



# wwPDB EM Model Validation Summary Report ⓘ

Mar 17, 2020 – 01:36 PM EDT

PDB ID : 6TSK  
EMDB ID : EMD-10564  
Title : Beta-galactosidase in complex with L-ribose  
Authors : Saur, M.; Hartshorn, M.J.; Dong, J.; Reeks, J.; Bunkoczi, G.; Jhoti, H.; Williams, P.A.  
Deposited on : 2019-12-20  
Resolution : 2.30 Å(reported)

This is a wwPDB EM Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.0 (224370), CSD as540be (2019)  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.8

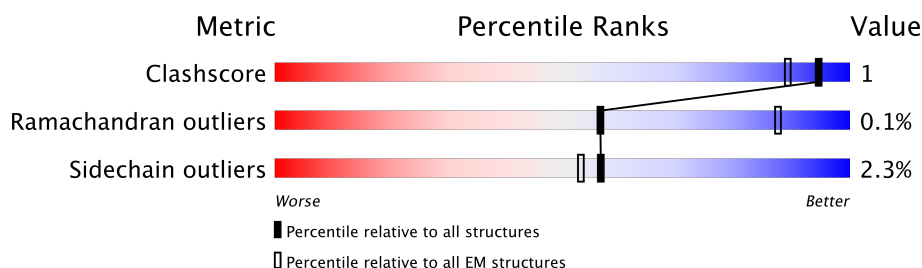
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 2.30 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	1040	92% 5% .
1	B	1040	92% 5% .
1	C	1040	92% 5% .
1	D	1040	92% 5% .

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 33632 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Beta-galactosidase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1015	Total	C	N	O	S	0	0
			8156	5159	1445	1514	38		
1	B	1015	Total	C	N	O	S	0	0
			8156	5159	1445	1514	38		
1	C	1015	Total	C	N	O	S	0	0
			8156	5159	1445	1514	38		
1	D	1015	Total	C	N	O	S	0	0
			8156	5159	1445	1514	38		

There are 100 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-16	MET	-	initiating methionine	UNP Q8VNN2
A	-15	GLY	-	expression tag	UNP Q8VNN2
A	-14	SER	-	expression tag	UNP Q8VNN2
A	-13	SER	-	expression tag	UNP Q8VNN2
A	-12	HIS	-	expression tag	UNP Q8VNN2
A	-11	HIS	-	expression tag	UNP Q8VNN2
A	-10	HIS	-	expression tag	UNP Q8VNN2
A	-9	HIS	-	expression tag	UNP Q8VNN2
A	-8	HIS	-	expression tag	UNP Q8VNN2
A	-7	HIS	-	expression tag	UNP Q8VNN2
A	-6	SER	-	expression tag	UNP Q8VNN2
A	-5	SER	-	expression tag	UNP Q8VNN2
A	-4	GLY	-	expression tag	UNP Q8VNN2
A	-3	LEU	-	expression tag	UNP Q8VNN2
A	-2	VAL	-	expression tag	UNP Q8VNN2
A	-1	PRO	-	expression tag	UNP Q8VNN2
A	0	ARG	-	expression tag	UNP Q8VNN2
A	1	GLY	-	expression tag	UNP Q8VNN2
A	2	SER	-	expression tag	UNP Q8VNN2
A	3	HIS	-	expression tag	UNP Q8VNN2
A	4	MET	-	expression tag	UNP Q8VNN2
A	5	LEU	-	expression tag	UNP Q8VNN2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	GLU	-	expression tag	UNP Q8VNN2
A	7	ASP	-	expression tag	UNP Q8VNN2
A	8	PRO	-	expression tag	UNP Q8VNN2
B	-16	MET	-	initiating methionine	UNP Q8VNN2
B	-15	GLY	-	expression tag	UNP Q8VNN2
B	-14	SER	-	expression tag	UNP Q8VNN2
B	-13	SER	-	expression tag	UNP Q8VNN2
B	-12	HIS	-	expression tag	UNP Q8VNN2
B	-11	HIS	-	expression tag	UNP Q8VNN2
B	-10	HIS	-	expression tag	UNP Q8VNN2
B	-9	HIS	-	expression tag	UNP Q8VNN2
B	-8	HIS	-	expression tag	UNP Q8VNN2
B	-7	HIS	-	expression tag	UNP Q8VNN2
B	-6	SER	-	expression tag	UNP Q8VNN2
B	-5	SER	-	expression tag	UNP Q8VNN2
B	-4	GLY	-	expression tag	UNP Q8VNN2
B	-3	LEU	-	expression tag	UNP Q8VNN2
B	-2	VAL	-	expression tag	UNP Q8VNN2
B	-1	PRO	-	expression tag	UNP Q8VNN2
B	0	ARG	-	expression tag	UNP Q8VNN2
B	1	GLY	-	expression tag	UNP Q8VNN2
B	2	SER	-	expression tag	UNP Q8VNN2
B	3	HIS	-	expression tag	UNP Q8VNN2
B	4	MET	-	expression tag	UNP Q8VNN2
B	5	LEU	-	expression tag	UNP Q8VNN2
B	6	GLU	-	expression tag	UNP Q8VNN2
B	7	ASP	-	expression tag	UNP Q8VNN2
B	8	PRO	-	expression tag	UNP Q8VNN2
C	-16	MET	-	initiating methionine	UNP Q8VNN2
C	-15	GLY	-	expression tag	UNP Q8VNN2
C	-14	SER	-	expression tag	UNP Q8VNN2
C	-13	SER	-	expression tag	UNP Q8VNN2
C	-12	HIS	-	expression tag	UNP Q8VNN2
C	-11	HIS	-	expression tag	UNP Q8VNN2
C	-10	HIS	-	expression tag	UNP Q8VNN2
C	-9	HIS	-	expression tag	UNP Q8VNN2
C	-8	HIS	-	expression tag	UNP Q8VNN2
C	-7	HIS	-	expression tag	UNP Q8VNN2
C	-6	SER	-	expression tag	UNP Q8VNN2
C	-5	SER	-	expression tag	UNP Q8VNN2
C	-4	GLY	-	expression tag	UNP Q8VNN2
C	-3	LEU	-	expression tag	UNP Q8VNN2

