wwPDB X-ray Structure Validation Summary Report (i)

Jun 22, 2024 – 11:45 PM EDT

PDB ID	:	6GQK
Title	:	Crystal structure of human c-KIT kinase domain in complex with AZD3229-analogue (compound 23)
Authors	:	Schimpl, M.; Hardy, C.J.; Ogg, D.J.; Overman, R.C.; Packer, M.J.; Kettle, J.G.; Anjum, R.; Barry, E.; Bhavsar, D.; Brown, C.; Campbell, A.; Goldberg, K.; Grondine, M.; Guichard, S.; Hunt, T.; Jones, O.; Li, X.; Moleva, O.; Pearson, S.; Shao, W.; Smith, A.; Smith, J.; Stead, D.; Stokes, S.; Tucker, M.; Ye, Y.
Deposited on	:	2018-06-07
Resolution	:	2.31 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

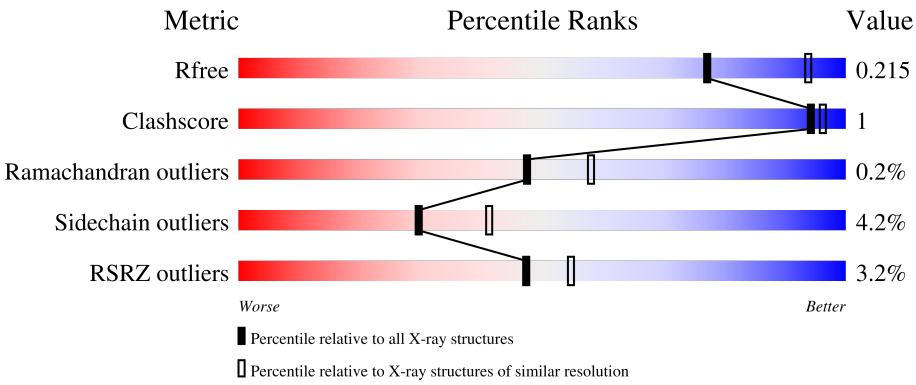
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)

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 2.31 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	5974 (2.34-2.30)
Clashscore	141614	6604 (2.34-2.30)
Ramachandran outliers	138981	6523 (2.34-2.30)
Sidechain outliers	138945	6523 (2.34-2.30)
RSRZ outliers	127900	5855 (2.34-2.30)

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 of 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	328	2%, 86%, 7%, 7%
1	B	328	3%, 84%, 5%, 11%

Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.37.1

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5064 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 Mast/stem cell growth factor receptor Kit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	305	2441	1570	402	453	16	0	2	0
1	B	293	2337	1501	388	432	16	0	2	0

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	547	GLY	-	expression tag	UNP P10721
A	548	SER	-	expression tag	UNP P10721
A	549	HIS	-	expression tag	UNP P10721
A	550	MET	-	expression tag	UNP P10721
A	563	SER	ILE	engineered mutation	UNP P10721
A	569	SER	VAL	engineered mutation	UNP P10721
A	609	GLN	TYR	engineered mutation	UNP P10721
A	631	SER	LEU	engineered mutation	UNP P10721
A	651	GLU	MET	engineered mutation	UNP P10721
A	662	HIS	ILE	engineered mutation	UNP P10721
A	688	GLU	-	linker	UNP P10721
A	689	PHE	-	linker	UNP P10721
A	690	VAL	-	linker	UNP P10721
A	691	PRO	-	linker	UNP P10721
A	692	TYR	-	linker	UNP P10721
A	693	LYS	-	linker	UNP P10721
A	694	VAL	-	linker	UNP P10721
A	695	ALA	-	linker	UNP P10721
A	696	PRO	-	linker	UNP P10721
A	757	GLU	-	linker	UNP P10721
A	758	ASP	-	linker	UNP P10721
A	759	LEU	-	linker	UNP P10721
A	760	TYR	-	linker	UNP P10721
A	761	LYS	-	linker	UNP P10721
A	762	ASP	-	linker	UNP P10721

Continued on next page...

