Full wwPDB X-ray Structure Validation Report (i)

Mar 13, 2023 – 03:32 PM JST

PDB ID	:	8IJY
Title	:	Synechococcus elongatus 6-4 photolyase with an 8-HDF as the antenna chromophore and a covalently linked FAD as the catalytic cofactor
Deposited on	:	2023-02-28
Resolution	:	2.29 Å(reported)

This wwPDB validation report is for manuscript review

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

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
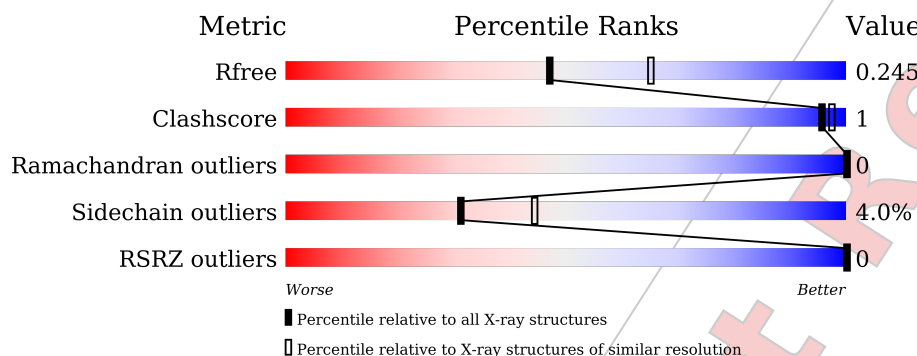http://www.wwpdb.org/validation/2017/FAQs#types.

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 2.29 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	Percentile Ranks	Value
Rfree		0.245
Clashscore		1
Ramachandran outliers		0
Sidechain outliers		4.0%
RSRZ outliers		0

Worse *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	514	92% 5% ·

Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 4336 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 Deoxyribodipyrimidine photolyase-related protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	501	4109	2635	724	733	17	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	504	ALA	-	expression tag	UNP Q31P18
A	505	ALA	-	expression tag	UNP Q31P18
A	506	ALA	-	expression tag	UNP Q31P18
A	507	LEU	-	expression tag	UNP Q31P18
A	508	GLU	-	expression tag	UNP Q31P18
A	509	HIS	-	expression tag	UNP Q31P18
A	510	HIS	-	expression tag	UNP Q31P18
A	511	HIS	-	expression tag	UNP Q31P18
A	512	HIS	-	expression tag	UNP Q31P18
A	513	HIS	-	expression tag	UNP Q31P18
A	514	HIS	-	expression tag	UNP Q31P18

- Molecule 2 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).

SF4
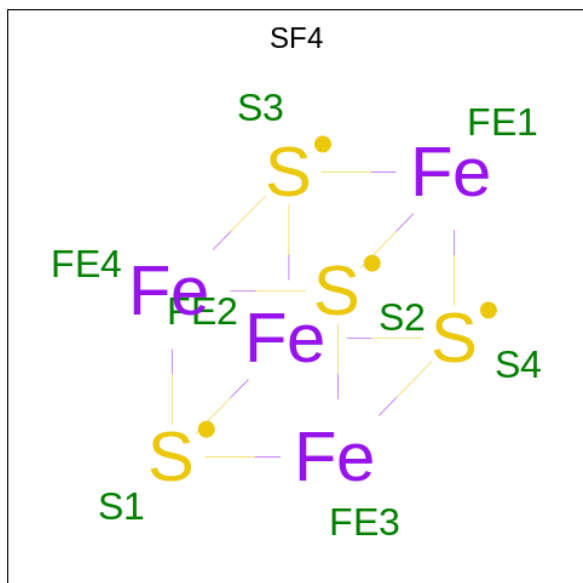
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	Fe	S	0	0
			8	4	4		

- Molecule 3 is 8-HYDROXY-10-(D-RIBO-2,3,4,5-TETRAHYDROXYPENTYL)-5-DEAZAI SOALLOXAZINE (three-letter code: HDF) (formula: $C_{16}H_{17}N_3O_7$) (labeled as "Ligand of Interest" by depositor).

