



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 06:18 am BST

PDB ID : 2E0I  
Title : Crystal structure of archaeal photolyase from *Sulfolobus tokodaii* with two FAD molecules: Implication of a novel light-harvesting cofactor  
Authors : Fujihashi, M.; Numoto, N.; Kobayashi, Y.; Mizushima, A.; Tsujimura, M.; Nakamura, A.; Kawarabayashi, Y.; Miki, K.  
Deposited on : 2006-10-10  
Resolution : 2.80 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](mailto: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 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) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
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) : 2.11

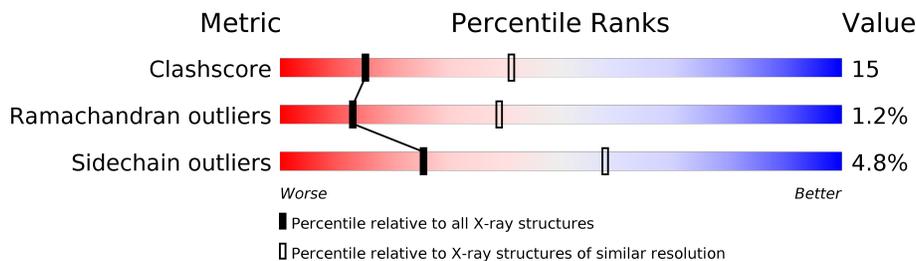
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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  $\geq 3$ , 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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	440	
1	B	440	
1	C	440	
1	D	440	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 15166 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 432aa long hypothetical deoxyribodipyrimidine photolyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	428	3622	2343	613	653	13	0	0	0
1	B	431	3643	2357	617	656	13	0	0	0
1	C	428	3622	2343	613	653	13	0	0	0
1	D	430	3636	2352	616	655	13	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

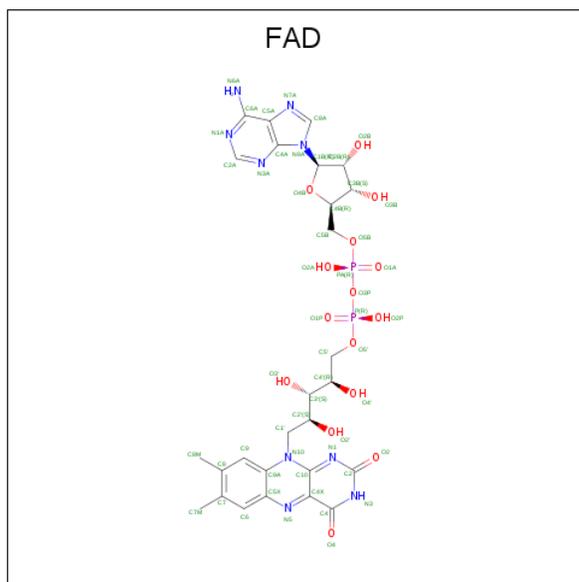
Chain	Residue	Modelled	Actual	Comment	Reference
A	433	LEU	-	EXPRESSION TAG	UNP Q973K9
A	434	GLU	-	EXPRESSION TAG	UNP Q973K9
A	435	HIS	-	EXPRESSION TAG	UNP Q973K9
A	436	HIS	-	EXPRESSION TAG	UNP Q973K9
A	437	HIS	-	EXPRESSION TAG	UNP Q973K9
A	438	HIS	-	EXPRESSION TAG	UNP Q973K9
A	439	HIS	-	EXPRESSION TAG	UNP Q973K9
A	440	HIS	-	EXPRESSION TAG	UNP Q973K9
B	433	LEU	-	EXPRESSION TAG	UNP Q973K9
B	434	GLU	-	EXPRESSION TAG	UNP Q973K9
B	435	HIS	-	EXPRESSION TAG	UNP Q973K9
B	436	HIS	-	EXPRESSION TAG	UNP Q973K9
B	437	HIS	-	EXPRESSION TAG	UNP Q973K9
B	438	HIS	-	EXPRESSION TAG	UNP Q973K9
B	439	HIS	-	EXPRESSION TAG	UNP Q973K9
B	440	HIS	-	EXPRESSION TAG	UNP Q973K9
C	433	LEU	-	EXPRESSION TAG	UNP Q973K9
C	434	GLU	-	EXPRESSION TAG	UNP Q973K9
C	435	HIS	-	EXPRESSION TAG	UNP Q973K9
C	436	HIS	-	EXPRESSION TAG	UNP Q973K9
C	437	HIS	-	EXPRESSION TAG	UNP Q973K9

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	438	HIS	-	EXPRESSION TAG	UNP Q973K9
C	439	HIS	-	EXPRESSION TAG	UNP Q973K9
C	440	HIS	-	EXPRESSION TAG	UNP Q973K9
D	433	LEU	-	EXPRESSION TAG	UNP Q973K9
D	434	GLU	-	EXPRESSION TAG	UNP Q973K9
D	435	HIS	-	EXPRESSION TAG	UNP Q973K9
D	436	HIS	-	EXPRESSION TAG	UNP Q973K9
D	437	HIS	-	EXPRESSION TAG	UNP Q973K9
D	438	HIS	-	EXPRESSION TAG	UNP Q973K9
D	439	HIS	-	EXPRESSION TAG	UNP Q973K9
D	440	HIS	-	EXPRESSION TAG	UNP Q973K9

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula:  $C_{27}H_{33}N_9O_{15}P_2$ ).



