W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 05:59 pm BST

PDB ID	:	4BBE
Title	:	Aminoalkylpyrimidine Inhibitor Complexes with JAK2
Authors	:	Li, J.
Deposited on	:	2012-09-21
Resolution	:	1.90 Å(reported)

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

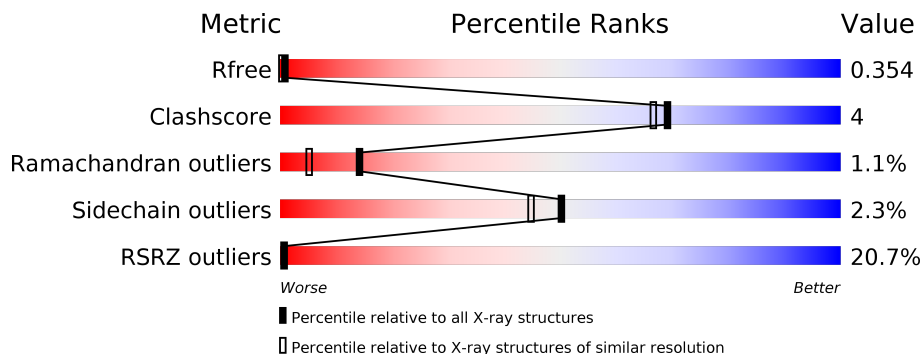|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.90 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		0.354
Clashscore		4
Ramachandran outliers		1.1%
Sidechain outliers		2.3%
RSRZ outliers		20.7%

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}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 on 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	298	22% / 86% / 9% • •
1	B	298	19% / 88% / 8% •
1	C	298	17% / 85% / 10% • •
1	D	298	21% / 88% / 7% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9717 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.
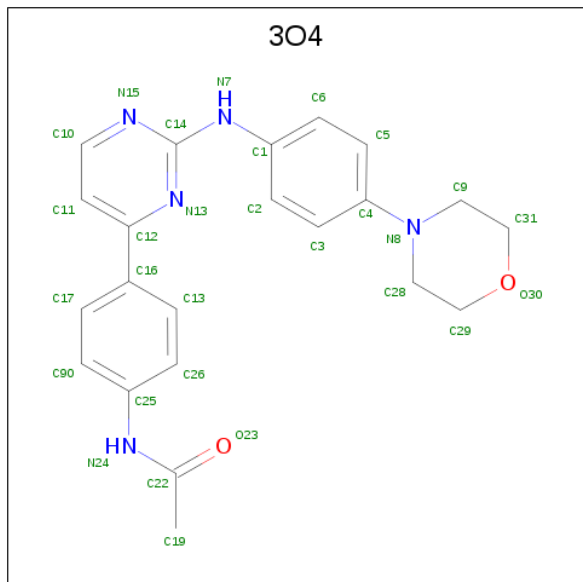
- Molecule 1 is a protein called TYROSINE-PROTEIN KINASE JAK2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	288	Total 2379	C 1515	N 415	O 436	S 13	0	0	0
1	B	288	Total 2379	C 1515	N 415	O 436	S 13	0	0	0
1	C	288	Total 2379	C 1515	N 415	O 436	S 13	0	0	0
1	D	288	Total 2379	C 1515	N 415	O 436	S 13	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	837	GLY	-	expression tag	UNP O60674
A	838	SER	-	expression tag	UNP O60674
A	976	ASN	ASP	engineered mutation	UNP O60674
A	1133	GLU	-	expression tag	UNP O60674
A	1134	PHE	-	expression tag	UNP O60674
B	837	GLY	-	expression tag	UNP O60674
B	838	SER	-	expression tag	UNP O60674
B	976	ASN	ASP	engineered mutation	UNP O60674
B	1133	GLU	-	expression tag	UNP O60674
B	1134	PHE	-	expression tag	UNP O60674
C	837	GLY	-	expression tag	UNP O60674
C	838	SER	-	expression tag	UNP O60674
C	976	ASN	ASP	engineered mutation	UNP O60674
C	1133	GLU	-	expression tag	UNP O60674
C	1134	PHE	-	expression tag	UNP O60674
D	837	GLY	-	expression tag	UNP O60674
D	838	SER	-	expression tag	UNP O60674
D	976	ASN	ASP	engineered mutation	UNP O60674
D	1133	GLU	-	expression tag	UNP O60674
D	1134	PHE	-	expression tag	UNP O60674

- Molecule 2 is N-[4-[2-[(4-morpholin-4-ylphenyl)amino]pyrimidin-4-yl]phenyl]ethanamide (three-letter code: 3O4) (formula: $C_{22}H_{23}N_5O_2$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 29	C 22	N 5	O 2	0	0
2	B	1	Total 29	C 22	N 5	O 2	0	0
2	C	1	Total 29	C 22	N 5	O 2	0	0
2	D	1	Total 29	C 22	N 5	O 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	22	Total 22	O 22	0	0
3	B	18	Total 18	O 18	0	0
3	C	28	Total 28	O 28	0	0
3	D	17	Total 17	O 17	0	0

