Full wwPDB X-ray Structure Validation Report ⓘ

May 7, 2025 – 04:09 am BST

PDB ID	:	9GL7 / pdb_00009gl7
Title	:	EGFR Exon20 insertion mutant NPG bound with (S)-3-((3-chloro-2-methoxyphenyl)amino)-2-(3-((tetrahydrofuran-2-yl)methoxy)pyridin-4-yl)-1,5,6,7-tetrahydro-4H-pyrrolo[3,2-c]pyridin-4-one
Authors	:	Hilbert, B.J.; Brooijmans, N.; Milgram, B.C.; Pagliarini, R.A.
Deposited on	:	2024-08-27
Resolution	:	1.88 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

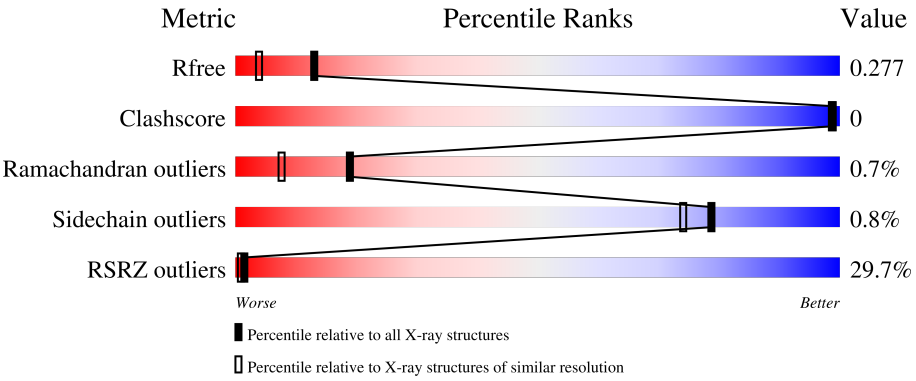
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 1.88 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}	164625	1090 (1.88-1.88)
Clashscore	180529	1144 (1.88-1.88)
Ramachandran outliers	177936	1135 (1.88-1.88)
Sidechain outliers	177891	1135 (1.88-1.88)
RSRZ outliers	164620	1090 (1.88-1.88)

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	338	26% 86% · 11%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5054 atoms, of which 2483 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.

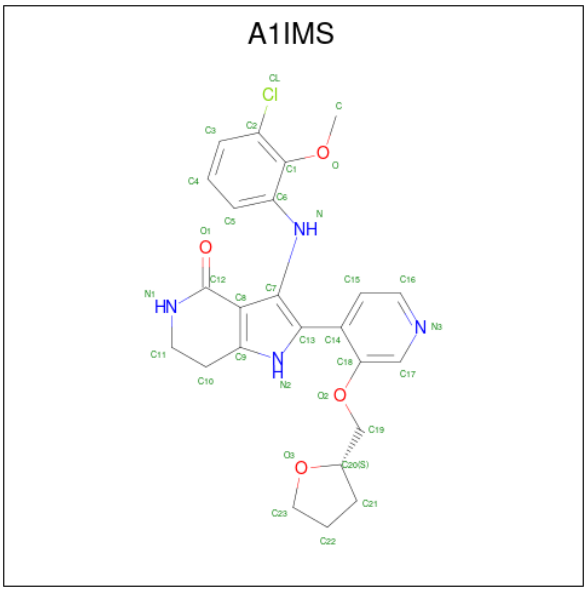
- Molecule 1 is a protein called Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	300	4914	1567	2483	415	433	16	2483	4	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	694	MET	-	initiating methionine	UNP P00533
A	770A	ASN	-	insertion	UNP P00533
A	770B	PRO	-	insertion	UNP P00533
A	770C	GLY	-	insertion	UNP P00533
A	948	ARG	VAL	engineered mutation	UNP P00533
A	1023	HIS	-	expression tag	UNP P00533
A	1024	HIS	-	expression tag	UNP P00533
A	1025	HIS	-	expression tag	UNP P00533
A	1026	HIS	-	expression tag	UNP P00533
A	1027	HIS	-	expression tag	UNP P00533
A	1028	HIS	-	expression tag	UNP P00533

