



Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2023 – 04:17 PM JST

PDB ID : 8GSX
Title : X-ray structure of Clostridium perfringens pili protein B collagen-binding domains
Authors : Kamitori, S.; Yamada, M.; Tamai, E.
Deposited on : 2022-09-07
Resolution : 2.86 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 : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.33
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) : 2.33

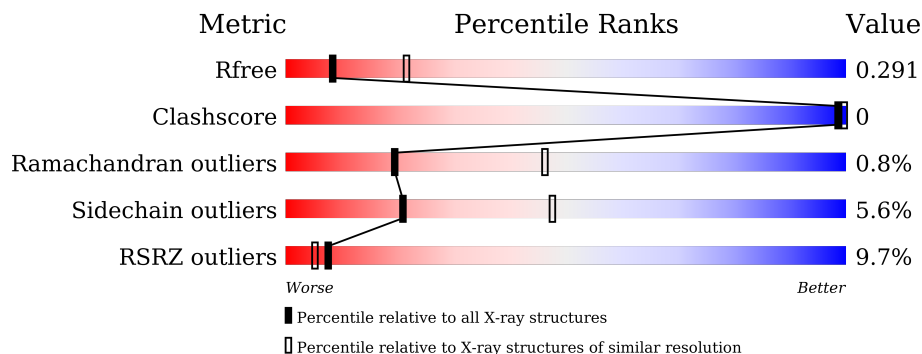
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.86 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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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 $\leq 5\%$. 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	335	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2097 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.

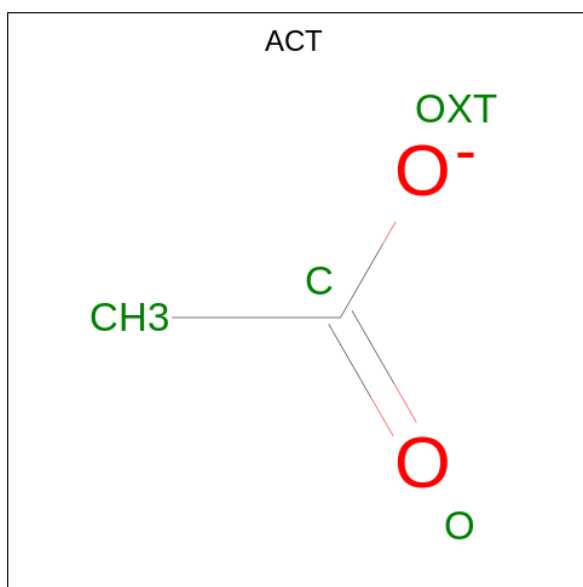
- Molecule 1 is a protein called pili protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	259	2072	1307	339	421	5	0	0	0

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	16	MET	-	initiating methionine	UNP Q8XP11
A	17	ASN	-	expression tag	UNP Q8XP11
A	18	HIS	-	expression tag	UNP Q8XP11
A	19	LYS	-	expression tag	UNP Q8XP11
A	20	VAL	-	expression tag	UNP Q8XP11
A	21	HIS	-	expression tag	UNP Q8XP11
A	22	HIS	-	expression tag	UNP Q8XP11
A	23	HIS	-	expression tag	UNP Q8XP11
A	24	HIS	-	expression tag	UNP Q8XP11
A	25	HIS	-	expression tag	UNP Q8XP11
A	26	HIS	-	expression tag	UNP Q8XP11
A	27	ILE	-	expression tag	UNP Q8XP11
A	28	GLU	-	expression tag	UNP Q8XP11
A	29	GLY	-	expression tag	UNP Q8XP11
A	30	ARG	-	expression tag	UNP Q8XP11
A	31	HIS	-	expression tag	UNP Q8XP11
A	32	MET	-	expression tag	UNP Q8XP11

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0

- Molecule 3 is water.

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

4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α , β , γ	127.39Å 127.39Å 163.16Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.83 – 2.86 48.78 – 2.86	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.83-2.86) 99.6 (48.78-2.86)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.88 (at 2.86Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.252 , 0.286 0.255 , 0.291	Depositor DCC
R_{free} test set	913 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å ²)	90.5	Xtrriage
Anisotropy	0.014	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 55.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	2097	wwPDB-VP
Average B, all atoms (Å ²)	99.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.44% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: ACT

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	0.69	0/2111	0.96	3/2853 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	65	ARG	NE-CZ-NH2	5.93	123.26	120.30
1	A	131	ARG	NE-CZ-NH1	5.68	123.14	120.30
1	A	202	ARG	NE-CZ-NH2	5.08	122.84	120.30

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	2072	0	1996	2	0
2	A	20	0	15	0	0
3	A	5	0	0	0	0
All	All	2097	0	2011	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:233:ASP:OD1	1:A:294:TYR:OH	2.30	0.50
1:A:248:ILE:HD12	1:A:271:ILE:HD11	1.99	0.44

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	251/335 (75%)	230 (92%)	19 (8%)	2 (1%)	19 46

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	171	ALA
1	A	42	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	232/302 (77%)	219 (94%)	13 (6%)	21 47

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	54	ASN
1	A	94	GLU
1	A	108	LYS
1	A	131	ARG
1	A	150	LEU
1	A	159	ILE
1	A	172	GLN
1	A	202	ARG
1	A	256	GLU
1	A	278	LEU
1	A	296	THR
1	A	314	ASP
1	A	328	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	135	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

5 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	ACT	A	403	-	3,3,3	0.99	0	3,3,3	0.74	0
2	ACT	A	401	-	3,3,3	0.95	0	3,3,3	0.91	0
2	ACT	A	405	-	3,3,3	1.07	0	3,3,3	0.72	0
2	ACT	A	402	-	3,3,3	1.03	0	3,3,3	0.70	0
2	ACT	A	404	-	3,3,3	0.97	0	3,3,3	0.80	0

There are no bond length outliers.