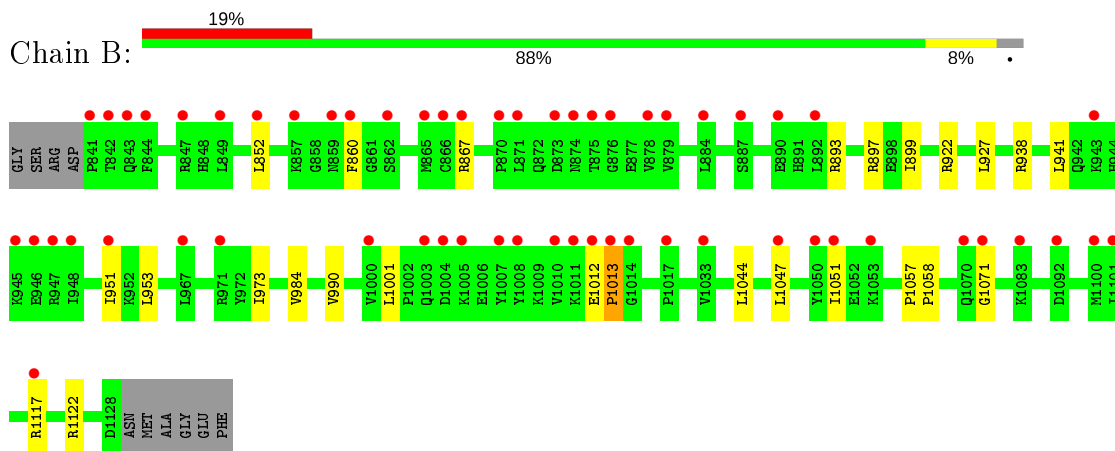# 3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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.
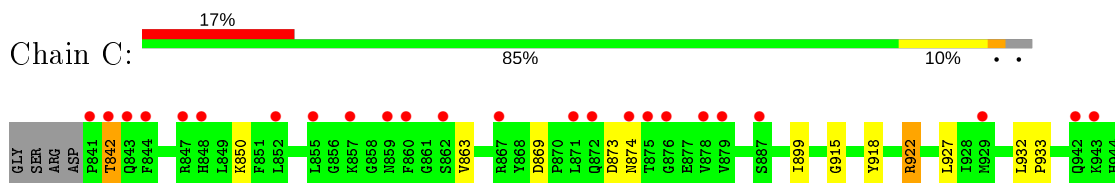
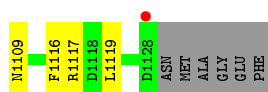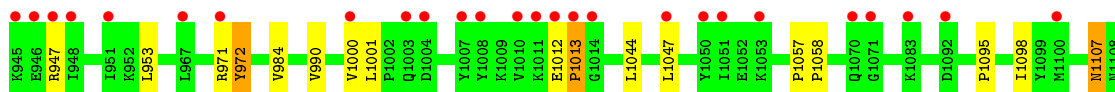- Molecule 1: TYROSINE-PROTEIN KINASE JAK2

- Molecule 1: TYROSINE-PROTEIN KINASE JAK2

- Molecule 1: TYROSINE-PROTEIN KINASE JAK2

K945 E946 R947 I948 I951 K952 L953 L967 R971 Y972 V984 V990 V1000 L1001 P1002 Q1003 D1004 Y1007 Y1008 K1009 V1010 K1011 E1012 P1013 G1014 L1044 L1047 Y1050 I1051 E1052 K1053 P1057 P1058 Q1070 G1071 K1083 D1092 P1095 I1098 Y1099 M1100 N1107 N1108

N1109 F1116 R1117 D1118 L1119 D1128 ASN MET ALA GLY PHE

● Molecule 1: TYROSINE-PROTEIN KINASE JAK2

Chain D:

21%

88%　　　　　　　　7%

GLY SER ARG ASP P841 T842 Q843 F844 E845 E846 R847 H848 L852 K857 G858 N859 F860 G861 S862 V863 R867 P870 L871 Q872 D873 N874 T875 V878 V879 A880 V881 L884 Q885 H886 S887 T888 R897 E898 I899 I910 Y918 S919 A920 G921 R922 L925 K926 L927 L932

P933 R938 L941 Q942 K943 H944 K945 E946 R947 I948 D949 H950 I951 K952 L953 L967 R971 R980 L983 I992 V1000 L1001 D1004 Y1007 Y1008 K1009 V1010 K1011 E1012 P1013 G1014 E1015 F1019 L1047 I1051 E1052 P1057 P1058 M1062 I1065 Q1070 G1071

D1092 M1100 W1106 N1107 N1108 N1109 D1128 ASN MET ALA GLY GLU PHE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	68.96Å 76.28Å 87.74Å 84.02° 66.87° 63.13°	Depositor
Resolution (Å)	41.24 – 1.90	Depositor
	41.23 – 1.90	EDS
% Data completeness	93.5 (41.24-1.90)	Depositor
(in resolution range)	93.5 (41.23-1.90)	EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.82 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.344 , 0.359	Depositor
	0.339 , 0.354	DCC
R_{free} test set	5357 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	26.5	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 9.3	EDS
L-test for twinning [2]	$<|L|> = 0.52, <L^2> = 0.35$	Xtriage
Estimated twinning fraction	0.478 for h,h-k,h-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	9717	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 63.07 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0234e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: 3O4

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	0.31	0/2432	0.48	0/3274
1	B	0.31	0/2432	0.46	0/3274
1	C	0.32	0/2432	0.48	0/3274
1	D	0.31	0/2432	0.46	0/3274
All	All	0.31	0/9728	0.47	0/13096

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	2379	0	2369	19	0
1	B	2379	0	2369	12	0
1	C	2379	0	2369	17	0
1	D	2379	0	2369	15	0
2	A	29	0	23	3	0
2	B	29	0	23	4	0
2	C	29	0	23	3	0
2	D	29	0	23	2	0
3	A	22	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	18	0	0	0	0
3	C	28	0	0	0	0
3	D	17	0	0	0	0
All	All	9717	0	9568	70	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 4.

