wwPDB X-ray Structure Validation Summary Report (i)

Oct 5, 2023 – 06:05 AM EDT

PDB ID	:	6VYP
Title	:	Crystal structure of the LSD1/CoREST histone demethylase bound to its nucleosome substrate
Authors	:	Kim, S.; Zhu, J.; Eek, P.; Yennawar, N.; Song, T.
Deposited on	:	2020-02-27
Resolution	:	4.99 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	:	FAILED
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 4.99 Å.

There are no overall percentile quality scores available for this entry.

MolProbity and EDS failed to run properly - the sequence quality summary graphics cannot be shown.

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 52996 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 Histone H3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	110	Total	C	N	O	S	0	0	0
			893	559	174	156	4			
1	E	110	Total	C	N	O	S	0	0	0
			889	557	174	154	4			
1	a	94	Total	C	N	O	S	0	0	0
			774	489	147	135	3			
1	e	103	Total	C	N	O	S	0	0	0
			836	525	158	149	4			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	MET	LYS	engineered mutation	UNP A0A310TTQ1
E	4	MET	LYS	engineered mutation	UNP A0A310TTQ1
a	4	MET	LYS	engineered mutation	UNP A0A310TTQ1
e	4	MET	LYS	engineered mutation	UNP A0A310TTQ1

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	81	Total	C	N	O	S	0	0	0
			648	410	126	111	1			
2	F	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			
2	b	80	Total	C	N	O	S	0	0	0
			638	401	125	111	1			
2	f	78	Total	C	N	O	S	0	0	0
			619	391	120	107	1			

- Molecule 3 is a protein called Histone H2A type 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	103	Total 795	C 501	N 155	O 139	0	0	0
3	G	105	Total 809	C 510	N 158	O 141	0	0	0
3	c	103	Total 795	C 501	N 155	O 139	0	0	0
3	g	104	Total 804	C 507	N 157	O 140	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	99	ARG	GLY	engineered mutation	UNP P06897
C	123	SER	ALA	engineered mutation	UNP P06897
G	99	ARG	GLY	engineered mutation	UNP P06897
G	123	SER	ALA	engineered mutation	UNP P06897
c	99	ARG	GLY	engineered mutation	UNP P06897
c	123	SER	ALA	engineered mutation	UNP P06897
g	99	ARG	GLY	engineered mutation	UNP P06897
g	123	SER	ALA	engineered mutation	UNP P06897

- Molecule 4 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	92	Total 719	C 453	N 129	O 135	S 2	0	0	0
4	H	93	Total 726	C 457	N 130	O 137	S 2	0	0	0
4	d	91	Total 708	C 447	N 125	O 134	S 2	0	0	0
4	h	93	Total 726	C 457	N 130	O 137	S 2	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	29	THR	SER	engineered mutation	UNP P02281
H	29	THR	SER	engineered mutation	UNP P02281
d	29	THR	SER	engineered mutation	UNP P02281
h	29	THR	SER	engineered mutation	UNP P02281

- Molecule 5 is a DNA chain called DNA (191-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	I	191	Total	C	N	O	P	0	0	0
			3895	1849	710	1145	191			
5	i	191	Total	C	N	O	P	0	0	0
			3895	1849	710	1145	191			

- Molecule 6 is a DNA chain called DNA (191-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	J	191	Total	C	N	O	P	0	0	0
			3936	1863	735	1147	191			
6	j	191	Total	C	N	O	P	0	0	0
			3936	1863	735	1147	191			