There are no bond angle outliers.

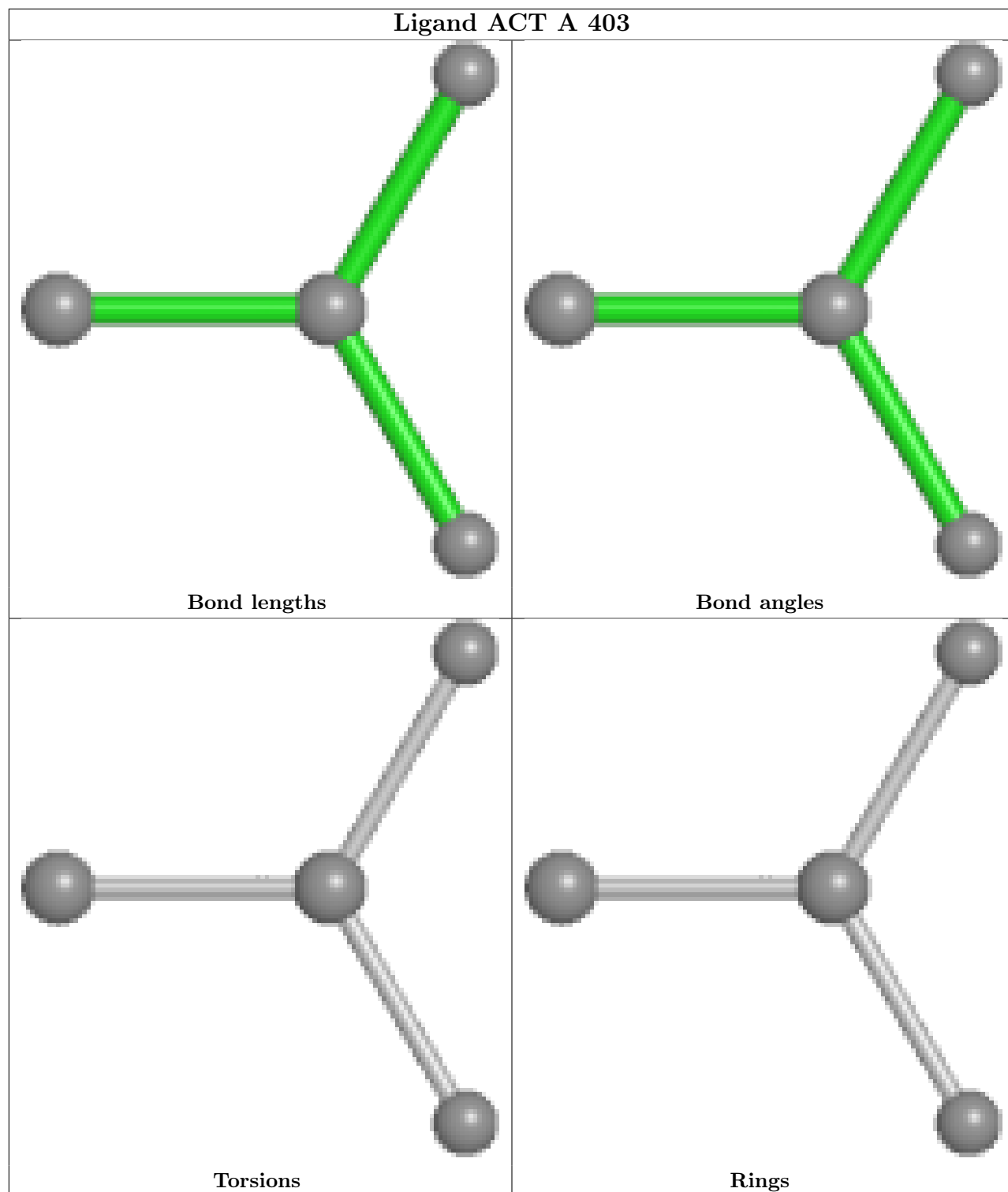
There are no chirality outliers.

