Full wwPDB X-ray Structure Validation Report ⓘ

Feb 3, 2025 – 06:06 pm GMT

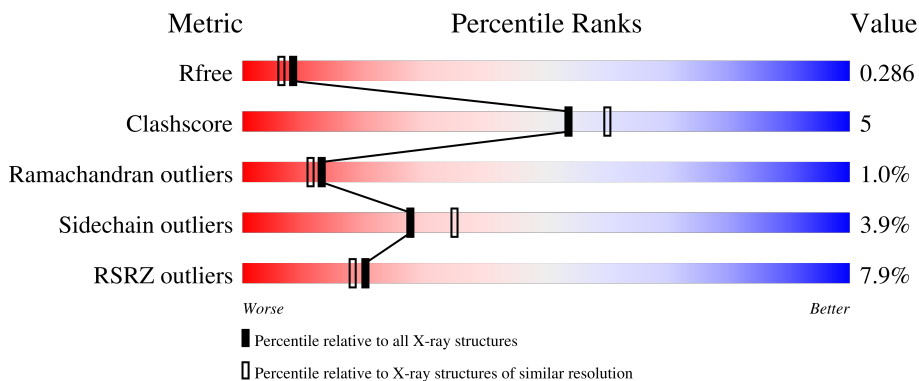|		
PDB ID	:	9GDV
Title	:	Highly optimized CNS penetrant inhibitors of EGFR Exon20 Insertion Mutations
Authors	:	Hargreaves, D.
Deposited on	:	2024-08-06
Resolution	:	2.22 Å(reported)

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 2.22 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	Percentile Ranks	Value
Rfree		0.286
Clashscore		5
Ramachandran outliers		1.0%
Sidechain outliers		3.9%
RSRZ outliers		7.9%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	7167 (2.24-2.20)
Clashscore	180529	8096 (2.24-2.20)
Ramachandran outliers	177936	8010 (2.24-2.20)
Sidechain outliers	177891	8011 (2.24-2.20)
RSRZ outliers	164620	7166 (2.24-2.20)

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	330	4% / 66% / 13% / 21%
1	B	330	9% / 68% / 12% / • / 19%

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

2   Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4417 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	262	Total	C	N	O	S	0	0	0
			2110	1363	358	374	15			
1	B	267	Total	C	N	O	S	0	0	0
			2140	1380	364	381	15			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	693	GLY	-	expression tag	UNP P00533
A	694	SER	-	expression tag	UNP P00533
A	948	ARG	VAL	engineered mutation	UNP P00533
B	693	GLY	-	expression tag	UNP P00533
B	694	SER	-	expression tag	UNP P00533
B	948	ARG	VAL	engineered mutation	UNP P00533

- Molecule 2 is {N}-[2-[2-(dimethylamino)ethyl-methyl-amino]-4-methoxy-5-[[4-(1-methyl lindol-3-yl)pyrimidin-2-yl]amino]phenyl]propanamide (three-letter code: Q6K) (formula: $C_{28}H_{35}N_7O_2$) (labeled as "Ligand of Interest" by depositor).
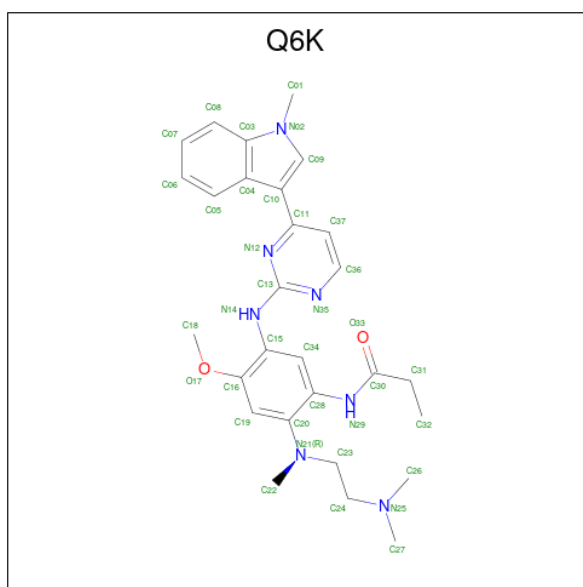
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			37	28	7	2		
2	B	1	Total	C	N	O	0	0
			37	28	7	2		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total 5	O 4	S 1	0	0
3	B	1	Total 5	O 4	S 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	46	Total 46	O 46	0	0
4	B	27	Total 27	O 27	0	0