- Molecule 7 is a protein called Lysine-specific histone demethylase 1A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	M	666	Total	C	N	O	S	0	0	0
			5205	3319	902	964	20			
7	m	666	Total	C	N	O	S	0	0	0
			5205	3319	902	964	20			
7	K	666	Total	C	N	O	S	0	0	0
			5205	3319	902	964	20			
7	k	666	Total	C	N	O	S	0	0	0
			5205	3319	902	964	20			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	169	GLY	-	expression tag	UNP O60341
M	170	SER	-	expression tag	UNP O60341
M	608	ALA	ARG	engineered mutation	UNP O60341
M	717	ALA	ASN	engineered mutation	UNP O60341
M	721	ALA	ASP	engineered mutation	UNP O60341
m	169	GLY	-	expression tag	UNP O60341
m	170	SER	-	expression tag	UNP O60341
m	608	ALA	ARG	engineered mutation	UNP O60341
m	717	ALA	ASN	engineered mutation	UNP O60341
m	721	ALA	ASP	engineered mutation	UNP O60341
K	169	GLY	-	expression tag	UNP O60341
K	170	SER	-	expression tag	UNP O60341
K	608	ALA	ARG	engineered mutation	UNP O60341
K	717	ALA	ASN	engineered mutation	UNP O60341
K	721	ALA	ASP	engineered mutation	UNP O60341

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
k	169	GLY	-	expression tag	UNP O60341
k	170	SER	-	expression tag	UNP O60341
k	608	ALA	ARG	engineered mutation	UNP O60341
k	717	ALA	ASN	engineered mutation	UNP O60341
k	721	ALA	ASP	engineered mutation	UNP O60341

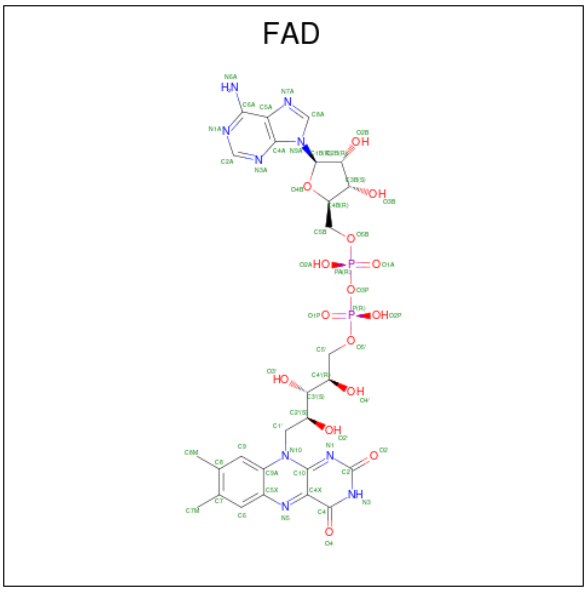
- Molecule 8 is a protein called REST corepressor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	N	133	Total	C	N	O	S	0	0	0
			1076	676	194	203	3			
8	n	133	Total	C	N	O	S	0	0	0
			1076	676	194	203	3			
8	L	133	Total	C	N	O	S	0	0	0
			1076	676	194	203	3			
8	l	133	Total	C	N	O	S	0	0	0
			1076	676	194	203	3			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
N	284	GLY	-	expression tag	UNP Q9UKL0
N	285	SER	-	expression tag	UNP Q9UKL0
n	284	GLY	-	expression tag	UNP Q9UKL0
n	285	SER	-	expression tag	UNP Q9UKL0
L	284	GLY	-	expression tag	UNP Q9UKL0
L	285	SER	-	expression tag	UNP Q9UKL0
l	284	GLY	-	expression tag	UNP Q9UKL0
l	285	SER	-	expression tag	UNP Q9UKL0

- Molecule 9 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula: $C_{27}H_{33}N_9O_{15}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
9	M	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
9	m	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
9	K	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
9	k	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

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

3 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	103.77Å 335.77Å 174.63Å 90.00° 91.07° 90.00°	Depositor
Resolution (Å)	20.17 – 4.99	Depositor
% Data completeness (in resolution range)	99.3 (20.17-4.99)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.25 (at 4.97Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.229 , 0.277	Depositor
Wilson B-factor (Å^2)	309.0	Xtriage
Anisotropy	0.119	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.165 for h,-k,-l	Xtriage
Total number of atoms	52996	wwPDB-VP
Average B, all atoms (Å^2)	415.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.43% 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.

4 Model quality

4.1 Standard geometry

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

4.2 Too-close contacts

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

4.3 Torsion angles

4.3.1 Protein backbone

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

4.3.2 Protein sidechains

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

4.3.3 RNA

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

4.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

4.5 Carbohydrates

There are no monosaccharides in this entry.