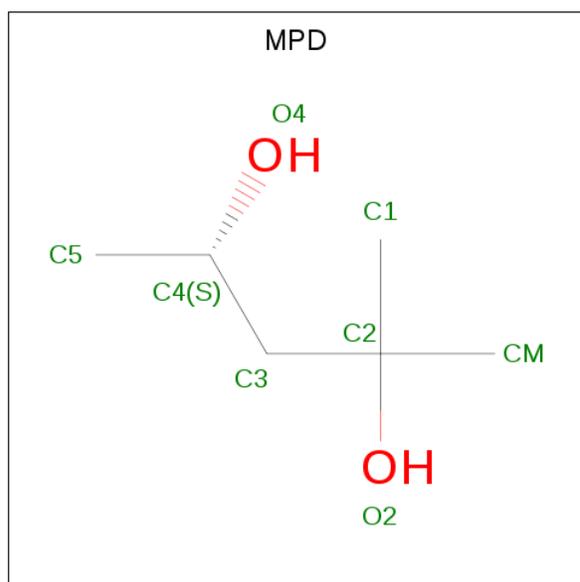
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
2	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	B	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	B	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	C	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	D	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	D	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			8	6	2		
3	A	1	Total	C	O	0	0
			8	6	2		
3	B	1	Total	C	O	0	0
			8	6	2		
3	C	1	Total	C	O	0	0
			8	6	2		
3	C	1	Total	C	O	0	0
			8	6	2		
3	D	1	Total	C	O	0	0
			8	6	2		

- Molecule 4 is water.

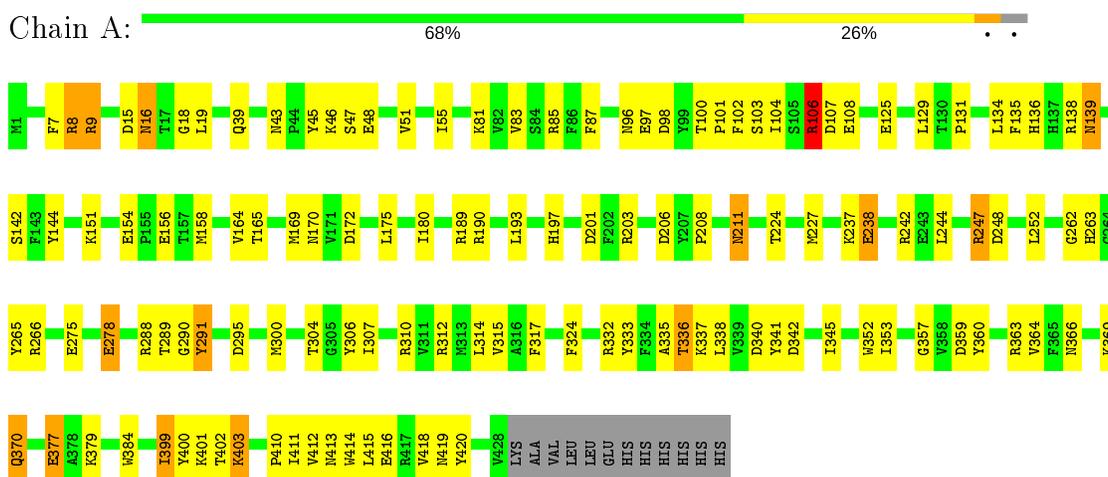
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	41	Total 41	O 41	0	0
4	B	38	Total 38	O 38	0	0
4	C	47	Total 47	O 47	0	0
4	D	45	Total 45	O 45	0	0

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

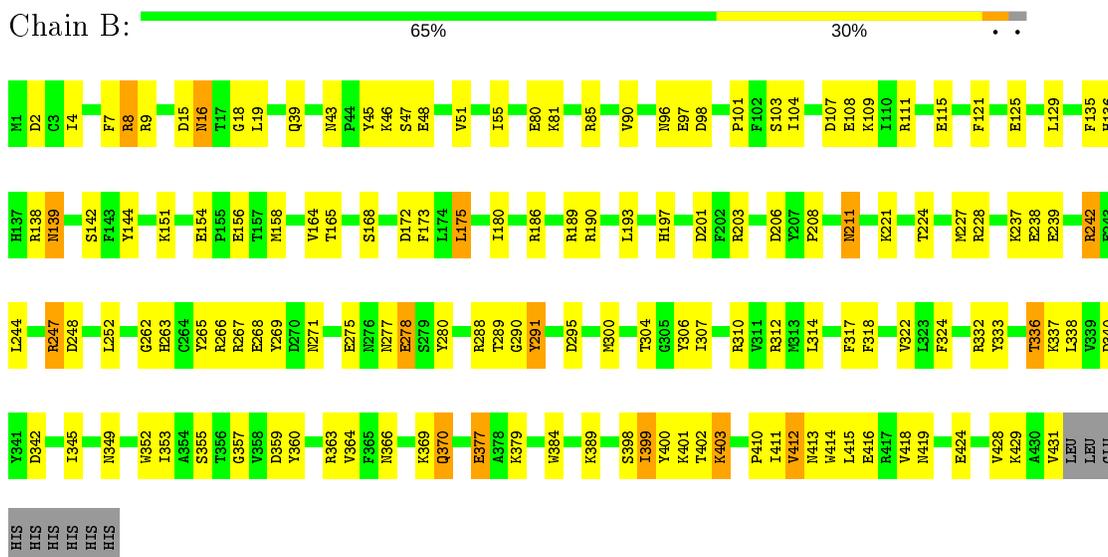
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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.

Note EDS was not executed.

- Molecule 1: 432aa long hypothetical deoxyribodipyrimidine photolyase



- Molecule 1: 432aa long hypothetical deoxyribodipyrimidine photolyase



- Molecule 1: 432aa long hypothetical deoxyribodipyrimidine photolyase



## 4 Data and refinement statistics

Xtrriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.88Å 114.72Å 167.68Å 90.00° 91.49° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80	Depositor
% Data completeness (in resolution range)	97.1 (50.00-2.80)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.12	Depositor
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.223 , 0.247	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	15166	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

## 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: MPD, FAD

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	0.46	0/3727	0.76	9/5040 (0.2%)
1	B	0.45	0/3748	0.69	6/5068 (0.1%)
1	C	0.46	0/3727	0.67	4/5040 (0.1%)
1	D	0.45	0/3741	0.67	5/5058 (0.1%)
All	All	0.46	0/14943	0.70	24/20206 (0.1%)

There are no bond length outliers.