3 Residue-property plots ⓘ

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:
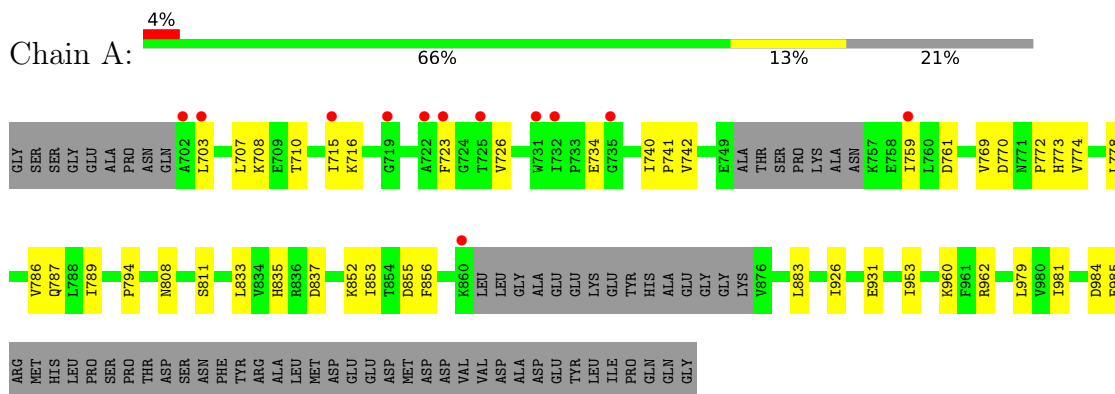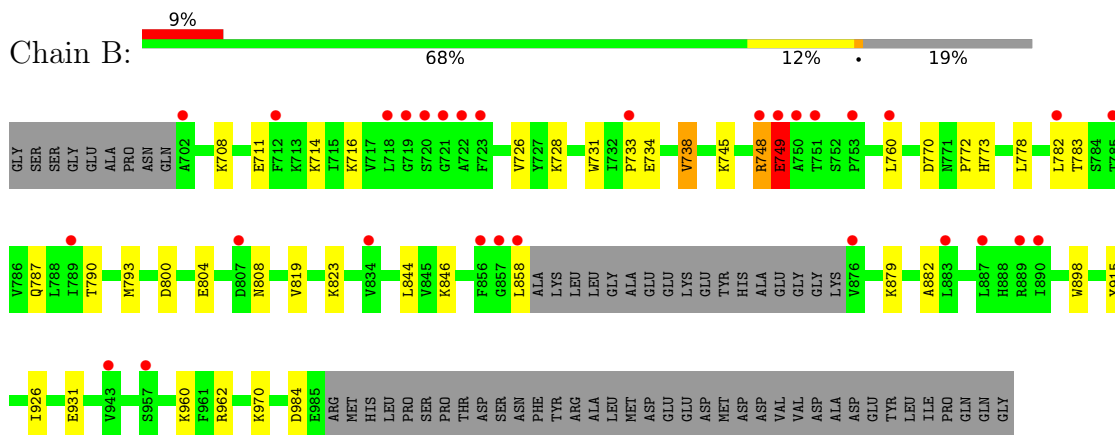
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.03Å 82.31Å 89.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.52 – 2.22 60.52 – 2.22	Depositor EDS
% Data completeness (in resolution range)	77.0 (60.52-2.22) 77.0 (60.52-2.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.87 (at 2.22Å)	Xtriage
Refinement program	BUSTER 2.11.8 (26-JUL-2023)	Depositor
R, R_{free}	0.232 , 0.293 0.226 , 0.286	Depositor DCC
R_{free} test set	7545 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	44.8	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 40.3	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	0.93	EDS
Total number of atoms	4417	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.26% 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: Q6K, SO4

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	0.39	0/2155	0.57	0/2913
1	B	0.37	0/2186	0.54	0/2959
All	All	0.38	0/4341	0.56	0/5872