- Molecule 2 is 3-[(3-chloranyl-2-methoxy-phenyl)amino]-2-[3-[[(2S)-oxolan-2-yl]methoxy]pyridin-4-yl]-1,5,6,7-tetrahydropyrrolo[3,2-c]pyridin-4-one (CCD ID: A1IMS) (formula: $C_{24}H_{25}ClN_4O_4$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	0	0
			33	24	1	4	4		

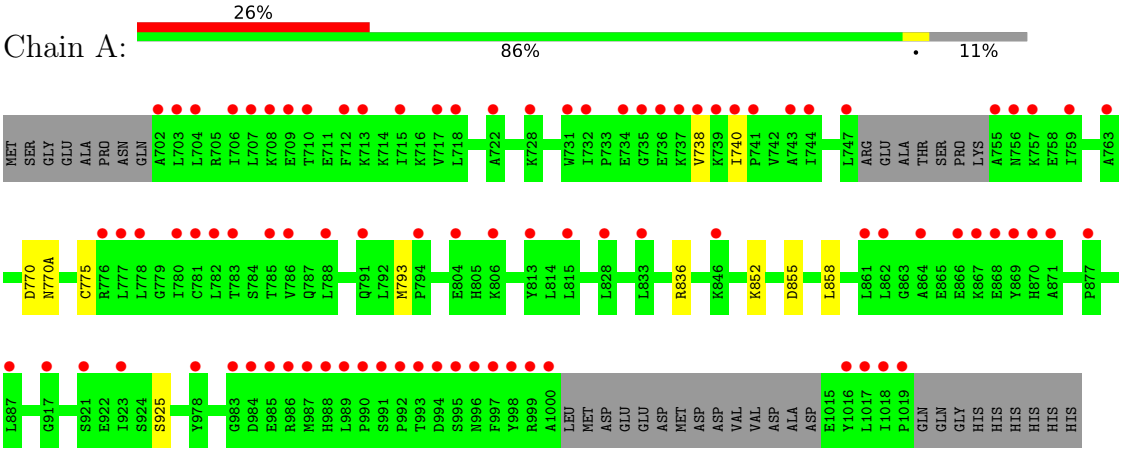
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	106	Total	O	0	2
			107	107		

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



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	39.74Å 105.78Å 174.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.22 – 1.88 34.22 – 1.88	Depositor EDS
% Data completeness (in resolution range)	53.9 (34.22-1.88) 53.9 (34.22-1.88)	Depositor EDS
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.57 (at 1.88Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.268 , 0.287 0.256 , 0.277	Depositor DCC
R_{free} test set	794 reflections (4.81%)	wwPDB-VP
Wilson B-factor (Å^2)	30.4	Xtriage
Anisotropy	0.147	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 46.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	5054	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 6.39% 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: A1IMS

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	0.79	0/2503	1.01	3/3382 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	770(A)	ASN	CA-CB-CG	-7.38	105.22	112.60
1	A	740	ILE	N-CA-C	5.65	113.46	107.76
1	A	770	ASP	CA-CB-CG	-5.29	107.31	112.60

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	2431	2483	2464	2	0
2	A	33	0	0	0	0
3	A	107	0	0	0	0
All	All	2571	2483	2464	2	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 0.

All (2) 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:793:MET:SD	1:A:852:LYS:HD2	2.58	0.43
1:A:836:ARG:HD3	1:A:858:LEU:O	2.18	0.43

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	298/338 (88%)	288 (97%)	8 (3%)	2 (1%)	19 7

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	855	ASP
1	A	738	VAL

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	268/297 (90%)	266 (99%)	2 (1%)	81 77

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	775	CYS
1	A	925	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	791	GLN
1	A	842	ASN
1	A	870	HIS
1	A	988	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 oligosaccharides 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	$\#\|Z\| > 2$
2	A1IMS	A	2001	-	33,37,37	1.02	2 (6%)	35,52,52	1.10	3 (8%)

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	A1IMS	A	2001	-	-	3/11/32/32	0/5/5/5

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2001	A1IMS	C14-C13	-3.10	1.46	1.49
2	A	2001	A1IMS	C17-C18	2.36	1.42	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	2001	A1IMS	O2-C18-C17	-2.96	118.68	124.77
2	A	2001	A1IMS	C19-O2-C18	-2.40	112.94	118.27
2	A	2001	A1IMS	O2-C18-C14	2.13	122.17	116.61

There are no chirality outliers.