Continued from previous page...

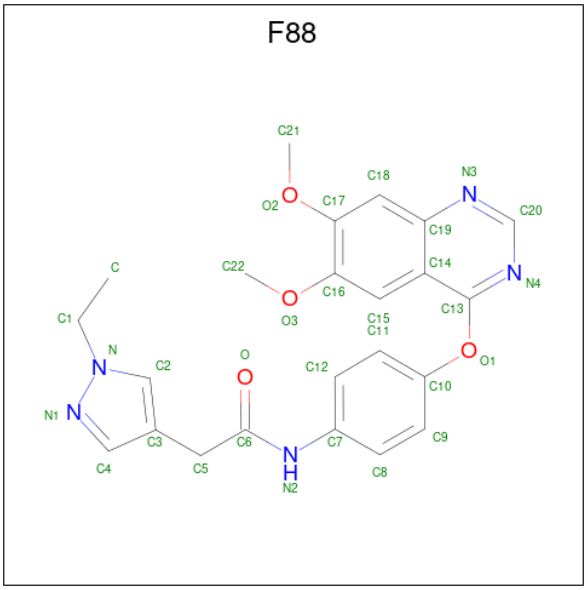
Chain	Residue	Modelled	Actual	Comment	Reference
A	763	PHE	-	linker	UNP P10721
A	764	LEU	-	linker	UNP P10721
A	765	THR	-	linker	UNP P10721
A	766	LEU	-	linker	UNP P10721
A	767	GLU	-	linker	UNP P10721
A	768	HIS	-	linker	UNP P10721
A	769	LEU	-	linker	UNP P10721
A	770	LEU	-	linker	UNP P10721
A	804	ASN	ARG	engineered mutation	UNP P10721
A	825	ASP	VAL	engineered mutation	UNP P10721
A	844	SER	CYS	engineered mutation	UNP P10721
A	890	SER	LEU	engineered mutation	UNP P10721
A	894	TYR	HIS	engineered mutation	UNP P10721
A	912	ASP	LEU	engineered mutation	UNP P10721
A	923	ASP	LEU	engineered mutation	UNP P10721
B	547	GLY	-	expression tag	UNP P10721
B	548	SER	-	expression tag	UNP P10721
B	549	HIS	-	expression tag	UNP P10721
B	550	MET	-	expression tag	UNP P10721
B	563	SER	ILE	engineered mutation	UNP P10721
B	569	SER	VAL	engineered mutation	UNP P10721
B	609	GLN	TYR	engineered mutation	UNP P10721
B	631	SER	LEU	engineered mutation	UNP P10721
B	651	GLU	MET	engineered mutation	UNP P10721
B	662	HIS	ILE	engineered mutation	UNP P10721
B	688	GLU	-	linker	UNP P10721
B	689	PHE	-	linker	UNP P10721
B	690	VAL	-	linker	UNP P10721
B	691	PRO	-	linker	UNP P10721
B	692	TYR	-	linker	UNP P10721
B	693	LYS	-	linker	UNP P10721
B	694	VAL	-	linker	UNP P10721
B	755	ALA	-	linker	UNP P10721
B	756	PRO	-	linker	UNP P10721
B	757	GLU	-	linker	UNP P10721
B	758	ASP	-	linker	UNP P10721
B	759	LEU	-	linker	UNP P10721
B	760	TYR	-	linker	UNP P10721
B	761	LYS	-	linker	UNP P10721
B	762	ASP	-	linker	UNP P10721
B	763	PHE	-	linker	UNP P10721
B	764	LEU	-	linker	UNP P10721

