wwPDB X-ray Structure Validation Summary Report (i)

Dec 4, 2023 – 12:23 am GMT

PDB ID	:	2JIV
Title	:	Crystal structure of EGFR kinase domain T790M mutation in compex with HKI-272
Authors	:	Yun, C.-H.; Mengwasser, K.E.; Toms, A.V.; Li, Y.; Woo, M.S.; Greulich, H.; Wong, K.-K.; Meyerson, M.; Eck, M.J.
Deposited on	:	2007-07-02
Resolution	:	3.50 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:

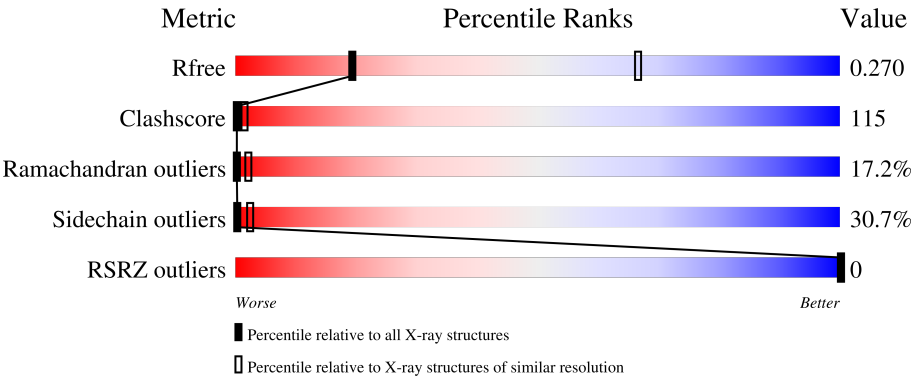
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.50 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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	9% 41% 28% · 18%
1	B	328	12% 46% 23% · 17%

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	HKI	A	1797	-	-	X	-
2	HKI	B	1797	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4427 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.

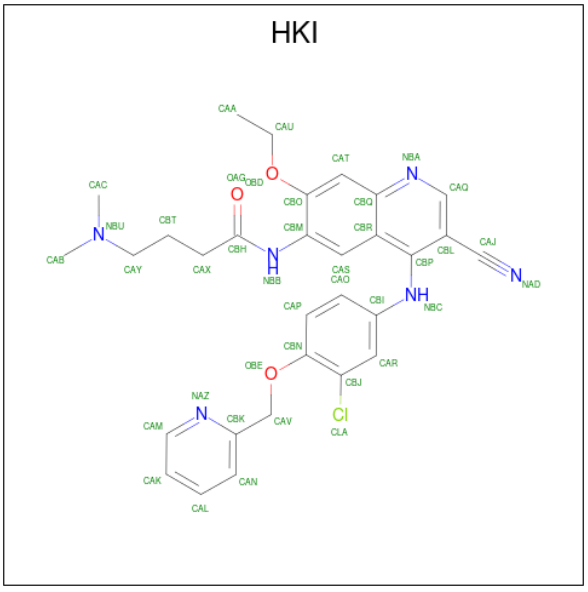
- Molecule 1 is a protein called EPIDERMAL GROWTH FACTOR RECEPTOR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	270	Total	C	N	O	S	0	0	0
			2137	1384	364	373	16			
1	B	272	Total	C	N	O	S	0	1	0
			2146	1391	363	376	16			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	790	MET	THR	engineered mutation	UNP P00533
B	790	MET	THR	engineered mutation	UNP P00533

- Molecule 2 is N-(4-{[3-chloro-4-(pyridin-2-ylmethoxy)phenyl]amino}-3-cyano-7-ethoxyquinolin-6-yl)-4-(dimethylamino)butanamide (three-letter code: HKI) (formula: $C_{30}H_{31}ClN_6O_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	0	0
			40	30	1	6	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	Cl	N	O	0	0
			40	30	1	6	3		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Cl	0	0
			2	2		

- Molecule 4 is water.