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	9	ARG	NE-CZ-NH1	-14.54	113.03	120.30
1	A	106	ARG	NE-CZ-NH2	13.42	127.01	120.30
1	A	9	ARG	NE-CZ-NH2	12.53	126.56	120.30
1	A	106	ARG	NE-CZ-NH1	-12.31	114.15	120.30
1	B	242	ARG	NE-CZ-NH2	-10.09	115.26	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	3622	0	3504	104	0
1	B	3643	0	3531	119	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3622	0	3504	96	0
1	D	3636	0	3522	120	0
2	A	106	0	62	6	0
2	B	106	0	62	6	0
2	C	106	0	62	5	0
2	D	106	0	62	7	0
3	A	16	0	28	5	0
3	B	8	0	14	3	0
3	C	16	0	28	3	0
3	D	8	0	14	2	0
4	A	41	0	0	2	0
4	B	38	0	0	3	0
4	C	47	0	0	1	0
4	D	45	0	0	3	0
All	All	15166	0	14393	442	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 15.

The worst 5 of 442 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:314:LEU:HD21	1:A:370:GLN:HG2	1.42	1.00
1:C:314:LEU:HD21	1:C:370:GLN:HG2	1.48	0.95
1:B:268:GLU:HG3	1:B:429:LYS:HE3	1.48	0.95
1:D:411:ILE:O	1:D:412:VAL:HG13	1.69	0.93
1:C:289:THR:HG22	1:C:295:ASP:OD2	1.73	0.89

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	426/440 (97%)	396 (93%)	26 (6%)	4 (1%)	17	46
1	B	429/440 (98%)	399 (93%)	24 (6%)	6 (1%)	11	34
1	C	426/440 (97%)	399 (94%)	23 (5%)	4 (1%)	17	46
1	D	428/440 (97%)	397 (93%)	25 (6%)	6 (1%)	11	34
All	All	1709/1760 (97%)	1591 (93%)	98 (6%)	20 (1%)	13	39

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	136	HIS
1	A	403	LYS
1	B	136	HIS
1	B	168	SER
1	C	136	HIS

### 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	392/403 (97%)	372 (95%)	20 (5%)	24	55
1	B	394/403 (98%)	376 (95%)	18 (5%)	27	60
1	C	392/403 (97%)	375 (96%)	17 (4%)	29	62
1	D	393/403 (98%)	373 (95%)	20 (5%)	24	55
All	All	1571/1612 (98%)	1496 (95%)	75 (5%)	25	58

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	338	LEU
1	C	158	MET
1	D	336	THR
1	B	353	ILE
1	C	8	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	271	ASN
1	C	96	ASN
1	D	263	HIS
1	B	419	ASN
1	C	139	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](#)

14 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	FAD	D	1700	-	51,58,58	1.55	5 (9%)	60,89,89	1.64	12 (20%)
2	FAD	C	1500	-	51,58,58	1.57	6 (11%)	60,89,89	1.63	10 (16%)
3	MPD	A	3005	-	7,7,7	0.52	0	9,10,10	0.45	0
2	FAD	C	1600	-	51,58,58	1.55	5 (9%)	60,89,89	1.60	11 (18%)
2	FAD	A	1100	-	51,58,58	1.55	5 (9%)	60,89,89	1.62	10 (16%)
2	FAD	A	1200	-	51,58,58	1.56	5 (9%)	60,89,89	1.61	11 (18%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	MPD	D	3004	-	7,7,7	0.55	0	9,10,10	0.48	0
2	FAD	B	1300	-	51,58,58	1.56	5 (9%)	60,89,89	1.61	10 (16%)
2	FAD	D	1800	-	51,58,58	1.55	5 (9%)	60,89,89	1.62	11 (18%)
3	MPD	A	3001	-	7,7,7	0.52	0	9,10,10	0.46	0
3	MPD	C	3003	-	7,7,7	0.58	0	9,10,10	0.52	0
2	FAD	B	1400	-	51,58,58	1.55	5 (9%)	60,89,89	1.69	11 (18%)
3	MPD	C	3006	-	7,7,7	0.51	0	9,10,10	0.52	0
3	MPD	B	3002	-	7,7,7	0.59	0	9,10,10	0.38	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	FAD	D	1700	-	-	3/30/50/50	0/6/6/6
2	FAD	C	1500	-	-	4/30/50/50	0/6/6/6
3	MPD	A	3005	-	-	0/5/5/5	-
2	FAD	C	1600	-	-	6/30/50/50	0/6/6/6
2	FAD	A	1100	-	-	3/30/50/50	0/6/6/6
2	FAD	A	1200	-	-	8/30/50/50	0/6/6/6
3	MPD	D	3004	-	-	0/5/5/5	-
2	FAD	B	1300	-	-	4/30/50/50	0/6/6/6
2	FAD	D	1800	-	-	5/30/50/50	0/6/6/6
3	MPD	A	3001	-	-	0/5/5/5	-
3	MPD	C	3003	-	-	0/5/5/5	-
2	FAD	B	1400	-	-	6/30/50/50	0/6/6/6
3	MPD	C	3006	-	-	0/5/5/5	-
3	MPD	B	3002	-	-	0/5/5/5	-

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1300	FAD	C10-N1	5.77	1.40	1.33
2	D	1700	FAD	C10-N1	5.74	1.40	1.33
2	C	1500	FAD	C10-N1	5.72	1.40	1.33
2	A	1100	FAD	C10-N1	5.70	1.40	1.33
2	C	1600	FAD	C10-N1	5.69	1.40	1.33

The worst 5 of 86 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1400	FAD	C4-N3-C2	6.34	120.50	115.14
2	A	1200	FAD	C4-N3-C2	5.97	120.18	115.14
2	D	1800	FAD	C4-N3-C2	5.77	120.02	115.14
2	C	1600	FAD	C4-N3-C2	5.68	119.94	115.14
2	D	1700	FAD	C4-N3-C2	5.56	119.84	115.14

There are no chirality outliers.