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	765	THR	-	linker	UNP P10721
B	766	LEU	-	linker	UNP P10721
B	767	GLU	-	linker	UNP P10721
B	768	HIS	-	linker	UNP P10721
B	769	LEU	-	linker	UNP P10721
B	770	LEU	-	linker	UNP P10721
B	804	ASN	ARG	engineered mutation	UNP P10721
B	825	ASP	VAL	engineered mutation	UNP P10721
B	844	SER	CYS	engineered mutation	UNP P10721
B	890	SER	LEU	engineered mutation	UNP P10721
B	894	TYR	HIS	engineered mutation	UNP P10721
B	912	ASP	LEU	engineered mutation	UNP P10721
B	923	ASP	LEU	engineered mutation	UNP P10721

- Molecule 2 is {N}-[4-(6,7-dimethoxyquinazolin-4-yl)oxyphenyl]-2-(1-ethylpyrazol-4-yl)ethanamide (three-letter code: F88) (formula: $C_{23}H_{23}N_5O_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	32	23	5	4	0	0
			Total	C	N	O		
2	B	1	32	23	5	4	0	0

- Molecule 3 is water.

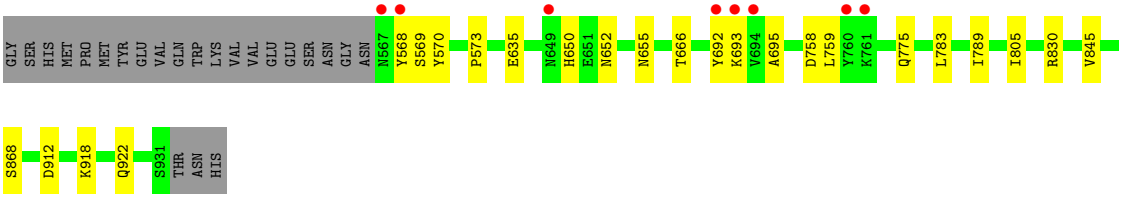
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	135	135	135	0	0
			Total	O		
3	B	87	87	87	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

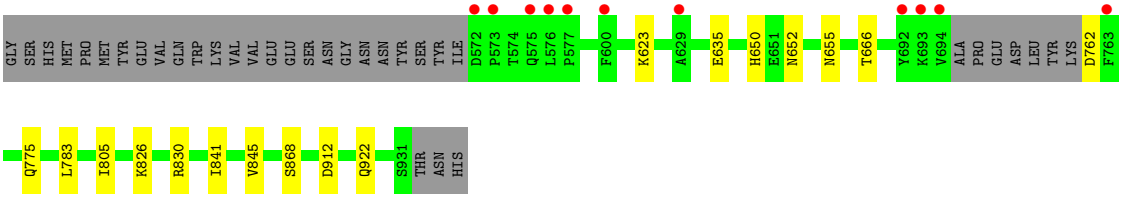
• Molecule 1: Mast/stem cell growth factor receptor Kit

Chain A:



• Molecule 1: Mast/stem cell growth factor receptor Kit

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	88.19Å 91.65Å 92.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.94 – 2.31 63.67 – 2.31	Depositor EDS
% Data completeness (in resolution range)	99.5 (64.94-2.31) 99.5 (63.67-2.31)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 2.32Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.187 , 0.220 0.186 , 0.215	Depositor DCC
R_{free} test set	1744 reflections (5.25%)	wwPDB-VP
Wilson B-factor (Å^2)	44.7	Xtriage
Anisotropy	0.188	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 66.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.015 for -h,l,k 0.024 for -l,-k,-h 0.014 for k,h,-l 0.000 for k,l,h 0.000 for l,h,k	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5064	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.97% 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: F88

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	0.53	0/2513	0.62	0/3401
1	B	0.51	0/2404	0.62	0/3250
All	All	0.52	0/4917	0.62	0/6651

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	2441	0	2401	7	0
1	B	2337	0	2305	5	0
2	A	32	0	0	0	0
2	B	32	0	0	0	0
3	A	135	0	0	0	0
3	B	87	0	0	0	0
All	All	5064	0	4706	12	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 1.