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	VAL	-	expression tag	UNP Q8VNN2
C	-1	PRO	-	expression tag	UNP Q8VNN2
C	0	ARG	-	expression tag	UNP Q8VNN2
C	1	GLY	-	expression tag	UNP Q8VNN2
C	2	SER	-	expression tag	UNP Q8VNN2
C	3	HIS	-	expression tag	UNP Q8VNN2
C	4	MET	-	expression tag	UNP Q8VNN2
C	5	LEU	-	expression tag	UNP Q8VNN2
C	6	GLU	-	expression tag	UNP Q8VNN2
C	7	ASP	-	expression tag	UNP Q8VNN2
C	8	PRO	-	expression tag	UNP Q8VNN2
D	-16	MET	-	initiating methionine	UNP Q8VNN2
D	-15	GLY	-	expression tag	UNP Q8VNN2
D	-14	SER	-	expression tag	UNP Q8VNN2
D	-13	SER	-	expression tag	UNP Q8VNN2
D	-12	HIS	-	expression tag	UNP Q8VNN2
D	-11	HIS	-	expression tag	UNP Q8VNN2
D	-10	HIS	-	expression tag	UNP Q8VNN2
D	-9	HIS	-	expression tag	UNP Q8VNN2
D	-8	HIS	-	expression tag	UNP Q8VNN2
D	-7	HIS	-	expression tag	UNP Q8VNN2
D	-6	SER	-	expression tag	UNP Q8VNN2
D	-5	SER	-	expression tag	UNP Q8VNN2
D	-4	GLY	-	expression tag	UNP Q8VNN2
D	-3	LEU	-	expression tag	UNP Q8VNN2
D	-2	VAL	-	expression tag	UNP Q8VNN2
D	-1	PRO	-	expression tag	UNP Q8VNN2
D	0	ARG	-	expression tag	UNP Q8VNN2
D	1	GLY	-	expression tag	UNP Q8VNN2
D	2	SER	-	expression tag	UNP Q8VNN2
D	3	HIS	-	expression tag	UNP Q8VNN2
D	4	MET	-	expression tag	UNP Q8VNN2
D	5	LEU	-	expression tag	UNP Q8VNN2
D	6	GLU	-	expression tag	UNP Q8VNN2
D	7	ASP	-	expression tag	UNP Q8VNN2
D	8	PRO	-	expression tag	UNP Q8VNN2

