wwPDB X-ray Structure Validation Summary Report (i)

Oct 10, 2023 – 11:09 AM EDT

PDB ID	:	5EC5
Title	:	Crystal structure of lysenin pore
Authors	:	Podobnik, M.; Savory, P.; Rojko, N.; Kisovec, M.; Bruce, M.; Jayasinghe, L.; Anderluh, G.
Deposited on	:	2015-10-20
Resolution	:	3.10 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i) 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 (1)) were used in the production of this report:

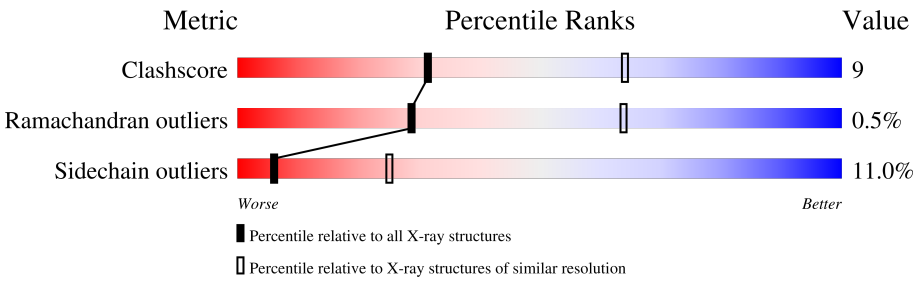
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.10 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))
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)

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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	298	76% 18% • •
1	B	298	77% 17% • 5%
1	C	298	72% 21% • •
1	D	298	68% 23% • •
1	E	298	74% 19% • 5%
1	F	298	74% 20% • 5%
1	G	298	74% 18% • • •
1	H	298	70% 22% • 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	I	298	73% 19% 5%
1	J	298	75% 20%
1	K	298	74% 20%
1	L	298	74% 18% 5%
1	M	298	72% 22%
1	N	298	76% 18%
1	O	298	74% 21%
1	P	298	72% 24%
1	R	298	74% 18% 5%
1	S	298	73% 20% 5%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	MBO	A	302	-	-	X	-
2	MBO	B	301	-	-	X	-
2	MBO	C	301	-	-	X	-
2	MBO	D	301	-	-	X	-
2	MBO	E	301	-	-	X	-
2	MBO	F	302	-	-	X	-
2	MBO	G	301	-	-	X	-
2	MBO	H	301	-	-	X	-
2	MBO	H	302	-	-	X	-
2	MBO	I	302	-	-	X	-
2	MBO	K	301	-	-	X	-
2	MBO	K	302	-	-	X	-
2	MBO	L	301	-	-	X	-
2	MBO	L	302	-	-	X	-
2	MBO	M	301	-	-	X	-
2	MBO	O	301	-	-	X	-
2	MBO	P	302	-	-	X	-
2	MBO	R	301	-	-	X	-
2	MBO	S	301	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 41248 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 Lysenin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	290	2301	1459	390	444	8	0	0	0
1	B	284	2260	1432	384	436	8	0	0	0
1	C	286	2268	1436	386	438	8	0	0	0
1	D	285	2265	1435	385	437	8	0	0	0
1	E	283	2252	1428	383	433	8	0	0	0
1	F	283	2252	1428	383	433	8	0	0	0
1	G	290	2301	1459	390	444	8	0	0	0
1	H	283	2252	1428	383	433	8	0	0	0
1	I	282	2248	1426	382	432	8	0	0	0
1	J	290	2301	1459	390	444	8	0	0	0
1	K	290	2301	1459	390	444	8	0	0	0
1	L	284	2260	1434	384	434	8	0	0	0
1	M	289	2297	1457	389	443	8	0	0	0
1	N	289	2297	1457	389	443	8	0	0	0
1	O	289	2297	1457	389	443	8	0	0	0
1	P	289	2297	1457	389	443	8	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	R	283	2252	1428	383	433	8	0	0	0
			Total	C	N	O	S			
1	S	283	2252	1428	383	433	8	0	0	0

There are 108 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP O18423
A	84	GLN	GLU	engineered mutation	UNP O18423
A	85	LYS	GLU	engineered mutation	UNP O18423
A	92	GLN	GLU	engineered mutation	UNP O18423
A	97	SER	GLU	engineered mutation	UNP O18423
A	126	GLY	ASP	engineered mutation	UNP O18423
B	0	GLY	-	expression tag	UNP O18423
B	84	GLN	GLU	engineered mutation	UNP O18423
B	85	LYS	GLU	engineered mutation	UNP O18423
B	92	GLN	GLU	engineered mutation	UNP O18423
B	97	SER	GLU	engineered mutation	UNP O18423
B	126	GLY	ASP	engineered mutation	UNP O18423
C	0	GLY	-	expression tag	UNP O18423
C	84	GLN	GLU	engineered mutation	UNP O18423
C	85	LYS	GLU	engineered mutation	UNP O18423
C	92	GLN	GLU	engineered mutation	UNP O18423
C	97	SER	GLU	engineered mutation	UNP O18423
C	126	GLY	ASP	engineered mutation	UNP O18423
D	0	GLY	-	expression tag	UNP O18423
D	84	GLN	GLU	engineered mutation	UNP O18423
D	85	LYS	GLU	engineered mutation	UNP O18423
D	92	GLN	GLU	engineered mutation	UNP O18423
D	97	SER	GLU	engineered mutation	UNP O18423
D	126	GLY	ASP	engineered mutation	UNP O18423
E	0	GLY	-	expression tag	UNP O18423
E	84	GLN	GLU	engineered mutation	UNP O18423
E	85	LYS	GLU	engineered mutation	UNP O18423
E	92	GLN	GLU	engineered mutation	UNP O18423
E	97	SER	GLU	engineered mutation	UNP O18423
E	126	GLY	ASP	engineered mutation	UNP O18423
F	0	GLY	-	expression tag	UNP O18423
F	84	GLN	GLU	engineered mutation	UNP O18423
F	85	LYS	GLU	engineered mutation	UNP O18423
F	92	GLN	GLU	engineered mutation	UNP O18423

Continued on next page...

Continued from previous page...

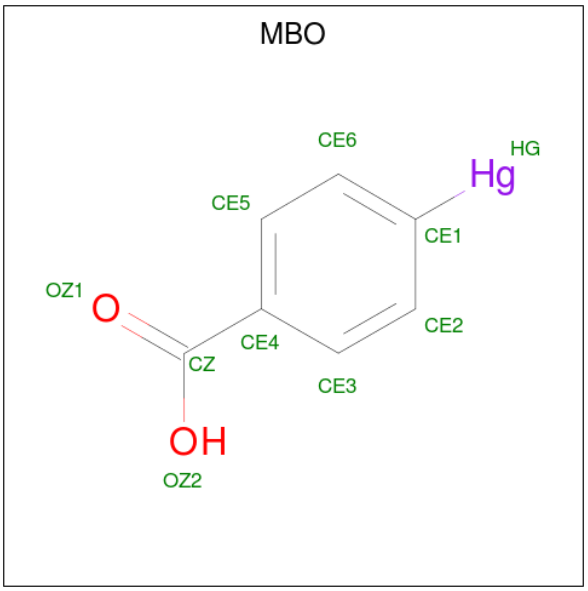
Chain	Residue	Modelled	Actual	Comment	Reference
F	97	SER	GLU	engineered mutation	UNP O18423
F	126	GLY	ASP	engineered mutation	UNP O18423
G	0	GLY	-	expression tag	UNP O18423
G	84	GLN	GLU	engineered mutation	UNP O18423
G	85	LYS	GLU	engineered mutation	UNP O18423
G	92	GLN	GLU	engineered mutation	UNP O18423
G	97	SER	GLU	engineered mutation	UNP O18423
G	126	GLY	ASP	engineered mutation	UNP O18423
H	0	GLY	-	expression tag	UNP O18423
H	84	GLN	GLU	engineered mutation	UNP O18423
H	85	LYS	GLU	engineered mutation	UNP O18423
H	92	GLN	GLU	engineered mutation	UNP O18423
H	97	SER	GLU	engineered mutation	UNP O18423
H	126	GLY	ASP	engineered mutation	UNP O18423
I	0	GLY	-	expression tag	UNP O18423
I	84	GLN	GLU	engineered mutation	UNP O18423
I	85	LYS	GLU	engineered mutation	UNP O18423
I	92	GLN	GLU	engineered mutation	UNP O18423
I	97	SER	GLU	engineered mutation	UNP O18423
I	126	GLY	ASP	engineered mutation	UNP O18423
J	0	GLY	-	expression tag	UNP O18423
J	84	GLN	GLU	engineered mutation	UNP O18423
J	85	LYS	GLU	engineered mutation	UNP O18423
J	92	GLN	GLU	engineered mutation	UNP O18423
J	97	SER	GLU	engineered mutation	UNP O18423
J	126	GLY	ASP	engineered mutation	UNP O18423
K	0	GLY	-	expression tag	UNP O18423
K	84	GLN	GLU	engineered mutation	UNP O18423
K	85	LYS	GLU	engineered mutation	UNP O18423
K	92	GLN	GLU	engineered mutation	UNP O18423
K	97	SER	GLU	engineered mutation	UNP O18423
K	126	GLY	ASP	engineered mutation	UNP O18423
L	0	GLY	-	expression tag	UNP O18423
L	84	GLN	GLU	engineered mutation	UNP O18423
L	85	LYS	GLU	engineered mutation	UNP O18423
L	92	GLN	GLU	engineered mutation	UNP O18423
L	97	SER	GLU	engineered mutation	UNP O18423
L	126	GLY	ASP	engineered mutation	UNP O18423
M	0	GLY	-	expression tag	UNP O18423
M	84	GLN	GLU	engineered mutation	UNP O18423
M	85	LYS	GLU	engineered mutation	UNP O18423
M	92	GLN	GLU	engineered mutation	UNP O18423

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	97	SER	GLU	engineered mutation	UNP O18423
M	126	GLY	ASP	engineered mutation	UNP O18423
N	0	GLY	-	expression tag	UNP O18423
N	84	GLN	GLU	engineered mutation	UNP O18423
N	85	LYS	GLU	engineered mutation	UNP O18423
N	92	GLN	GLU	engineered mutation	UNP O18423
N	97	SER	GLU	engineered mutation	UNP O18423
N	126	GLY	ASP	engineered mutation	UNP O18423
O	0	GLY	-	expression tag	UNP O18423
O	84	GLN	GLU	engineered mutation	UNP O18423
O	85	LYS	GLU	engineered mutation	UNP O18423
O	92	GLN	GLU	engineered mutation	UNP O18423
O	97	SER	GLU	engineered mutation	UNP O18423
O	126	GLY	ASP	engineered mutation	UNP O18423
P	0	GLY	-	expression tag	UNP O18423
P	84	GLN	GLU	engineered mutation	UNP O18423
P	85	LYS	GLU	engineered mutation	UNP O18423
P	92	GLN	GLU	engineered mutation	UNP O18423
P	97	SER	GLU	engineered mutation	UNP O18423
P	126	GLY	ASP	engineered mutation	UNP O18423
R	0	GLY	-	expression tag	UNP O18423
R	84	GLN	GLU	engineered mutation	UNP O18423
R	85	LYS	GLU	engineered mutation	UNP O18423
R	92	GLN	GLU	engineered mutation	UNP O18423
R	97	SER	GLU	engineered mutation	UNP O18423
R	126	GLY	ASP	engineered mutation	UNP O18423
S	0	GLY	-	expression tag	UNP O18423
S	84	GLN	GLU	engineered mutation	UNP O18423
S	85	LYS	GLU	engineered mutation	UNP O18423
S	92	GLN	GLU	engineered mutation	UNP O18423
S	97	SER	GLU	engineered mutation	UNP O18423
S	126	GLY	ASP	engineered mutation	UNP O18423