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 ⓘ

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	2110	0	2160	25	0
1	B	2140	0	2192	21	0
2	A	37	0	0	1	0
2	B	37	0	0	0	0
3	A	10	0	0	0	0
3	B	10	0	0	1	0
4	A	46	0	0	0	0
4	B	27	0	0	0	0
All	All	4417	0	4352	42	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 5.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

All (42) 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:960:LYS:NZ	1:A:962:ARG:HH12	1.67	0.92
1:A:960:LYS:NZ	1:A:962:ARG:NH1	2.27	0.81
1:A:960:LYS:HZ1	1:A:962:ARG:HH12	1.36	0.70
1:B:926:ILE:HG23	1:B:931:GLU:HB2	1.75	0.68
1:A:960:LYS:HZ1	1:A:962:ARG:NH1	1.91	0.68
1:B:960:LYS:NZ	1:B:962:ARG:HH22	1.92	0.67
1:A:707:LEU:HD12	1:A:789:ILE:HD13	1.78	0.65
1:B:726:VAL:HG22	1:B:745:LYS:HG2	1.79	0.64
1:A:770:ASP:OD2	1:B:773:HIS:HE1	1.80	0.63
1:A:981:ILE:HB	1:A:984:ASP:HB2	1.85	0.58
1:A:960:LYS:HZ3	1:A:962:ARG:NH1	1.99	0.58
1:B:960:LYS:HZ1	1:B:962:ARG:HH22	1.50	0.58
1:A:926:ILE:HG23	1:A:931:GLU:HB2	1.85	0.57
1:A:759:ILE:HD11	1:A:786:VAL:HG11	1.88	0.54
1:A:773:HIS:HE1	1:B:770:ASP:OD2	1.90	0.54
1:A:772:PRO:HD3	1:B:772:PRO:HD3	1.90	0.54
1:B:846:LYS:NZ	3:B:1102:SO4:O3	2.29	0.52
1:A:835:HIS:HD2	1:A:837:ASP:H	1.56	0.51
1:B:800:ASP:O	1:B:804:GLU:HG3	2.11	0.51
1:A:794:PRO:HB2	1:B:738:VAL:HG11	1.92	0.51
1:A:883:LEU:HD23	1:A:953:ILE:HD12	1.94	0.49
1:B:708:LYS:O	1:B:711:GLU:HG2	2.12	0.49
1:A:769:VAL:HG11	1:A:774:VAL:HG11	1.94	0.49
1:B:793:MET:HG3	1:B:844:LEU:HD13	1.94	0.49
1:A:772:PRO:O	1:A:852:LYS:HE3	2.14	0.47
1:A:715:ILE:HG22	1:A:716:LYS:HG3	1.98	0.46
1:B:716:LYS:HD3	1:B:728:LYS:HD3	1.96	0.46
1:A:833:LEU:HD13	1:A:856:PHE:CZ	2.51	0.46
1:B:760:LEU:HD13	1:B:782:LEU:HD11	1.99	0.44
1:B:748:ARG:O	1:B:749:GLU:HB2	2.17	0.44
1:A:769:VAL:HG11	1:A:774:VAL:CG1	2.47	0.43
1:B:879:LYS:HG2	1:B:915:TYR:HD2	1.83	0.43
1:A:835:HIS:CE1	1:A:853:ILE:CG2	3.02	0.42
1:A:811:SER:HB3	1:A:979:LEU:HB2	2.01	0.42
1:B:731:TRP:CZ2	1:B:733:PRO:HB3	2.54	0.42
1:A:740:ILE:HG22	1:A:742:VAL:HG13	2.01	0.42
1:B:960:LYS:NZ	1:B:962:ARG:NH2	2.63	0.42
1:A:726:VAL:HG11	2:A:1101:Q6K:C04	2.49	0.42
1:B:819:VAL:O	1:B:823:LYS:HG3	2.20	0.42
1:B:882:ALA:HA	1:B:898:TRP:CD2	2.55	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:740:ILE:HA	1:A:741:PRO:HD3	1.97	0.41
1:B:714:LYS:NZ	1:B:787:GLN:HE22	2.19	0.40

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	256/330 (78%)	248 (97%)	8 (3%)	0	100	100
1	B	263/330 (80%)	251 (95%)	7 (3%)	5 (2%)	6	4
All	All	519/660 (79%)	499 (96%)	15 (3%)	5 (1%)	13	11