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

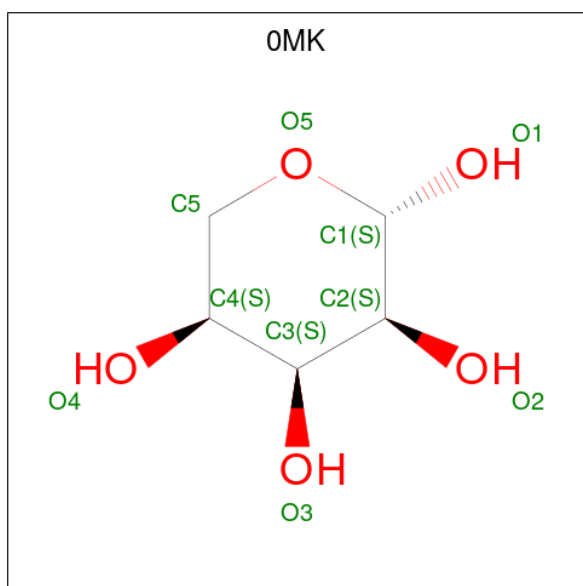
Mol	Chain	Residues	Atoms	AltConf
2	B	1	Total Mg 1 1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		AltConf
2	A	1	Total	Mg	0
			1	1	
2	D	1	Total	Mg	0
			1	1	
2	C	1	Total	Mg	0
			1	1	

- Molecule 3 is L-ribofuranose (three-letter code: 0MK) (formula: C<sub>5</sub>H<sub>10</sub>O<sub>5</sub>) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms			AltConf
3	A	1	Total	C	O	0
			10	5	5	
3	B	1	Total	C	O	0
			10	5	5	
3	C	1	Total	C	O	0
			10	5	5	
3	D	1	Total	C	O	0
			10	5	5	

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		AltConf
4	A	241	Total	O	0
			241	241	
4	B	241	Total	O	0
			241	241	

*Continued on next page...*

*Continued from previous page...*

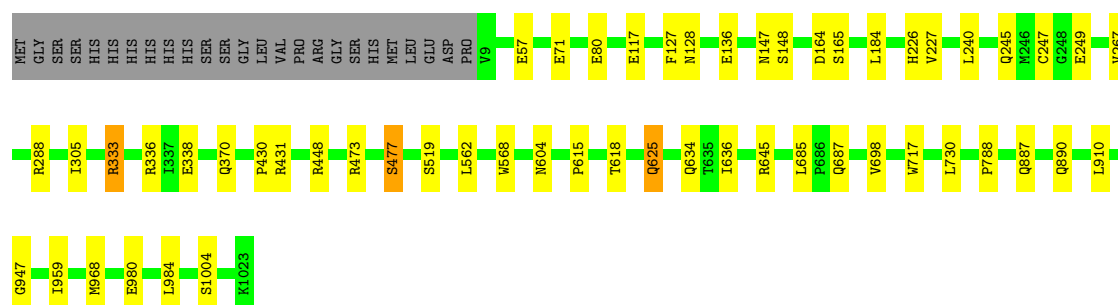
Mol	Chain	Residues	Atoms		AltConf
4	C	241	Total 241	O 241	0
4	D	241	Total 241	O 241	0

### 3 Residue-property plots [i](#)

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. 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. 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.

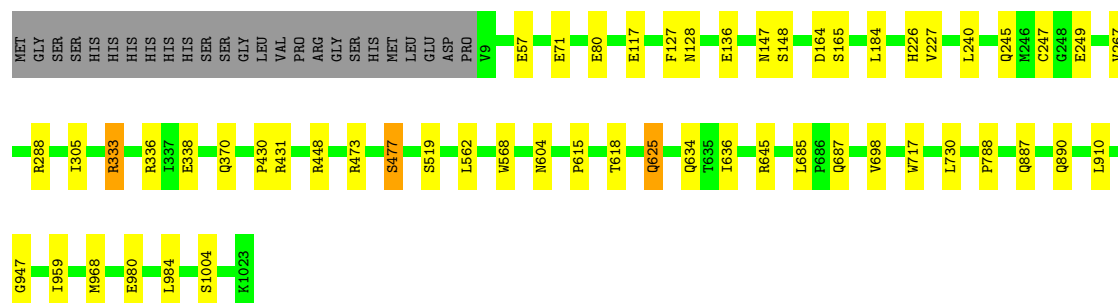
- Molecule 1: Beta-galactosidase

Chain A: 