4.6 Ligand geometry

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond

length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
9	FAD	K	900	-	53,58,58	0.48	0	68,89,89	0.54	2 (2%)
9	FAD	m	900	-	53,58,58	0.46	0	68,89,89	0.53	2 (2%)
9	FAD	k	900	-	53,58,58	0.54	0	68,89,89	0.60	1 (1%)
9	FAD	M	900	-	53,58,58	0.45	0	68,89,89	0.51	2 (2%)

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
9	FAD	K	900	-	-	13/30/50/50	0/6/6/6
9	FAD	m	900	-	-	15/30/50/50	0/6/6/6
9	FAD	k	900	-	-	18/30/50/50	0/6/6/6
9	FAD	M	900	-	-	15/30/50/50	0/6/6/6

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
9	m	900	FAD	P-O3P-PA	-2.60	123.90	132.83
9	K	900	FAD	P-O3P-PA	-2.52	124.17	132.83
9	M	900	FAD	P-O3P-PA	-2.43	124.50	132.83
9	M	900	FAD	C5A-C6A-N6A	2.29	123.83	120.35
9	K	900	FAD	C5A-C6A-N6A	2.28	123.82	120.35

There are no chirality outliers.

5 of 61 torsion outliers are listed below:

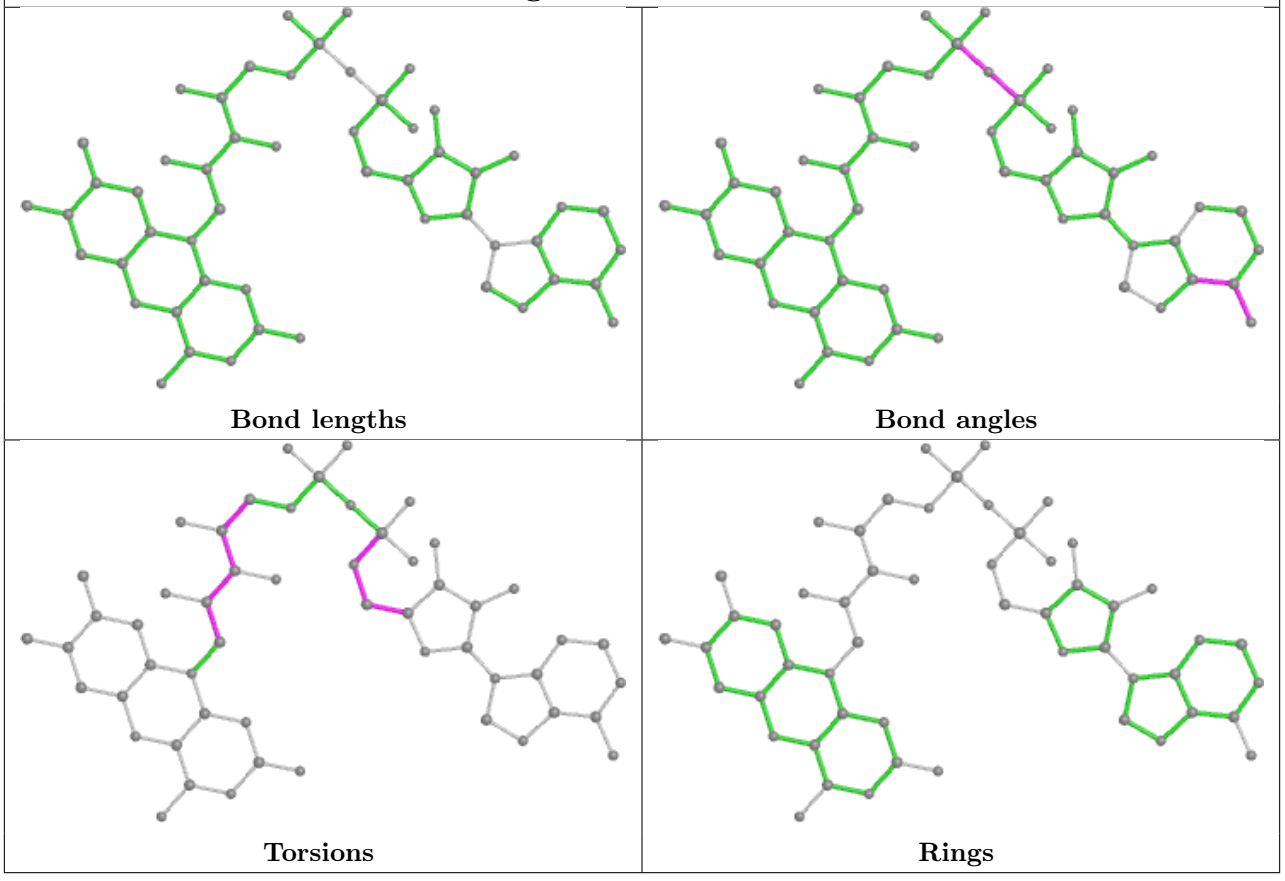
Mol	Chain	Res	Type	Atoms
9	M	900	FAD	C5B-O5B-PA-O1A
9	M	900	FAD	C5B-O5B-PA-O2A
9	M	900	FAD	C3'-C4'-C5'-O5'
9	M	900	FAD	O4'-C4'-C5'-O5'
9	M	900	FAD	C5'-O5'-P-O1P