The worst 5 of 70 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:B:2229:3O4:H2	2:B:2229:3O4:N13	2.00	0.76
1:D:947:ARG:HA	1:D:947:ARG:HH11	1.51	0.75
1:C:971:ARG:N	1:C:972:TYR:HB2	2.05	0.72
1:A:944:HIS:N	1:A:945:LYS:HB2	2.06	0.71
2:D:2229:3O4:N13	2:D:2229:3O4:H2	2.08	0.68

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	286/298 (96%)	264 (92%)	18 (6%)	4 (1%)	11	3
1	B	286/298 (96%)	270 (94%)	14 (5%)	2 (1%)	22	12
1	C	286/298 (96%)	267 (93%)	15 (5%)	4 (1%)	11	3
1	D	286/298 (96%)	269 (94%)	14 (5%)	3 (1%)	15	6
All	All	1144/1192 (96%)	1070 (94%)	61 (5%)	13 (1%)	14	5

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	945	LYS
1	A	1013	PRO
1	B	1013	PRO
1	C	842	THR
1	C	1013	PRO

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	263/271 (97%)	257 (98%)	6 (2%)	50	45
1	B	263/271 (97%)	257 (98%)	6 (2%)	50	45
1	C	263/271 (97%)	257 (98%)	6 (2%)	50	45
1	D	263/271 (97%)	257 (98%)	6 (2%)	50	45
All	All	1052/1084 (97%)	1028 (98%)	24 (2%)	50	45

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	1117	ARG
1	C	869	ASP
1	D	1011	LYS
1	B	1122	ARG
1	C	850	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	944	HIS
1	D	1107	ASN
1	D	955	GLN
1	A	1112	GLN
1	C	1107	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 carbohydrates in this entry.

5.6   Ligand geometry (i)

4 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	3O4	D	2229	-	32,32,32	1.45	3 (9%)	43,43,43	2.15	7 (16%)
2	3O4	B	2229	-	32,32,32	1.48	3 (9%)	43,43,43	2.12	6 (13%)
2	3O4	C	2229	-	32,32,32	1.48	3 (9%)	43,43,43	2.09	7 (16%)
2	3O4	A	2229	-	32,32,32	1.47	3 (9%)	43,43,43	2.08	7 (16%)

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	3O4	D	2229	-	-	4/16/24/24	0/4/4/4
2	3O4	B	2229	-	-	0/16/24/24	0/4/4/4
2	3O4	C	2229	-	-	0/16/24/24	0/4/4/4
2	3O4	A	2229	-	-	0/16/24/24	0/4/4/4

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2229	3O4	C16-C12	-5.61	1.40	1.48
2	A	2229	3O4	C16-C12	-5.59	1.40	1.48
2	C	2229	3O4	C16-C12	-5.56	1.40	1.48
2	D	2229	3O4	C16-C12	-5.48	1.40	1.48
2	C	2229	3O4	C25-N24	-3.56	1.34	1.41

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	2229	3O4	N15-C14-N13	-7.18	119.75	126.55
2	B	2229	3O4	N15-C14-N13	-7.16	119.76	126.55
2	A	2229	3O4	N15-C14-N13	-7.16	119.76	126.55
2	C	2229	3O4	C10-N15-C14	7.06	121.72	115.45
2	C	2229	3O4	N15-C14-N13	-6.97	119.94	126.55

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	2229	3O4	C3-C4-N8-C28
2	D	2229	3O4	C5-C4-N8-C9
2	D	2229	3O4	C3-C4-N8-C9
2	D	2229	3O4	C5-C4-N8-C28