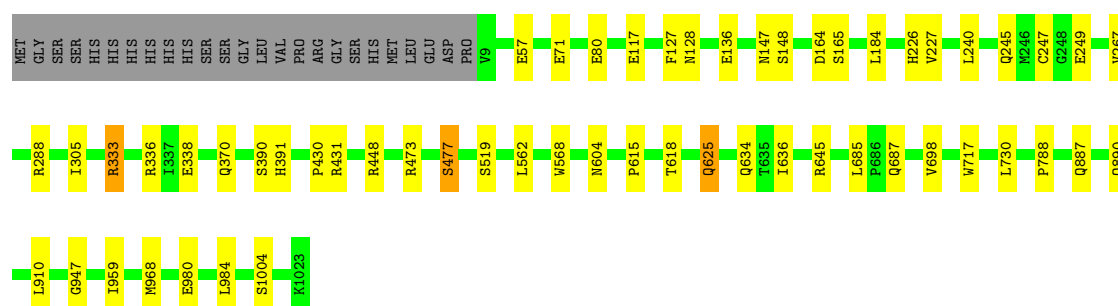
- Molecule 1: Beta-galactosidase

Chain B: 



- Molecule 1: Beta-galactosidase

Chain C: 



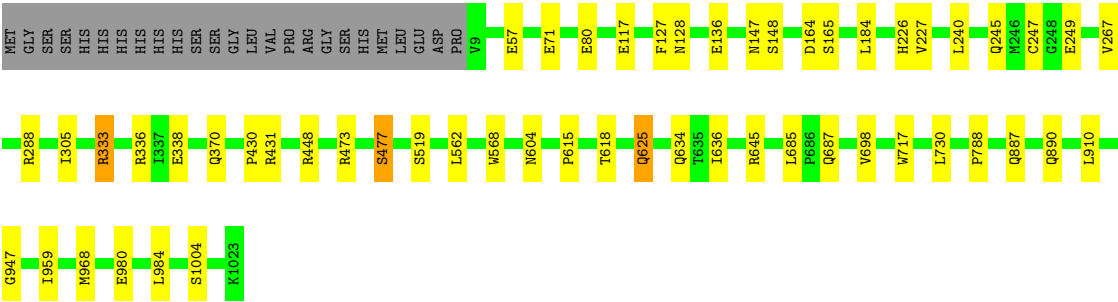


● Molecule 1: Beta-galactosidase

Chain D: 

92%

5%



## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, D2	Depositor
Number of particles used	40086	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	66.83	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	FEI FALCON III (4k x 4k)	Depositor

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, 0MK

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	0.53	0/8398	0.74	1/11458 (0.0%)
1	B	0.53	0/8398	0.74	1/11458 (0.0%)
1	C	0.53	0/8398	0.74	1/11458 (0.0%)
1	D	0.53	0/8398	0.74	1/11458 (0.0%)
All	All	0.53	0/33592	0.74	4/45832 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	4
1	B	0	4
1	C	0	4
1	D	0	4
All	All	0	16

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	625	GLN	CB-CA-C	-7.75	94.90	110.40
1	B	625	GLN	CB-CA-C	-7.75	94.90	110.40
1	C	625	GLN	CB-CA-C	-7.75	94.90	110.40
1	D	625	GLN	CB-CA-C	-7.75	94.90	110.40

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	333	ARG	Sidechain
1	A	336	ARG	Sidechain
1	A	431	ARG	Sidechain
1	A	448	ARG	Sidechain
1	B	333	ARG	Sidechain

## 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	8156	0	7753	19	0
1	B	8156	0	7753	19	0
1	C	8156	0	7753	20	0
1	D	8156	0	7753	19	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	10	0	10	0	0
3	B	10	0	10	0	0
3	C	10	0	10	0	0
3	D	10	0	10	0	0
4	A	241	0	0	2	0
4	B	241	0	0	2	0
4	C	241	0	0	2	0
4	D	241	0	0	2	0
All	All	33632	0	31052	77	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 1.

The worst 5 of 77 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:C:890:GLN:HE22	1:C:947:GLY:HA3	1.45	0.82
1:A:890:GLN:HE22	1:A:947:GLY:HA3	1.45	0.82
1:B:890:GLN:HE22	1:B:947:GLY:HA3	1.45	0.80

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:890:GLN:HE22	1:D:947:GLY:HA3	1.45	0.79
1:D:227:VAL:HG13	1:D:240:LEU:HD11	1.82	0.61

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 PDB entries followed by that with respect to all EM entries.