There are no ring outliers.

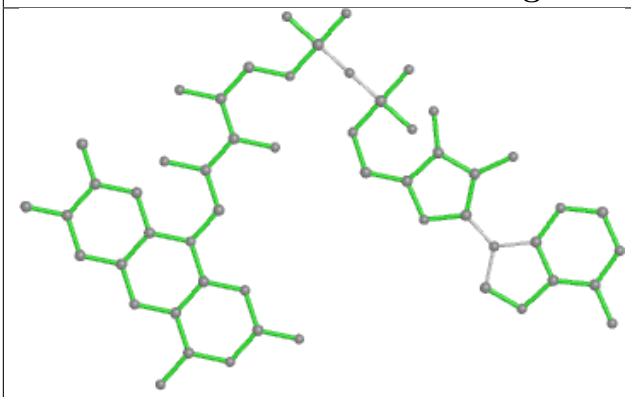
No monomer is involved in short contacts.

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 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.

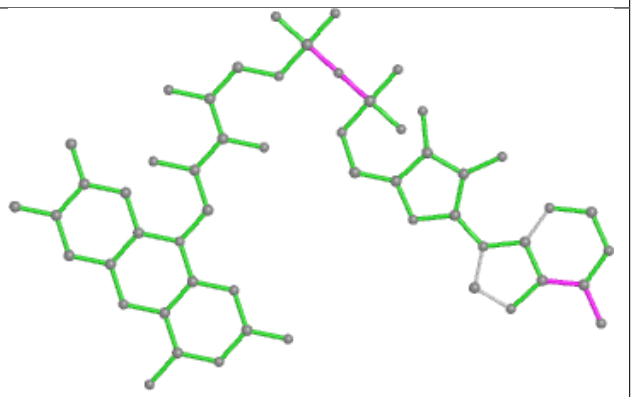
Ligand FAD K 900



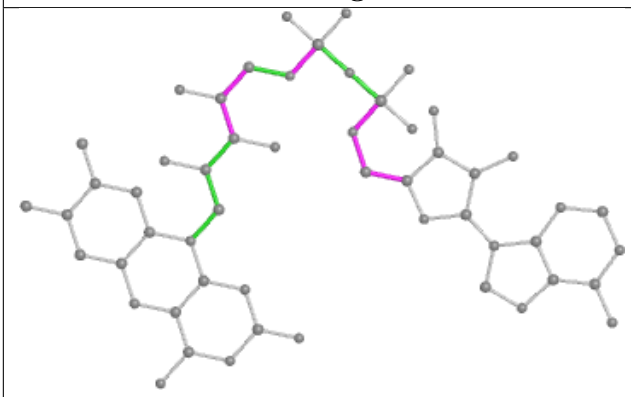
Ligand FAD m 900



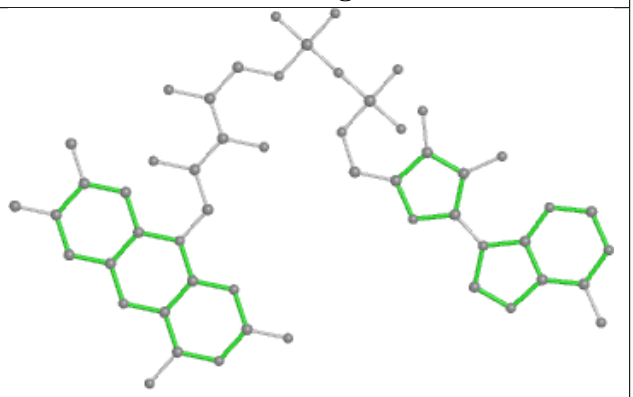
Bond lengths