5 of 39 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	1600	FAD	N10-C1'-C2'-O2'
2	C	1600	FAD	N10-C1'-C2'-C3'
2	A	1200	FAD	N10-C1'-C2'-O2'
2	A	1200	FAD	N10-C1'-C2'-C3'
2	D	1800	FAD	N10-C1'-C2'-O2'

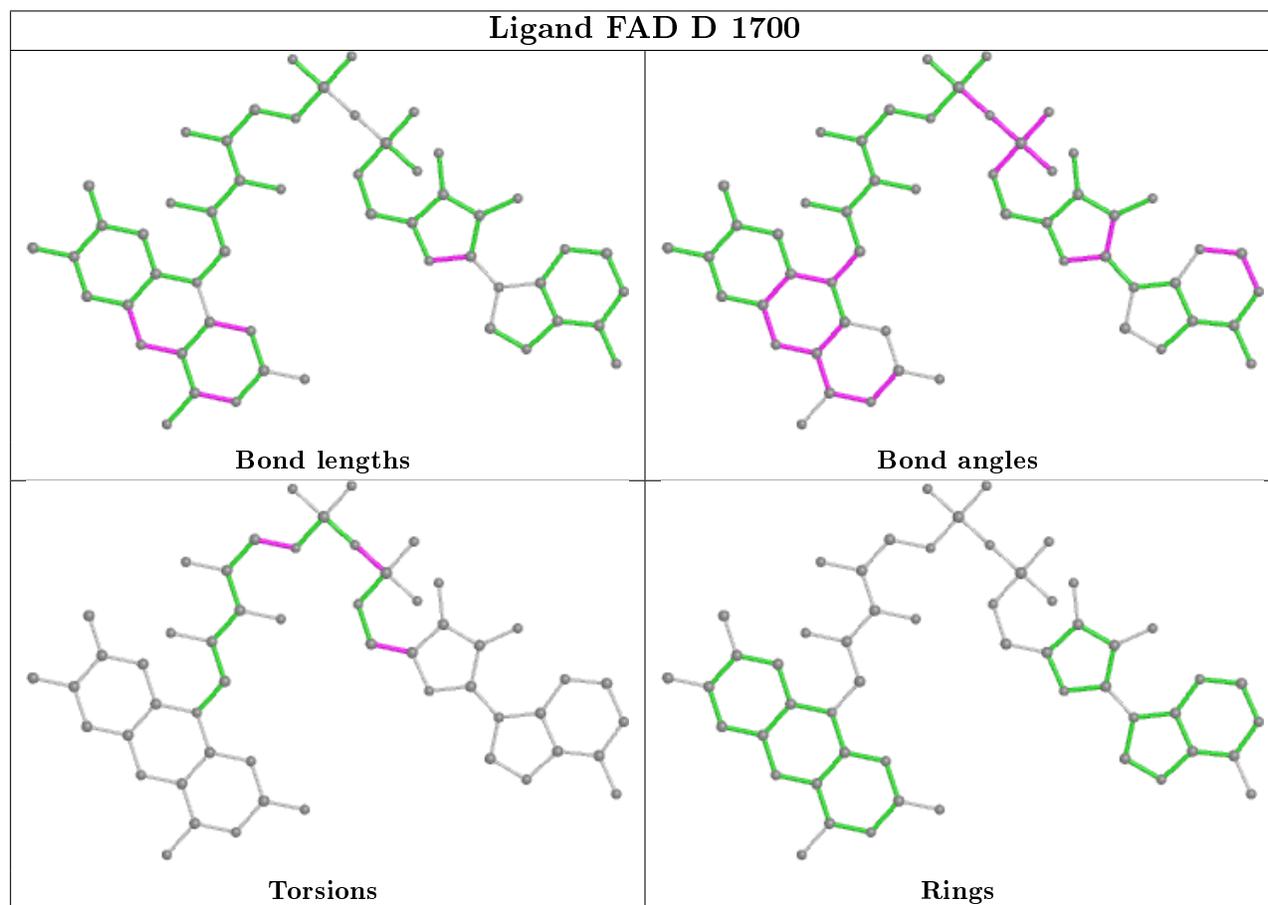
There are no ring outliers.

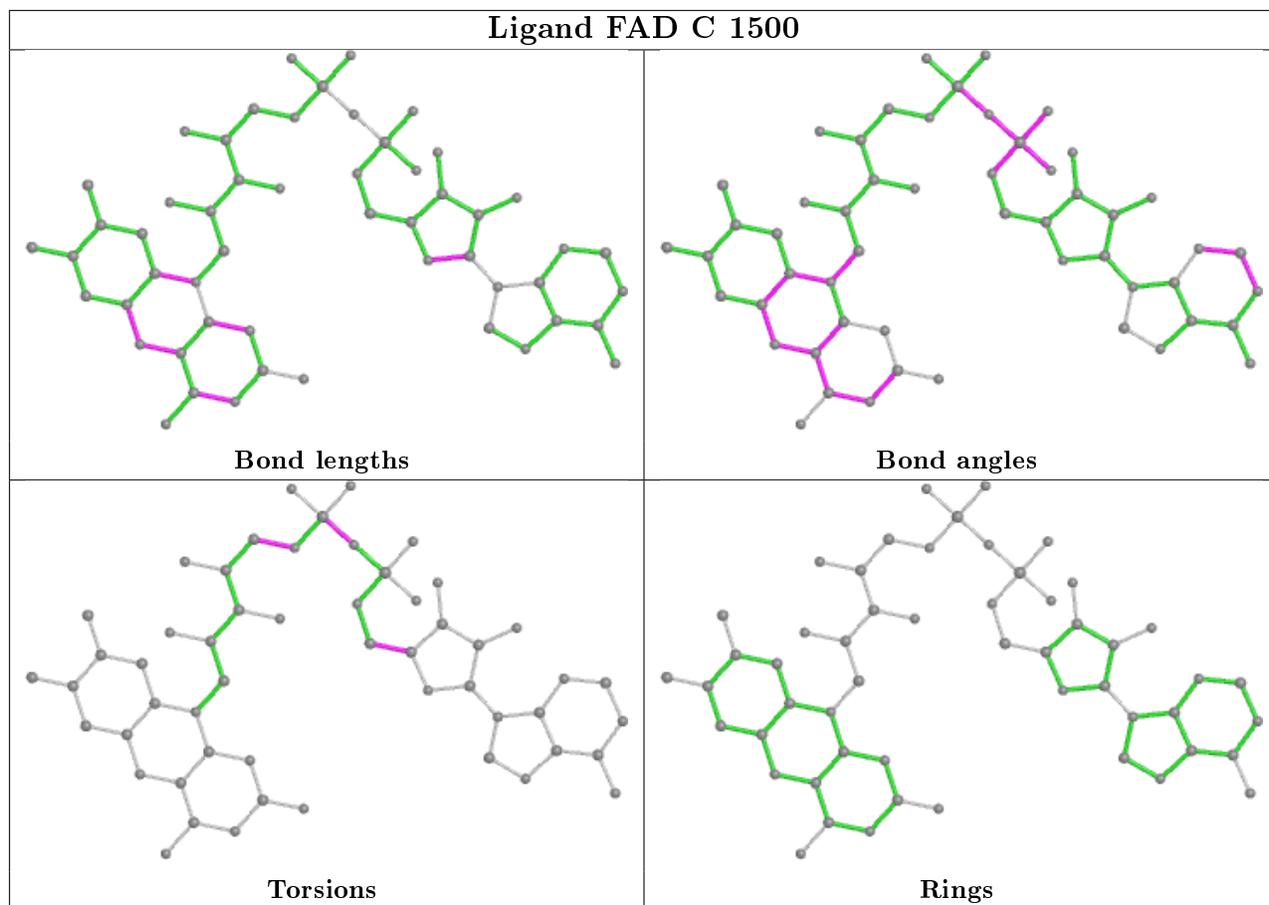
14 monomers are involved in 33 short contacts:

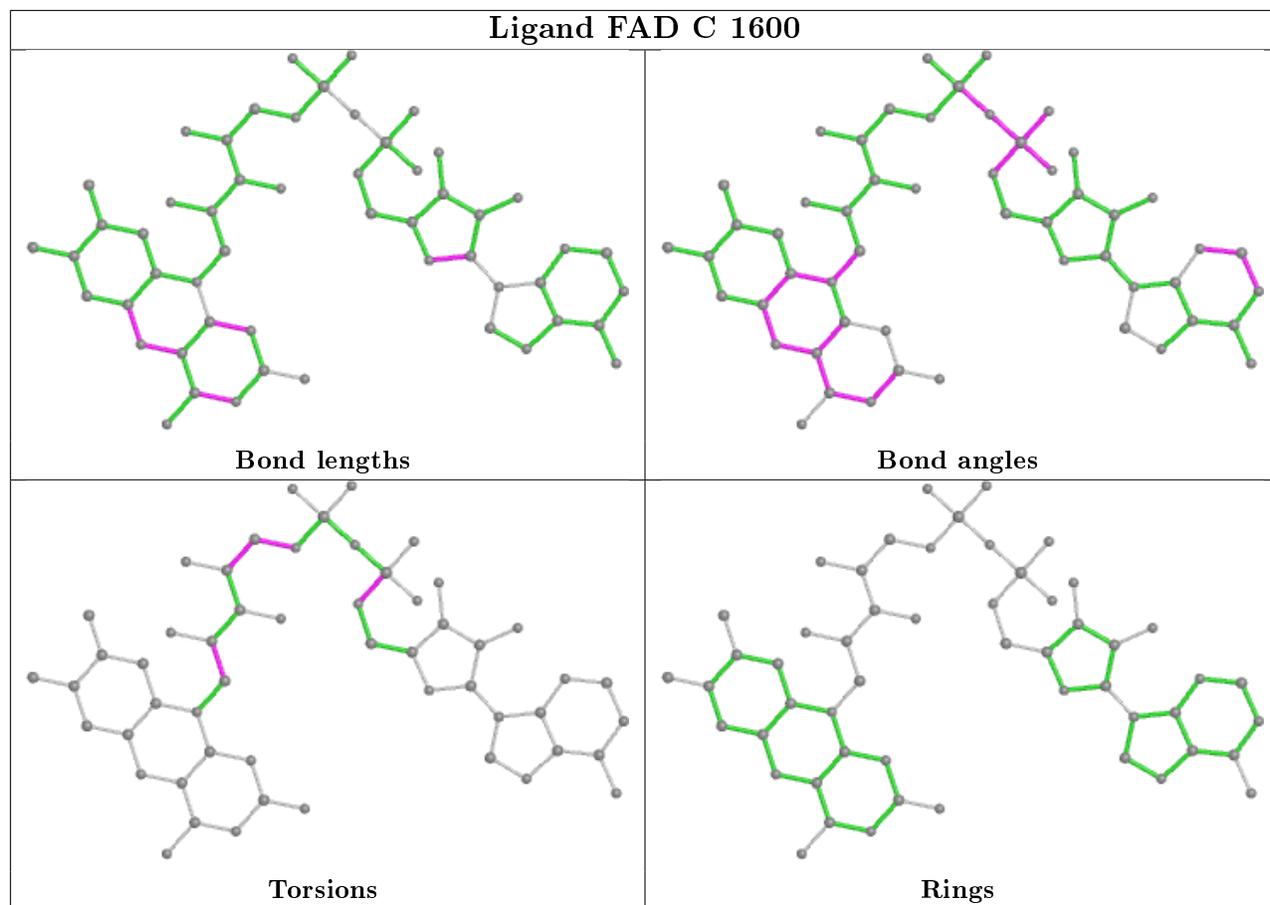
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	1700	FAD	5	0
2	C	1500	FAD	3	0
3	A	3005	MPD	1	0
2	C	1600	FAD	2	0
2	A	1100	FAD	3	0
2	A	1200	FAD	3	0
3	D	3004	MPD	2	0
2	B	1300	FAD	3	0
2	D	1800	FAD	2	0
3	A	3001	MPD	4	0
3	C	3003	MPD	2	0
2	B	1400	FAD	3	0
3	C	3006	MPD	1	0
3	B	3002	MPD	3	0