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	748	ARG
1	B	749	GLU
1	B	734	GLU
1	B	984	ASP
1	B	783	THR

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	232/289 (80%)	221 (95%)	11 (5%)	22	27
1	B	235/289 (81%)	228 (97%)	7 (3%)	36	46
All	All	467/578 (81%)	449 (96%)	18 (4%)	27	35

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	703	LEU
1	A	708	LYS
1	A	710	THR
1	A	723	PHE
1	A	734	GLU
1	A	761	ASP
1	A	778	LEU
1	A	787	GLN
1	A	808	ASN
1	A	855	ASP
1	A	985	GLU
1	B	738	VAL
1	B	749	GLU
1	B	778	LEU
1	B	790	THR
1	B	808	ASN
1	B	858	LEU
1	B	970	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	773	HIS
1	A	805	HIS
1	A	816	ASN
1	A	835	HIS
1	A	849	GLN
1	A	982	GLN
1	B	773	HIS
1	B	787	GLN
1	B	816	ASN

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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

6 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	Q6K	B	1101	1	38,40,40	0.65	0	50,56,56	0.72	2 (4%)
2	Q6K	A	1101	1	38,40,40	0.64	0	50,56,56	0.67	2 (4%)
3	SO4	A	1102	-	4,4,4	0.18	0	6,6,6	0.47	0
3	SO4	B	1102	-	4,4,4	0.15	0	6,6,6	0.57	0
3	SO4	A	1103	-	4,4,4	0.16	0	6,6,6	0.73	0
3	SO4	B	1103	-	4,4,4	0.14	0	6,6,6	0.87	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	Q6K	B	1101	1	-	8/21/25/25	0/4/4/4
2	Q6K	A	1101	1	-	3/21/25/25	0/4/4/4

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1101	Q6K	C08-C03-N02	-2.29	130.17	132.14
2	B	1101	Q6K	C08-C03-N02	-2.17	130.28	132.14
2	A	1101	Q6K	C11-C10-C04	2.15	128.11	123.90
2	B	1101	Q6K	C11-C10-C04	2.15	128.10	123.90

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1101	Q6K	C28-C20-N21-C22
2	A	1101	Q6K	C15-C16-O17-C18
2	A	1101	Q6K	C19-C16-O17-C18
2	B	1101	Q6K	C23-C24-N25-C26
2	B	1101	Q6K	C28-C20-N21-C23
2	B	1101	Q6K	C23-C24-N25-C27
2	B	1101	Q6K	C19-C20-N21-C22
2	A	1101	Q6K	C24-C23-N21-C20
2	B	1101	Q6K	C24-C23-N21-C20
2	B	1101	Q6K	C24-C23-N21-C22
2	B	1101	Q6K	C19-C20-N21-C23

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	Q6K	1	0
3	B	1102	SO4	1	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 Q6K B 1101