HDF
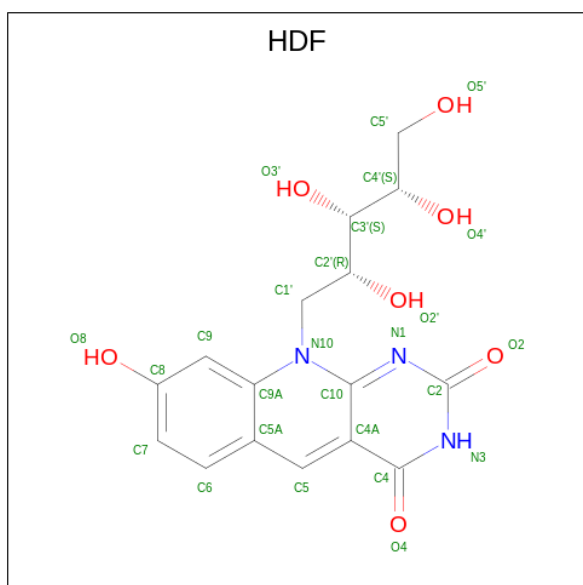
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			26	16	3	7		

- Molecule 4 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula:

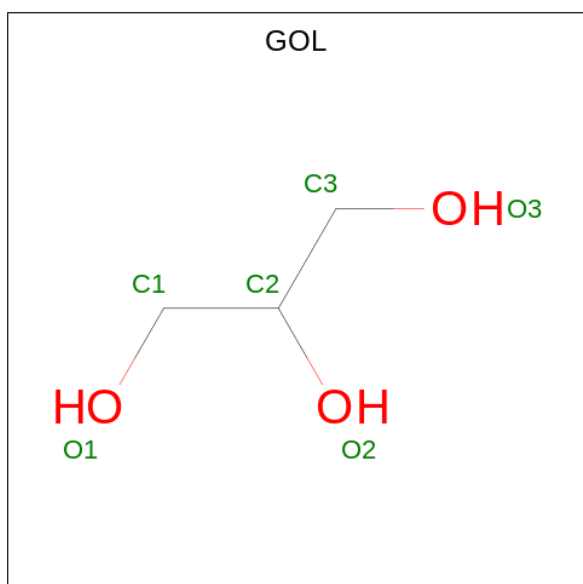$C_{27}H_{33}N_9O_{15}P_2$) (labeled as "Ligand of Interest" by depositor).

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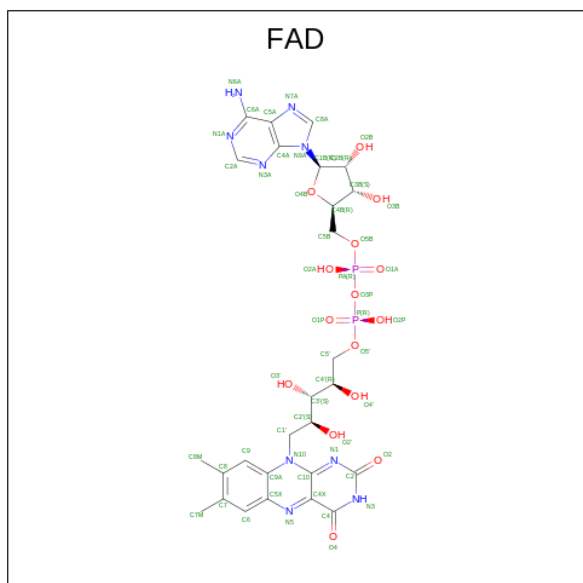
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total 53	C 27	N 9	O 15	P 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 6	C 3	O 3	0	0
5	A	1	Total 6	C 3	O 3	0	0

- Molecule 6 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (three-letter code: TRS) (formula: $C_4H_{12}NO_3$).

TRS
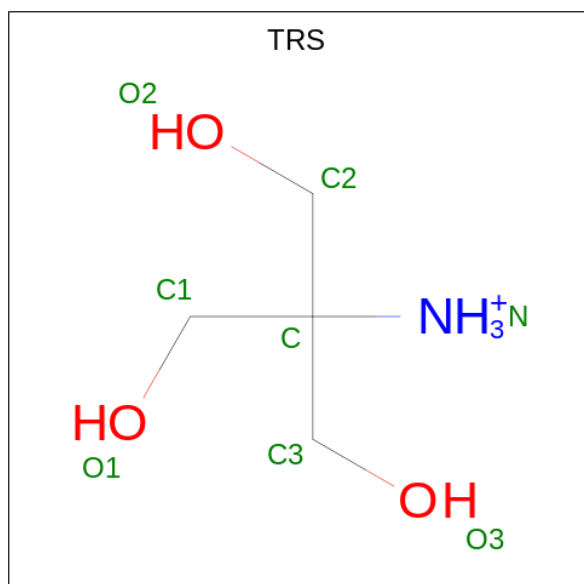
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total 8	C 4	N 1	O 3	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	120	Total 120	O 120	0	0