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	1013/1040 (97%)	983 (97%)	29 (3%)	1 (0%)	53	66
1	B	1013/1040 (97%)	983 (97%)	29 (3%)	1 (0%)	53	66
1	C	1013/1040 (97%)	982 (97%)	30 (3%)	1 (0%)	53	66
1	D	1013/1040 (97%)	983 (97%)	29 (3%)	1 (0%)	53	66
All	All	4052/4160 (97%)	3931 (97%)	117 (3%)	4 (0%)	56	66

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	164	ASP
1	B	164	ASP
1	C	164	ASP
1	D	164	ASP

### 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 PDB entries followed by that with respect to all EM entries.

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	868/890 (98%)	848 (98%)	20 (2%)	53	71
1	B	868/890 (98%)	848 (98%)	20 (2%)	53	71
1	C	868/890 (98%)	848 (98%)	20 (2%)	53	71
1	D	868/890 (98%)	848 (98%)	20 (2%)	53	71
All	All	3472/3560 (98%)	3392 (98%)	80 (2%)	57	71

5 of 80 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	687	GLN
1	C	128	ASN
1	D	370	GLN
1	B	730	LEU
1	C	57	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 40 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	890	GLN
1	C	365	GLN
1	D	702	GLN
1	C	266	GLN
1	C	604	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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 for Mogul analysis.

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	$\# Z  > 2$
3	0MK	A	3002	-	10,10,10	0.86	0	14,14,14	0.69	0
3	0MK	B	3002	-	10,10,10	0.86	0	14,14,14	0.69	0
3	0MK	C	3002	-	10,10,10	0.86	0	14,14,14	0.68	0
3	0MK	D	3002	-	10,10,10	0.86	0	14,14,14	0.69	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
3	0MK	A	3002	-	-	-	0/1/1/1
3	0MK	B	3002	-	-	-	0/1/1/1
3	0MK	C	3002	-	-	-	0/1/1/1
3	0MK	D	3002	-	-	-	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

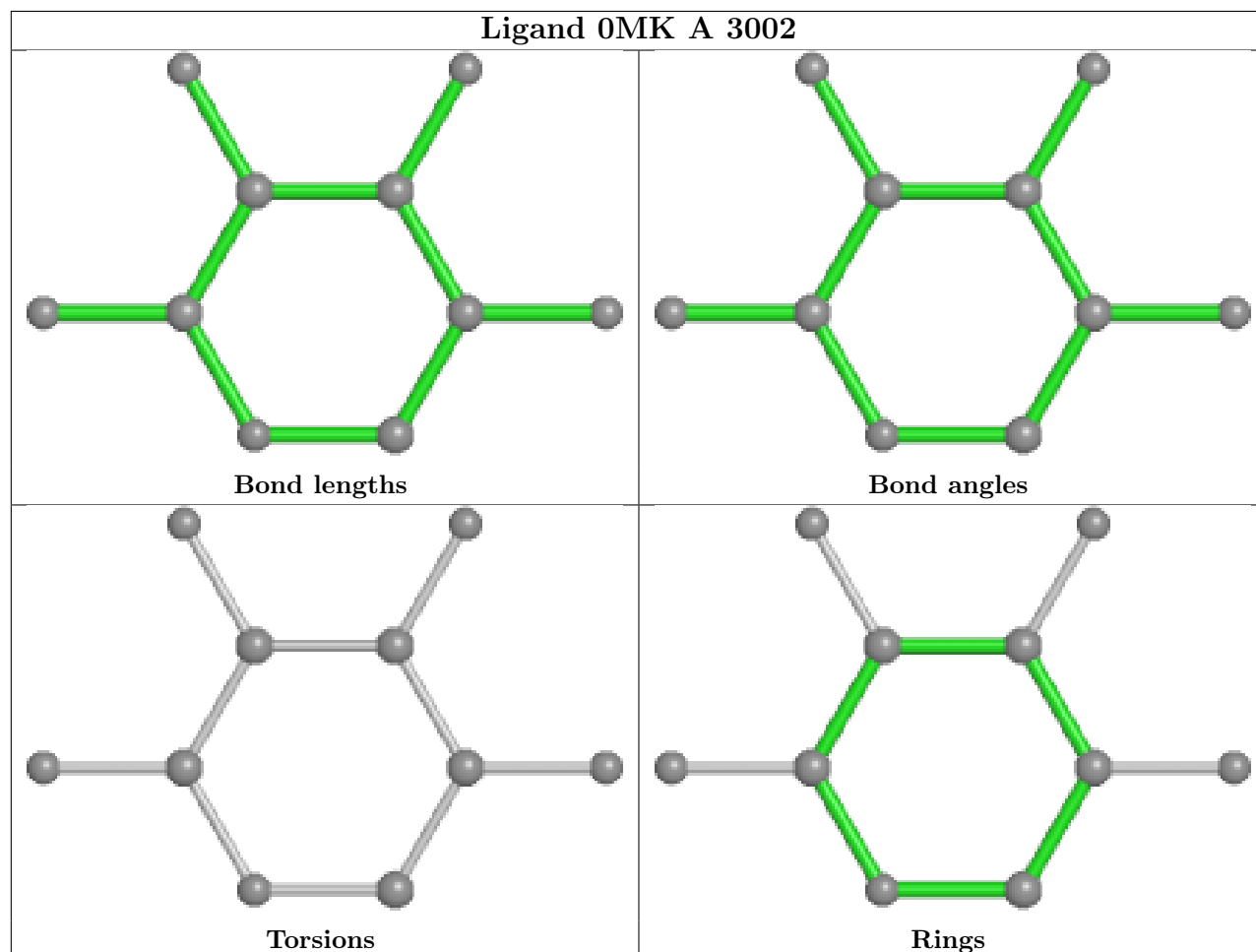
There are no torsion outliers.