- Molecule 2 is MERCURIBENZOIC ACID (three-letter code: MBO) (formula: $C_7H_5HgO_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	Hg	O		
2	A	1	10	7	1	2	0	0
2	A	1	10	7	1	2	0	0
2	B	1	10	7	1	2	0	0
2	B	1	10	7	1	2	0	0
2	C	1	10	7	1	2	0	0
2	D	1	10	7	1	2	0	0
2	E	1	10	7	1	2	0	0
2	F	1	10	7	1	2	0	0
2	G	1	10	7	1	2	0	0
2	H	1	10	7	1	2	0	0
2	H	1	10	7	1	2	0	0
2	I	1	10	7	1	2	0	0
2	K	1	10	7	1	2	0	0
2	K	1	10	7	1	2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	Hg	O		
2	L	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	L	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	M	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	M	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	O	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	P	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	R	1	10	7	1	2	0	0
			Total	C	Hg	O		
2	S	1	10	7	1	2	0	0

- Molecule 3 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Hg		
3	C	1	1	1	0	0
			Total	Hg		
3	D	1	1	1	0	0
			Total	Hg		
3	E	1	1	1	0	0
			Total	Hg		
3	F	1	1	1	0	0
			Total	Hg		
3	G	1	1	1	0	0
			Total	Hg		
3	I	1	1	1	0	0
			Total	Hg		
3	J	2	2	2	0	0
			Total	Hg		
3	N	2	2	2	0	0
			Total	Hg		
3	O	1	1	1	0	0
			Total	Hg		
3	P	1	1	1	0	0
			Total	Hg		
3	R	1	1	1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Hg		
3	S	1	1	1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	7	7	7	0	0
4	B	2	2	2	0	0
4	C	2	2	2	0	0
4	D	7	7	7	0	0
4	E	5	5	5	0	0
4	F	5	5	5	0	0
4	G	2	2	2	0	0
4	H	2	2	2	0	0
4	I	4	4	4	0	0
4	J	4	4	4	0	0
4	K	8	8	8	0	0
4	L	2	2	2	0	0
4	M	2	2	2	0	0
4	N	2	2	2	0	0
4	O	2	2	2	0	0
4	P	1	1	1	0	0
4	R	3	3	3	0	0
4	S	1	1	1	0	0

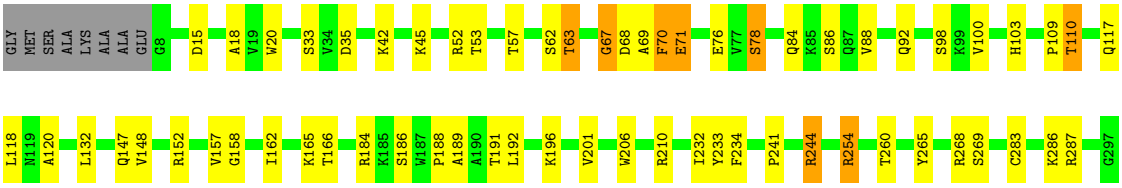
3 Residue-property plots (i)

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. 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.

Note EDS failed to run properly.

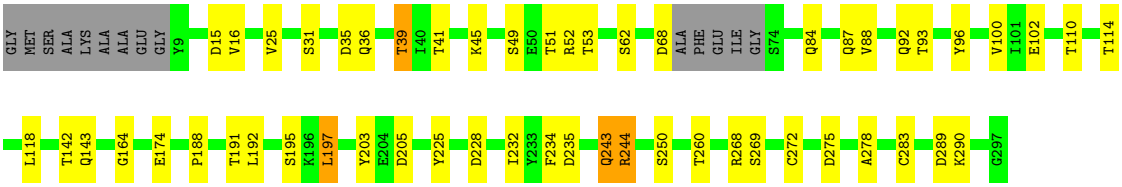
• Molecule 1: Lysenin

Chain A:



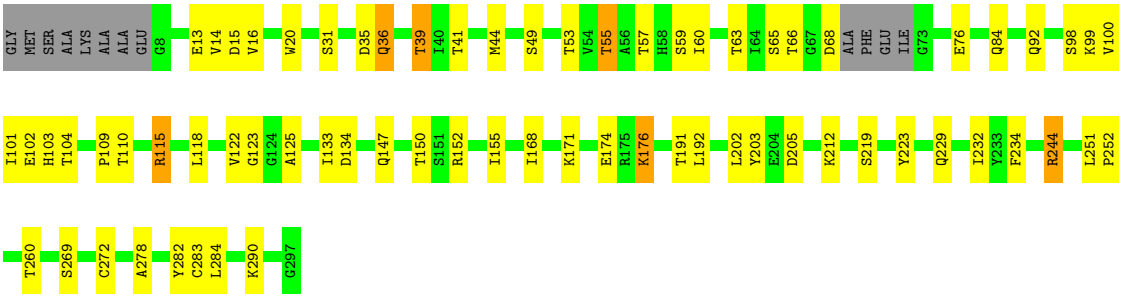
• Molecule 1: Lysenin

Chain B:



• Molecule 1: Lysenin

Chain C:



• Molecule 1: Lysenin

Chain D:

GLY MET SER ALA LYS ALA ALA E7 V14 D15 V16 W20 V25 V34 D35 Q36 K37 I38 T39 I40 T41 V47 N48 S49 R52 T53 V54 T55 A56 T57 H58 S59 I64 S65 T66 GLY ASP ALA PHE GLU ILE G73 S74 S80 Q84 V88 S89 M90 T91 Q92

T93 E94 S98 K99 V100 I101 E102 H103 T114 R115 W116 Q117 L118 M130 I133 D134 I145 P146 Q147 V148 I149 I155 R159 I162 K165 E174 R175 K176 E177 R184 P188 T191 L192 L197 V201 R210 I211 M217 Y218 Y225 D228

Q229 D235 P241 K242 Q243 R244 N248 H255 T260 Y265 S269 G270 L271 C272 V281 Y282 C283 K286 K290 E294 G297

- Molecule 1: Lysenin

Chain E: 74% 19% • 5%

GLY MET SER ALA LYS ALA ALA ALA GLU GLY Y9 E10 Q11 I12 E13 V14 D15 V16 V19 V25 Y26 E27 N28 R29 S33 V34 D35 Q36 K37 I40 T41 K45 T51 R52 T53 V54 T55 T66 GLY ASP ALA PHE GLU ILE G73 S74 V75 S78 Q84 V88 S89

M90 T91 Q92 K99 V100 I101 E102 H103 T104 I105 P109 T110 T114 L118 I145 V148 R152 I155 T166 R184 R185 S186 W187 P188 T191 K199 F200 V201 L214 M217 D228 Q229 I232 Y233 F234 P241 R244 T260 Y265 F266

S269 C283 E294 V295 V296 G297

- Molecule 1: Lysenin

Chain F: 74% 20% • 5%

GLY MET SER ALA LYS ALA ALA ALA GLU G8 Y9 E10 E13 V14 D15 V16 W20 Y24 S33 V34 D35 Q36 K37 K42 S49 E50 T51 R52 T55 S62 T66 GLY ASP ALA PHE GLU ILE GLY S74 S78 Y79 S80 S86 Q87 V88 S89 M90 E94

S98 K99 V100 H103 P109 T110 T142 Q147 V148 R152 I155 R159 K171 R184 T191 L197 V201 S227 D228 Q229 G230 G231 I232 Q236 G237 T238 D239 R244 K249 S250 T260 K264 Y265 F266 T267 R268 S269 V281 Y282 C283 L284

D285 K286 R287 E288 G297

- Molecule 1: Lysenin

Chain G: 74% 18% • • •

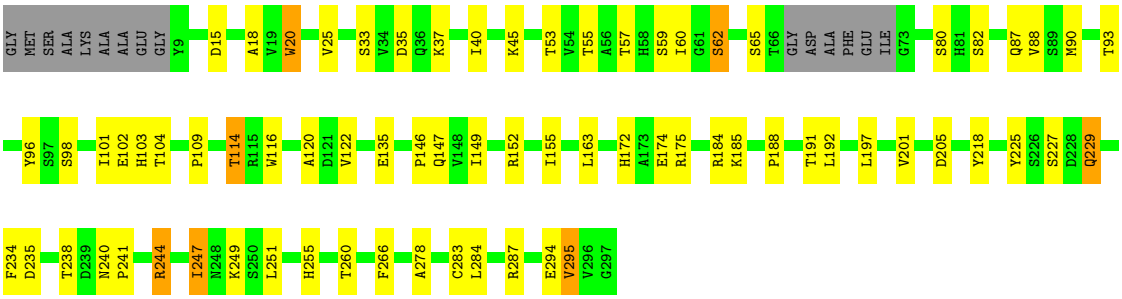
GLY MET SER ALA LYS ALA ALA ALA GLU G8 V14 D15 V16 V17 A18 V19 W20 Y24 R29 G30 S31 D35 Q36 K37 T41 V47 N48 S49 T53 T57 H58 S59 G67 D68 A69 F70 E71 T72 Q84 V88 S89 M90 T91 Q92 T93 S98 K99 V100 I101

E102 H103 T114 R115 L118 N119 A120 D134 Q147 V148 I155 R159 I162 K165 H172 A173 E174 R175 L192 Y203 R210 S219 Y225 S226 S227 D228 Q229 T238 R244 W245 A246 I247 N248 T260 K264 Y265 F266 T267 R268 Y273

A278 C283 L284 D289 G297

- Molecule 1: Lysenin

Chain H:

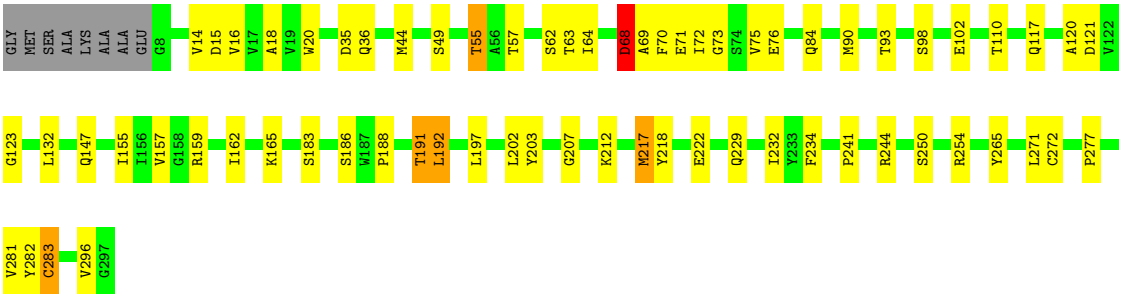


- Molecule 1: Lysenin

Chain I:

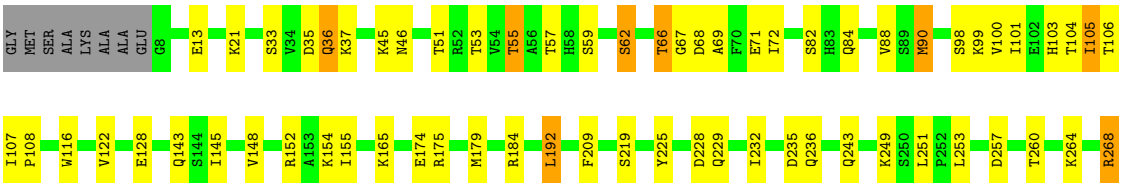
- Molecule 1: Lysenin

Chain J:



- Molecule 1: Lysenin

Chain K:



S269 Y273 K286 R287 K290 G297

- Molecule 1: Lysenin

Chain L: 74% 18% • 5%

GLY MET SER ALA LYS ALA ALA ALA GLU GLU GLY Y9 E13 V16 W20 K21 V25 S33 V34 D35 Q36 T39 I40 T41 K45 M46 V47 T53 V54 T55 A56 T57 T66 GLY ASP ALA PHE GLU I72 G73 Q84 K85 V88 S89 M90 E94 S98 K99 V100

I101 E102 H103 T104 I107 P108 P109 T110 S111 T114 R115 V122 L132 I133 D134 T142 Q147 I155 E174 R184 P188 L192 S195 K196 L197 D205 W206 S219 D228 F234 T238 R244 W245 N248 T260 L271 Y282 C283

L284 D285 K286 G297

- Molecule 1: Lysenin

Chain M: 72% 22% • •

GLY MET SER ALA LYS ALA ALA ALA GLU GLU GLY Y9 V14 D15 A18 V19 W20 V25 S31 T32 S33 V34 D35 Q36 T39 I40 T41 K42 N46 V47 N48 S49 T53 V54 T55 A56 T57 G67 D68 A69 F70 G73 S74 V75 Q84 Q87 V88 S89 M90 T91 Q92

T93 E94 V95 Y96 S97 S98 K99 V100 T104 P109 T110 T114 R115 L118 G123 M130 D134 T142 Q147 V148 I149 T150 S151 R152 I155 I156 V157 G158 R159 H172 R175 R184 K185 L192 L202 M217 Y218 G230 N240 P241 K242

A246 R254 T260 K264 Y265 F266 T267 R268 C283 L284 D285 K286 R287 E288 D289 E294 G297

- Molecule 1: Lysenin

Chain N: 76% 18% • •

GLY MET SER ALA LYS ALA ALA ALA GLU GLU GLY Y9 E13 V14 D15 V16 S31 T32 S33 V34 D35 T39 G43 M44 K45 N48 S49 R52 T53 V54 T55 A56 T57 I64 G67 I72 V75 E76 Q84 K85 V88 S89 M90 T93 Y96 K99 V100

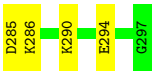
I105 P109 R115 L118 V122 G123 D134 T150 K154 I155 I156 V157 G158 R159 H172 A173 E174 R175 R184 K185 T191 L197 L202 R210 I211 S226 R244 I247 N248 K249 D257 T260 P277 Y282 C283 G297

- Molecule 1: Lysenin

Chain O: 74% 21% • •

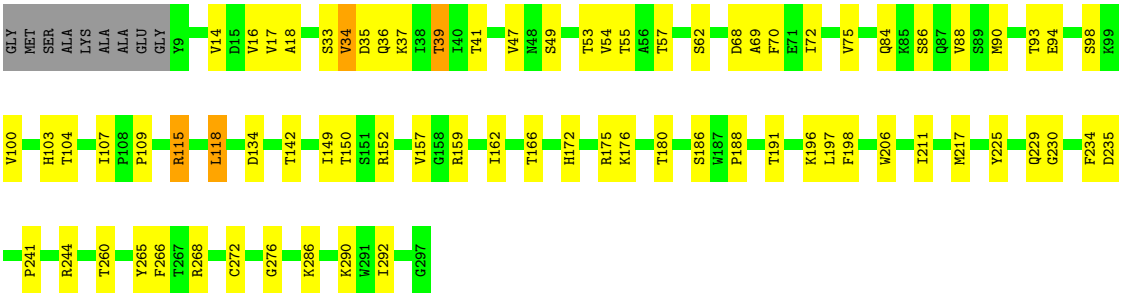
GLY MET SER ALA LYS ALA ALA ALA GLU GLU GLY Y9 E13 V14 D15 V16 V17 W20 K21 E22 V25 T32 S33 V34 D35 T39 I40 T41 K45 T53 V54 T55 S62 I63 I64 G67 D68 A69 F70 E71 I72 Q84 V88 S89 M90 T93 Y96 K99

V100 I101 E102 H103 T104 T110 T114 R115 W116 Q117 G123 D134 Q147 T150 V157 I162 L163 G164 K165 S186 T191 Y203 K212 T213 L214 M217 Y225 I232 D235 R244 R254 T260 S269 C272 A278 C283 L284



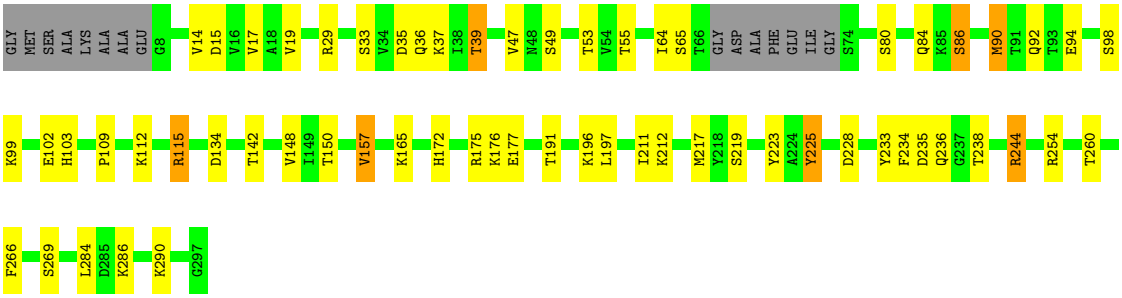
- Molecule 1: Lysenin

Chain P:



- Molecule 1: Lysenin

Chain R:



- Molecule 1: Lysenin

Chain S:

4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	192.99Å 192.99Å 493.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.10 – 3.10	Depositor
% Data completeness (in resolution range)	99.9 (49.10-3.10)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.20	Depositor
$< I/\sigma(I) >$ [1]	2.61 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.214 , 0.258	Depositor
Wilson B-factor (Å^2)	57.6	Xtriage
Anisotropy	0.541	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	41248	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.61% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MBO, HG

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	0.49	0/2352	0.73	0/3186
1	B	0.48	0/2309	0.70	0/3127
1	C	0.49	0/2317	0.69	0/3137
1	D	0.49	0/2314	0.73	1/3133 (0.0%)
1	E	0.50	0/2301	0.73	0/3116
1	F	0.55	0/2301	0.77	0/3116
1	G	0.56	0/2352	0.81	4/3186 (0.1%)
1	H	0.51	0/2301	0.75	0/3116
1	I	0.50	0/2297	0.73	1/3111 (0.0%)
1	J	0.49	0/2352	0.69	0/3186
1	K	0.51	0/2352	0.75	0/3186
1	L	0.48	0/2309	0.71	1/3127 (0.0%)
1	M	0.49	0/2348	0.70	0/3181
1	N	0.50	0/2348	0.72	1/3181 (0.0%)
1	O	0.50	0/2348	0.72	0/3181
1	P	0.51	0/2348	0.72	0/3181
1	R	0.49	0/2301	0.72	0/3116
1	S	0.49	0/2301	0.71	1/3116 (0.0%)
All	All	0.50	0/41851	0.73	9/56683 (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	1
1	S	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	118	LEU	CA-CB-CG	7.35	132.21	115.30
1	N	118	LEU	CA-CB-CG	6.95	131.29	115.30
1	G	115	ARG	NE-CZ-NH1	6.37	123.49	120.30
1	G	210	ARG	NE-CZ-NH1	6.13	123.37	120.30
1	G	118	LEU	CA-CB-CG	6.00	129.11	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	67	GLY	Peptide
1	S	221	TYR	Peptide

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	2301	0	2264	32	0
1	B	2260	0	2226	26	0
1	C	2268	0	2233	31	0
1	D	2265	0	2232	42	0
1	E	2252	0	2222	33	0
1	F	2252	0	2222	35	1
1	G	2301	0	2265	53	0
1	H	2252	0	2223	49	0
1	I	2248	0	2219	30	0
1	J	2301	0	2265	26	0
1	K	2301	0	2265	32	0
1	L	2260	0	2234	39	0
1	M	2297	0	2262	45	0
1	N	2297	0	2262	37	0
1	O	2297	0	2262	39	0
1	P	2297	0	2261	36	0
1	R	2252	0	2222	35	1
1	S	2252	0	2223	33	0
2	A	20	0	8	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	20	0	8	16	0
2	C	10	0	4	9	0
2	D	10	0	4	17	0
2	E	10	0	4	5	0
2	F	10	0	4	19	0
2	G	10	0	4	13	0
2	H	20	0	8	29	0
2	I	10	0	4	6	0
2	K	20	0	8	29	0
2	L	20	0	8	25	0
2	M	20	0	8	18	0
2	O	10	0	4	12	0
2	P	10	0	4	20	0
2	R	10	0	4	7	0
2	S	10	0	4	15	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
3	I	1	0	0	0	0
3	J	2	0	0	0	0
3	N	2	0	0	0	0
3	O	1	0	0	1	0
3	P	1	0	0	0	0
3	R	1	0	0	0	0
3	S	1	0	0	0	0
4	A	7	0	0	0	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	7	0	0	0	0
4	E	5	0	0	0	0
4	F	5	0	0	0	0
4	G	2	0	0	0	0
4	H	2	0	0	0	0
4	I	4	0	0	0	0
4	J	4	0	0	0	0
4	K	8	0	0	0	0
4	L	2	0	0	0	0
4	M	2	0	0	0	0
4	N	2	0	0	0	0
4	O	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	P	1	0	0	0	0
4	R	3	0	0	0	0
4	S	1	0	0	0	0
All	All	41248	0	40450	735	1

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 9.