3 Residue-property plots ⓘ

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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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: Deoxyribodipyrimidine photolyase-related protein

Chain A:
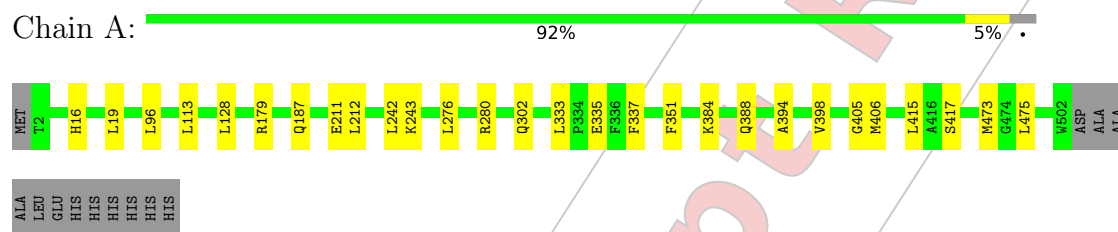
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	105.03Å 109.14Å 95.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	75.79 – 2.29 47.35 – 2.29	Depositor EDS
% Data completeness (in resolution range)	91.4 (75.79-2.29) 91.4 (47.35-2.29)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.94 (at 2.29Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.179 , 0.242 0.187 , 0.245	Depositor DCC
R_{free} test set	1166 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	30.9	Xtriage
Anisotropy	0.008	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 36.4	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.024 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4336	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.82% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: SF4, FAD, GOL, TRS, HDF

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	0.75	0/4235	0.81	0/5766

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	4109	0	3960	8	0
2	A	8	0	0	0	0
3	A	26	0	16	1	0
4	A	53	0	31	3	0
5	A	12	0	16	0	0
6	A	8	0	12	0	0
7	A	120	0	0	0	0
All	All	4336	0	4035	9	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.

All (9) 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:406:MET:HG2	1:A:415:LEU:HD22	1.93	0.51
1:A:398:VAL:HG21	4:A:603:FAD:C2	2.40	0.51
1:A:337:PHE:HA	1:A:351:PHE:CE1	2.52	0.45
1:A:388:GLN:HA	1:A:394:ALA:HB1	1.99	0.44
3:A:602:HDF:H9	3:A:602:HDF:H1'1	1.67	0.44
1:A:405:GLY:HA3	1:A:415:LEU:HD13	2.00	0.43
1:A:302:GLN:HB3	4:A:603:FAD:C8A	2.48	0.42
1:A:302:GLN:HB3	4:A:603:FAD:H8A	2.01	0.42
1:A:16:HIS:HB3	1:A:19:LEU:HG	2.03	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	499/514 (97%)	485 (97%)	14 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	430/440 (98%)	413 (96%)	17 (4%)	31	44

All (17) residues with a non-rotameric sidechain are listed below:

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Mol	Chain	Res	Type
1	A	96	LEU
1	A	113	LEU
1	A	128	LEU
1	A	179	ARG
1	A	187	GLN
1	A	211	GLU
1	A	212	LEU
1	A	242	LEU
1	A	243	LYS
1	A	276	LEU
1	A	280	ARG
1	A	333	LEU
1	A	335	GLU
1	A	384	LYS
1	A	417	SER
1	A	473	MET
1	A	475	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	158	GLN
1	A	165	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

6 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	$\#|Z| > 2$
5	GOL	A	604	-	5,5,5	0.27	0	5,5,5	0.56	0
4	FAD	A	603	1	53,58,58	1.54	8 (15%)	68,89,89	1.52	12 (17%)
3	HDF	A	602	-	25,28,28	1.89	5 (20%)	29,41,41	3.29	11 (37%)
2	SF4	A	601	1	0,12,12	-	-	-		
5	GOL	A	605	-	5,5,5	0.44	0	5,5,5	0.42	0
6	TRS	A	606	-	7,7,7	0.42	0	9,9,9	0.90	1 (11%)