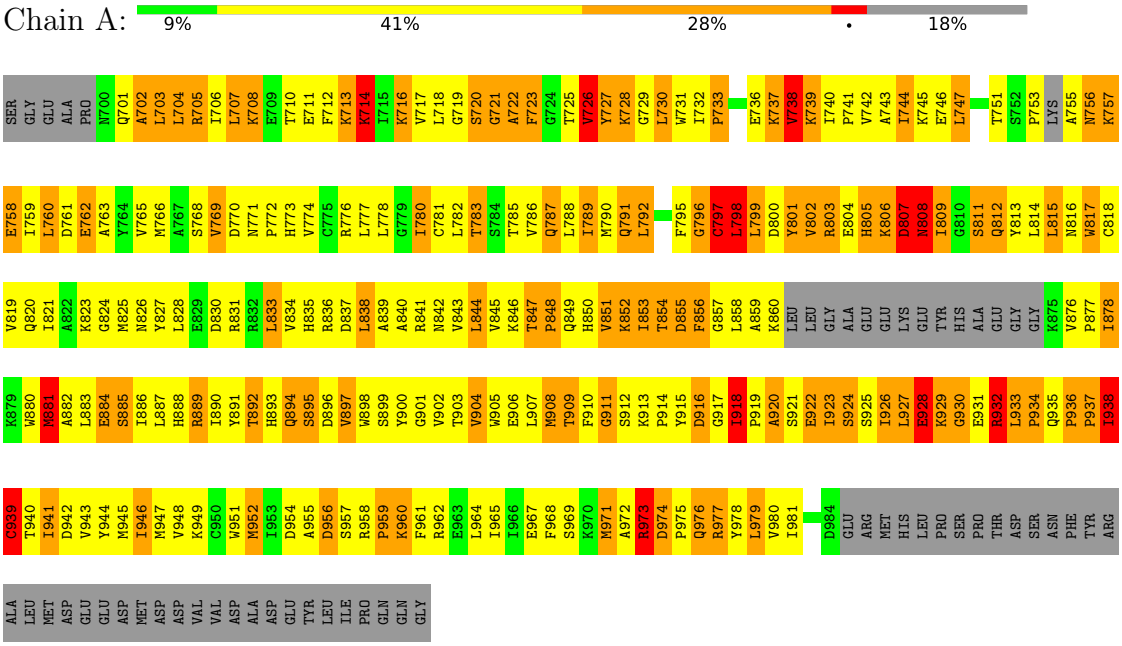
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	43	Total	O	0	0
			43	43		
4	B	19	Total	O	0	0
			19	19		

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.

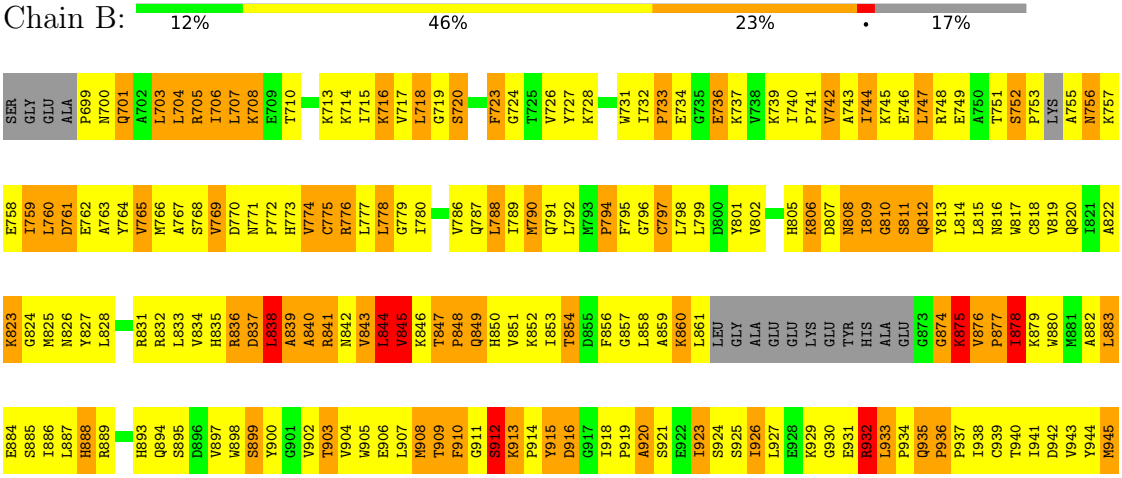
• Molecule 1: EPIDERMAL GROWTH FACTOR RECEPTOR

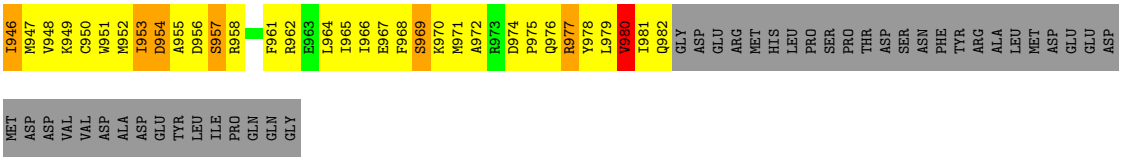
Chain A:



• Molecule 1: EPIDERMAL GROWTH FACTOR RECEPTOR

Chain B:




I946 M947 V948 K949 C950 W951 M952 I953 D954 A955 D956 S957 R958 F961 R962 E963 L964 I965 I966 E967 F968 S969 K970 M971 A972 R973 D974 P975 Q976 R977 Y978 L979 V980 I981 Q982 GLY ASP GLU ARG MET HIS LEU PRO SER PRO THR ASP SER ASN PHE TYR ARG ALA LEU MET ASP GLU GLU ASP
MET ASP ASP VAL VAL ASP ALA ASP GLU TYR LEU ILE PRO GLN GLN GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	56.09Å 98.99Å 73.33Å 90.00° 109.94° 90.00°	Depositor
Resolution (Å)	25.00 – 3.50 24.75 – 3.50	Depositor EDS
% Data completeness (in resolution range)	90.2 (25.00-3.50) 90.3 (24.75-3.50)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.54Å)	Xtriage
Refinement program	REFMAC 5.3.0026	Depositor
R, R_{free}	0.251 , 0.284 0.246 , 0.270	Depositor DCC
R_{free} test set	428 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	53.5	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 71.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	4427	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.33% 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

5.1 Standard geometry

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

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	0.44	1/2182 (0.0%)	0.55	1/2951 (0.0%)
1	B	0.42	0/2190	0.53	1/2967 (0.0%)
All	All	0.43	1/4372 (0.0%)	0.54	2/5918 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	797	CYS	C-N	5.24	1.46	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	753	PRO	N-CA-CB	5.91	110.40	103.30
1	B	699	PRO	N-CA-CB	5.62	110.05	103.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	932	ARG	Sidechain

5.2 Too-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 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	2137	0	2181	546	0
1	B	2146	0	2183	453	0
2	A	40	0	30	22	0
2	B	40	0	30	23	0
3	B	2	0	0	0	0
4	A	43	0	0	1	0
4	B	19	0	0	0	0
All	All	4427	0	4424	1012	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 115.

The worst 5 of 1012 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)
2:A:1797:HKI:H17	2:A:1797:HKI:CAX	1.49	1.28
2:A:1797:HKI:H23	2:A:1797:HKI:CAC	1.52	1.26
1:A:835:HIS:CE1	1:A:856:PHE:HB3	1.70	1.25
1:A:798:LEU:O	1:A:802:VAL:HG23	1.33	1.23
1:A:743:ALA:O	1:A:744:ILE:HG12	1.32	1.23

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	264/328 (80%)	150 (57%)	66 (25%)	48 (18%)	0 2
1	B	267/328 (81%)	168 (63%)	55 (21%)	44 (16%)	0 2
All	All	531/656 (81%)	318 (60%)	121 (23%)	92 (17%)	0 2

5 of 92 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	702	ALA
1	A	714	LYS
1	A	738	VAL
1	A	739	LYS
1	A	796	GLY

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	230/288 (80%)	153 (66%)	77 (34%)	0 1
1	B	230/288 (80%)	165 (72%)	65 (28%)	0 3
All	All	460/576 (80%)	318 (69%)	142 (31%)	0 2

5 of 142 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	875[A]	LYS
1	B	888	HIS
1	B	926	ILE
1	A	885	SER
1	A	884	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	756	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	816	ASN
1	B	849	GLN
1	B	826	ASN
1	A	835	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	HKI	B	1797	1	42,43,43	1.28	4 (9%)	56,58,58	1.87	12 (21%)
2	HKI	A	1797	1	42,43,43	1.28	4 (9%)	56,58,58	1.83	11 (19%)

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	HKI	B	1797	1	-	6/24/24/24	0/4/4/4
2	HKI	A	1797	1	-	3/24/24/24	0/4/4/4

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1797	HKI	CBM-NBB	-4.54	1.33	1.41
2	B	1797	HKI	CBM-NBB	-4.44	1.33	1.41
2	B	1797	HKI	CBI-NBC	-3.34	1.33	1.40
2	A	1797	HKI	CBI-NBC	-3.27	1.33	1.40
2	B	1797	HKI	CBP-CBR	-2.40	1.39	1.43

