Full wwPDB X-ray Structure Validation Report (i)

Dec 9, 2024 – 02:20 PM EST

PDB ID	:	9CZW
Title	:	HPK1 kinase domain T165E,S171E phosphomimetic mutant in complex with compound 13
Authors	:	Johnson, E.; Mc Tigue, M.
Deposited on	:	2024-08-05
Resolution	:	1.59 Å(reported)

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

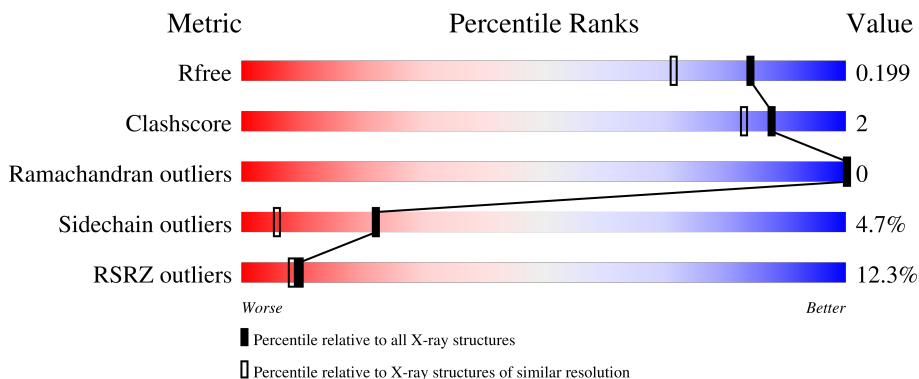|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 1.59 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		0.199
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		4.7%
RSRZ outliers		12.3%

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	4274 (1.60-1.60)
Clashscore	180529	4682 (1.60-1.60)
Ramachandran outliers	177936	4583 (1.60-1.60)
Sidechain outliers	177891	4582 (1.60-1.60)
RSRZ outliers	164620	4272 (1.60-1.60)

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 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	309	11% — 85% · 7% · 7%
1	B	309	12% — 85% · 7% · 6%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 5334 atoms, of which 52 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 Mitogen-activated protein kinase kinase kinase kinase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	288	Total	C	N	O	S	0	0	0
			2276	1464	393	408	11			
1	B	289	Total	C	N	O	S	0	0	0
			2284	1468	395	410	11			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q92918
A	0	SER	-	expression tag	UNP Q92918
A	165	GLU	THR	engineered mutation	UNP Q92918
A	171	GLU	SER	engineered mutation	UNP Q92918
B	-1	GLY	-	expression tag	UNP Q92918
B	0	SER	-	expression tag	UNP Q92918
B	165	GLU	THR	engineered mutation	UNP Q92918
B	171	GLU	SER	engineered mutation	UNP Q92918

- Molecule 2 is 6-(dimethylamino)-4-[(methylamino)methyl]-2-[6-(4-propyl-4H-1,2,4-triazol-3-yl)pyridin-2-yl]-2,3-dihydro-1H-pyrrolo[3,4-c]pyridin-1-one (three-letter code: A1A1C) (formula: $C_{21}H_{26}N_8O$) (labeled as "Ligand of Interest" by depositor).
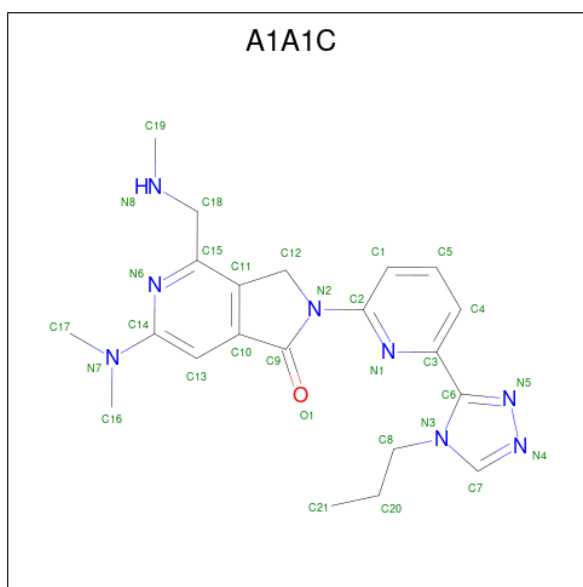
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	26	0
			56	21	26	8	1		
2	B	1	Total	C	H	N	O	26	0
			56	21	26	8	1		

• Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	321	Total	O	0	0
			321	321		
3	B	341	Total	O	0	0
			341	341		

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: Mitogen-activated protein kinase kinase kinase kinase 1

Chain A:

- Molecule 1: Mitogen-activated protein kinase kinase kinase kinase 1

Chain B:
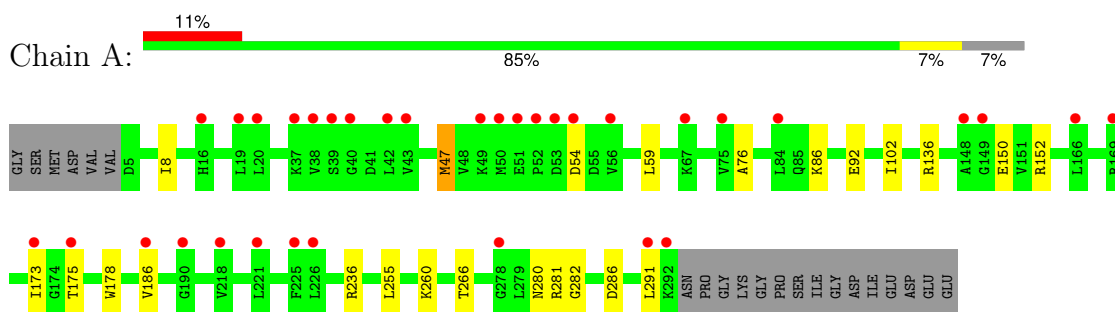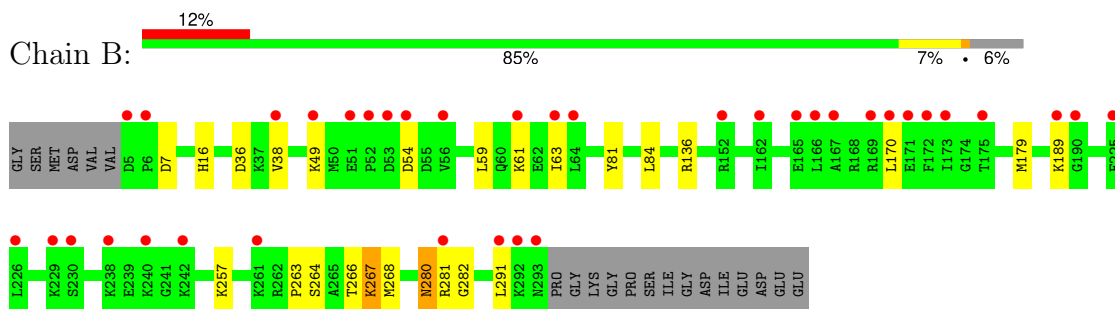
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	51.94Å 57.97Å 62.13Å 87.22° 86.09° 66.51°	Depositor
Resolution (Å)	61.97 – 1.59 61.97 – 1.59	Depositor EDS
% Data completeness (in resolution range)	81.5 (61.97-1.59) 80.7 (61.97-1.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.40 (at 1.46Å)	Xtriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.189 , 0.221 0.193 , 0.199	Depositor DCC
R_{free} test set	4522 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	18.4	Xtriage
Anisotropy	0.028	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 49.1	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5334	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.28% 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: A1A1C

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	0.52	0/2325	0.61	0/3143
1	B	0.54	1/2333 (0.0%)	0.62	0/3154
All	All	0.53	1/4658 (0.0%)	0.61	0/6297

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	179	MET	SD-CE	-5.88	1.45	1.77

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	2276	0	2330	12	0
1	B	2284	0	2336	6	0
2	A	30	26	0	0	0
2	B	30	26	0	0	0
3	A	321	0	0	7	0
3	B	341	0	0	1	0
All	All	5282	52	4666	18	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 2.