There are no ring outliers.
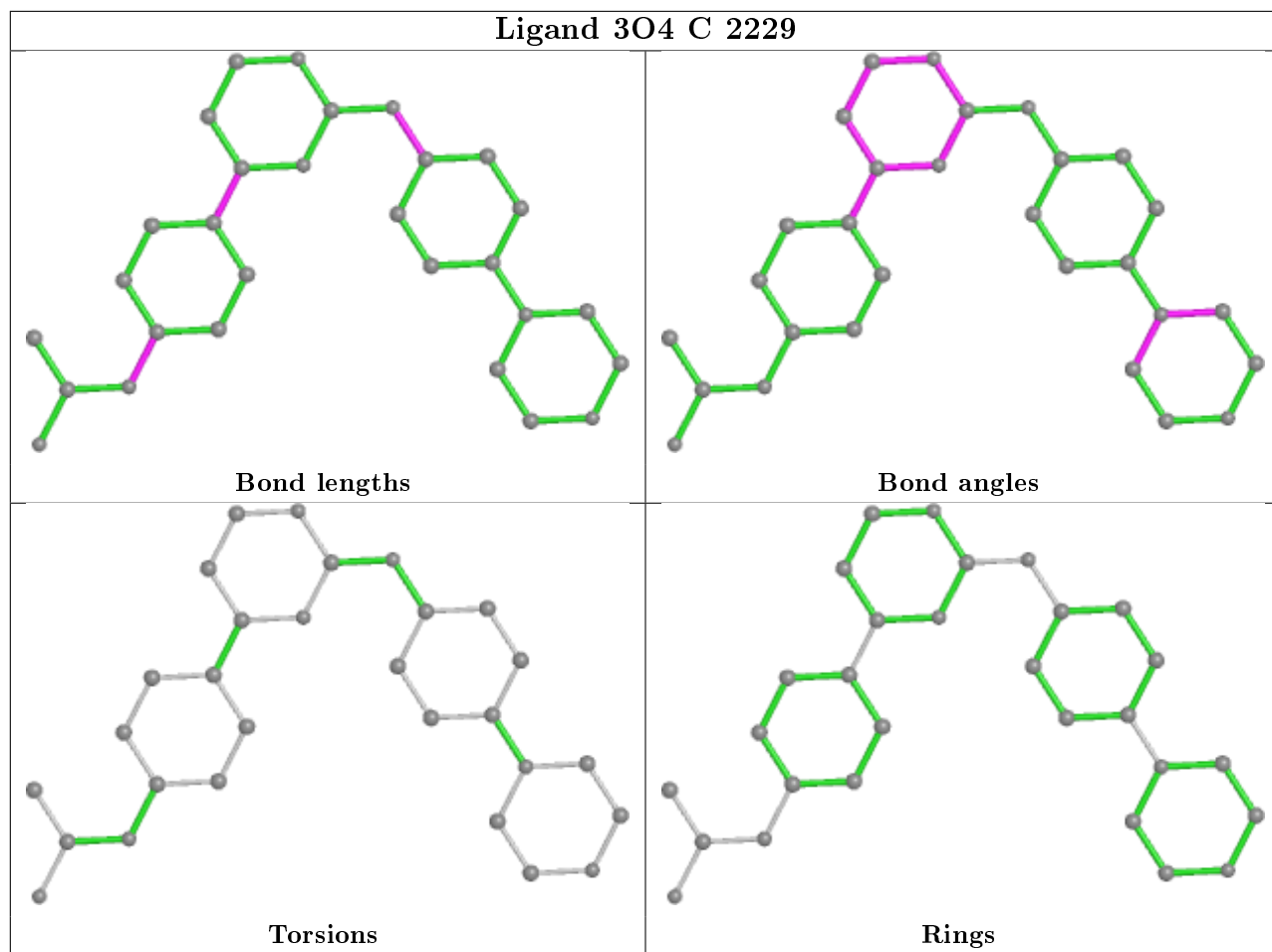
4 monomers are involved in 12 short contacts:

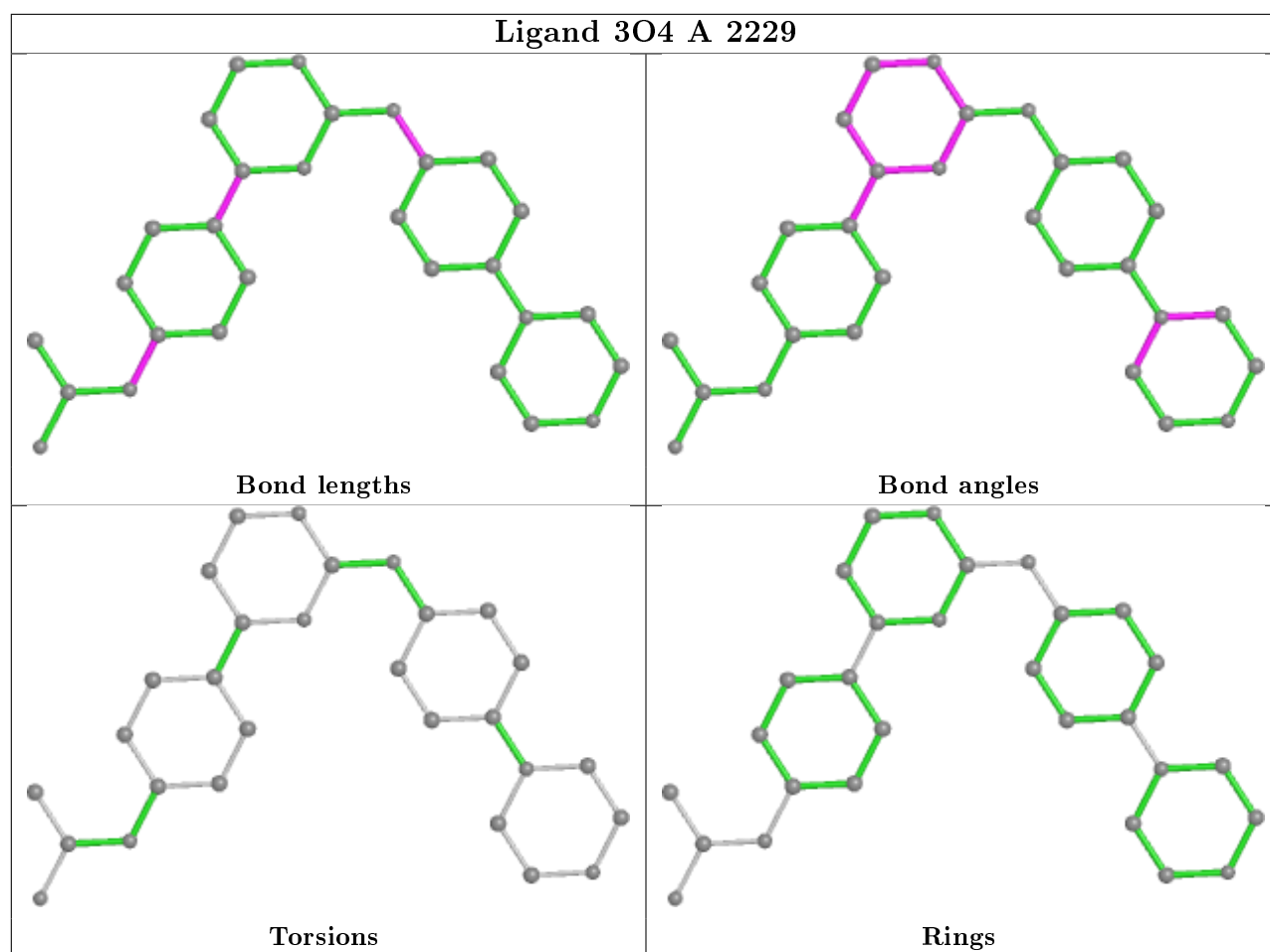| Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	2229	3O4	2	0
2	B	2229	3O4	4	0
2	C	2229	3O4	3	0
2	A	2229	3O4	3	0