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1797	HKI	CBL-CAQ-NBA	-6.53	120.83	125.68
2	A	1797	HKI	CBL-CAQ-NBA	-6.40	120.92	125.68
2	B	1797	HKI	CAV-OBE-CBN	-4.46	108.98	117.76
2	A	1797	HKI	CAV-OBE-CBN	-4.05	109.80	117.76
2	B	1797	HKI	OBE-CAV-CBK	4.01	120.21	109.42

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1797	HKI	CBT-CAY-NBU-CAC
2	A	1797	HKI	NBU-CAY-CBT-CAX
2	A	1797	HKI	CBO-CBM-NBB-CBH
2	B	1797	HKI	CBO-CBM-NBB-CBH
2	B	1797	HKI	CAA-CAU-OBD-CBO

There are no ring outliers.

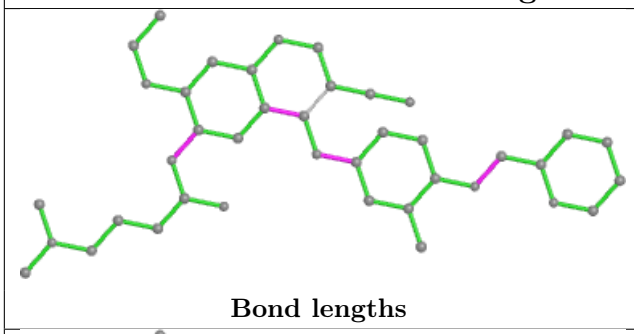
2 monomers are involved in 45 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1797	HKI	23	0
2	A	1797	HKI	22	0

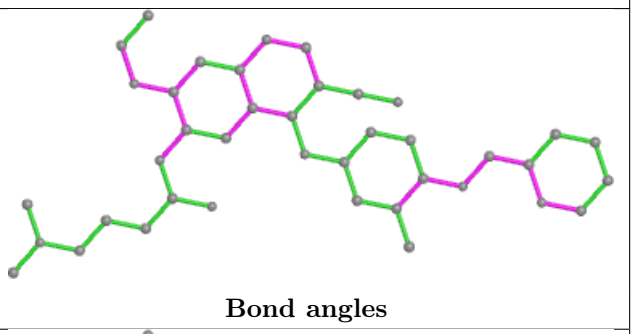
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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