All (18) 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:255:LEU:HB2	3:A:501:HOH:O	1.56	1.03
1:B:280:ASN:HD22	1:B:282:GLY:H	1.42	0.66
1:A:150:GLU:HB3	3:A:502:HOH:O	1.97	0.64
1:B:16:HIS:HD2	1:B:81:TYR:OH	1.84	0.60
1:A:280:ASN:HD22	1:A:282:GLY:H	1.52	0.56
1:A:102:ILE:HG12	1:A:291:LEU:HD11	1.88	0.54
1:A:173:ILE:HD11	3:A:689:HOH:O	2.08	0.53
1:A:47:MET:HE3	1:A:86:LYS:HG2	1.92	0.51
1:B:264:SER:H	1:B:267:LYS:HZ3	1.59	0.49
1:A:152:ARG:HD3	3:A:502:HOH:O	2.13	0.49
1:A:236:ARG:NH1	3:A:501:HOH:O	2.46	0.48
1:A:186:VAL:HG23	3:A:689:HOH:O	2.14	0.47
1:B:36:ASP:OD1	1:B:38:VAL:HG22	2.17	0.45
1:B:266:THR:HG23	3:B:725:HOH:O	2.18	0.44
1:A:266:THR:HG23	3:A:715:HOH:O	2.17	0.44
1:B:263:PRO:HG2	1:B:268:MET:SD	2.59	0.43
1:A:76:ALA:HB3	1:A:92:GLU:HB2	2.03	0.41
1:A:175:THR:HG21	1:A:178:TRP:HZ3	1.86	0.41

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	286/309 (93%)	282 (99%)	4 (1%)	0	100	100
1	B	287/309 (93%)	282 (98%)	5 (2%)	0	100	100
All	All	573/618 (93%)	564 (98%)	9 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	245/262 (94%)	237 (97%)	8 (3%)	33	12
1	B	246/262 (94%)	231 (94%)	15 (6%)	15	3
All	All	491/524 (94%)	468 (95%)	23 (5%)	22	6

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	ILE
1	A	47	MET
1	A	54	ASP
1	A	59	LEU
1	A	136	ARG
1	A	260	LYS
1	A	281	ARG
1	A	286	ASP
1	B	7	ASP
1	B	49	LYS
1	B	54	ASP
1	B	59	LEU
1	B	61	LYS
1	B	63	ILE
1	B	84	LEU
1	B	136	ARG
1	B	170	LEU
1	B	189	LYS
1	B	257	LYS
1	B	267	LYS
1	B	280	ASN
1	B	281	ARG
1	B	291	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	142	ASN
1	A	280	ASN
1	B	16	HIS
1	B	104	GLN
1	B	142	ASN
1	B	212	GLN
1	B	280	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 ⓘ

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	A1A1C	B	401	-	30,33,33	0.57	0	37,47,47	0.91	2 (5%)
2	A1A1C	A	401	-	30,33,33	0.56	1 (3%)	37,47,47	0.82	2 (5%)

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	A1A1C	B	401	-	-	2/17/30/30	0/4/4/4
2	A1A1C	A	401	-	-	1/17/30/30	0/4/4/4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	A1A1C	C18-N8	2.06	1.49	1.46

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	401	A1A1C	C13-C10-C11	-3.54	119.95	123.27
2	A	401	A1A1C	C13-C10-C11	-2.69	120.75	123.27
2	A	401	A1A1C	C12-C11-C15	2.48	133.70	128.95
2	B	401	A1A1C	C12-C11-C15	2.05	132.89	128.95

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	401	A1A1C	C11-C15-C18-N8
2	B	401	A1A1C	C11-C15-C18-N8
2	B	401	A1A1C	C21-C20-C8-N3