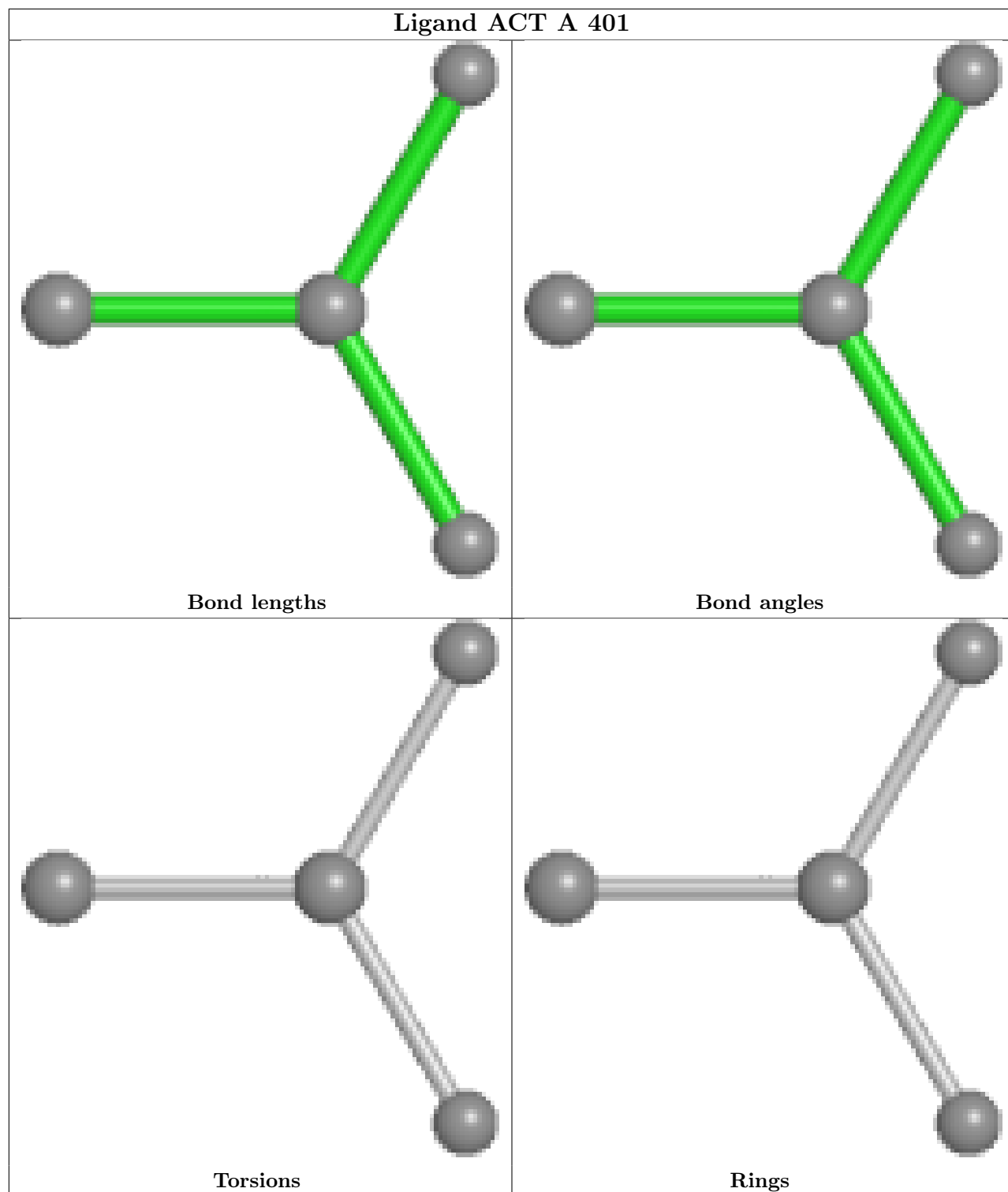
There are no torsion outliers.

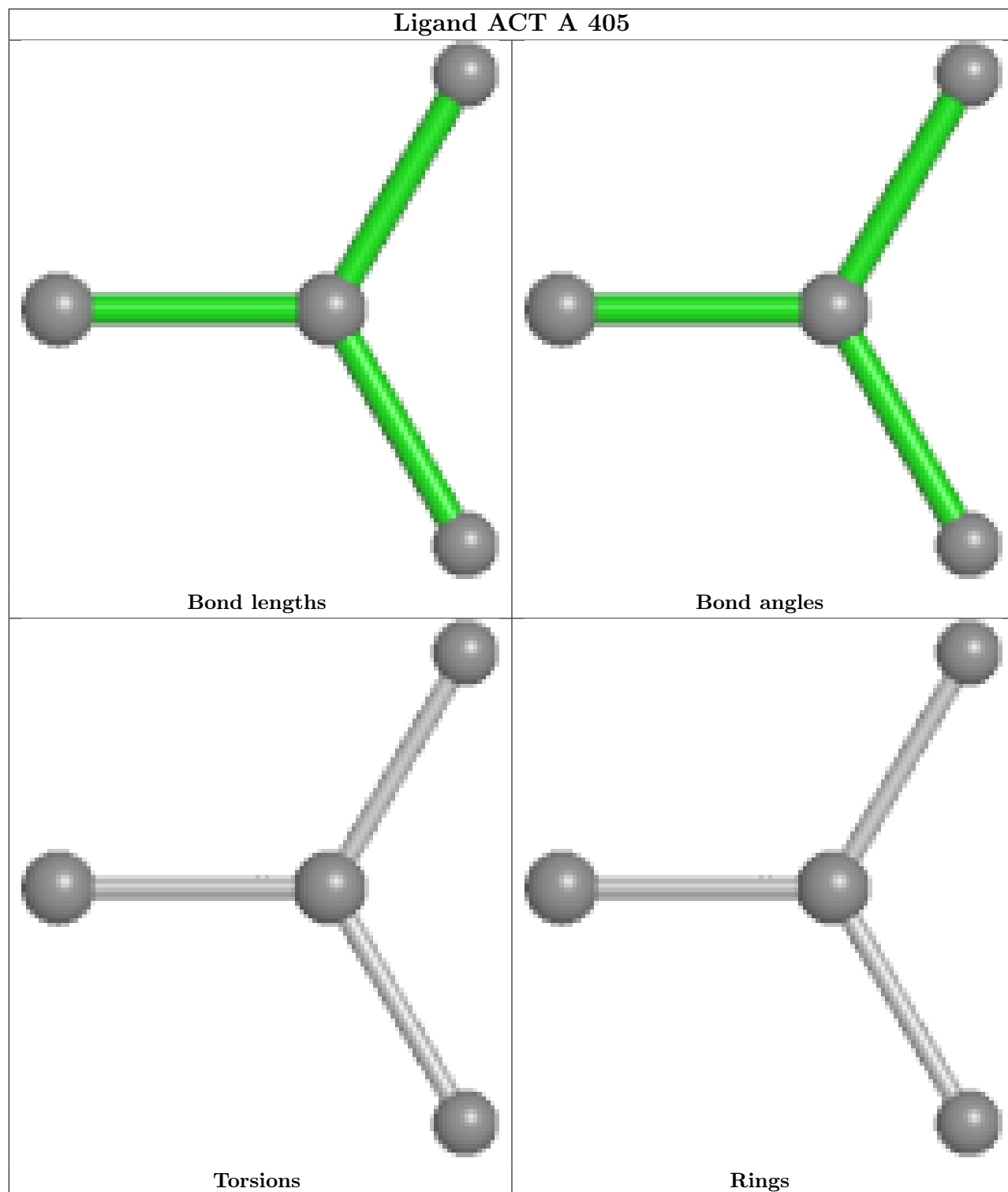
There are no ring outliers.