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 3O4 D 2229

Bond lengths	Bond angles

Torsions	Rings

Ligand 3O4 B 2229

Bond lengths

Bond angles

Torsions

Rings

Ligand 3O4 C 2229

Bond lengths

Bond angles

Torsions

Rings

Ligand 3O4 A 2229

Bond lengths

Bond angles

Torsions

Rings
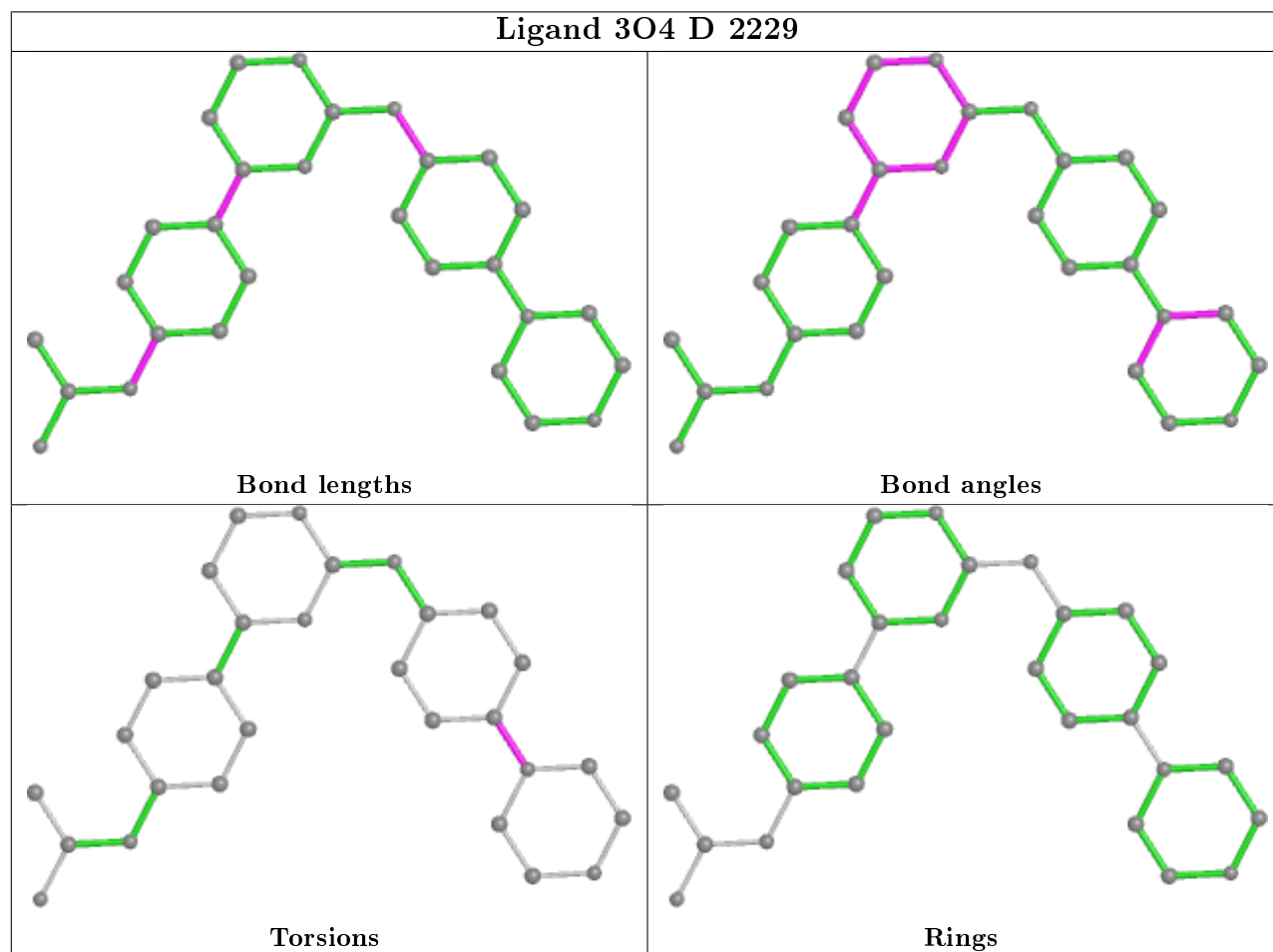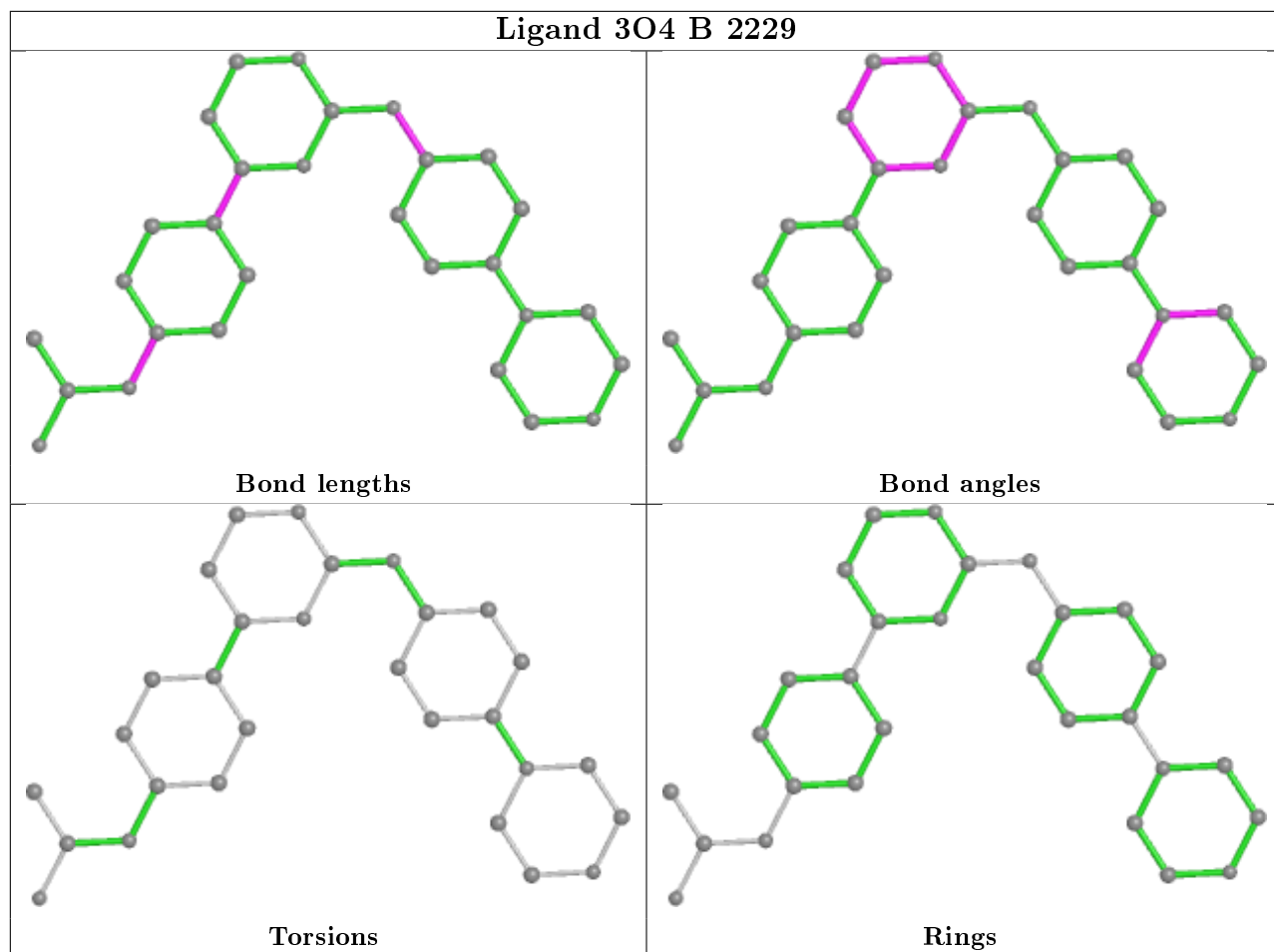
5.7 Other polymers ⓘ