The worst 5 of 735 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:H:301:MBO:HE6	2:H:301:MBO:CE1	2.06	1.24
2:S:301:MBO:HE2	2:S:301:MBO:CE3	2.06	1.24
2:B:301:MBO:HE3	2:B:301:MBO:CE4	2.06	1.24
2:L:302:MBO:HE6	2:L:302:MBO:CE5	2.06	1.24
2:O:301:MBO:HE6	2:O:301:MBO:CE1	2.06	1.24

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:239:ASP:OD2	1:R:236:GLN:OE1[7_556]	1.52	0.68

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	288/298 (97%)	273 (95%)	12 (4%)	3 (1%)	15 49
1	B	280/298 (94%)	269 (96%)	11 (4%)	0	100 100
1	C	282/298 (95%)	269 (95%)	13 (5%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	281/298 (94%)	272 (97%)	8 (3%)	1 (0%)	34	69
1	E	279/298 (94%)	267 (96%)	12 (4%)	0	100	100
1	F	279/298 (94%)	259 (93%)	18 (6%)	2 (1%)	22	57
1	G	288/298 (97%)	273 (95%)	11 (4%)	4 (1%)	11	40
1	H	279/298 (94%)	272 (98%)	7 (2%)	0	100	100
1	I	278/298 (93%)	269 (97%)	9 (3%)	0	100	100
1	J	288/298 (97%)	274 (95%)	11 (4%)	3 (1%)	15	49
1	K	288/298 (97%)	272 (94%)	13 (4%)	3 (1%)	15	49
1	L	280/298 (94%)	268 (96%)	12 (4%)	0	100	100
1	M	287/298 (96%)	265 (92%)	20 (7%)	2 (1%)	22	57
1	N	287/298 (96%)	270 (94%)	16 (6%)	1 (0%)	41	73
1	O	287/298 (96%)	274 (96%)	10 (4%)	3 (1%)	15	49
1	P	287/298 (96%)	270 (94%)	15 (5%)	2 (1%)	22	57
1	R	279/298 (94%)	270 (97%)	9 (3%)	0	100	100
1	S	279/298 (94%)	263 (94%)	15 (5%)	1 (0%)	34	69
All	All	5096/5364 (95%)	4849 (95%)	222 (4%)	25 (0%)	29	64

5 of 25 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	68	ASP
1	A	70	PHE
1	A	71	GLU
1	G	68	ASP
1	G	72	ILE

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	253/257 (98%)	225 (89%)	28 (11%)	6	24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	250/257 (97%)	228 (91%)	22 (9%)	10	36
1	C	250/257 (97%)	222 (89%)	28 (11%)	6	24
1	D	250/257 (97%)	212 (85%)	38 (15%)	3	12
1	E	249/257 (97%)	224 (90%)	25 (10%)	7	28
1	F	249/257 (97%)	216 (87%)	33 (13%)	4	16
1	G	253/257 (98%)	227 (90%)	26 (10%)	7	27
1	H	249/257 (97%)	222 (89%)	27 (11%)	6	25
1	I	249/257 (97%)	221 (89%)	28 (11%)	6	24
1	J	253/257 (98%)	223 (88%)	30 (12%)	5	20
1	K	253/257 (98%)	220 (87%)	33 (13%)	4	18
1	L	250/257 (97%)	227 (91%)	23 (9%)	9	33
1	M	253/257 (98%)	231 (91%)	22 (9%)	10	36
1	N	253/257 (98%)	226 (89%)	27 (11%)	6	26
1	O	253/257 (98%)	230 (91%)	23 (9%)	9	33
1	P	253/257 (98%)	222 (88%)	31 (12%)	4	19
1	R	249/257 (97%)	224 (90%)	25 (10%)	7	28
1	S	249/257 (97%)	223 (90%)	26 (10%)	7	27
All	All	4518/4626 (98%)	4023 (89%)	495 (11%)	6	25

5 of 495 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	19	VAL
1	P	292	ILE
1	K	33	SER
1	P	217	MET
1	S	37	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
1	I	229	GLN
1	L	84	GLN
1	R	92	GLN
1	J	28	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	K	143	GLN

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)

Of 36 ligands modelled in this entry, 14 are monoatomic - leaving 22 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	$\#\|Z\| > 2$
2	MBO	R	301	1	7,10,10	0.94	0	11,13,13	0.93	0
2	MBO	G	301	1	7,10,10	0.93	0	11,13,13	1.00	0
2	MBO	S	301	1	7,10,10	0.83	0	11,13,13	1.11	0
2	MBO	H	302	1	7,10,10	0.99	0	11,13,13	1.12	0
2	MBO	L	301	1	7,10,10	0.93	0	11,13,13	0.87	0
2	MBO	B	302	1	7,10,10	0.94	0	11,13,13	0.92	0
2	MBO	A	302	1	7,10,10	1.14	0	11,13,13	0.98	0
2	MBO	D	301	1	7,10,10	0.88	0	11,13,13	1.13	0
2	MBO	C	301	1	7,10,10	0.96	0	11,13,13	0.76	0
2	MBO	B	301	1	7,10,10	0.98	0	11,13,13	1.06	0
2	MBO	K	302	1	7,10,10	0.92	0	11,13,13	0.92	0
2	MBO	M	302	1	7,10,10	0.95	0	11,13,13	1.46	1 (9%)
2	MBO	A	301	1	7,10,10	1.01	0	11,13,13	0.87	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	MBO	E	301	1	7,10,10	0.96	0	11,13,13	0.97	0
2	MBO	H	301	1	7,10,10	0.97	0	11,13,13	1.02	0
2	MBO	O	301	1	7,10,10	0.99	0	11,13,13	1.10	0
2	MBO	F	302	1	7,10,10	0.96	0	11,13,13	1.16	0
2	MBO	I	302	1	7,10,10	1.02	0	11,13,13	1.02	0
2	MBO	K	301	1	7,10,10	0.97	0	11,13,13	1.21	0
2	MBO	L	302	1	7,10,10	0.97	0	11,13,13	0.91	0
2	MBO	M	301	1	7,10,10	0.90	0	11,13,13	1.09	0
2	MBO	P	302	1	7,10,10	1.00	0	11,13,13	0.85	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
2	MBO	R	301	1	-	4/4/4/4	0/1/1/1
2	MBO	G	301	1	-	0/4/4/4	0/1/1/1
2	MBO	S	301	1	-	4/4/4/4	0/1/1/1
2	MBO	H	302	1	-	4/4/4/4	0/1/1/1
2	MBO	L	301	1	-	0/4/4/4	0/1/1/1
2	MBO	B	302	1	-	4/4/4/4	0/1/1/1
2	MBO	A	302	1	-	4/4/4/4	0/1/1/1
2	MBO	D	301	1	-	4/4/4/4	0/1/1/1
2	MBO	C	301	1	-	4/4/4/4	0/1/1/1
2	MBO	B	301	1	-	4/4/4/4	0/1/1/1
2	MBO	K	302	1	-	4/4/4/4	0/1/1/1
2	MBO	M	302	1	-	4/4/4/4	0/1/1/1
2	MBO	A	301	1	-	4/4/4/4	0/1/1/1
2	MBO	E	301	1	-	0/4/4/4	0/1/1/1
2	MBO	H	301	1	-	4/4/4/4	0/1/1/1
2	MBO	O	301	1	-	4/4/4/4	0/1/1/1
2	MBO	F	302	1	-	4/4/4/4	0/1/1/1
2	MBO	I	302	1	-	4/4/4/4	0/1/1/1
2	MBO	K	301	1	-	4/4/4/4	0/1/1/1
2	MBO	L	302	1	-	4/4/4/4	0/1/1/1
2	MBO	M	301	1	-	4/4/4/4	0/1/1/1
2	MBO	P	302	1	-	0/4/4/4	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	M	302	MBO	CE5-CE4-CZ	-2.28	115.90	120.39

There are no chirality outliers.

5 of 72 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	302	MBO	CE3-CE4-CZ-OZ2
2	A	302	MBO	CE5-CE4-CZ-OZ2
2	B	301	MBO	CE3-CE4-CZ-OZ2
2	B	302	MBO	CE3-CE4-CZ-OZ2
2	B	302	MBO	CE5-CE4-CZ-OZ1

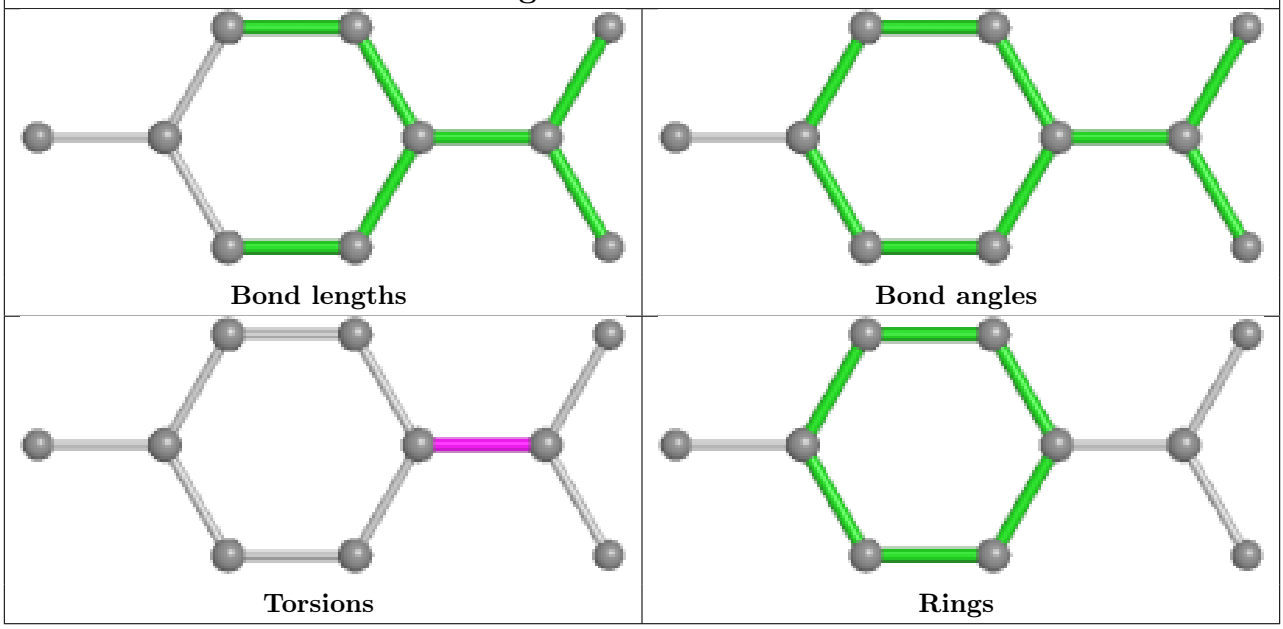
There are no ring outliers.