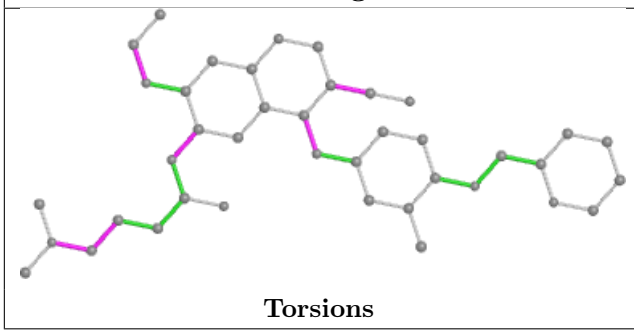
Ligand HKI B 1797



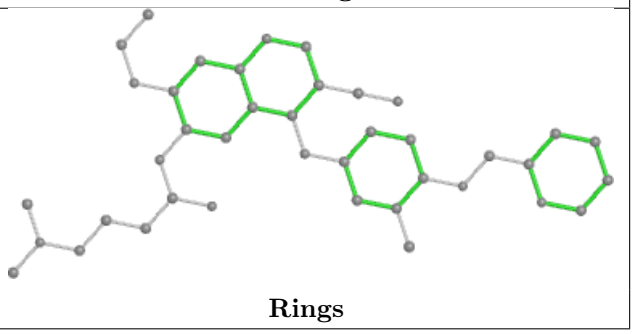
Bond lengths



Bond angles

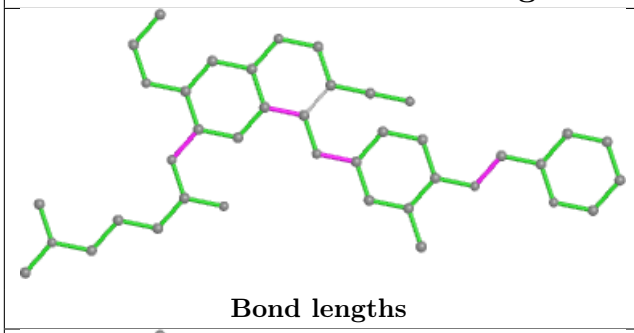


Torsions

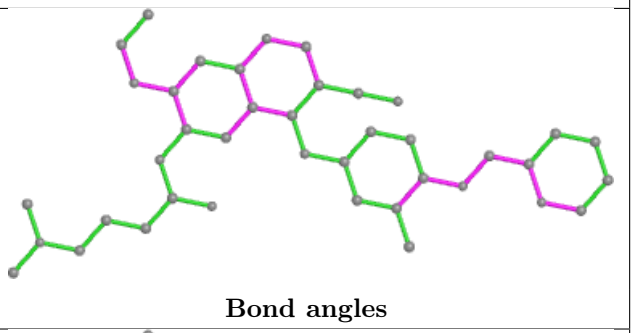


Rings

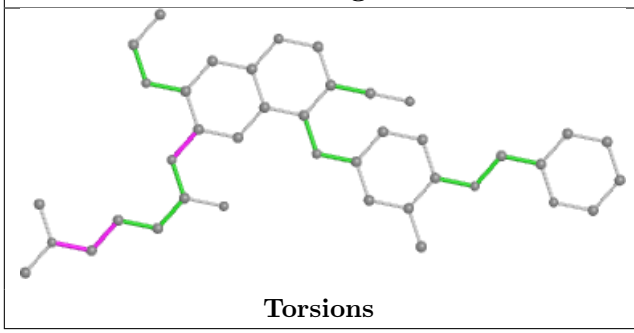
Ligand HKI A 1797



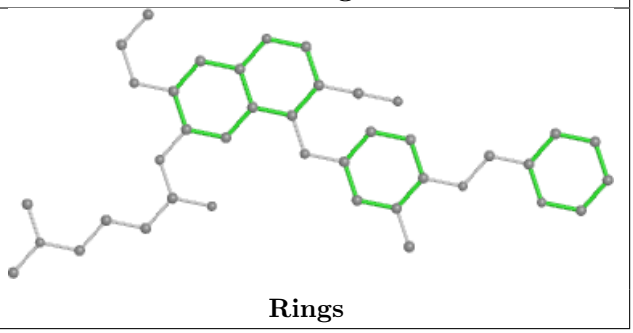
Bond lengths



Bond angles



Torsions



Rings

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

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^{th} 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	270/328 (82%)	-0.52	0	100	100	28, 47, 54, 61	0
1	B	272/328 (82%)	-0.52	0	100	100	31, 46, 55, 61	0
All	All	542/656 (82%)	-0.52	0	100	100	28, 47, 55, 61	0

There are no RSRZ outliers to report.

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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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.

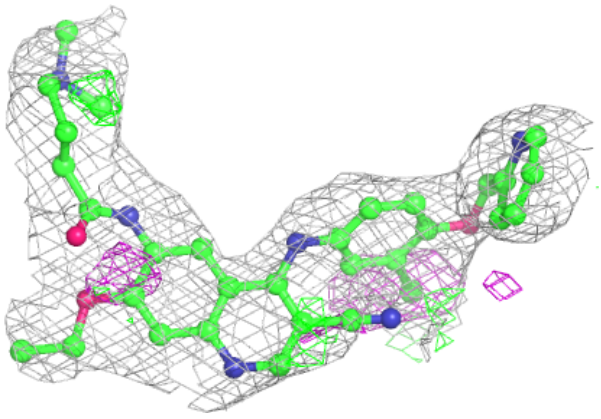
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	CL	B	1985	1/1	0.66	0.21	45,45,45,45	0
2	HKI	A	1797	40/40	0.88	0.22	29,35,38,38	0
2	HKI	B	1797	40/40	0.90	0.20	29,31,34,34	0
3	CL	B	1984	1/1	0.92	0.12	12,12,12,12	0

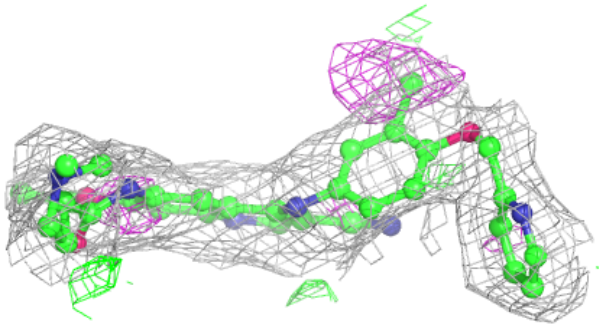
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers

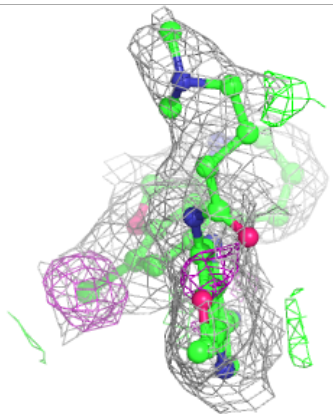
as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around HKI A 1797:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

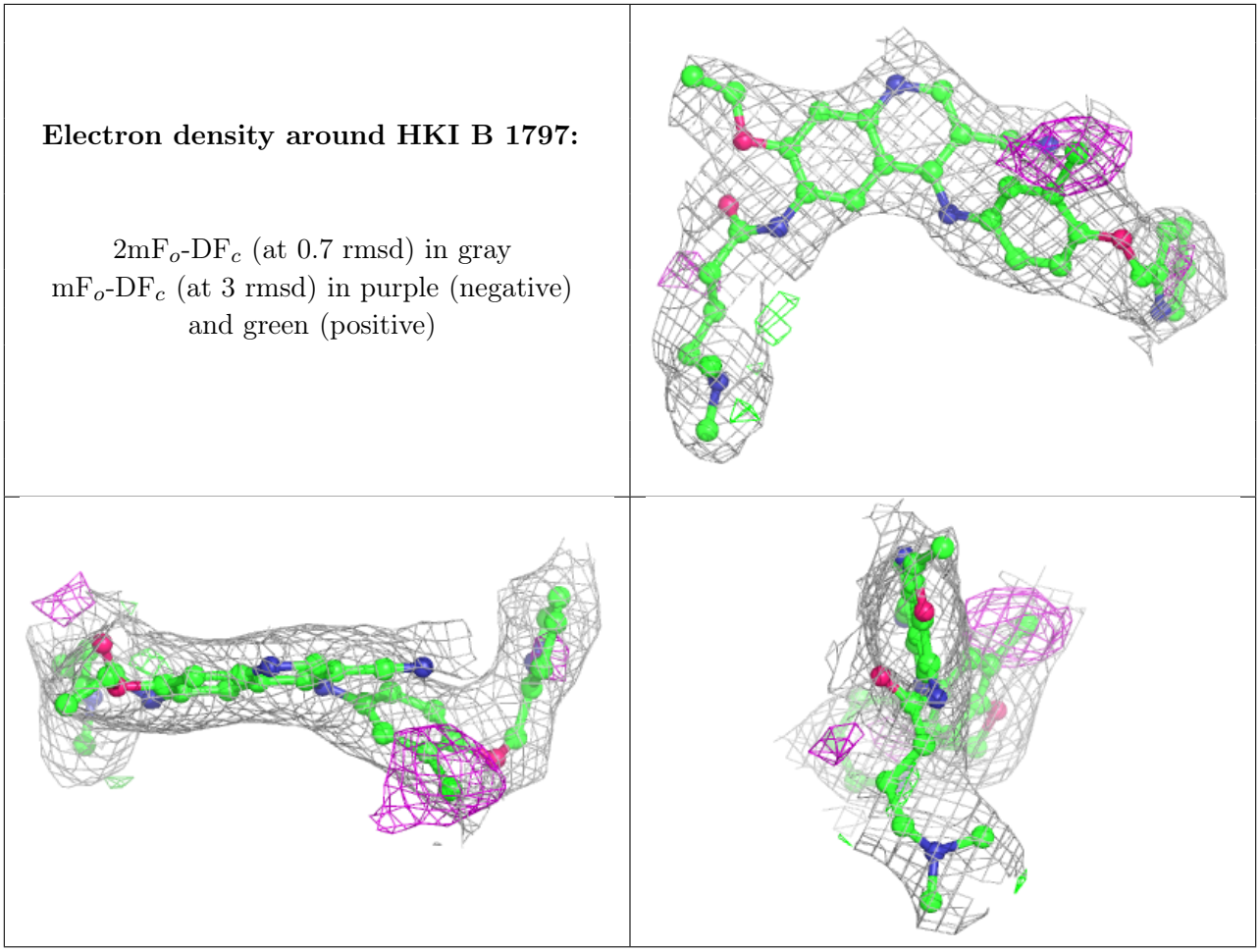






Electron density around HKI B 1797:

$2mF_o$-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.