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 (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(Å²)	Q<0.9
1	A	288/298 (96%)	1.46	67 (23%)	0	0	15, 22, 37, 39	3 (1%)
1	B	288/298 (96%)	1.45	58 (20%)	1	1	16, 23, 36, 37	3 (1%)
1	C	288/298 (96%)	1.39	51 (17%)	1	1	16, 23, 35, 37	3 (1%)
1	D	288/298 (96%)	1.48	62 (21%)	0	0	15, 22, 37, 39	3 (1%)
All	All	1152/1192 (96%)	1.45	238 (20%)	1	1	15, 23, 36, 39	12 (1%)

The worst 5 of 238 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1013	PRO	10.3
1	B	871	LEU	10.2
1	C	871	LEU	9.9
1	B	1013	PRO	9.1
1	A	871	LEU	8.7

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.
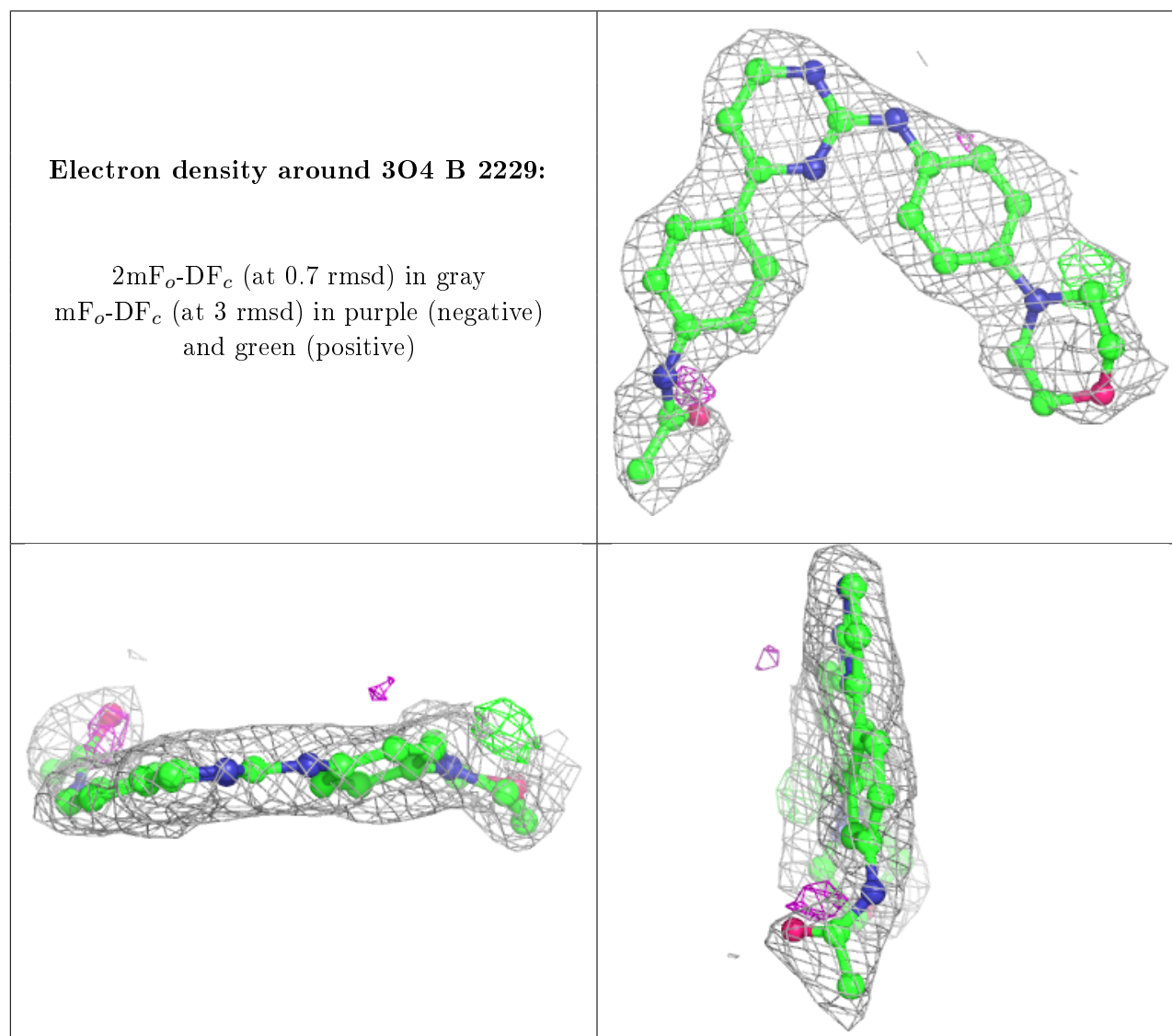
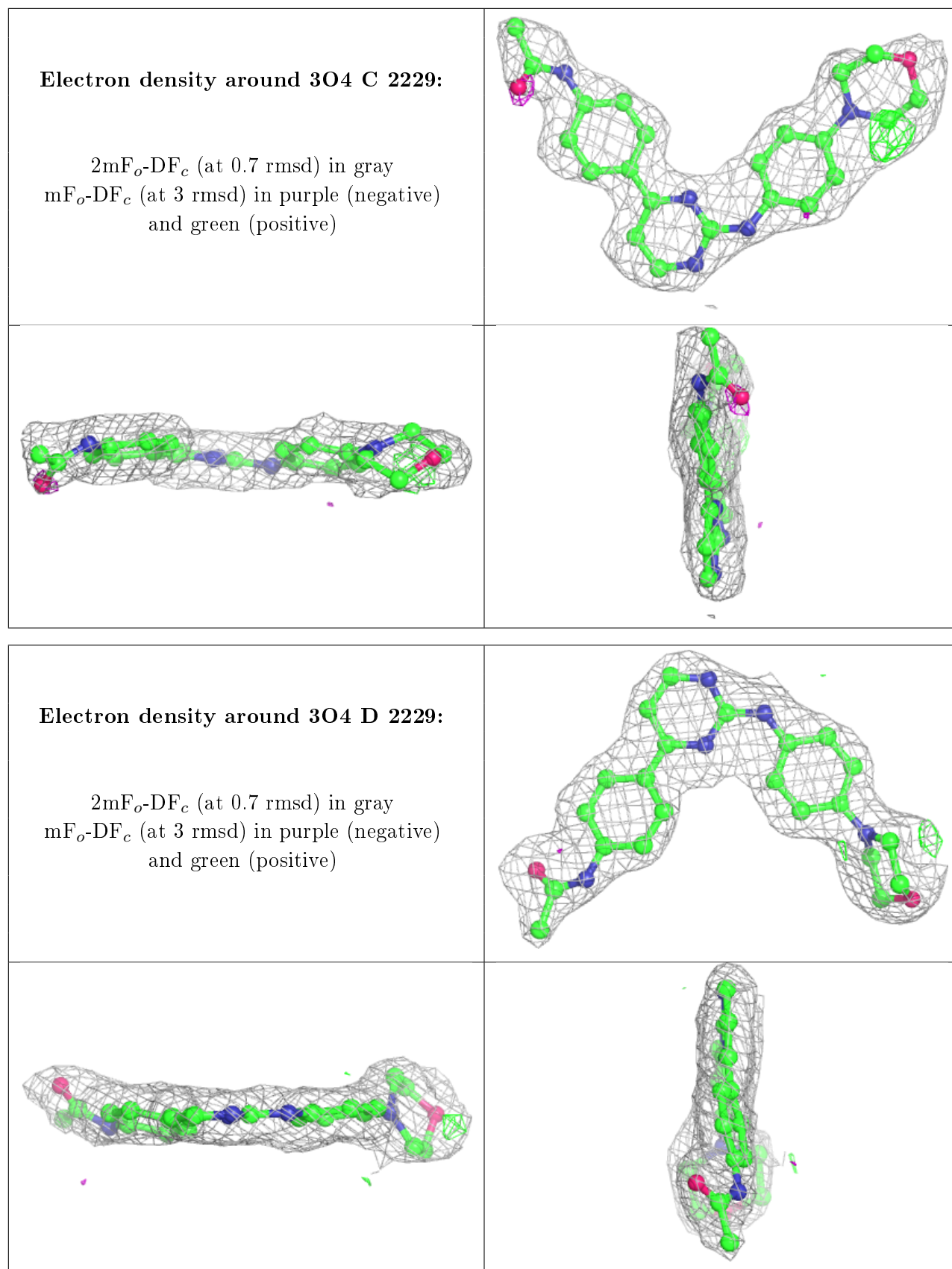## 6.3 Carbohydrates (i)