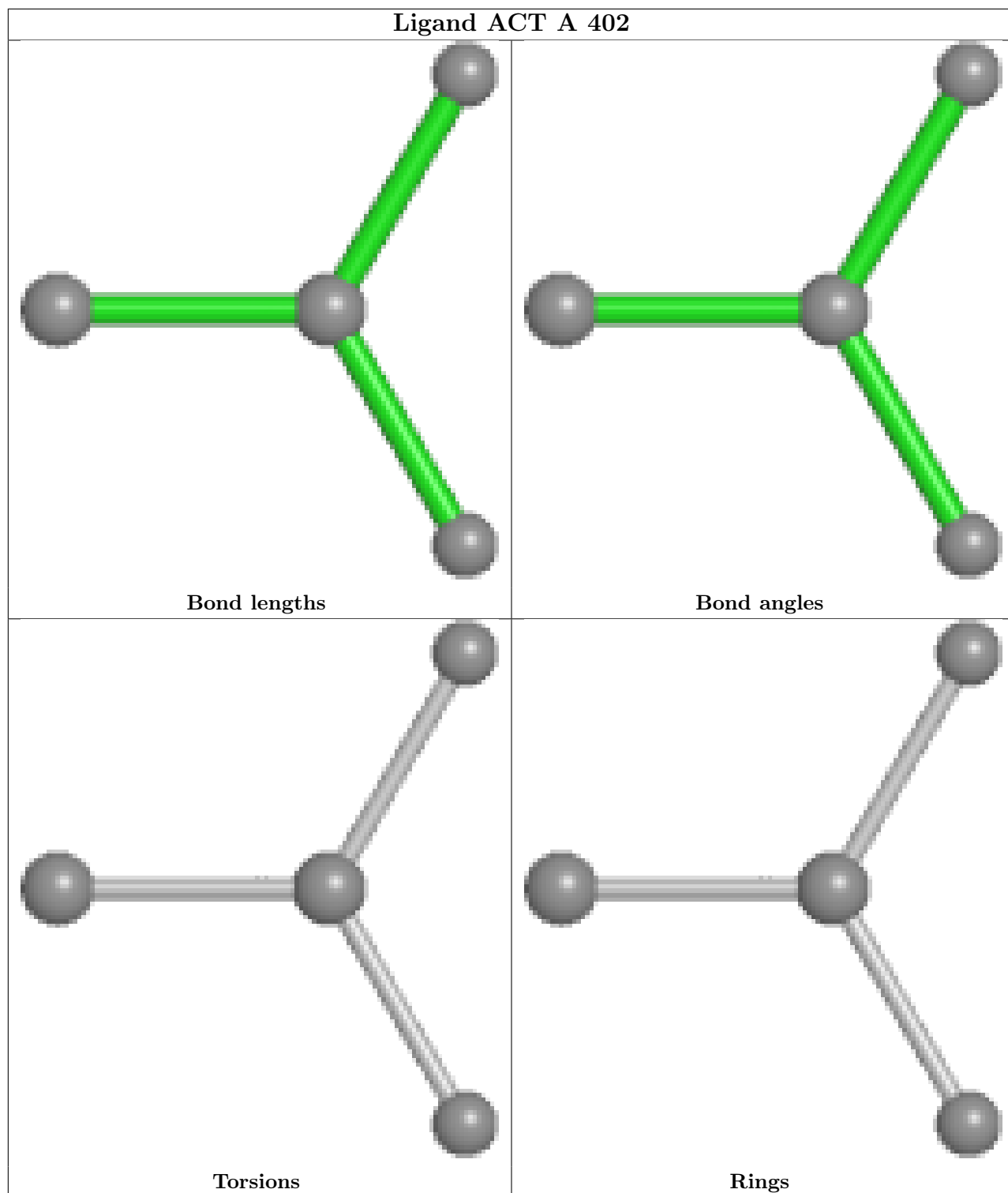
No monomer is involved in short contacts.