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 A1A1C B 401

Ligand A1A1C A 401
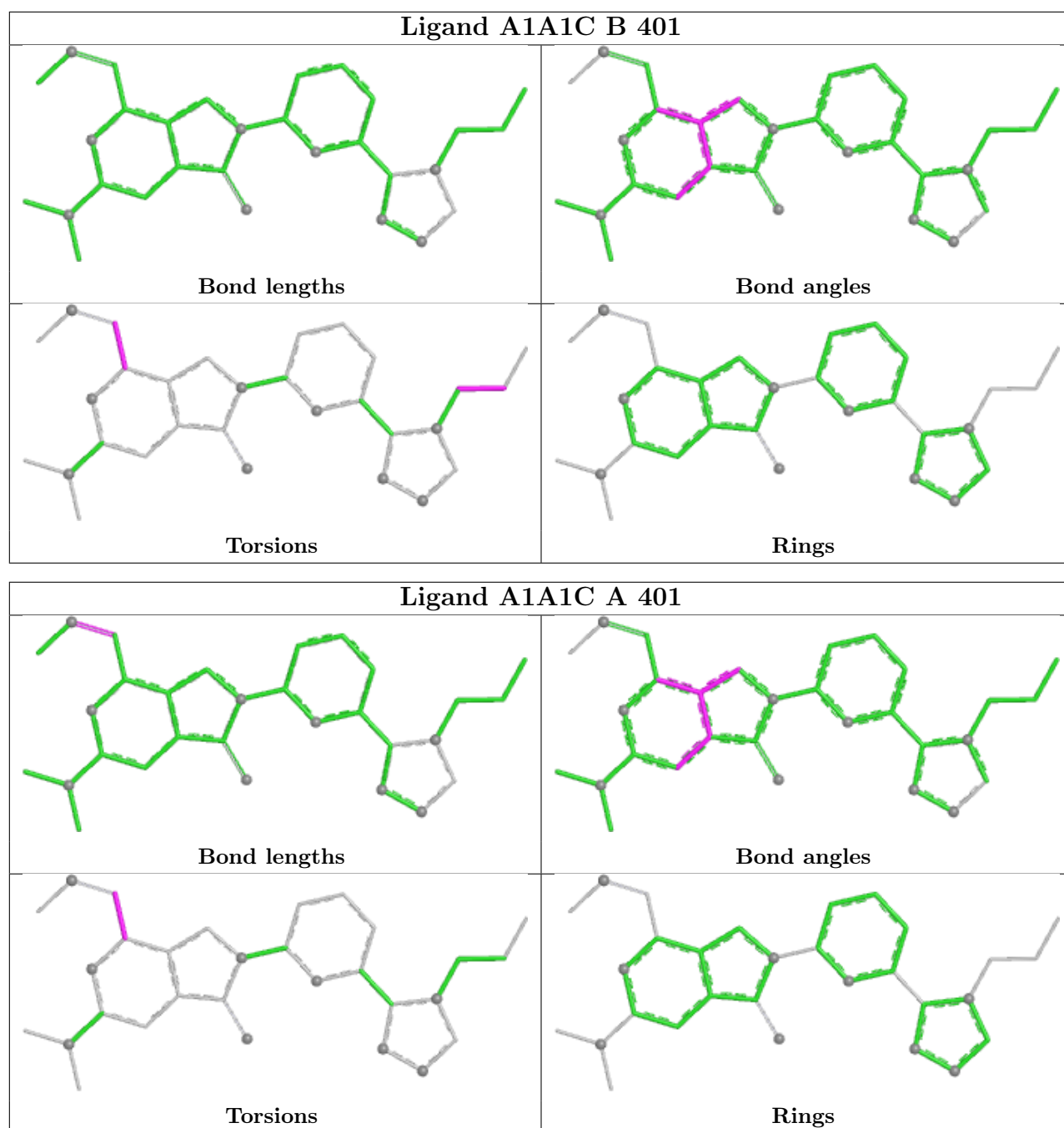
5.7 Other polymers ⓘ

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(Å²)	Q<0.9
1	A	288/309 (93%)	0.72	34 (11%)	10	9	14, 25, 52, 81	0
1	B	289/309 (93%)	0.58	37 (12%)	9	8	13, 22, 52, 89	0
All	All	577/618 (93%)	0.65	71 (12%)	9	8	13, 23, 52, 89	0