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
5	GOL	A	604	-	-	0/4/4/4	-
4	FAD	A	603	1	-	2/30/50/50	0/6/6/6
3	HDF	A	602	-	-	10/14/14/14	0/3/3/3
2	SF4	A	601	1	-	-	0/6/5/5
5	GOL	A	605	-	-	0/4/4/4	-
6	TRS	A	606	-	-	3/9/9/9	-

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	602	HDF	C4A-C10	6.90	1.48	1.41
4	A	603	FAD	C4X-N5	5.15	1.40	1.30
4	A	603	FAD	C9A-C5X	5.00	1.49	1.41
3	A	602	HDF	C5A-C9A	3.83	1.48	1.42
4	A	603	FAD	C5A-C4A	2.72	1.48	1.40
4	A	603	FAD	C2-N3	-2.69	1.32	1.39
4	A	603	FAD	C5X-N5	2.59	1.44	1.39
4	A	603	FAD	C8-C7	2.59	1.47	1.40
3	A	602	HDF	C2-N3	-2.59	1.33	1.38
4	A	603	FAD	C4-N3	-2.58	1.34	1.38

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	603	FAD	O4B-C1B	2.26	1.44	1.41
3	A	602	HDF	C2-N1	-2.25	1.33	1.38
3	A	602	HDF	C9A-N10	2.25	1.41	1.38

All (24) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	602	HDF	C4A-C4-N3	-8.69	118.34	124.40
3	A	602	HDF	C2-N3-C4	8.49	122.31	115.14
3	A	602	HDF	C1'-N10-C10	7.67	125.28	118.41
4	A	603	FAD	C4-C4X-N5	5.13	125.53	118.23
3	A	602	HDF	C5-C4A-C10	4.59	122.83	117.11
4	A	603	FAD	N3A-C2A-N1A	-4.11	122.25	128.68
3	A	602	HDF	C9A-N10-C10	-3.95	116.73	121.91
3	A	602	HDF	C5A-C9A-N10	3.39	122.03	118.71
4	A	603	FAD	C3B-C2B-C1B	3.20	105.79	100.98
3	A	602	HDF	O4'-C4'-C3'	-3.16	101.42	109.10
4	A	603	FAD	C4X-C10-N10	2.86	120.66	116.48
4	A	603	FAD	C10-N1-C2	2.85	122.60	116.90
4	A	603	FAD	C6-C5X-N5	2.71	123.25	118.51
3	A	602	HDF	C5'-C4'-C3'	2.67	118.19	112.41
3	A	602	HDF	C4A-C5-C5A	-2.56	118.47	122.27
4	A	603	FAD	C4X-C10-N1	-2.52	118.89	124.73
4	A	603	FAD	C4X-C4-N3	2.42	119.33	113.19
4	A	603	FAD	C9A-C5X-N5	-2.41	119.81	122.43
3	A	602	HDF	C4'-C3'-C2'	2.39	118.33	113.36
3	A	602	HDF	C10-C4A-C4	-2.36	117.45	119.99
4	A	603	FAD	C10-C4X-N5	-2.33	119.92	124.86
4	A	603	FAD	C4A-C5A-N7A	-2.30	107.00	109.40
6	A	606	TRS	O3-C3-C	2.11	117.67	111.00
4	A	603	FAD	C2A-N1A-C6A	2.10	122.34	118.75

There are no chirality outliers.

All (15) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	602	HDF	C2'-C1'-N10-C9A
3	A	602	HDF	C2'-C1'-N10-C10
3	A	602	HDF	N10-C1'-C2'-O2'
3	A	602	HDF	C1'-C2'-C3'-O3'
3	A	602	HDF	C1'-C2'-C3'-C4'
3	A	602	HDF	O2'-C2'-C3'-O3'