22 monomers are involved in 249 short contacts:

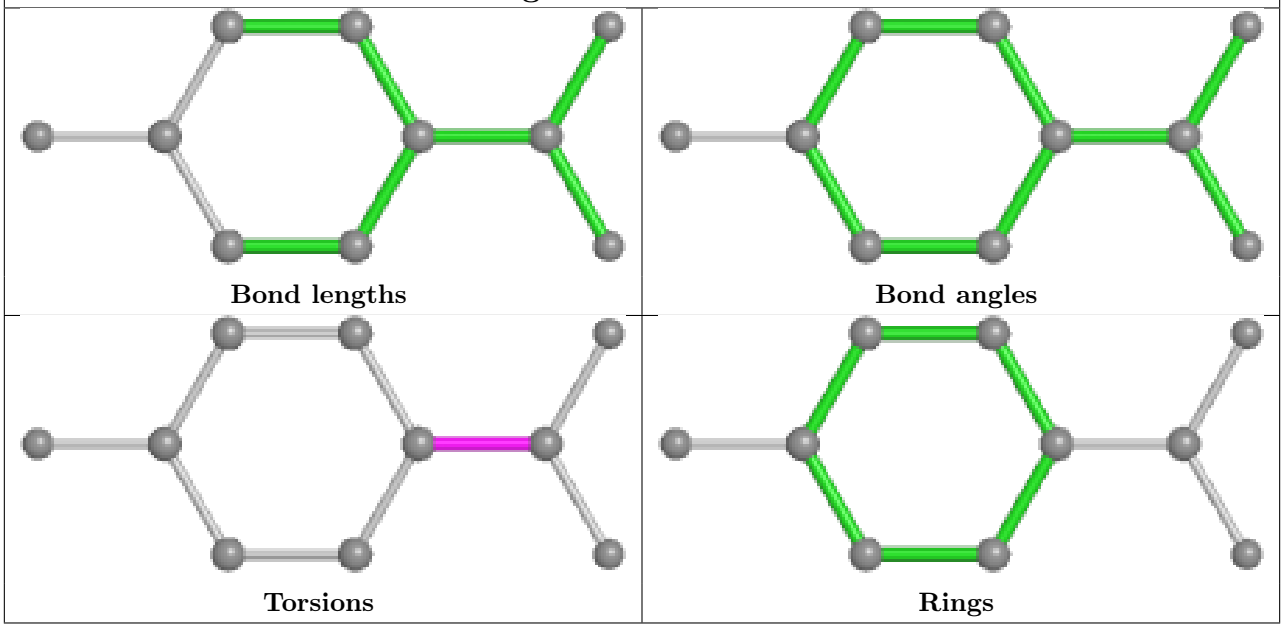
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	R	301	MBO	7	0
2	G	301	MBO	13	0
2	S	301	MBO	15	0
2	H	302	MBO	14	0
2	L	301	MBO	17	0
2	B	302	MBO	3	0
2	A	302	MBO	8	0
2	D	301	MBO	17	0
2	C	301	MBO	9	0
2	B	301	MBO	13	0
2	K	302	MBO	14	0
2	M	302	MBO	1	0
2	A	301	MBO	1	0
2	E	301	MBO	5	0
2	H	301	MBO	15	0
2	O	301	MBO	12	0
2	F	302	MBO	19	0
2	I	302	MBO	6	0
2	K	301	MBO	15	0
2	L	302	MBO	8	0
2	M	301	MBO	17	0
2	P	302	MBO	20	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.