Bond angles

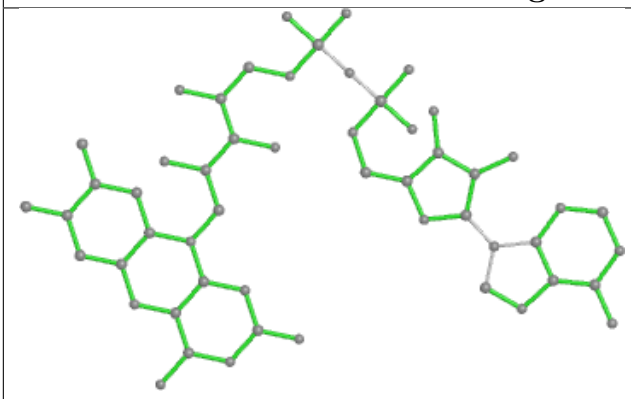


Torsions

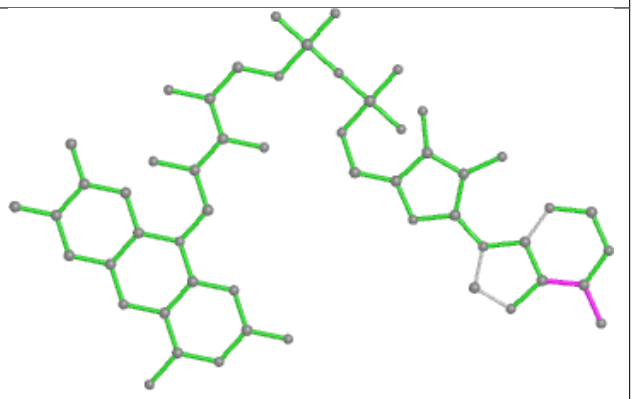


Rings

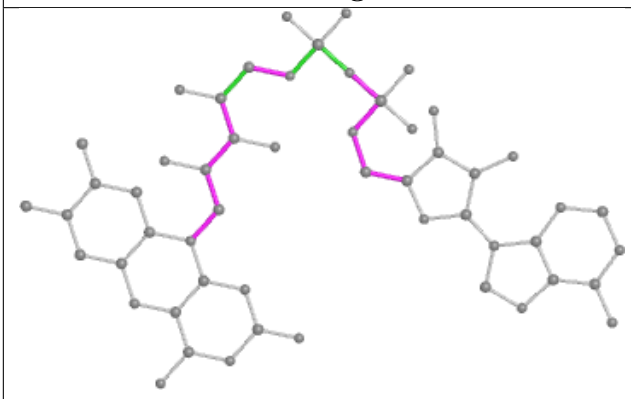
Ligand FAD k 900



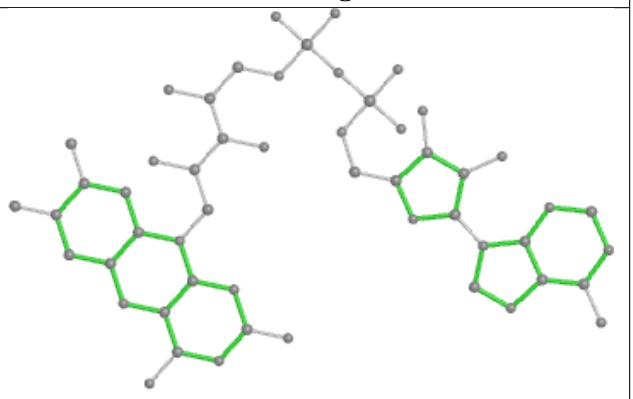
Bond lengths



Bond angles

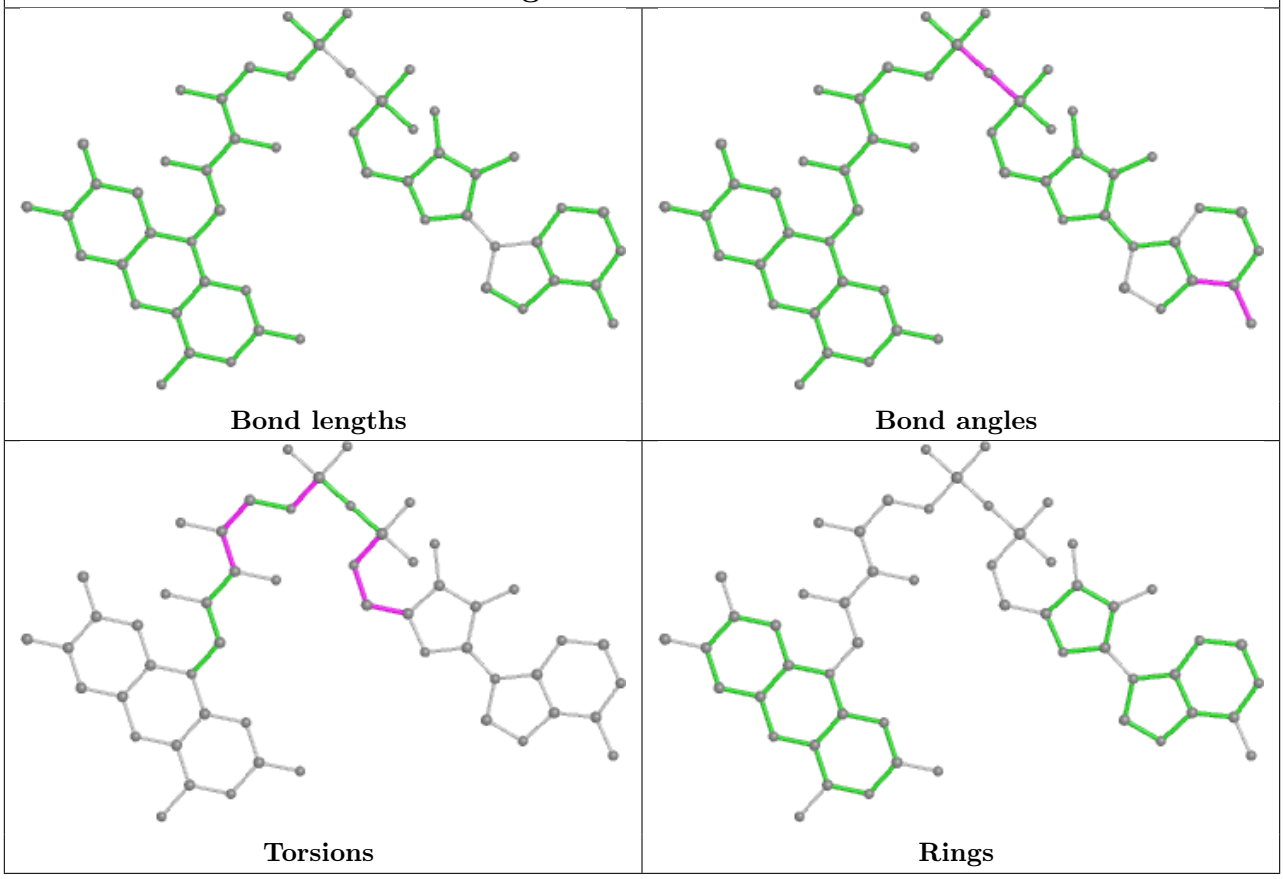


Torsions



Rings

Ligand FAD M 900



4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data

5.1 Protein, DNA and RNA chains

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

5.2 Non-standard residues in protein, DNA, RNA chains

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

5.3 Carbohydrates

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

5.4 Ligands

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

5.5 Other polymers

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