The worst 5 of 12 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:570:TYR:CE2	1:A:789:ILE:HD11	2.33	0.64
1:B:775:GLN:HE22	1:B:805:ILE:HA	1.64	0.63
1:A:650:HIS:HD2	1:A:652:ASN:H	1.47	0.63
1:A:775:GLN:HE22	1:A:805:ILE:HA	1.64	0.62
1:A:650:HIS:H	1:A:655:ASN:HD21	1.52	0.57

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	305/328 (93%)	299 (98%)	5 (2%)	1 (0%)	41	50
1	B	291/328 (89%)	286 (98%)	5 (2%)	0	100	100
All	All	596/656 (91%)	585 (98%)	10 (2%)	1 (0%)	47	58

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	692	TYR

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	267/286 (93%)	255 (96%)	12 (4%)	27	38
1	B	256/286 (90%)	246 (96%)	10 (4%)	32	45
All	All	523/572 (91%)	501 (96%)	22 (4%)	30	41

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	666	THR
1	B	830	ARG
1	B	783	LEU
1	B	845	VAL
1	A	830	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	655	ASN
1	B	775	GLN
1	A	655	ASN
1	A	775	GLN
1	B	587	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 ligands are 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	F88	B	1001	-	34,35,35	0.56	0	42,48,48	0.51	0
2	F88	A	1001	-	34,35,35	0.56	0	42,48,48	0.53	0

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	F88	B	1001	-	-	2/18/18/18	0/4/4/4
2	F88	A	1001	-	-	0/18/18/18	0/4/4/4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1001	F88	C12-C7-N2-C6
2	B	1001	F88	C8-C7-N2-C6

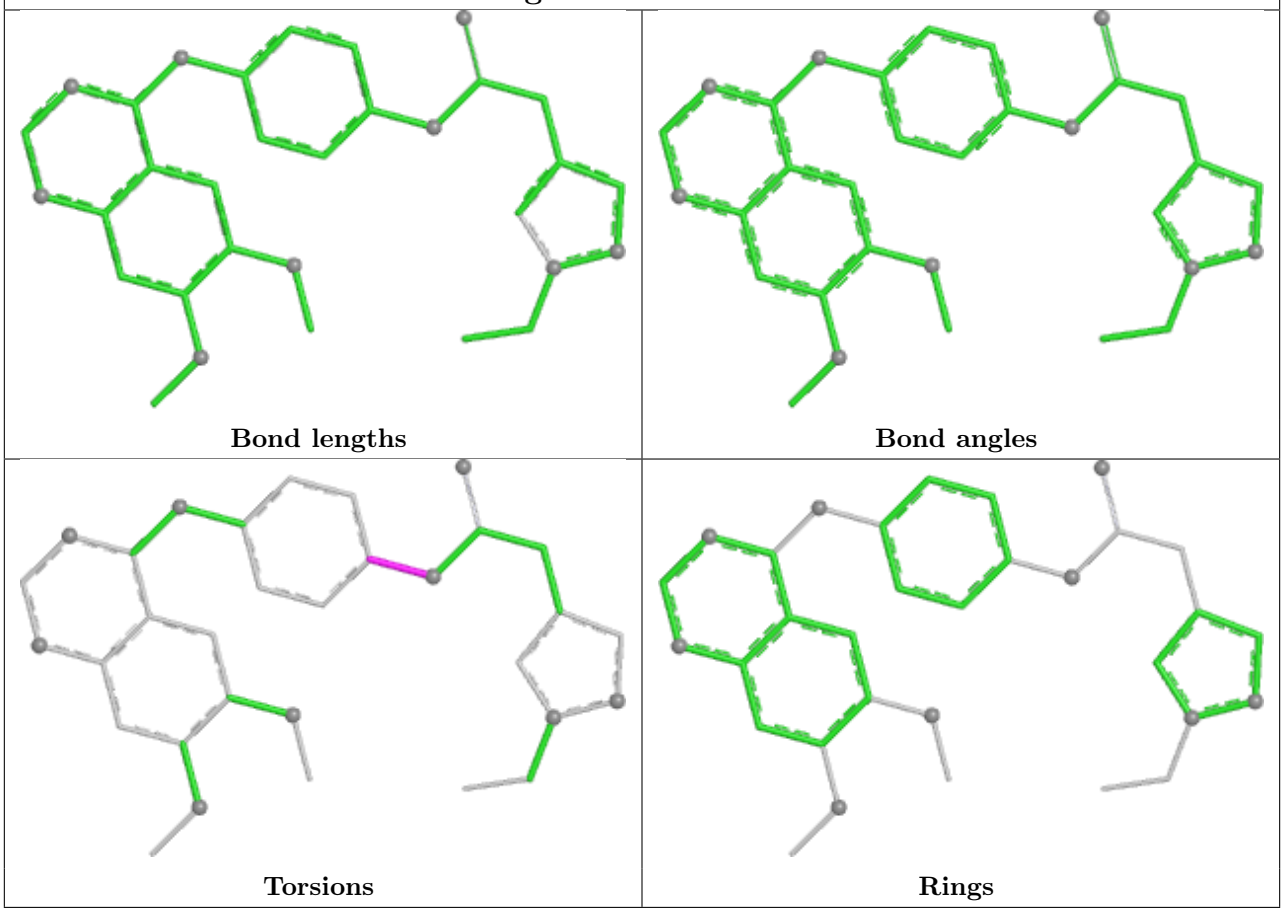
There are no ring outliers.