All (3) torsion outliers are listed below:

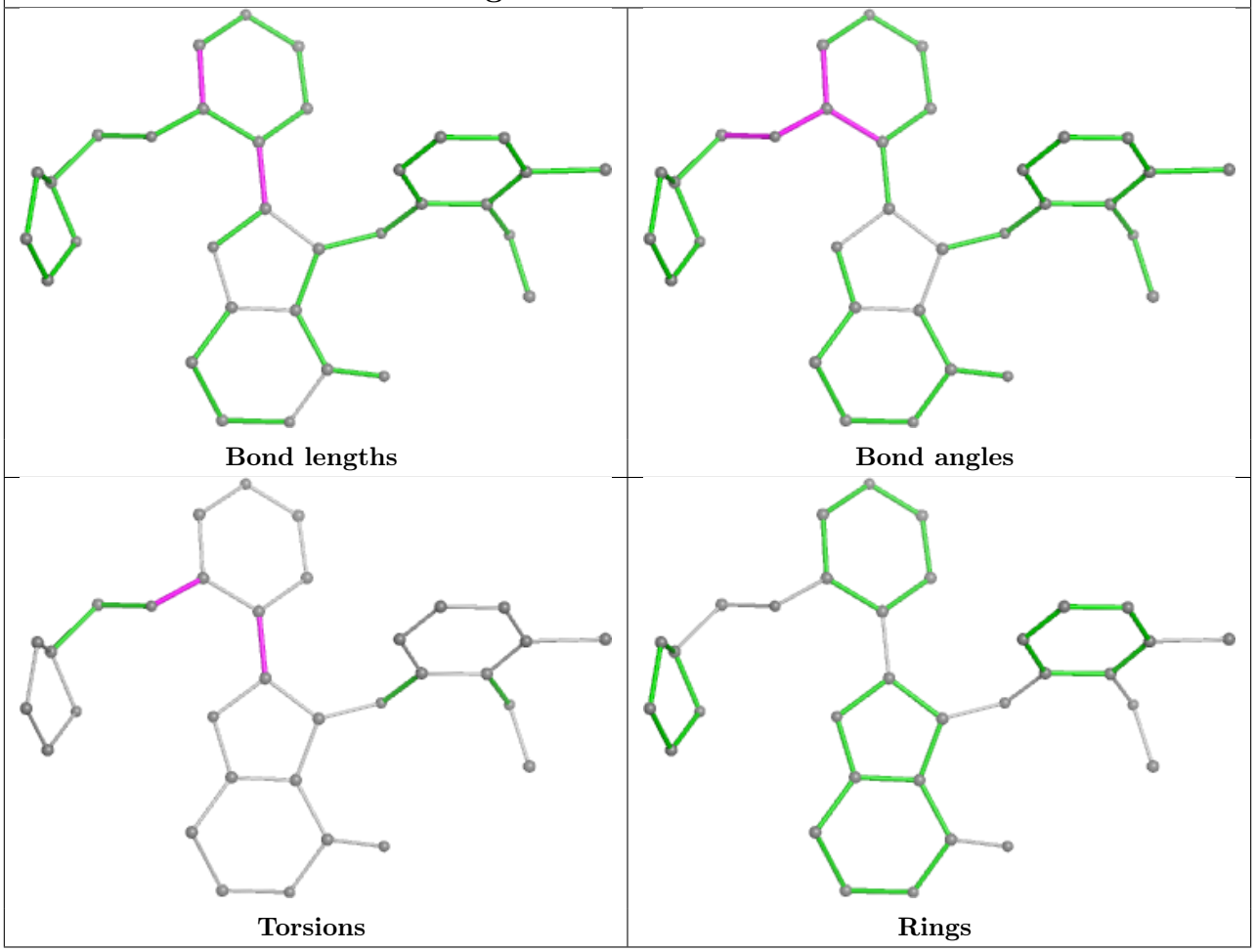
Mol	Chain	Res	Type	Atoms
2	A	2001	A1IMS	C7-C13-C14-C15
2	A	2001	A1IMS	C17-C18-O2-C19
2	A	2001	A1IMS	C14-C18-O2-C19