There are no ring outliers.

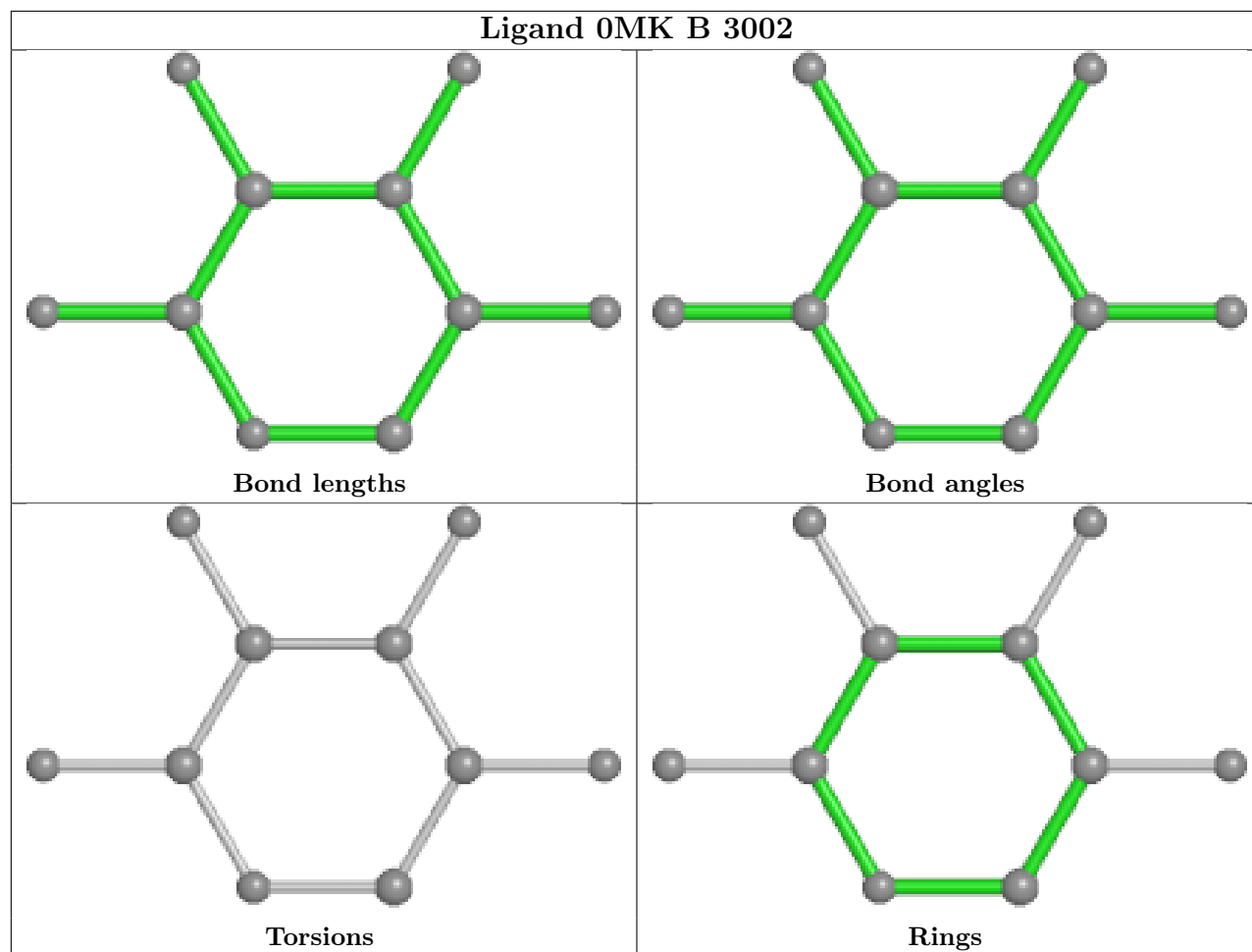
No monomer is involved in short contacts.