No monomer is involved in short contacts.

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 F88 B 1001



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	305/328 (92%)	0.01	8 (2%)	56	63	28, 45, 90, 146	0
1	B	293/328 (89%)	0.03	11 (3%)	40	47	34, 59, 97, 136	0
All	All	598/656 (91%)	0.02	19 (3%)	47	55	28, 52, 94, 146	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	694	VAL	8.5
1	B	694	VAL	8.1
1	A	567	ASN	5.5
1	A	760	TYR	5.4
1	A	692	TYR	5.3

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 monosaccharides in this entry.

6.4 Ligands (i)

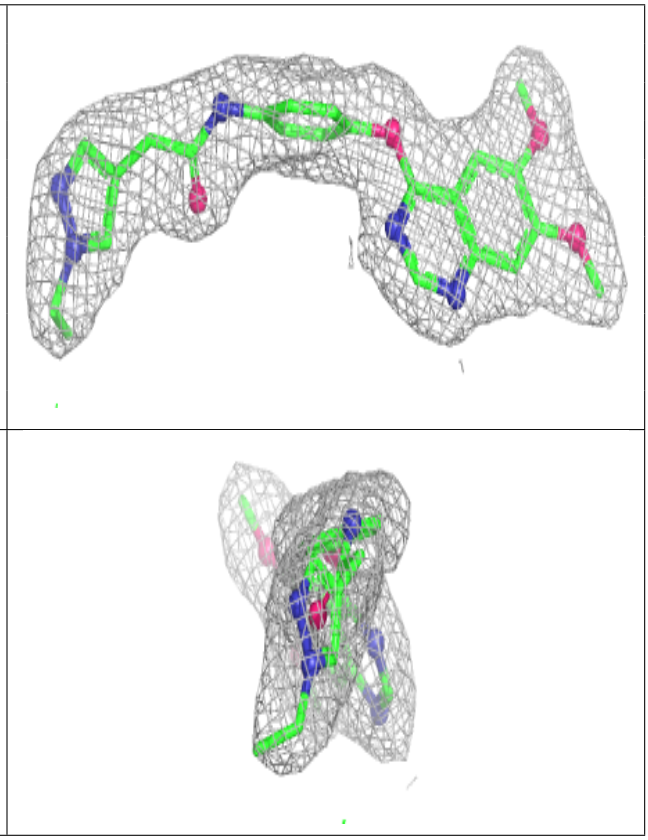
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	F88	A	1001	32/32	0.97	0.12	32,37,46,53	0
2	F88	B	1001	32/32	0.97	0.11	39,47,53,54	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 F88 B 1001:

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.