Bond lengths

Bond angles

Torsions

Rings

Ligand Q6K A 1101

Bond lengths

Bond angles

Torsions

Rings
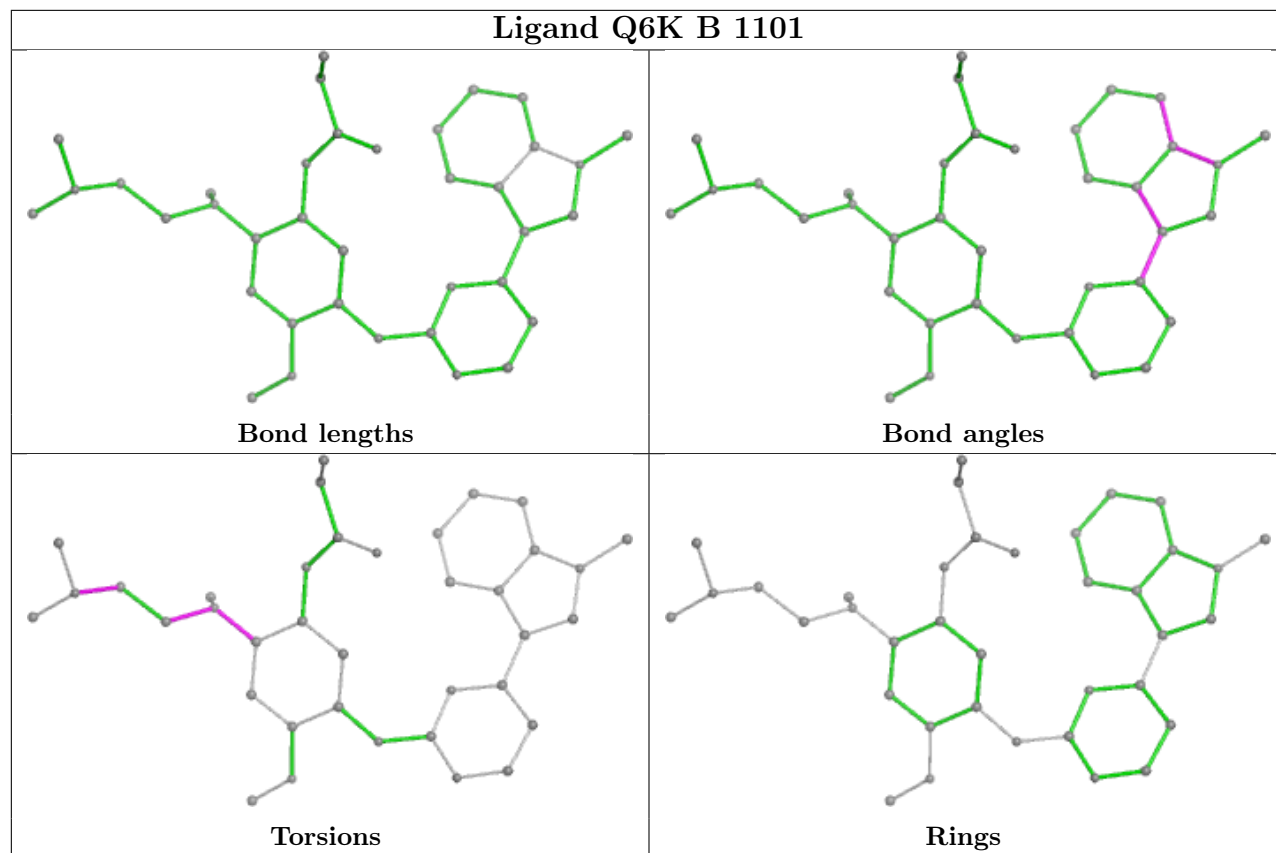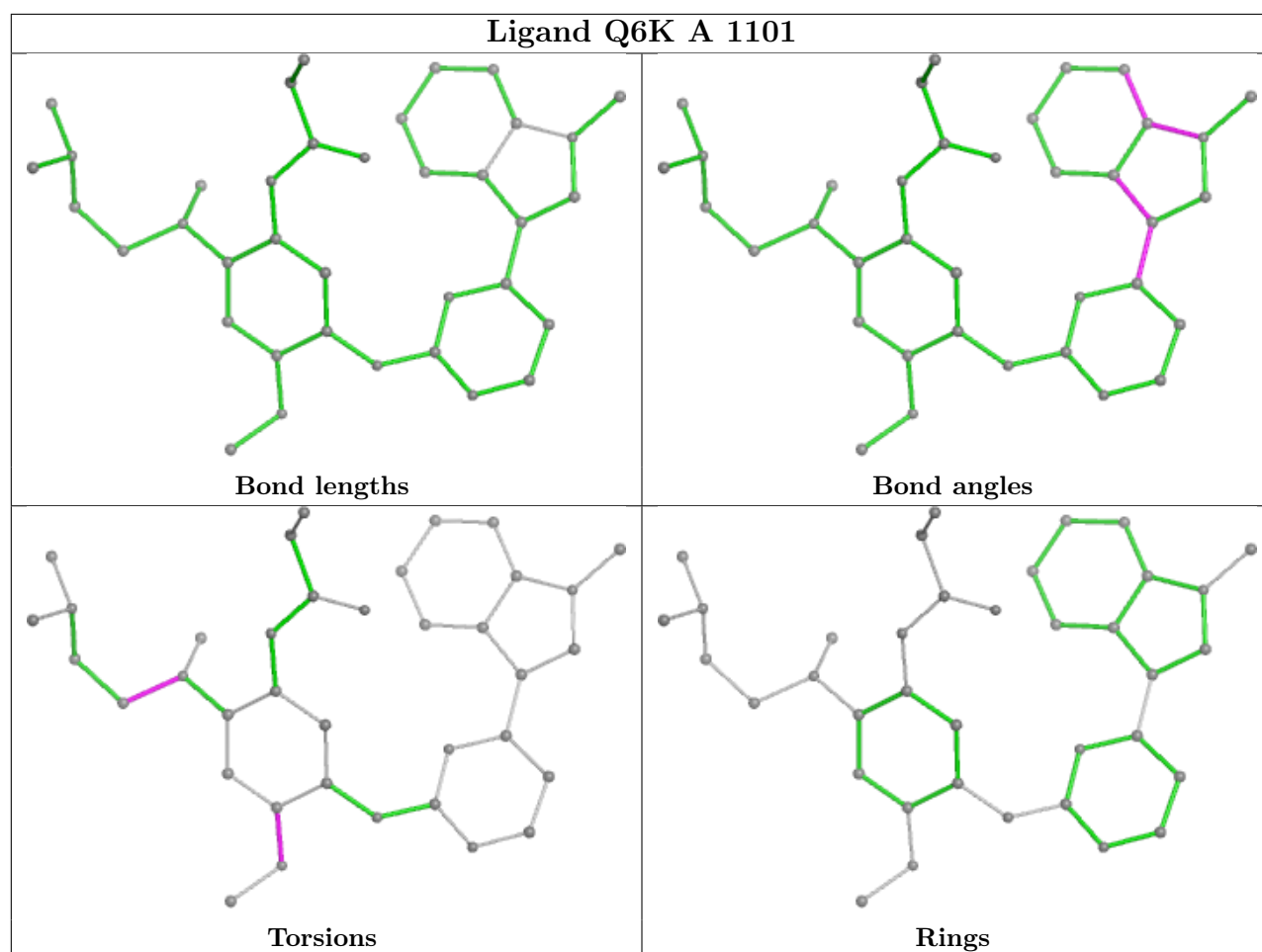
5.7　Other polymers ⓘ