All (71) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	292	LYS	4.9
1	A	38	VAL	4.6
1	A	291	LEU	4.5
1	B	5	ASP	4.4
1	B	173	ILE	4.3
1	B	190	GLY	4.2
1	B	293	ASN	4.0
1	A	148	ALA	4.0
1	A	50	MET	3.9
1	A	52	PRO	3.8
1	B	175	THR	3.7
1	B	229	LYS	3.5
1	B	172	PHE	3.5
1	B	6	PRO	3.4
1	B	169	ARG	3.4
1	B	291	LEU	3.3
1	B	170	LEU	3.3
1	B	63	ILE	3.2
1	A	56	VAL	3.1
1	B	240	LYS	3.0
1	B	166	LEU	3.0
1	A	218	VAL	3.0
1	B	53	ASP	3.0
1	A	39	SER	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	173	ILE	2.9
1	A	16	HIS	2.8
1	A	175	THR	2.8
1	A	84	LEU	2.8
1	B	51	GLU	2.7
1	B	162	ILE	2.7
1	A	225	PHE	2.6
1	B	52	PRO	2.5
1	B	292	LYS	2.5
1	A	53	ASP	2.5
1	A	186	VAL	2.5
1	A	49	LYS	2.5
1	B	230	SER	2.5
1	B	242	LYS	2.5
1	B	171	GLU	2.5
1	A	42	LEU	2.4
1	A	221	LEU	2.4
1	A	19	LEU	2.4
1	B	225	PHE	2.4
1	A	190	GLY	2.3
1	A	169	ARG	2.3
1	A	166	LEU	2.3
1	A	37	LYS	2.3
1	B	189	LYS	2.3
1	B	56	VAL	2.3
1	B	261	LYS	2.3
1	B	61	LYS	2.3
1	B	38	VAL	2.3
1	B	49	LYS	2.3
1	B	54	ASP	2.2
1	B	167	ALA	2.2
1	A	226	LEU	2.2
1	B	238	LYS	2.2
1	A	20	LEU	2.1
1	A	43	VAL	2.1
1	B	165	GLU	2.1
1	A	51	GLU	2.1
1	A	67	LYS	2.1
1	B	152	ARG	2.1
1	A	54	ASP	2.1
1	B	64	LEU	2.1
1	A	40	GLY	2.1

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	149	GLY	2.1
1	A	75	VAL	2.0
1	B	226	LEU	2.0
1	A	278	GLY	2.0
1	B	281	ARG	2.0

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 ⓘ

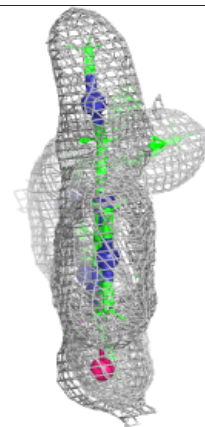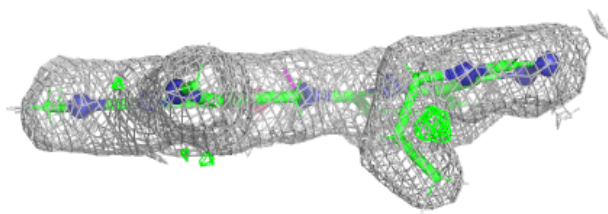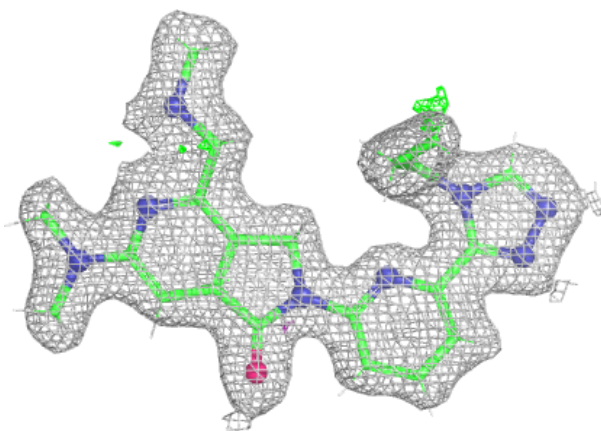There are no monosaccharides in this entry.

