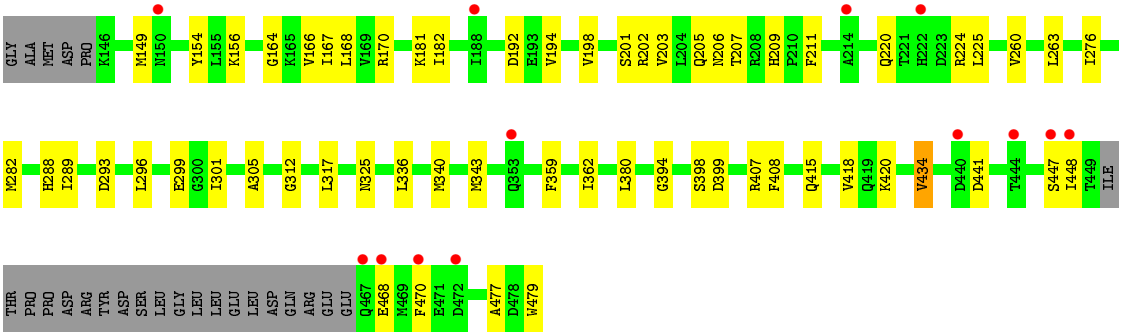
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	41	41	41	0	0
			Total	O		
4	C	3	3	3	0	0

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 the various outlier classes 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: RAC-BETA SERINE/THREONINE-PROTEIN KINASE

Chain A:



- Molecule 2: GLYCOGEN SYNTHASE KINASE-3 BETA

Chain C:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	44.95Å 61.23Å 135.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	67.42 – 2.72 55.81 – 2.72	Depositor EDS
% Data completeness (in resolution range)	97.5 (67.42-2.72) 97.5 (55.81-2.72)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.16 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.6.0062	Depositor
R, R_{free}	0.211 , 0.305 0.219 , 0.309	Depositor DCC
R_{free} test set	492 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	43.5	Xtriage
Anisotropy	0.465	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 40.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2771	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.59% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: TPO, X37

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	0.67	0/2639	0.74	1/3554 (0.0%)
2	C	0.68	0/80	0.80	0/105
All	All	0.67	0/2719	0.74	1/3659 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	170	ARG	NE-CZ-NH1	5.17	122.89	120.30

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	2591	0	2542	35	0
2	C	79	0	74	0	0
3	A	27	30	30	4	0
4	A	41	0	0	1	0
4	C	3	0	0	0	0
All	All	2741	30	2646	37	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 7.

The worst 5 of 37 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:167:ILE:HG22	1:A:168:LEU:O	1.74	0.87
1:A:220:GLN:HB3	1:A:225:LEU:HD23	1.64	0.78
1:A:380:LEU:HD22	1:A:408:PHE:CD2	2.20	0.77
1:A:301:ILE:HG23	1:A:305:ALA:HB3	1.67	0.74
3:A:1480:X37:H30	3:A:1480:X37:H2A	1.76	0.68

There are no symmetry-related clashes.

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	312/342 (91%)	288 (92%)	22 (7%)	2 (1%)	25 48
2	C	8/10 (80%)	7 (88%)	0	1 (12%)	0 0
All	All	320/352 (91%)	295 (92%)	22 (7%)	3 (1%)	17 38

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	4	ARG
1	A	468	GLU
1	A	293	ASP

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	274/300 (91%)	263 (96%)	11 (4%)	31	58
2	C	8/8 (100%)	7 (88%)	1 (12%)	4	10
All	All	282/308 (92%)	270 (96%)	12 (4%)	29	55

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	407	ARG
1	A	418	VAL
1	A	447	SER
1	A	398	SER
1	A	441	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	209	HIS
1	A	415	GLN
1	A	233	ASN
1	A	205	GLN
1	A	391	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue 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	$\#\|Z\| > 2$
1	TPO	A	309	1	8,10,11	0.98	0	10,14,16	1.35	2 (20%)

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
1	TPO	A	309	1	-	0/9/11/13	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	309	TPO	O3P-P-O2P	2.05	115.46	107.64
1	A	309	TPO	CG2-CB-CA	-2.02	109.17	113.16

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	$\#\|Z\| > 2$
3	X37	A	1480	-	25,30,30	0.96	1 (4%)	32,45,45	1.07	3 (9%)

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
3	X37	A	1480	-	-	4/14/27/27	0/4/4/4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1480	X37	C26-N25	-3.53	1.32	1.37

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	1480	X37	C3-C4-C9	3.37	114.83	110.05
3	A	1480	X37	C20-C4-N5	-2.57	105.03	109.95
3	A	1480	X37	C3-C4-N5	-2.35	105.44	109.95

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1480	X37	N23-C22-N1-C2
3	A	1480	X37	C31-C22-N1-C2
3	A	1480	X37	C20-C4-C9-C10
3	A	1480	X37	C12-C13-C16-C18

There are no ring outliers.