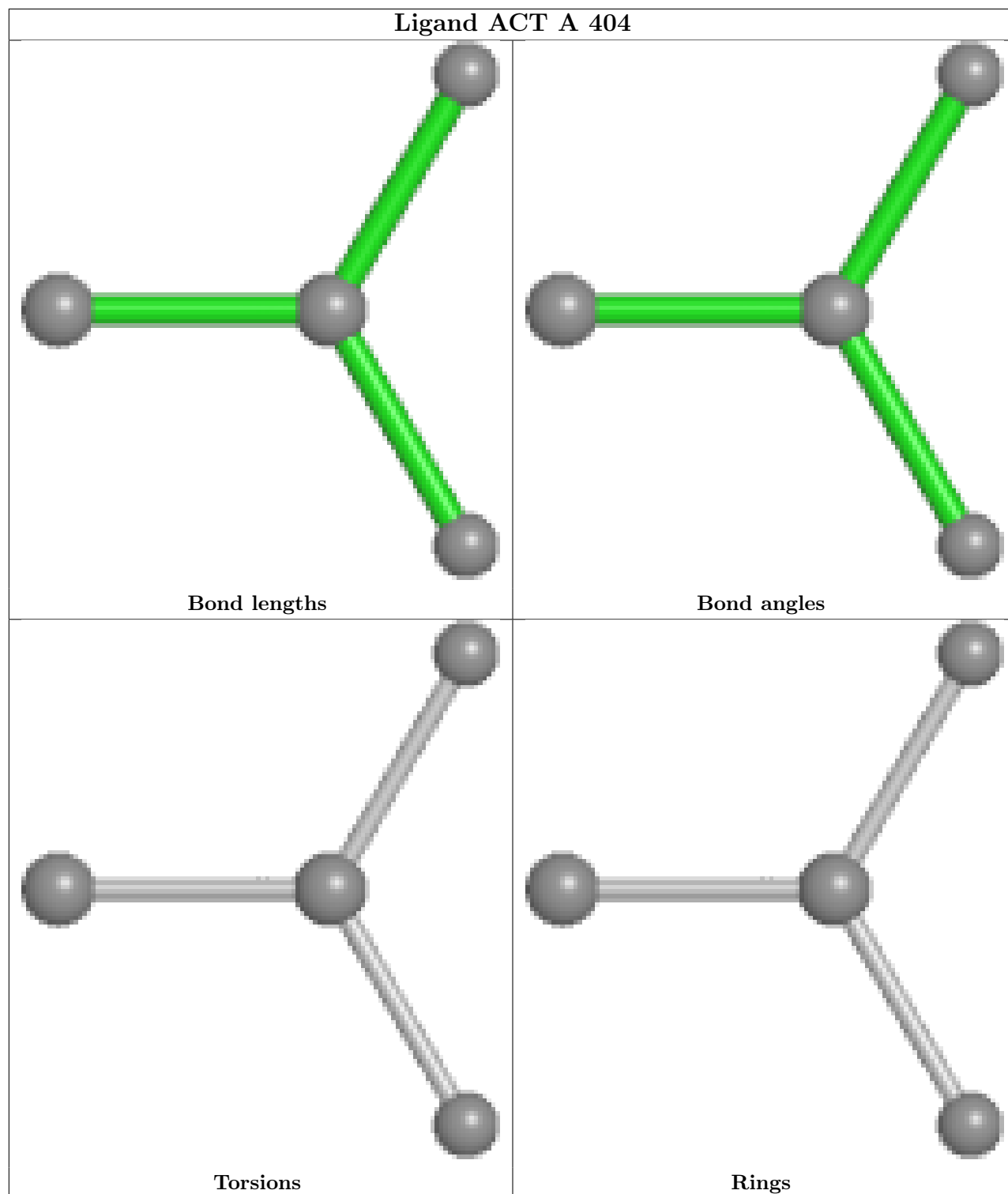
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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











5.7 Other polymers [\(i\)](#)

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	259/335 (77%)	0.67	25 (9%) 7 5	66, 92, 158, 160	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	114	LEU	6.0
1	A	104	ASP	4.7
1	A	83	VAL	4.7
1	A	106	ASP	4.4
1	A	39	GLN	4.3
1	A	128	MET	4.1
1	A	153	LYS	3.5
1	A	40	ASN	3.3
1	A	149	ASP	3.3
1	A	84	ILE	3.3
1	A	105	PHE	3.3
1	A	113	THR	3.2
1	A	112	VAL	3.0
1	A	108	LYS	2.8
1	A	66	ILE	2.7
1	A	150	LEU	2.7
1	A	130	ILE	2.6
1	A	152	GLY	2.4
1	A	44	VAL	2.4
1	A	126	GLY	2.3
1	A	42	GLY	2.3
1	A	329	ASP	2.2
1	A	107	VAL	2.2
1	A	68	TYR	2.1
1	A	151	ASN	2.1

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](#)