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, 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	262/330 (79%)	0.48	12 (4%)	38	35	27, 44, 73, 87	0
1	B	267/330 (80%)	0.96	30 (11%)	11	9	31, 56, 76, 95	0
All	All	529/660 (80%)	0.72	42 (7%)	20	18	27, 50, 75, 95	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	856	PHE	5.3
1	A	702	ALA	4.4
1	B	720	SER	4.3
1	B	857	GLY	4.1
1	B	702	ALA	4.0
1	B	723	PHE	3.9
1	B	782	LEU	3.9
1	B	883	LEU	3.5
1	B	718	LEU	3.2
1	A	723	PHE	3.1
1	B	722	ALA	3.1
1	B	719	GLY	3.1
1	B	887	LEU	3.1
1	A	715	ILE	2.9
1	B	943	VAL	2.9
1	B	753	PRO	2.8
1	B	890	ILE	2.7
1	B	858	LEU	2.7
1	B	957	SER	2.6
1	B	876	VAL	2.6
1	B	749	GLU	2.5
1	A	722	ALA	2.5
1	A	731	TRP	2.5
1	A	759	ILE	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	860	LYS	2.5
1	B	834	VAL	2.5
1	B	733	PRO	2.4
1	A	703	LEU	2.4
1	A	719	GLY	2.4
1	B	889	ARG	2.3
1	A	735	GLY	2.3
1	B	748	ARG	2.3
1	B	712	PHE	2.3
1	B	789	ILE	2.3
1	A	725	THR	2.2
1	B	807	ASP	2.2
1	A	732	ILE	2.2
1	B	750	ALA	2.2
1	B	721	GLY	2.1
1	B	760	LEU	2.1
1	B	751	THR	2.1
1	B	785	THR	2.1