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 A1IMS A 2001



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	300/338 (88%)	1.57	89 (29%) 1 1	6, 21, 39, 48	2 (0%)

All (89) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1000	ALA	6.8
1	A	737	LYS	6.0
1	A	997	PHE	5.6
1	A	806[A]	LYS	5.4
1	A	993	THR	5.0
1	A	992	PRO	4.7
1	A	755	ALA	4.6
1	A	738	VAL	4.5
1	A	759	ILE	4.5
1	A	785	THR	4.5
1	A	996	ASN	4.3
1	A	703	LEU	4.2
1	A	747	LEU	4.2
1	A	989	LEU	4.1
1	A	778	LEU	4.1
1	A	861	LEU	4.1
1	A	998	TYR	4.0
1	A	788	LEU	4.0
1	A	991	SER	3.8
1	A	734	GLU	3.7
1	A	1018	ILE	3.7
1	A	988	HIS	3.7
1	A	756	ASN	3.6
1	A	741	PRO	3.6
1	A	999	ARG	3.6
1	A	740	ILE	3.5
1	A	739	LYS	3.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	718	LEU	3.4
1	A	869	TYR	3.4
1	A	777	LEU	3.4
1	A	1019	PRO	3.3
1	A	710	THR	3.3
1	A	702	ALA	3.3
1	A	1016	TYR	3.3
1	A	722	ALA	3.2
1	A	707	LEU	3.2
1	A	706	ILE	3.2
1	A	917	GLY	3.2
1	A	986	ARG	3.2
1	A	780	ILE	3.1
1	A	708	LYS	3.1
1	A	783	THR	3.1
1	A	871	ALA	3.0
1	A	704	LEU	3.0
1	A	862	LEU	3.0
1	A	735	GLY	3.0
1	A	994	ASP	3.0
1	A	1017	LEU	3.0
1	A	985	GLU	2.8
1	A	867	LYS	2.8
1	A	870	HIS	2.8
1	A	923	ILE	2.8
1	A	995	SER	2.7
1	A	743	ALA	2.7
1	A	717	VAL	2.6
1	A	794	PRO	2.6
1	A	866	GLU	2.6
1	A	786	VAL	2.5
1	A	983	GLY	2.5
1	A	731	TRP	2.5
1	A	732	ILE	2.5
1	A	757	LYS	2.5
1	A	782	LEU	2.5
1	A	791	GLN	2.5
1	A	828	LEU	2.4
1	A	744	ILE	2.4
1	A	987	MET	2.3
1	A	864	ALA	2.3
1	A	713	LYS	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	978	TYR	2.3
1	A	877	PRO	2.3
1	A	715	ILE	2.2
1	A	804	GLU	2.2
1	A	990	PRO	2.2
1	A	921	SER	2.2
1	A	846	LYS	2.2
1	A	712	PHE	2.2
1	A	728	LYS	2.1
1	A	736	GLU	2.1
1	A	887	LEU	2.1
1	A	776	ARG	2.1
1	A	813	TYR	2.1
1	A	709	GLU	2.1
1	A	833	LEU	2.1
1	A	984	ASP	2.1
1	A	781	CYS	2.1
1	A	763	ALA	2.1
1	A	815	LEU	2.0
1	A	868	GLU	2.0

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, 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.