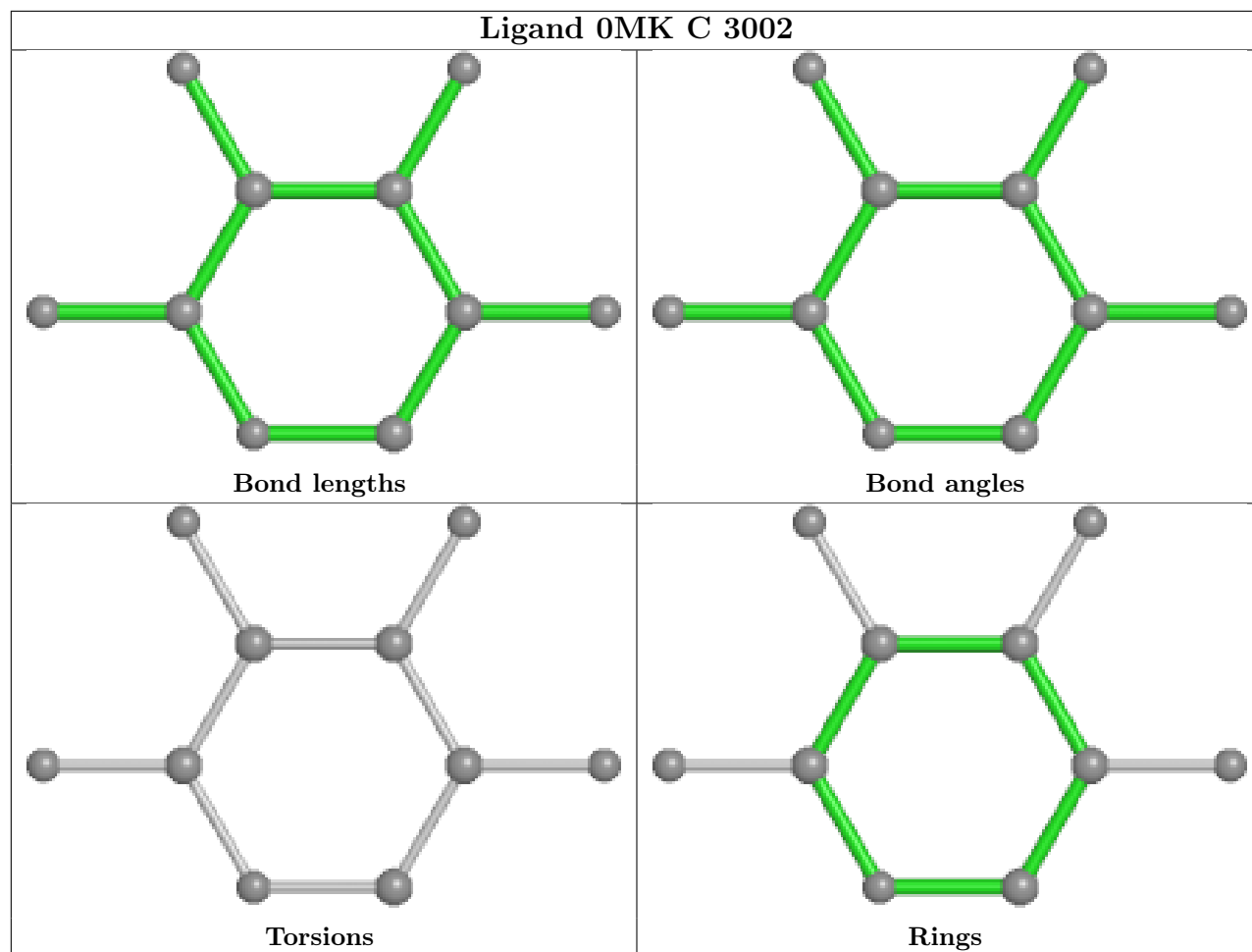
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight  $> 250$  and outliers as shown on the validation Tables will also be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