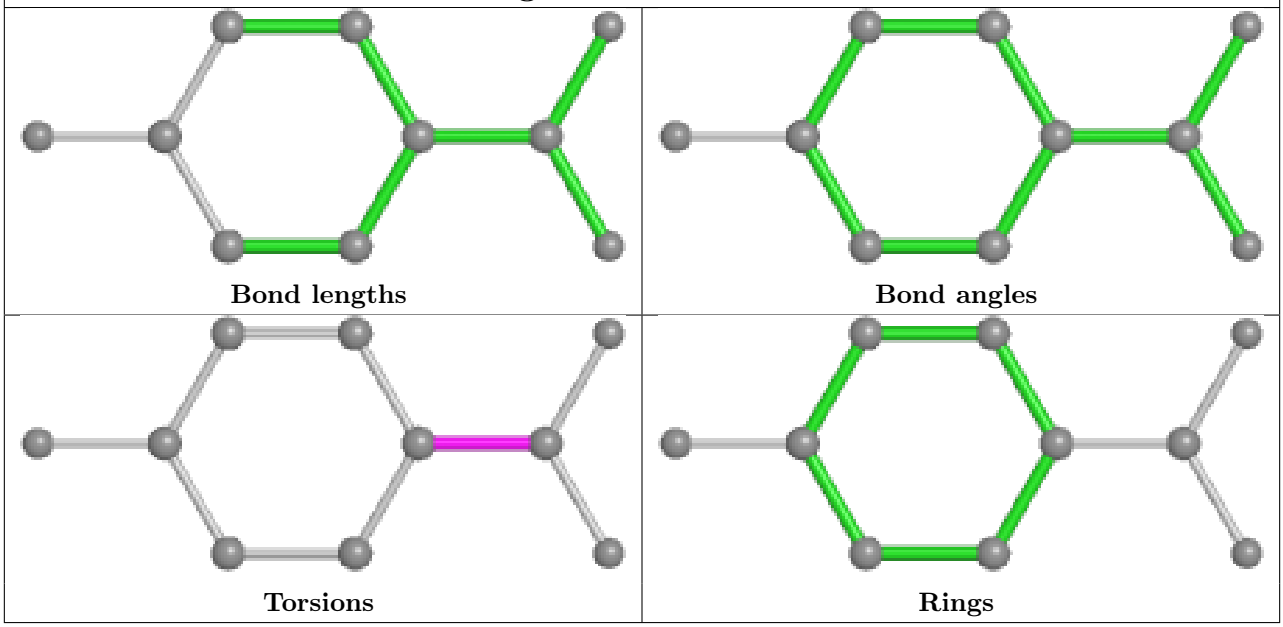
Ligand MBO R 301



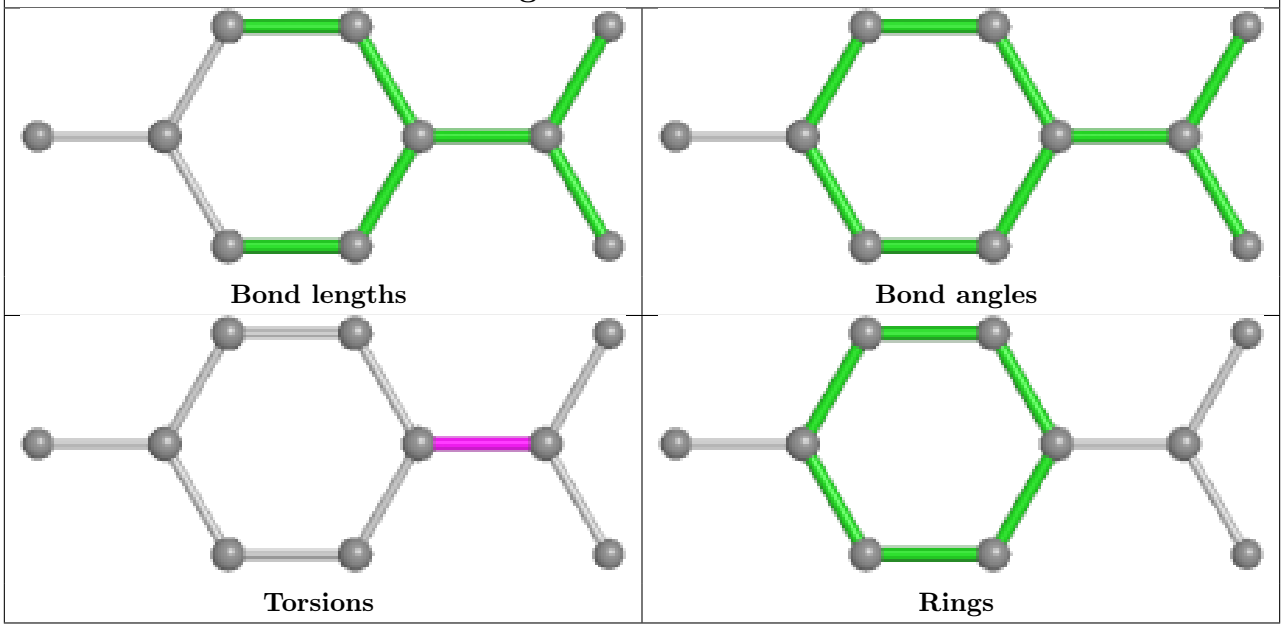
Ligand MBO S 301



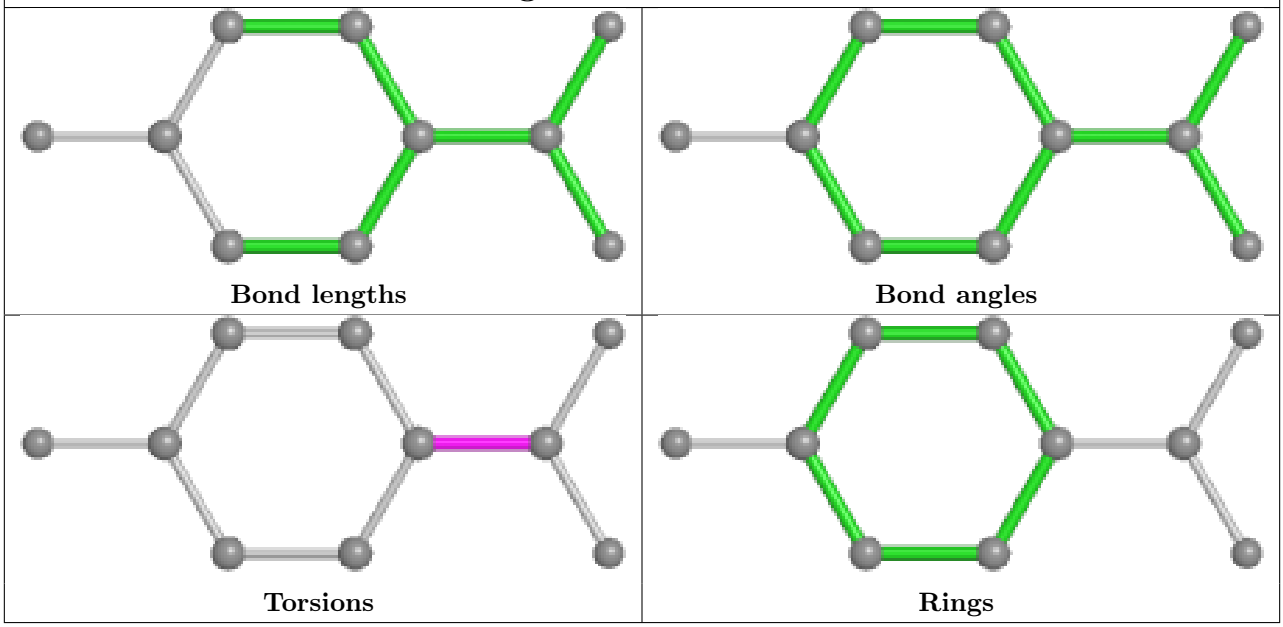
Ligand MBO H 302



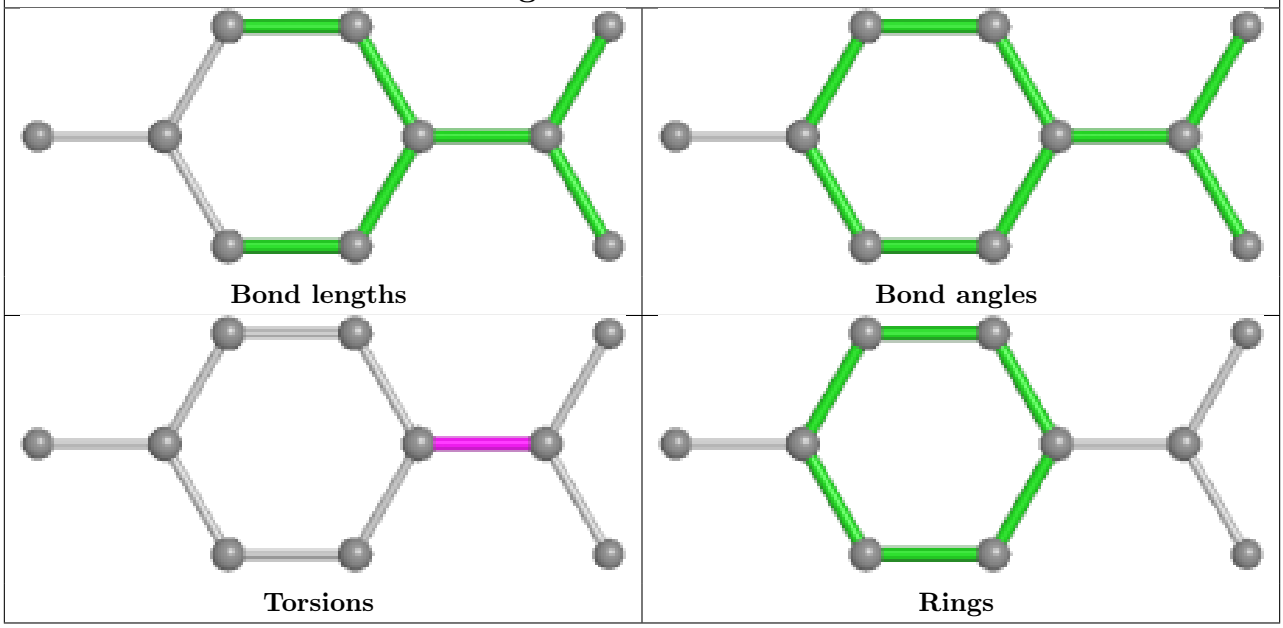
Ligand MBO B 302



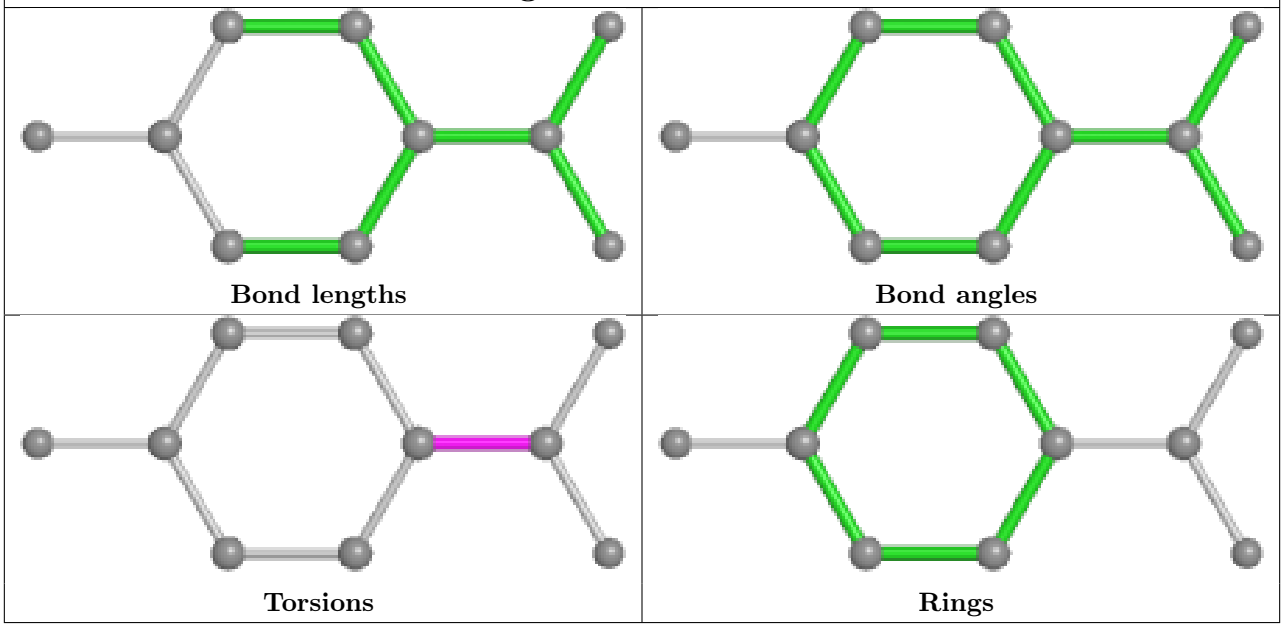
Ligand MBO A 302



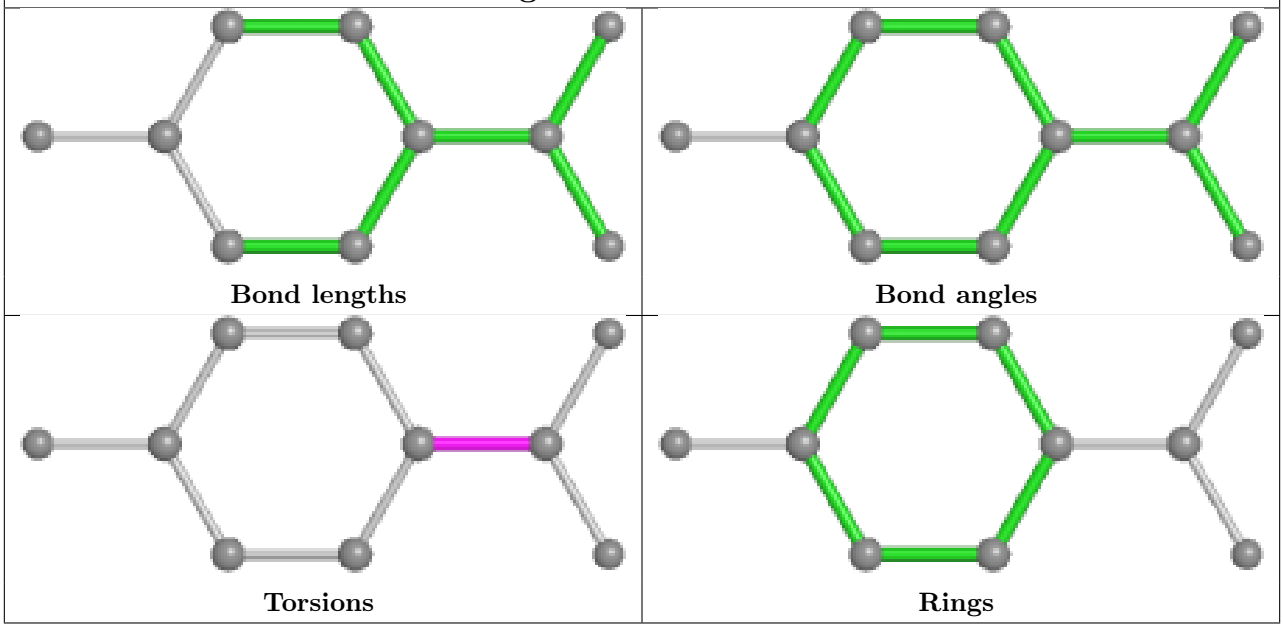
Ligand MBO D 301



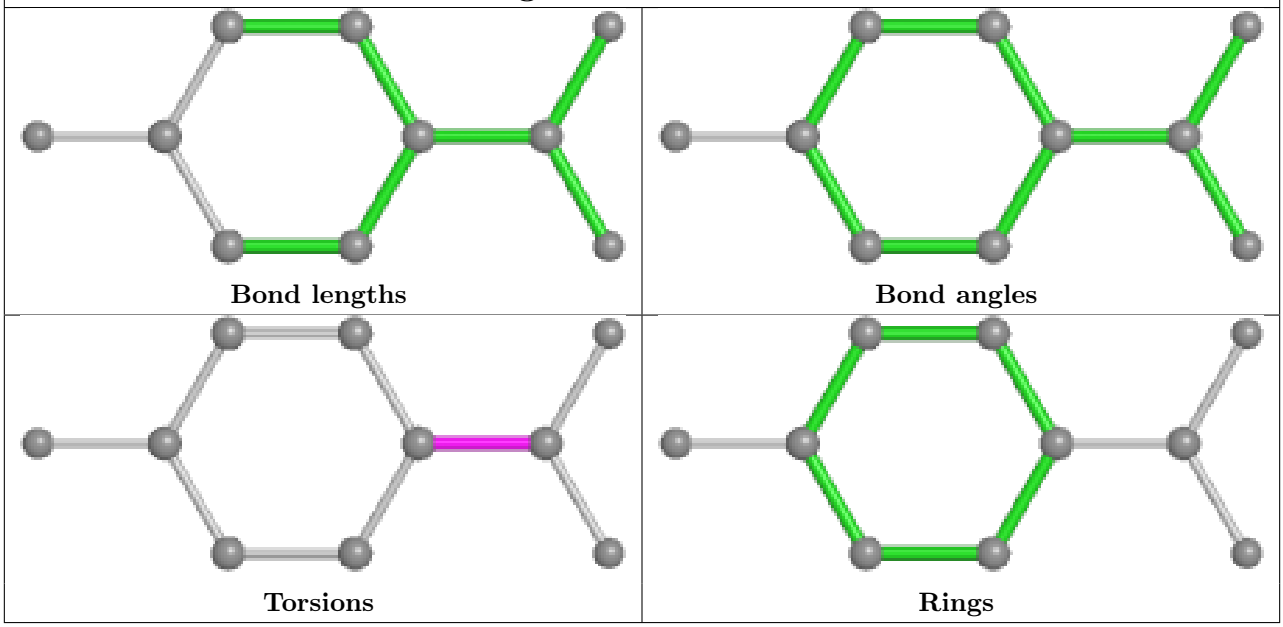