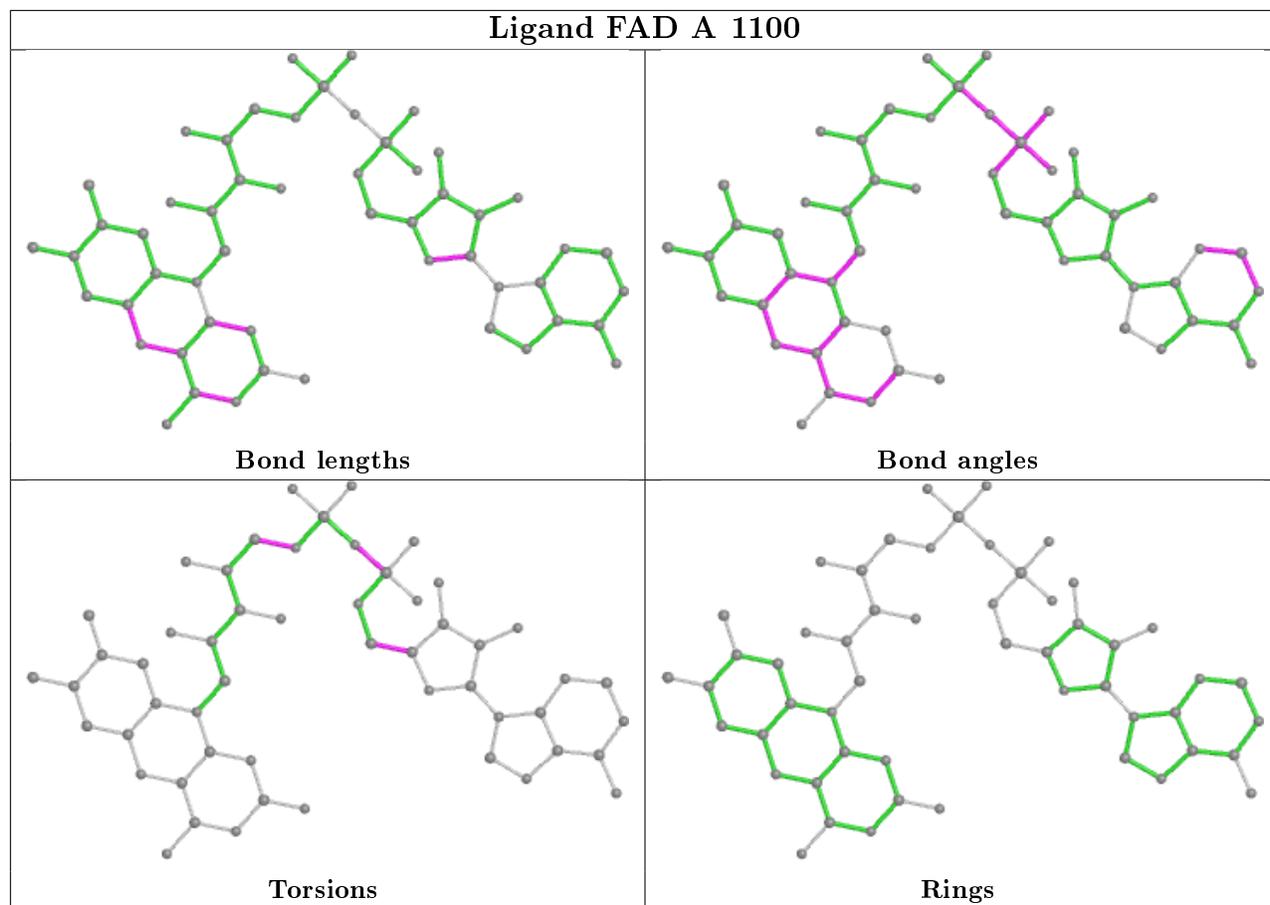
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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

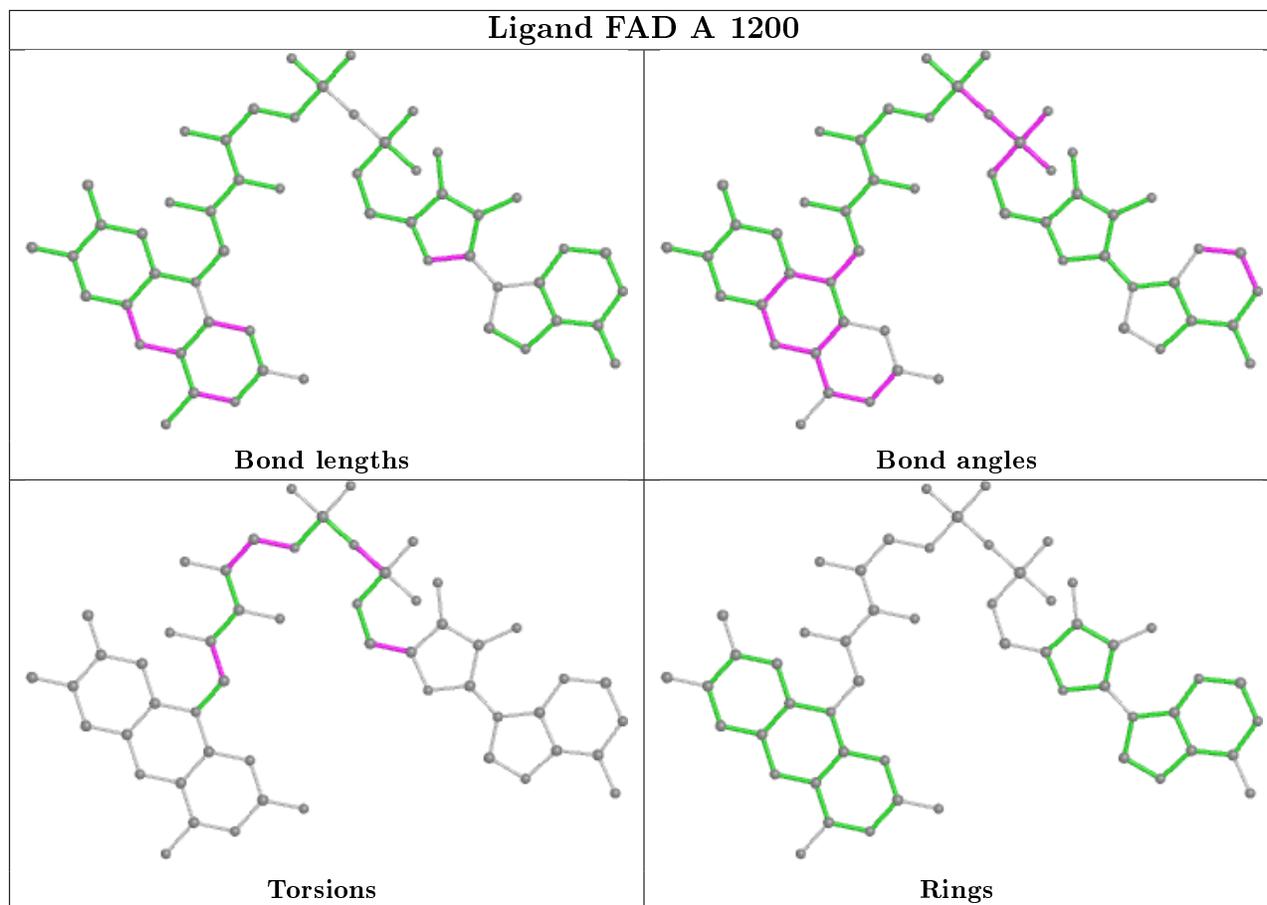
average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