There are no carbohydrates in this entry.
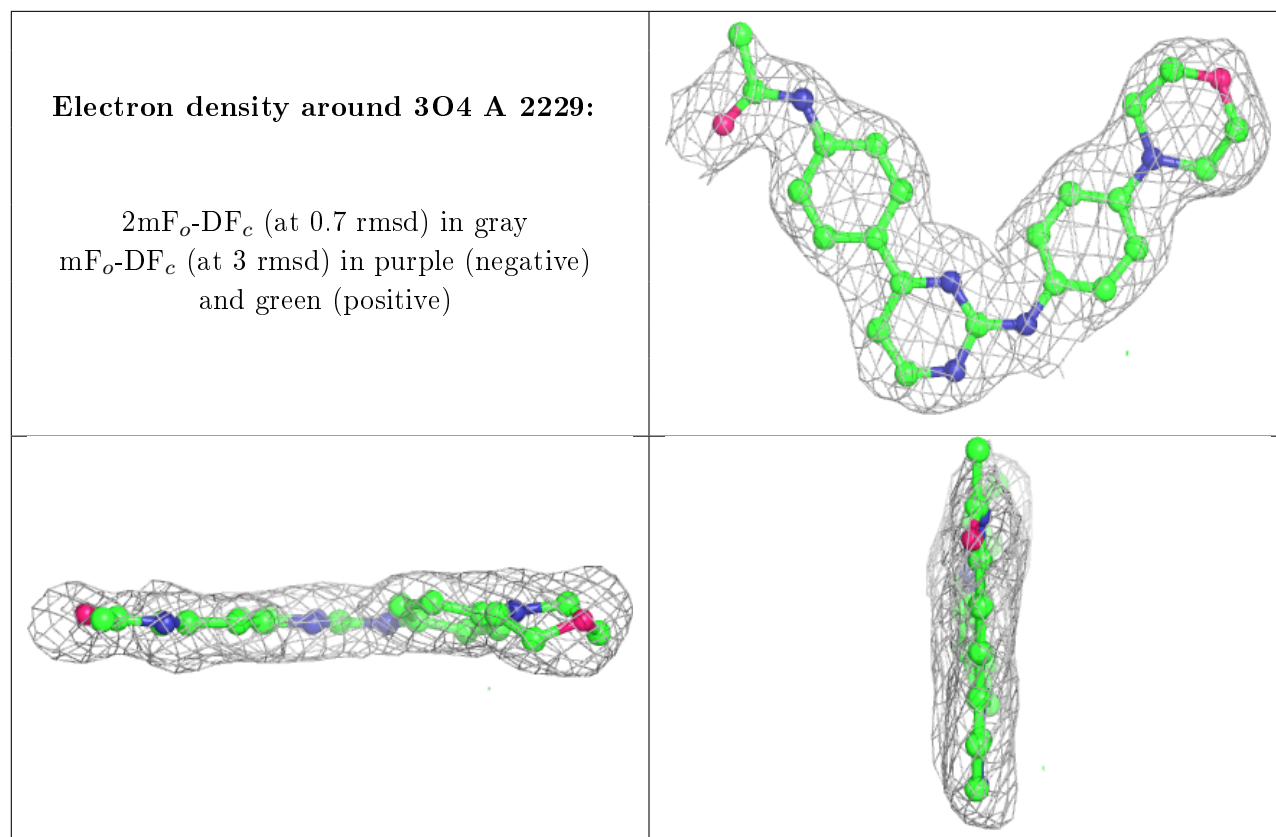
6.4 Ligands (i)

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	3O4	B	2229	29/29	0.80	0.20	32,32,32,32	0
2	3O4	C	2229	29/29	0.82	0.20	32,32,32,32	0
2	3O4	D	2229	29/29	0.84	0.18	30,30,30,30	0
2	3O4	A	2229	29/29	0.86	0.16	30,30,30,30	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 3O4 B 2229:

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

Electron density around 3O4 C 2229:

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

Electron density around 3O4 D 2229:

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

Electron density around 3O4 A 2229:

$2mF_o$-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.