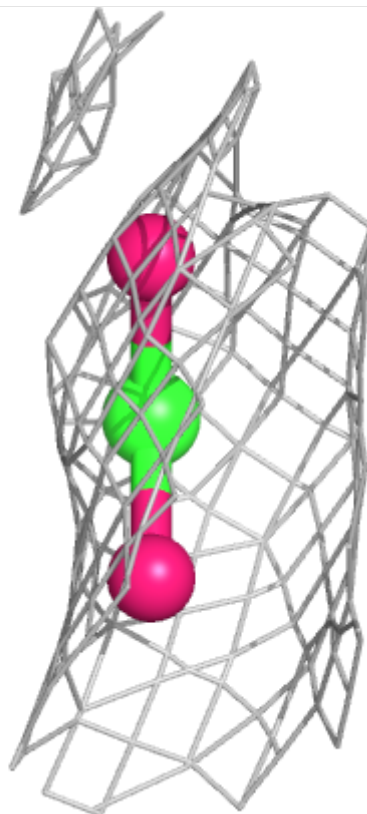
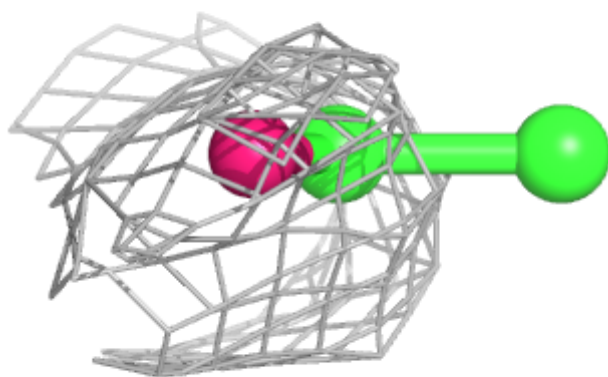
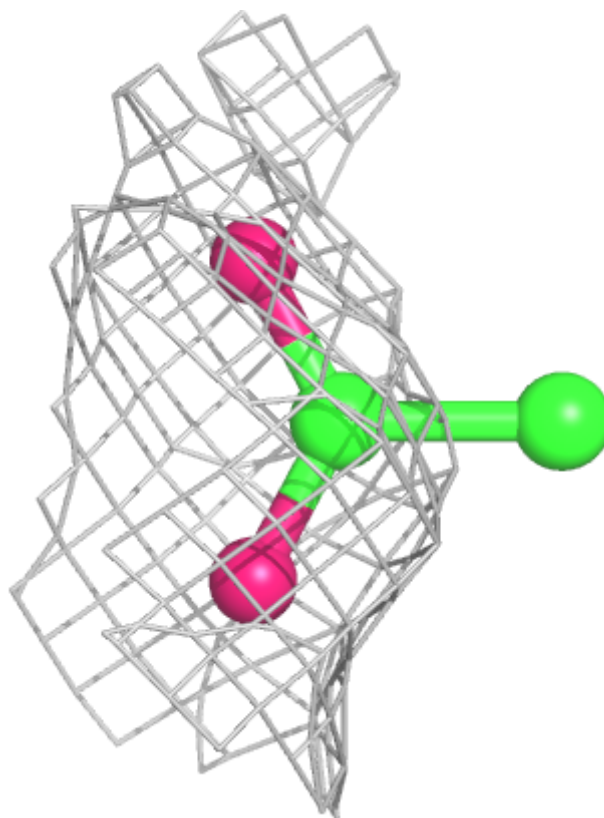
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists 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	ACT	A	404	4/4	0.49	0.40	135,136,139,142	0
2	ACT	A	403	4/4	0.76	0.33	102,108,112,113	0
2	ACT	A	401	4/4	0.83	0.30	92,103,107,114	0
2	ACT	A	405	4/4	0.90	0.20	100,101,102,102	0
2	ACT	A	402	4/4	0.91	0.19	96,107,110,115	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.

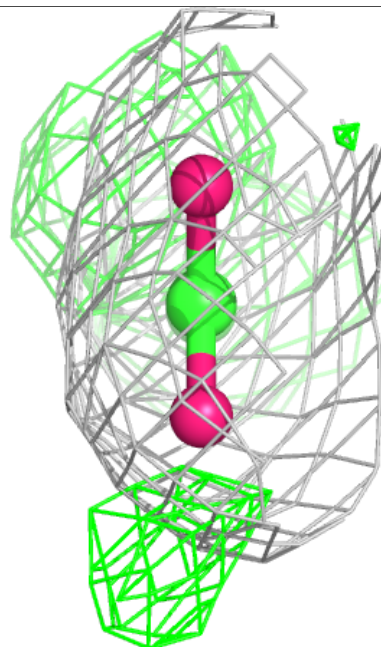
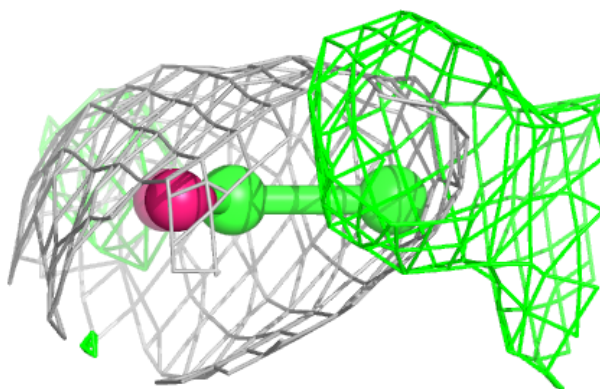
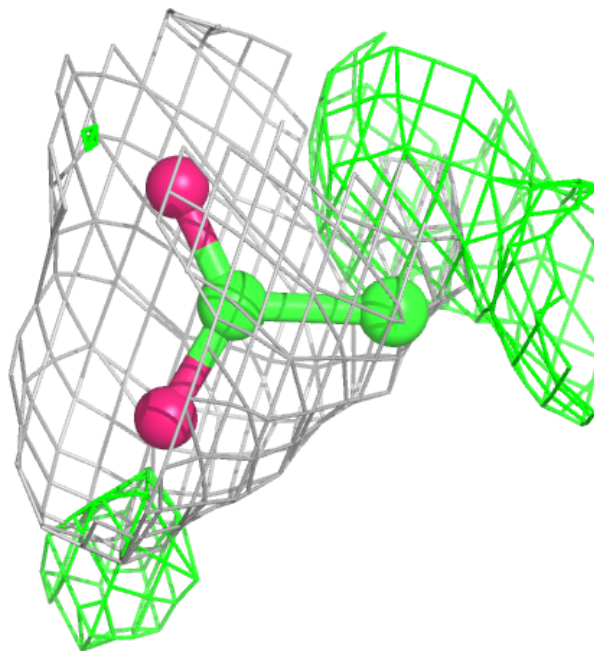
Electron density around ACT A 404:

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



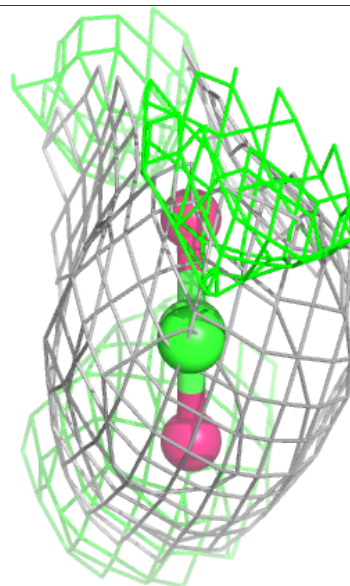
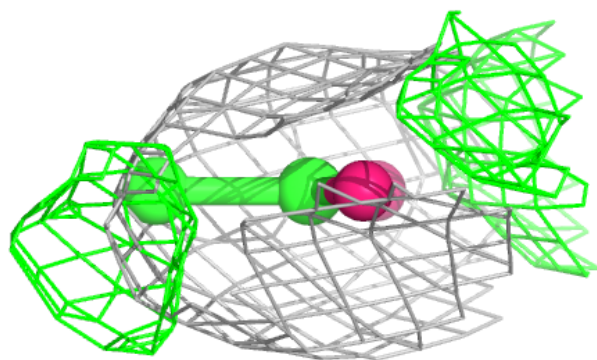
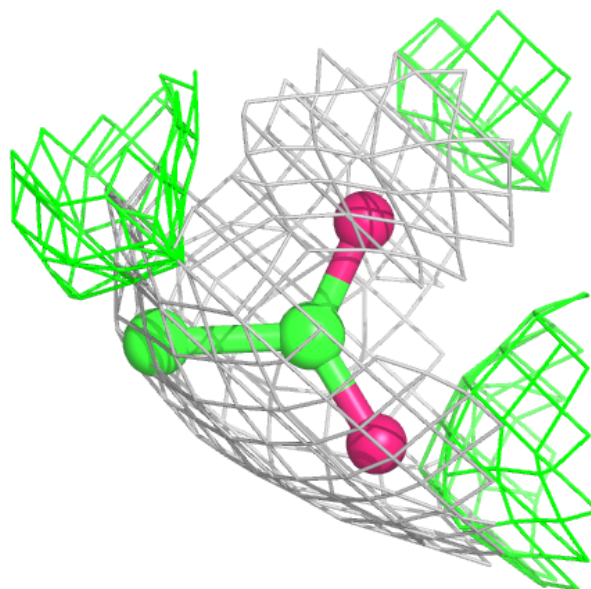
Electron density around ACT A 403:

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



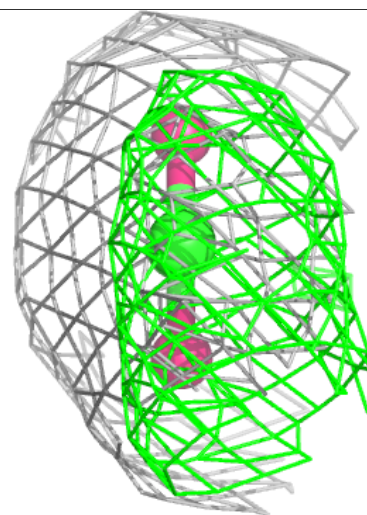
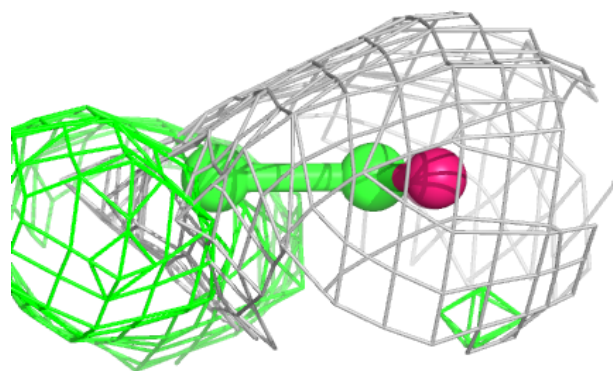
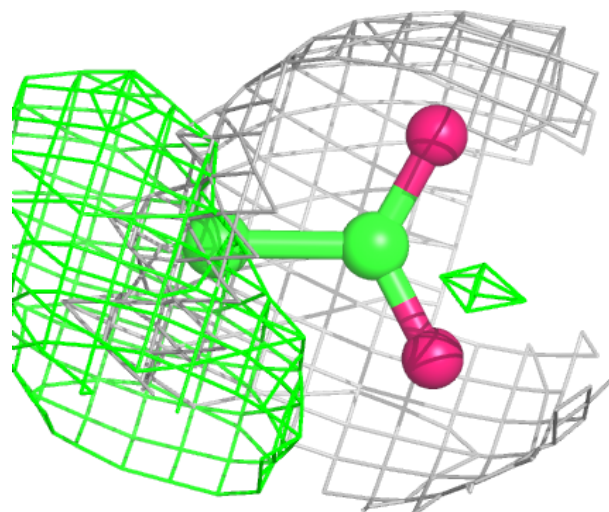
Electron density around ACT A 401:

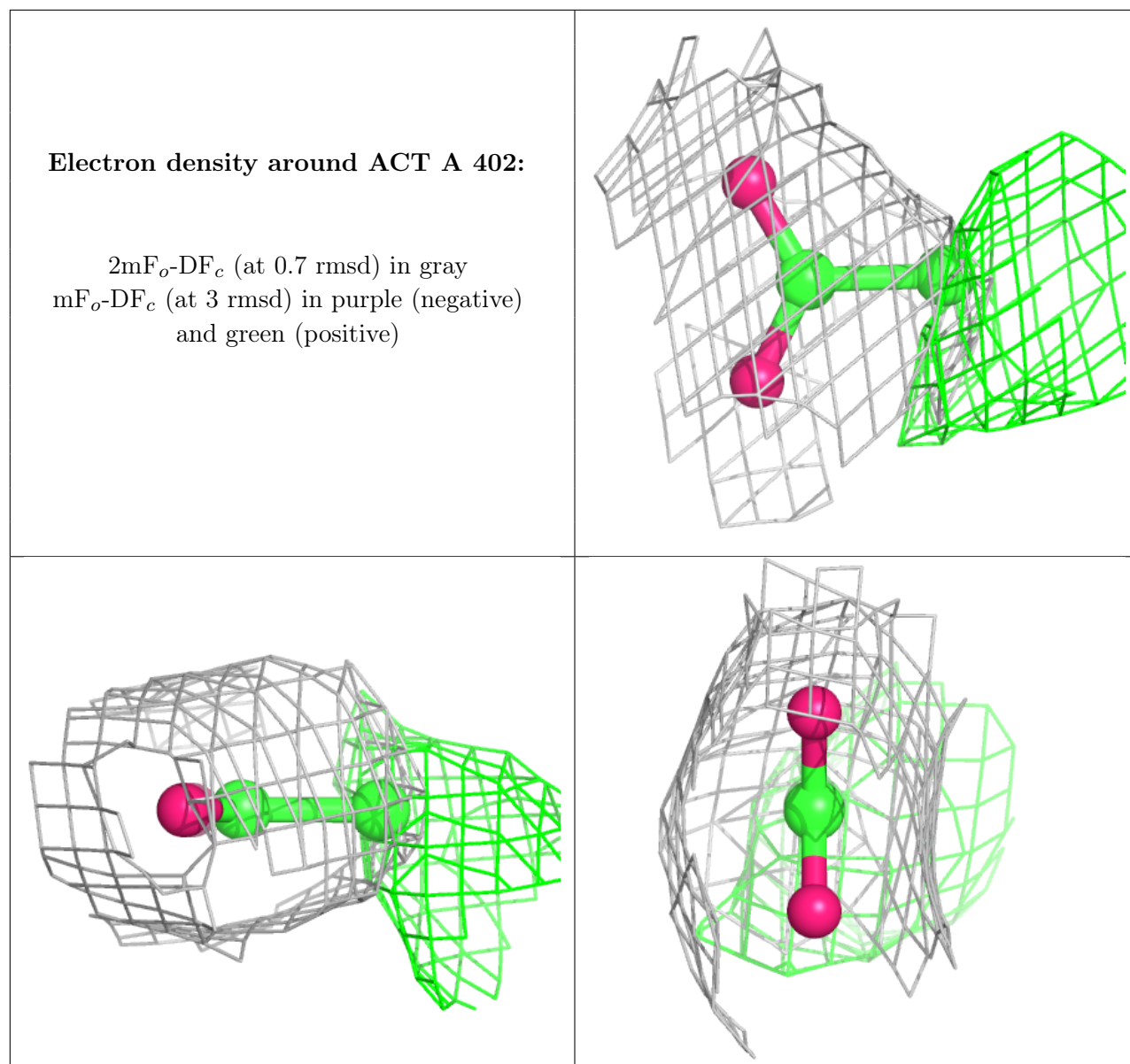
$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



Electron density around ACT A 405:

$2mF_o-DF_c$ (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.