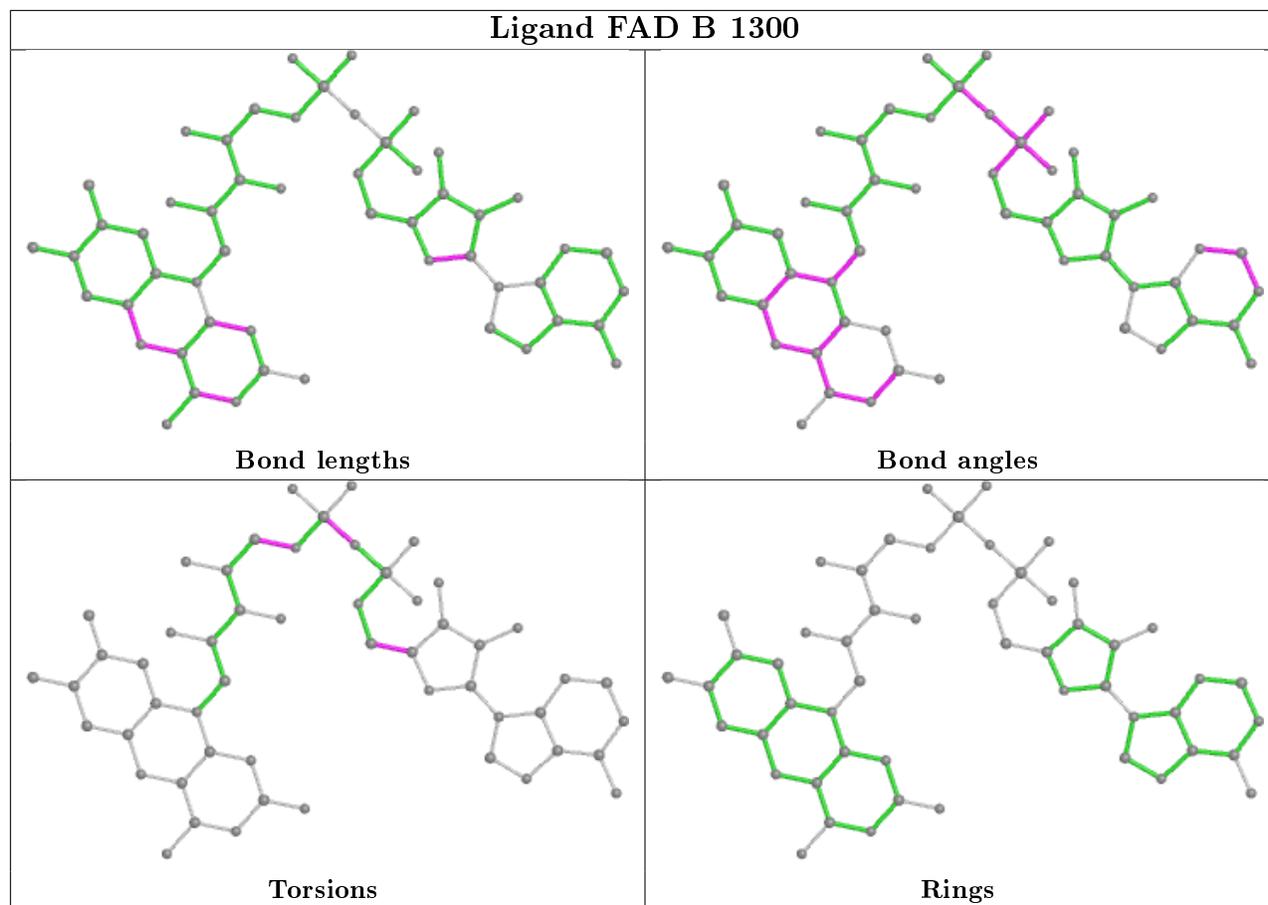


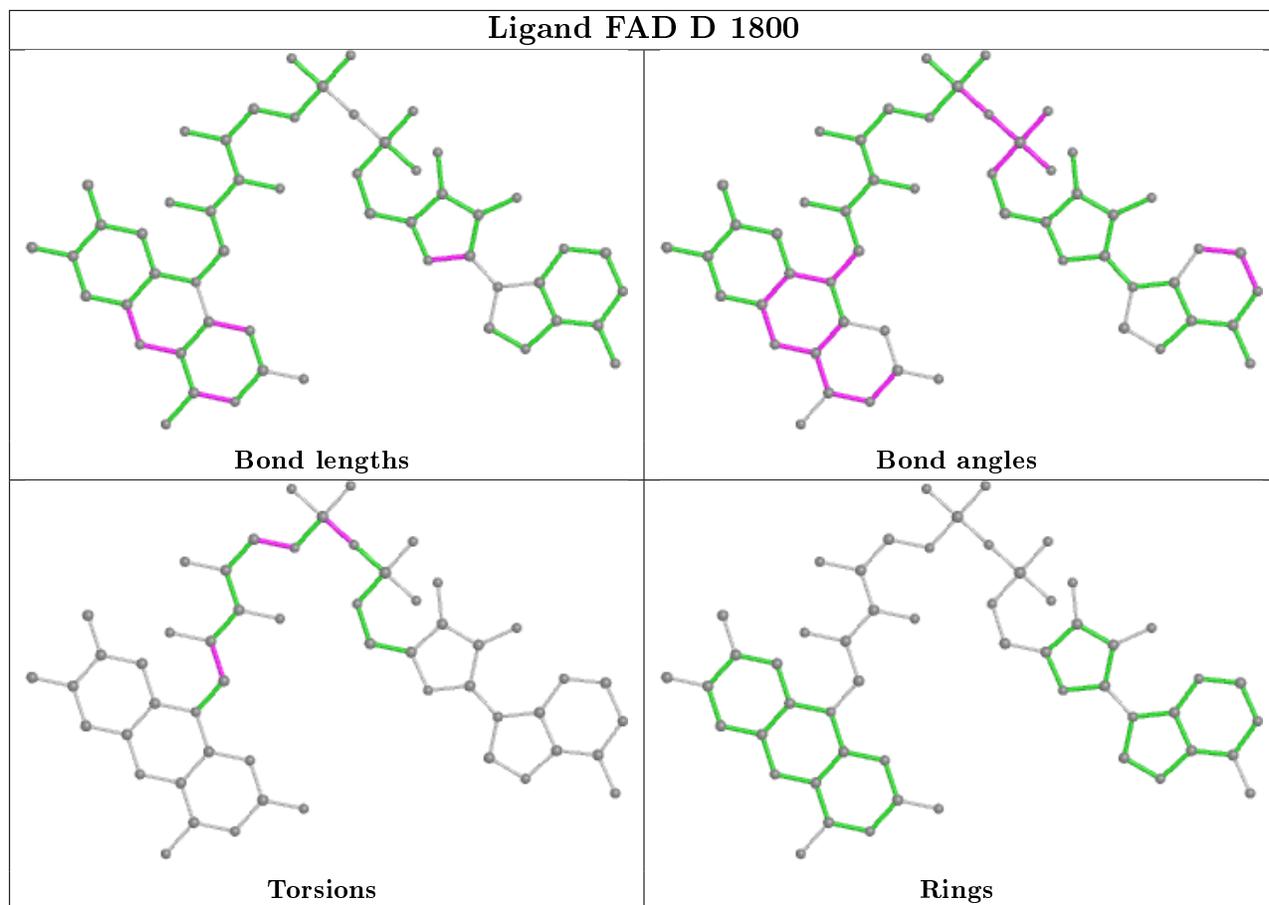