Ligand MBO C 301



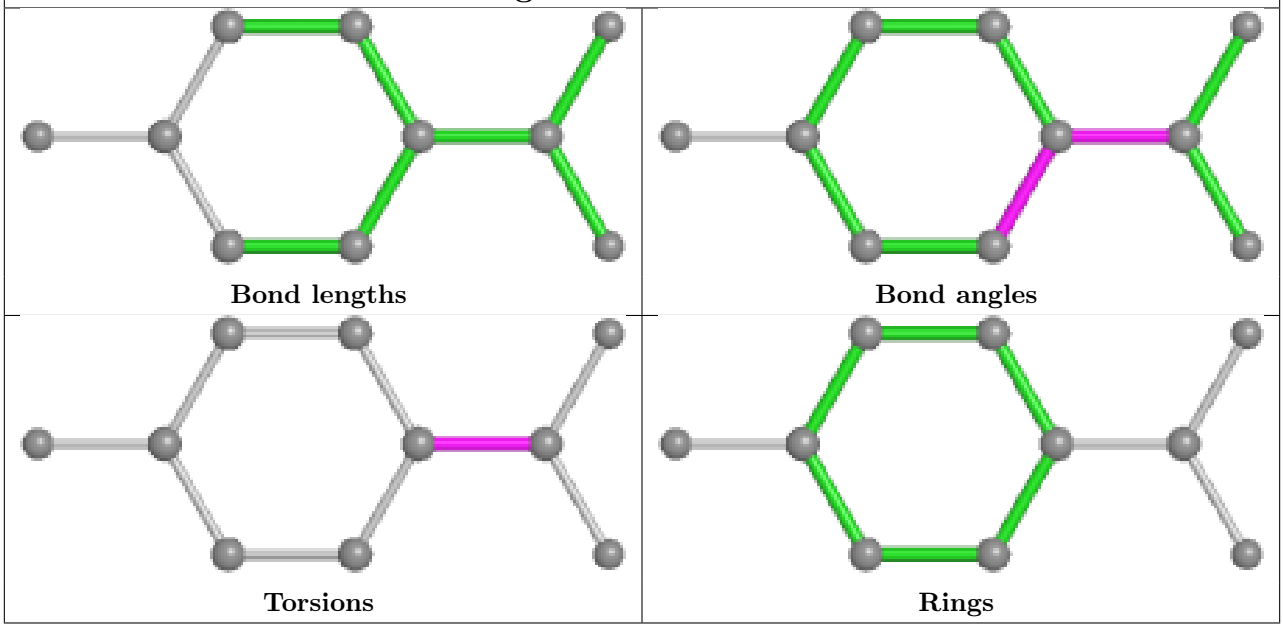
Ligand MBO B 301



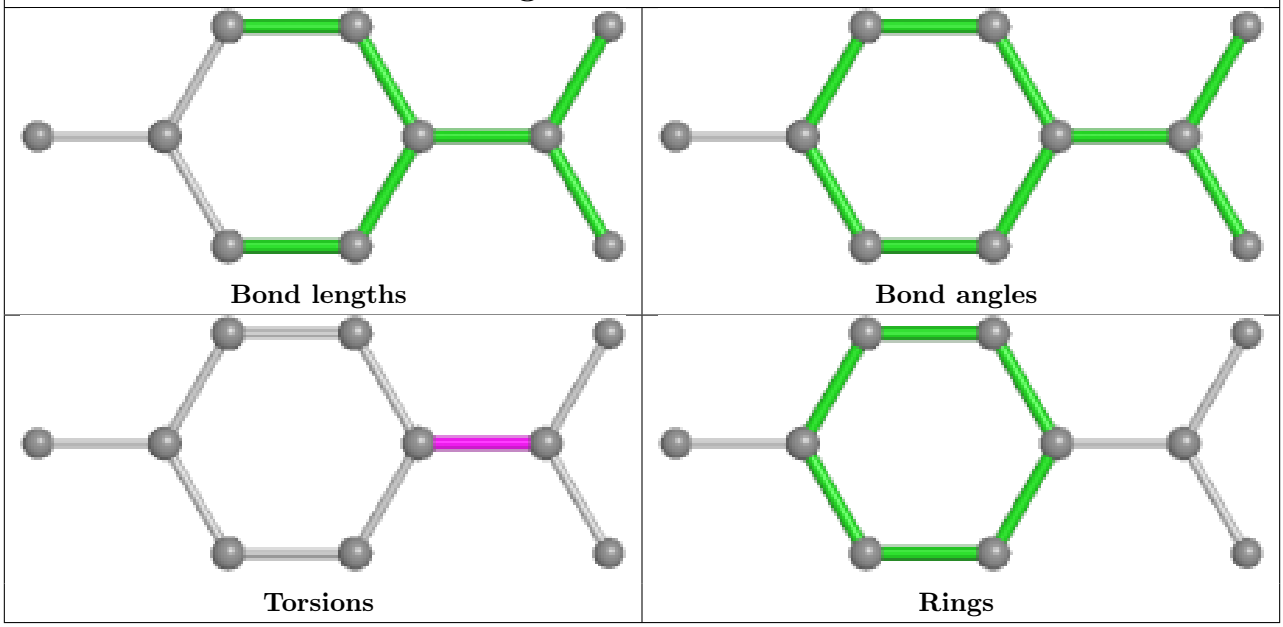
Ligand MBO K 302



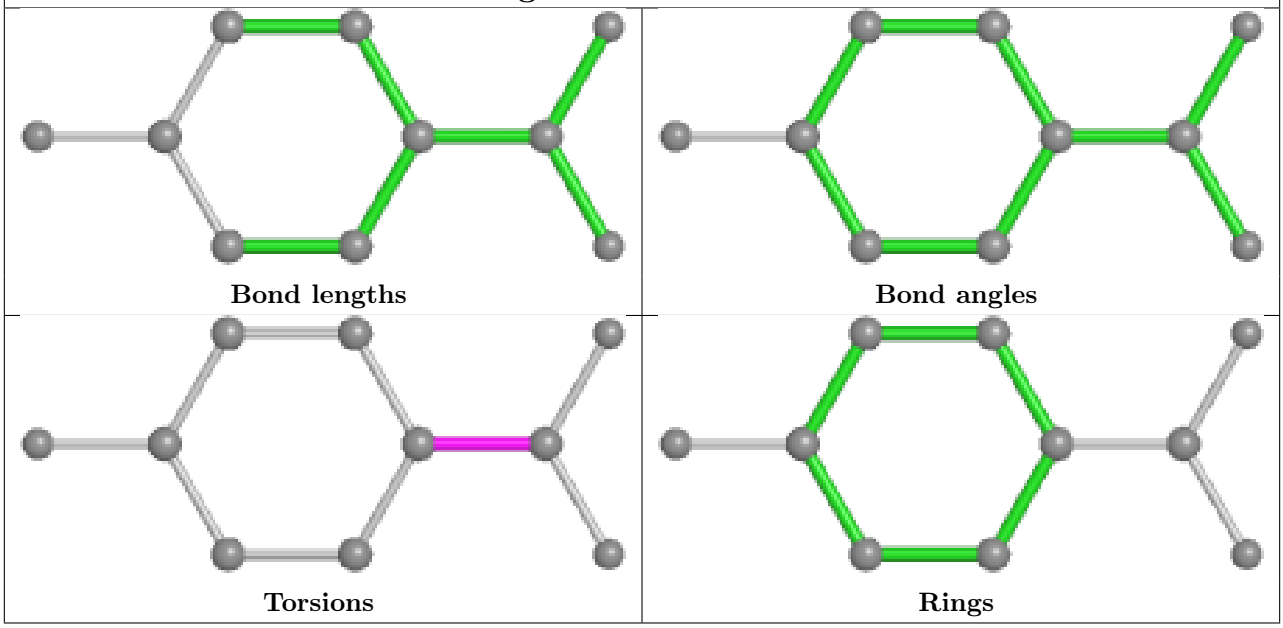
Ligand MBO M 302



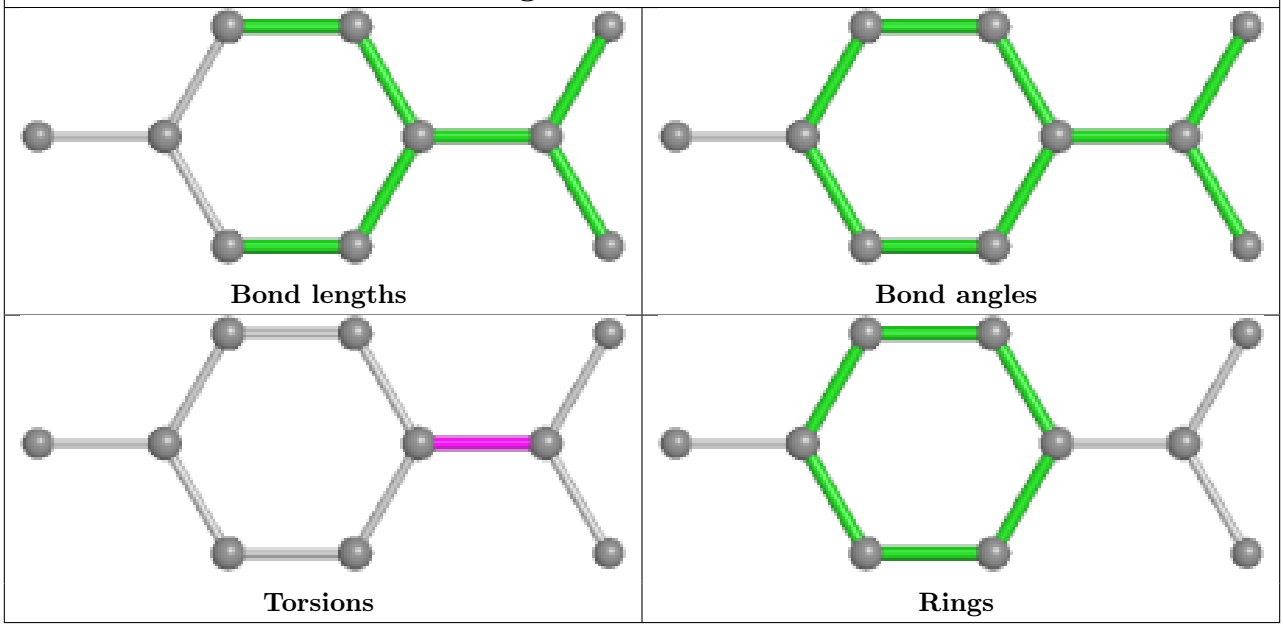
Ligand MBO A 301



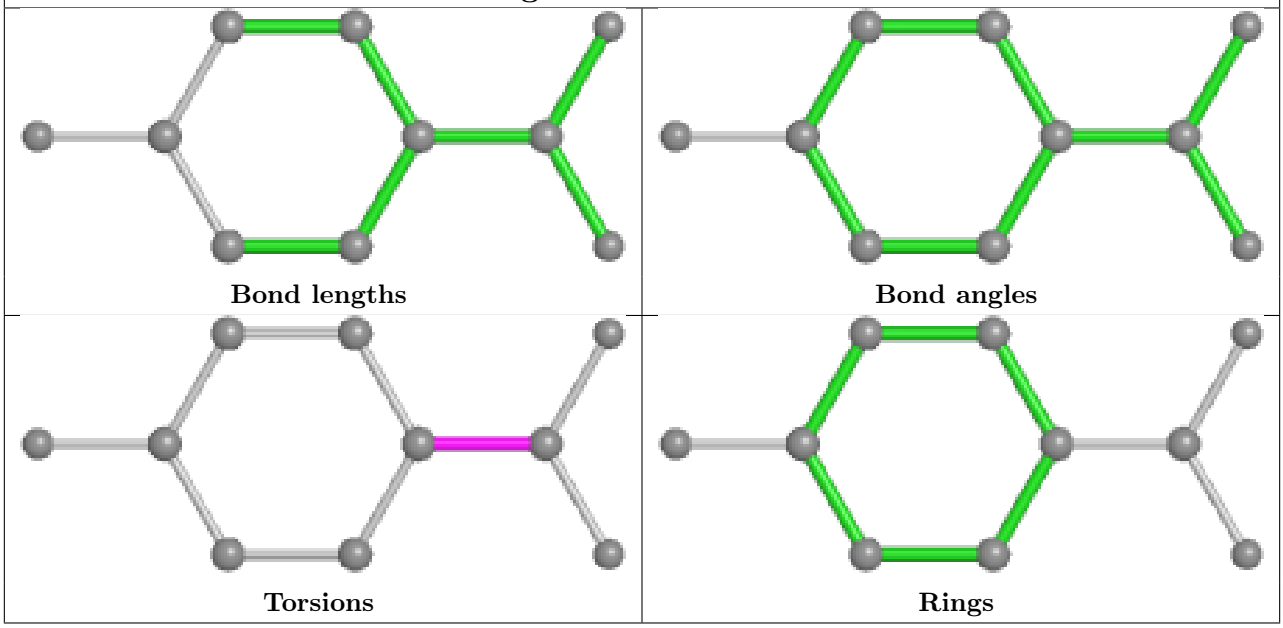
Ligand MBO H 301



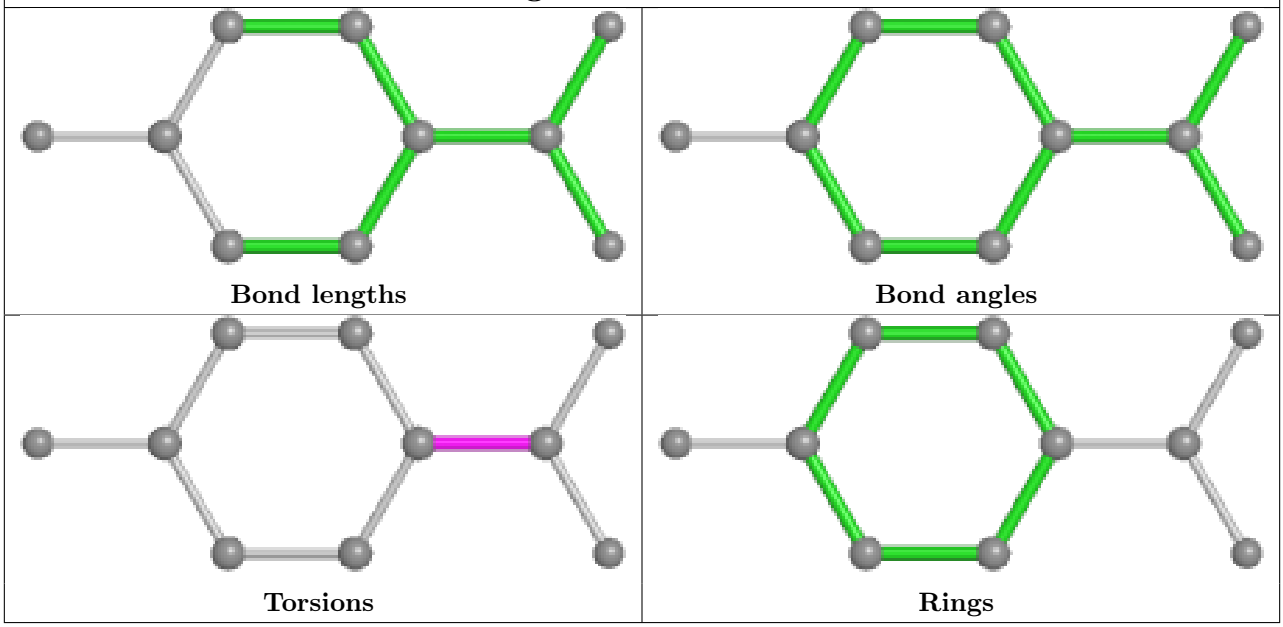
Ligand MBO O 301



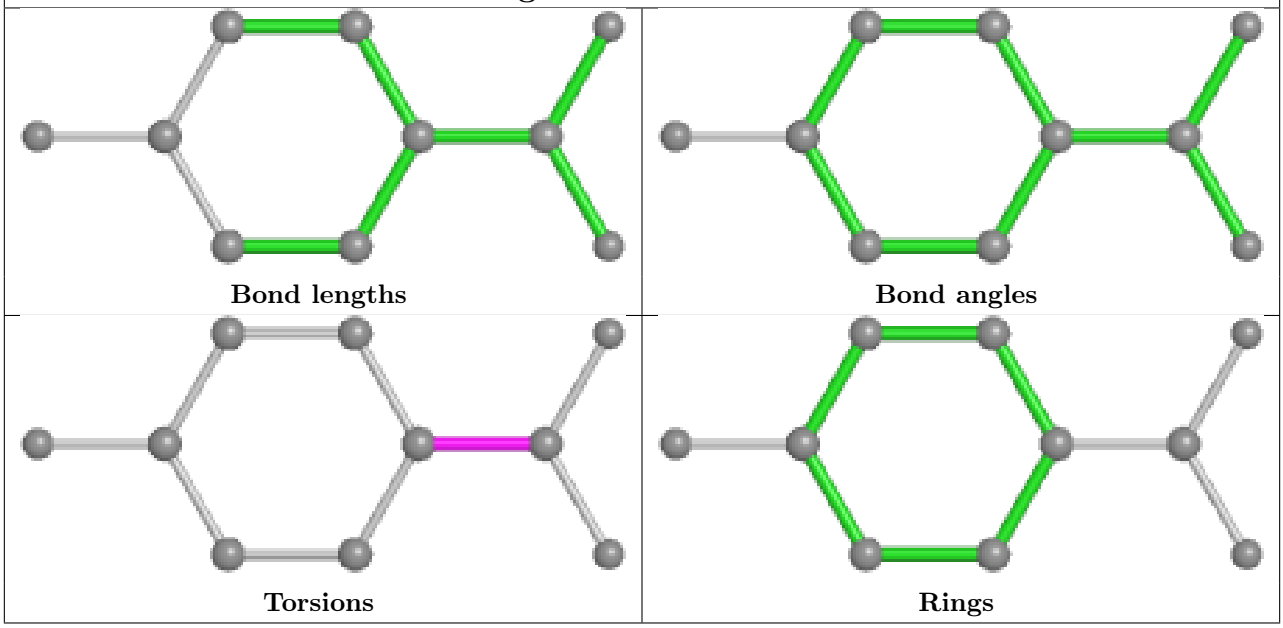