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.

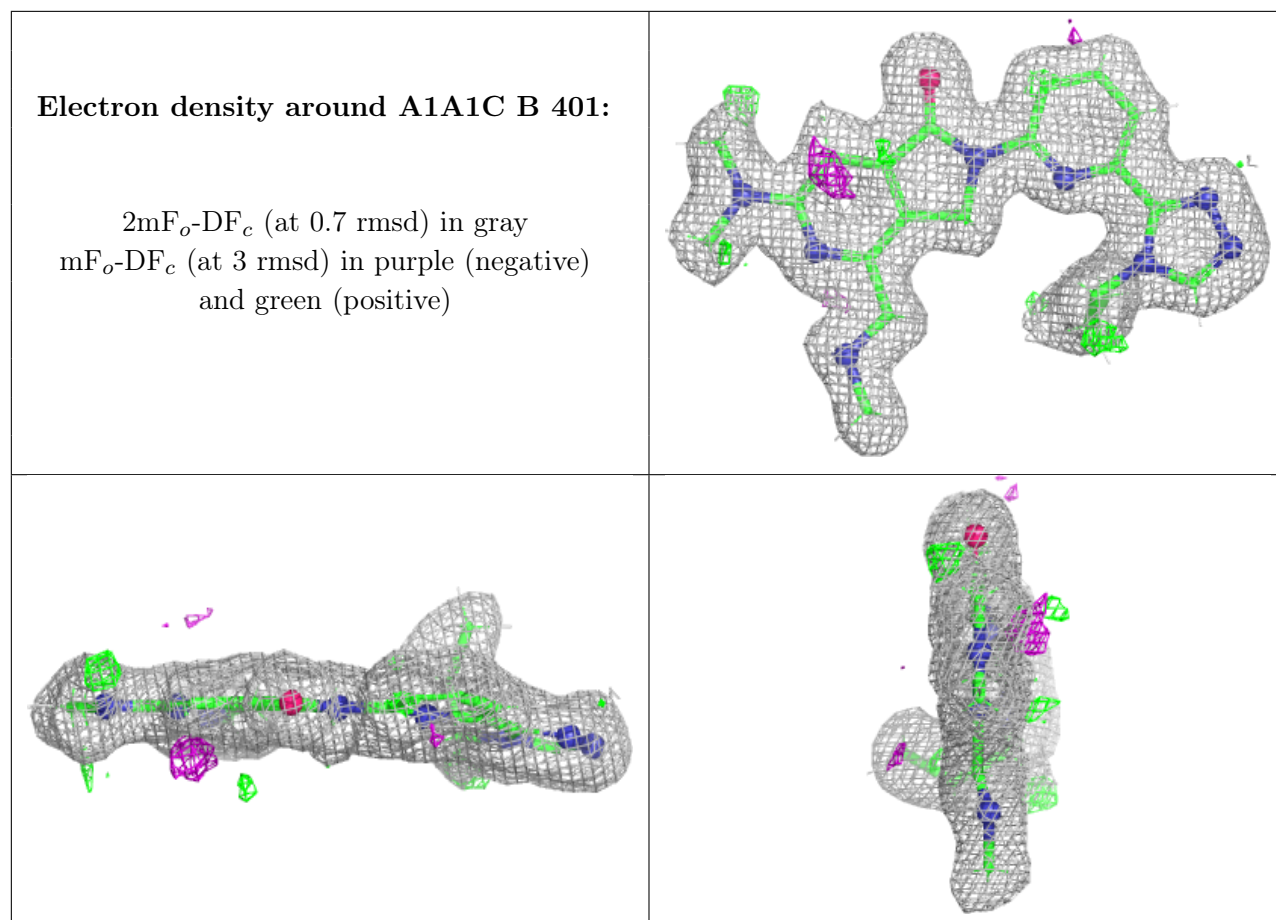| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
2	A1A1C	A	401	30/30	0.95	0.07	19,20,22,24	26
2	A1A1C	B	401	30/30	0.96	0.07	12,20,20,20	26

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Electron density around A1A1C A 401:

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

Electron density around A1A1C B 401:

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.