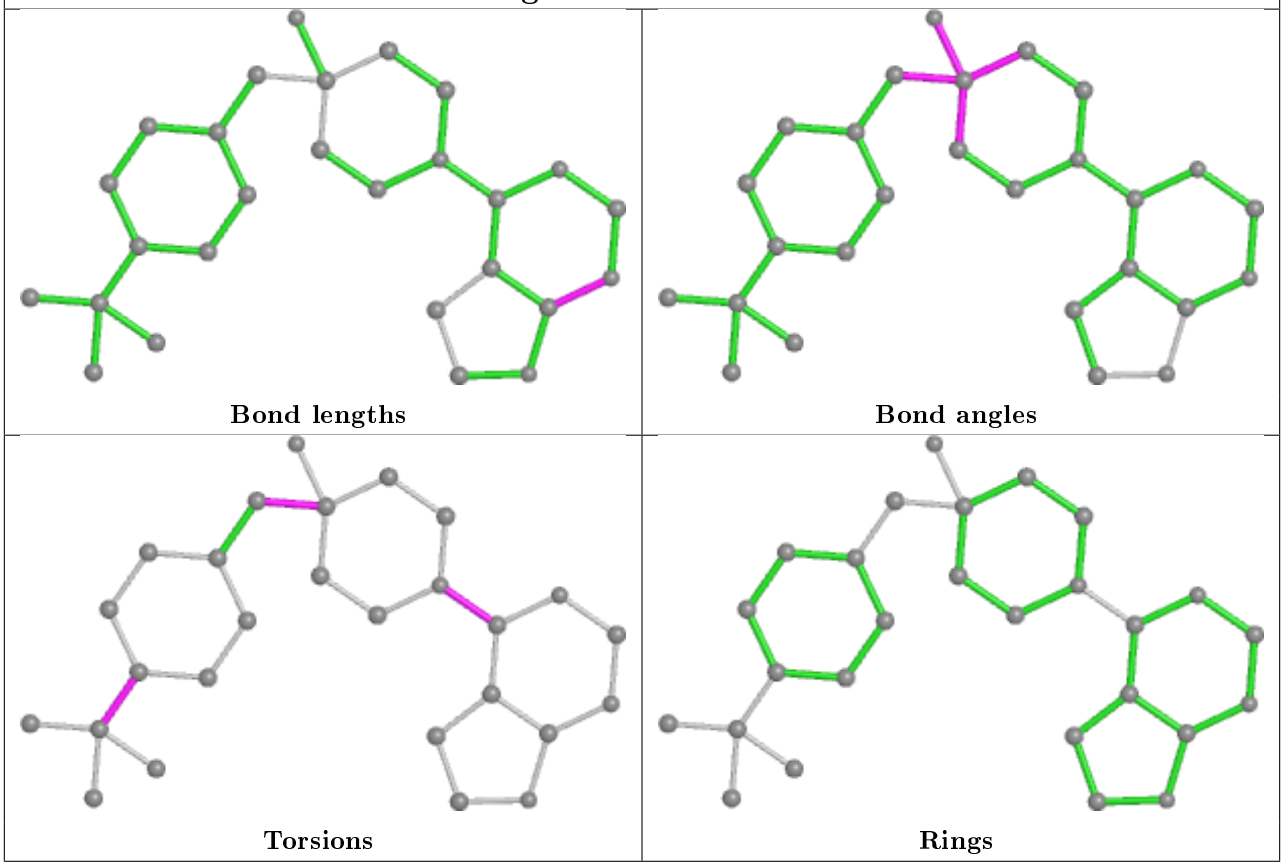
1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1480	X37	4	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will

also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand X37 A 1480



5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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^2)	Q<0.9
1	A	316/342 (92%)	0.23	13 (4%)	37	36	19, 35, 60, 109	0
2	C	10/10 (100%)	0.84	1 (10%)	7	5	46, 48, 74, 76	0
All	All	326/352 (92%)	0.25	14 (4%)	35	34	19, 36, 60, 109	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	12	GLU	4.5
1	A	448	ILE	4.2
1	A	468	GLU	4.0
1	A	444	THR	3.4
1	A	447	SER	3.1

6.2 Non-standard residues in protein, DNA, RNA chains (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. The B-factors column lists the minimum, median, 95th 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	B-factors(Å^2)	Q<0.9
1	TPO	A	309	11/12	0.96	0.17	29,34,37,39	0

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

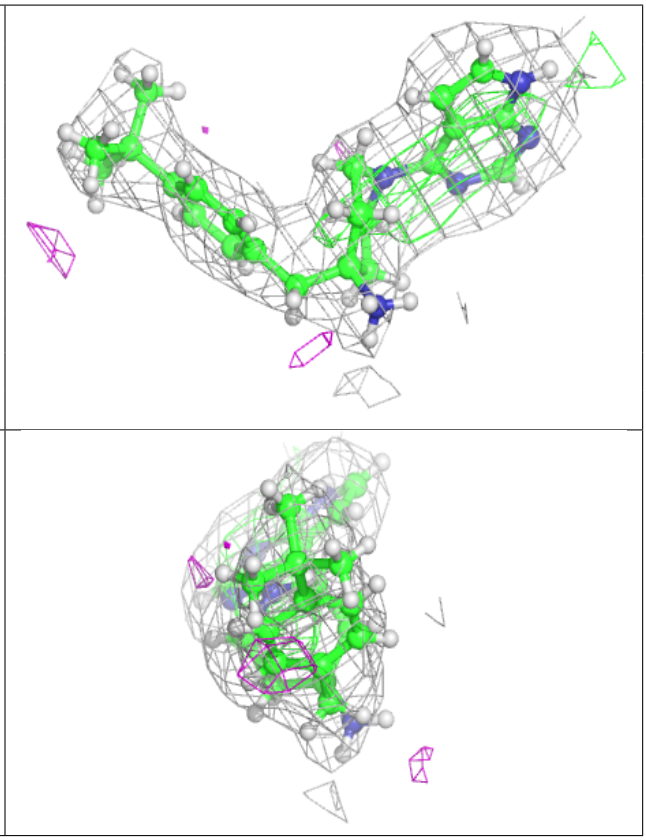
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	X37	A	1480	27/27	0.94	0.21	20,27,31,32	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around X37 A 1480:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.