Ligand MBO F 302



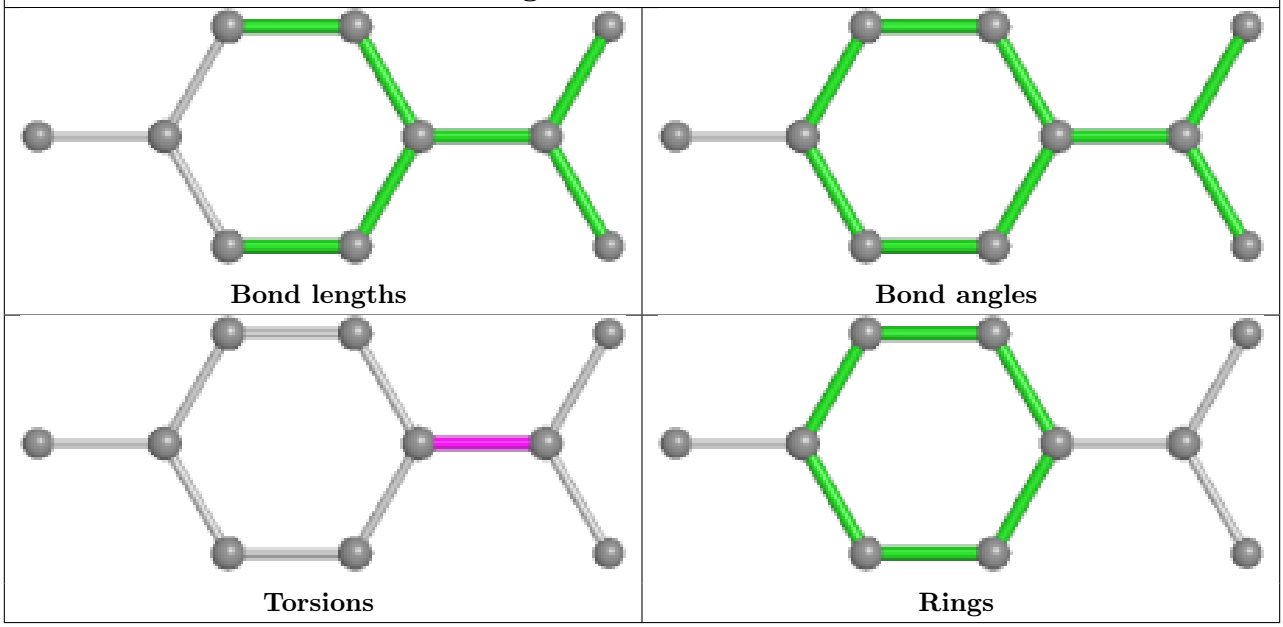
Ligand MBO I 302

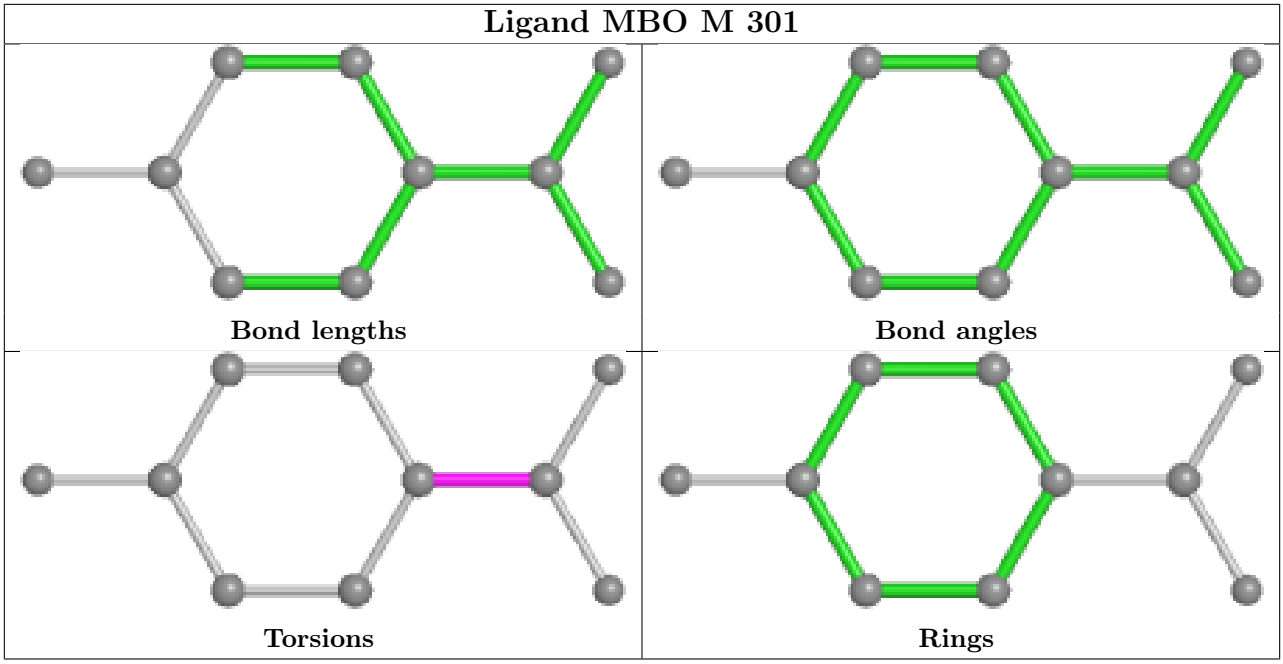


Ligand MBO K 301



Ligand MBO L 302





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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

6.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers (i)

EDS failed to run properly - this section is therefore empty.