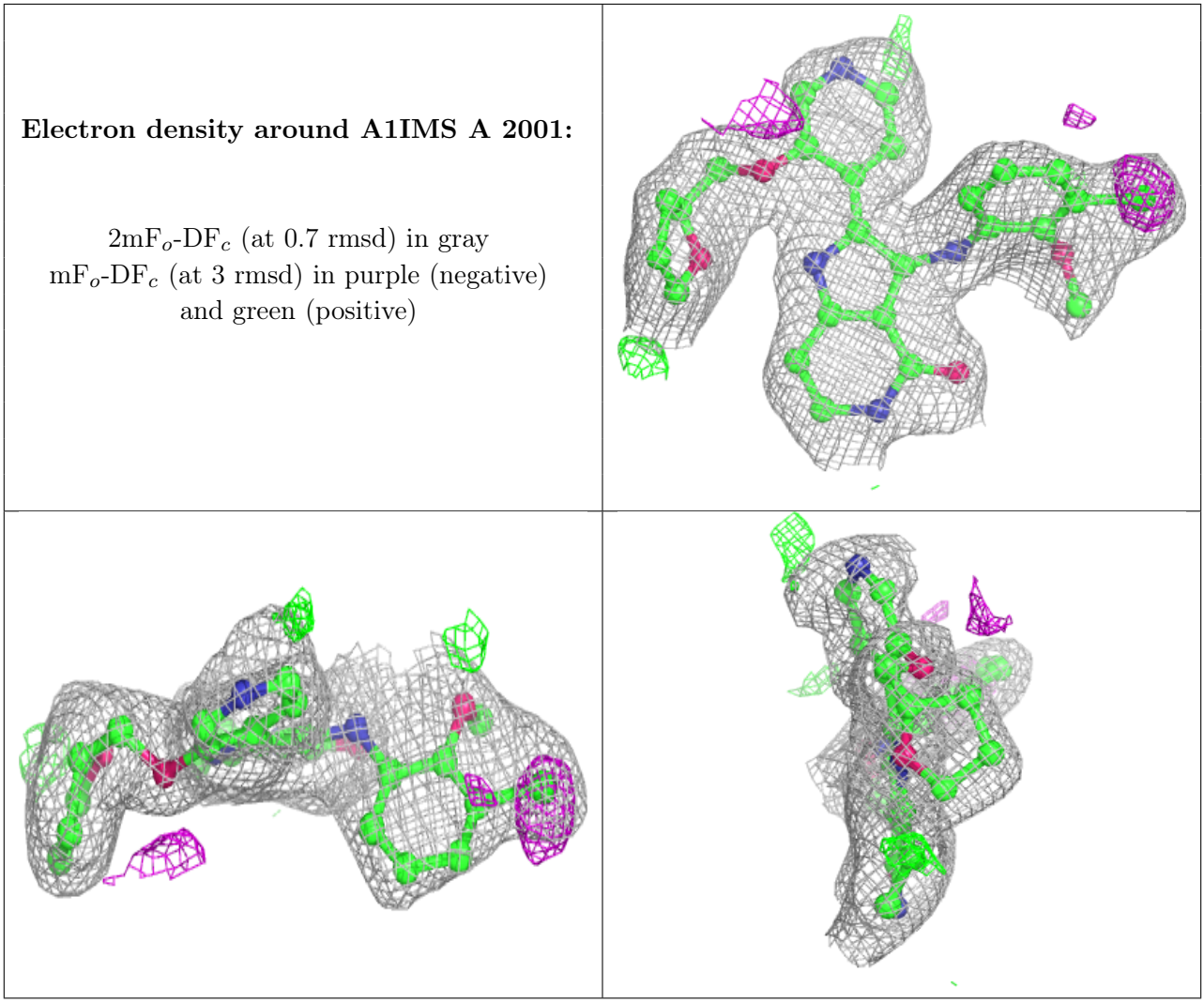
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	A1IMS	A	2001	33/33	0.84	0.13	36,39,41,42	0

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 A1IMS A 2001:

$2mF_o$-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.