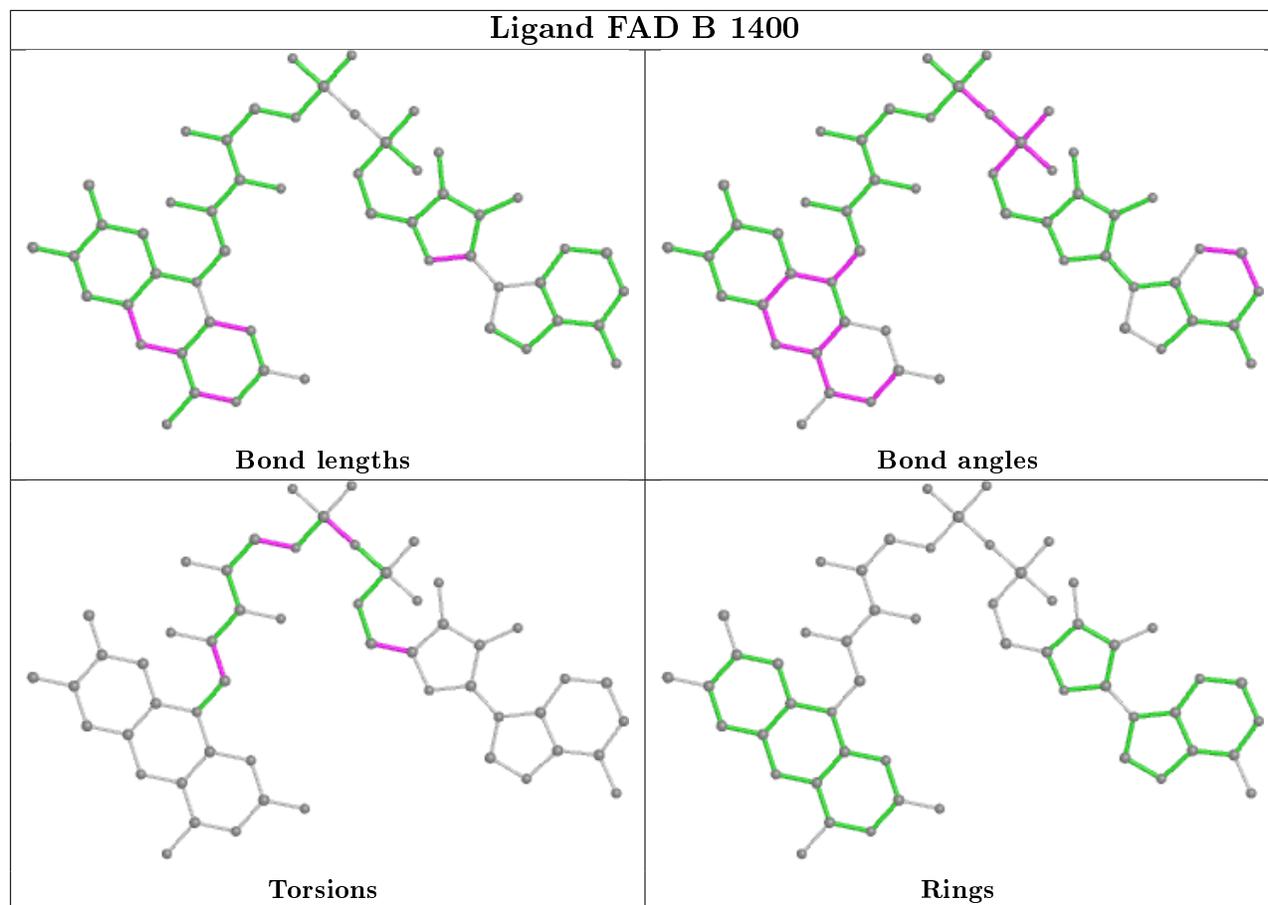












## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.