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, 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.
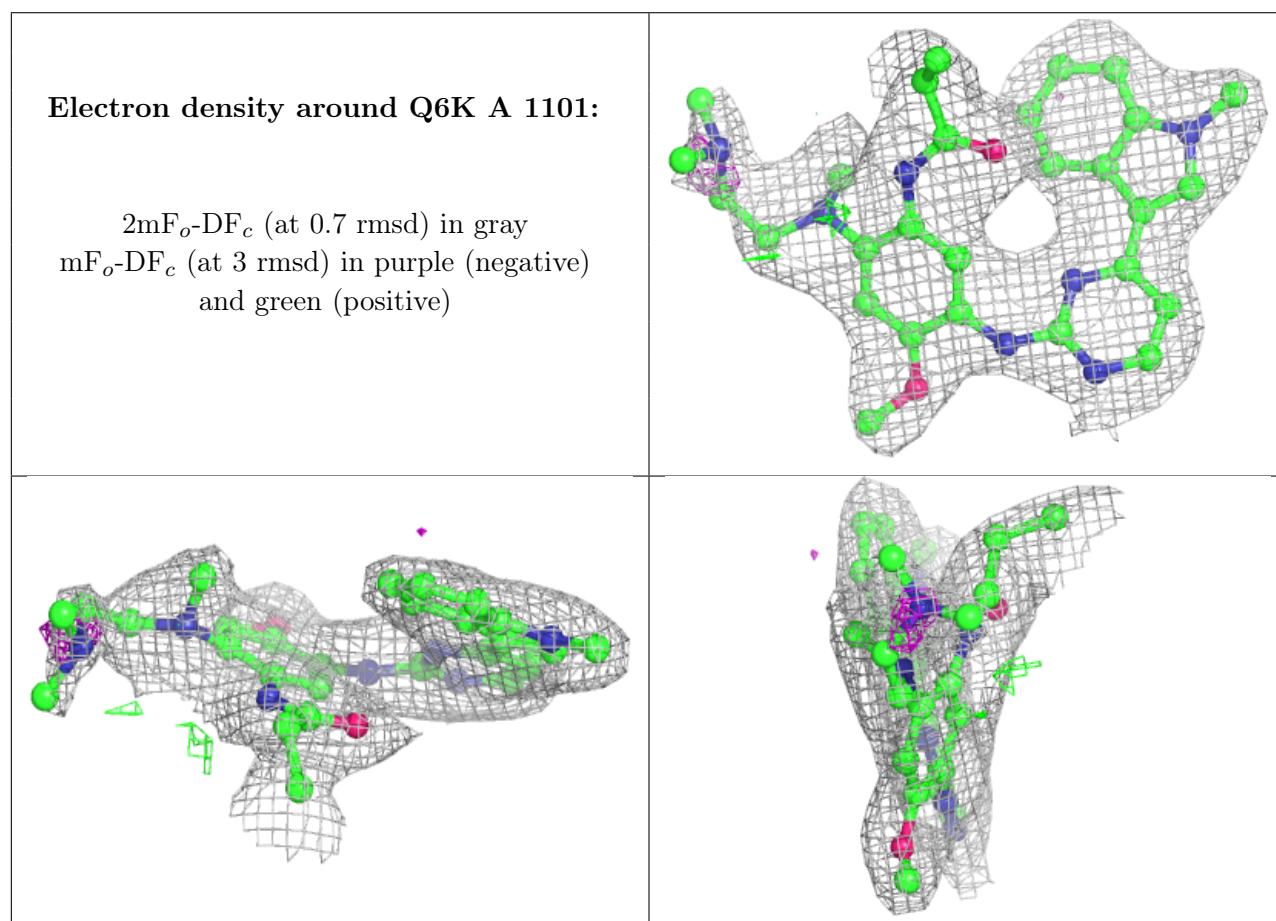
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	SO4	A	1102	5/5	0.80	0.10	89,89,89,90	0
2	Q6K	A	1101	37/37	0.89	0.10	32,37,46,47	0
2	Q6K	B	1101	37/37	0.90	0.10	42,43,50,50	0
3	SO4	A	1103	5/5	0.90	0.11	72,72,73,73	0
3	SO4	B	1102	5/5	0.93	0.10	78,78,78,78	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	SO4	B	1103	5/5	0.94	0.11	57,57,57,57	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 Q6K A 1101:

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

Electron density around Q6K B 1101:

$2mF_o$-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 ⓘ

There are no such residues in this entry.