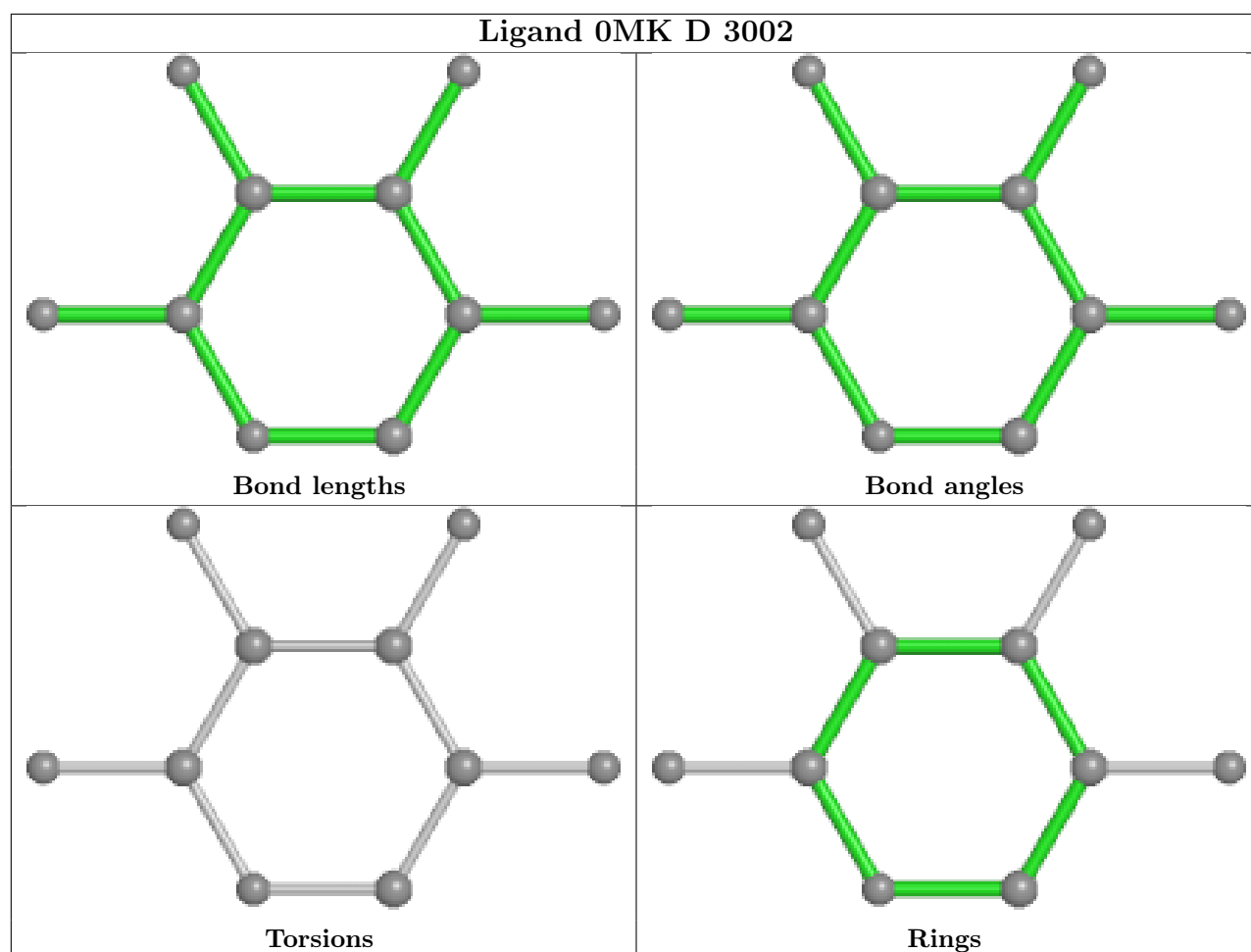
in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers 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 [i](#)

There are no chain breaks in this entry.