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
6	A	606	TRS	C3-C-C1-O1
6	A	606	TRS	N-C-C1-O1
3	A	602	HDF	O2'-C2'-C3'-C4'
3	A	602	HDF	O4'-C4'-C5'-O5'
6	A	606	TRS	C2-C-C1-O1
3	A	602	HDF	N10-C1'-C2'-C3'
4	A	603	FAD	C4'-C5'-O5'-P
4	A	603	FAD	C4B-C5B-O5B-PA
3	A	602	HDF	C3'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	603	FAD	3	0
3	A	602	HDF	1	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 FAD A 603

Bond lengths

Bond angles

Torsions

Rings

Ligand HDF A 602

Bond lengths

Bond angles

Torsions

Rings
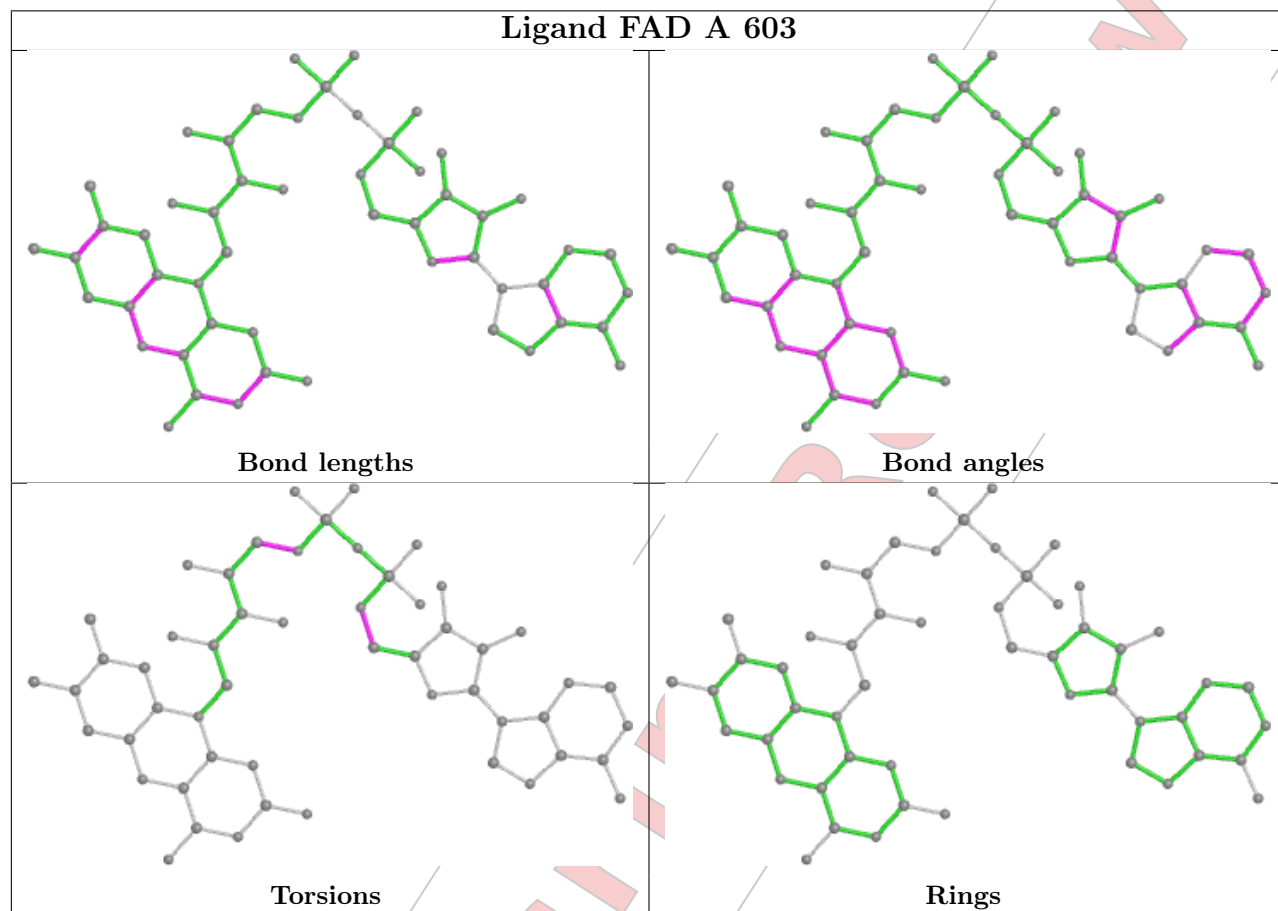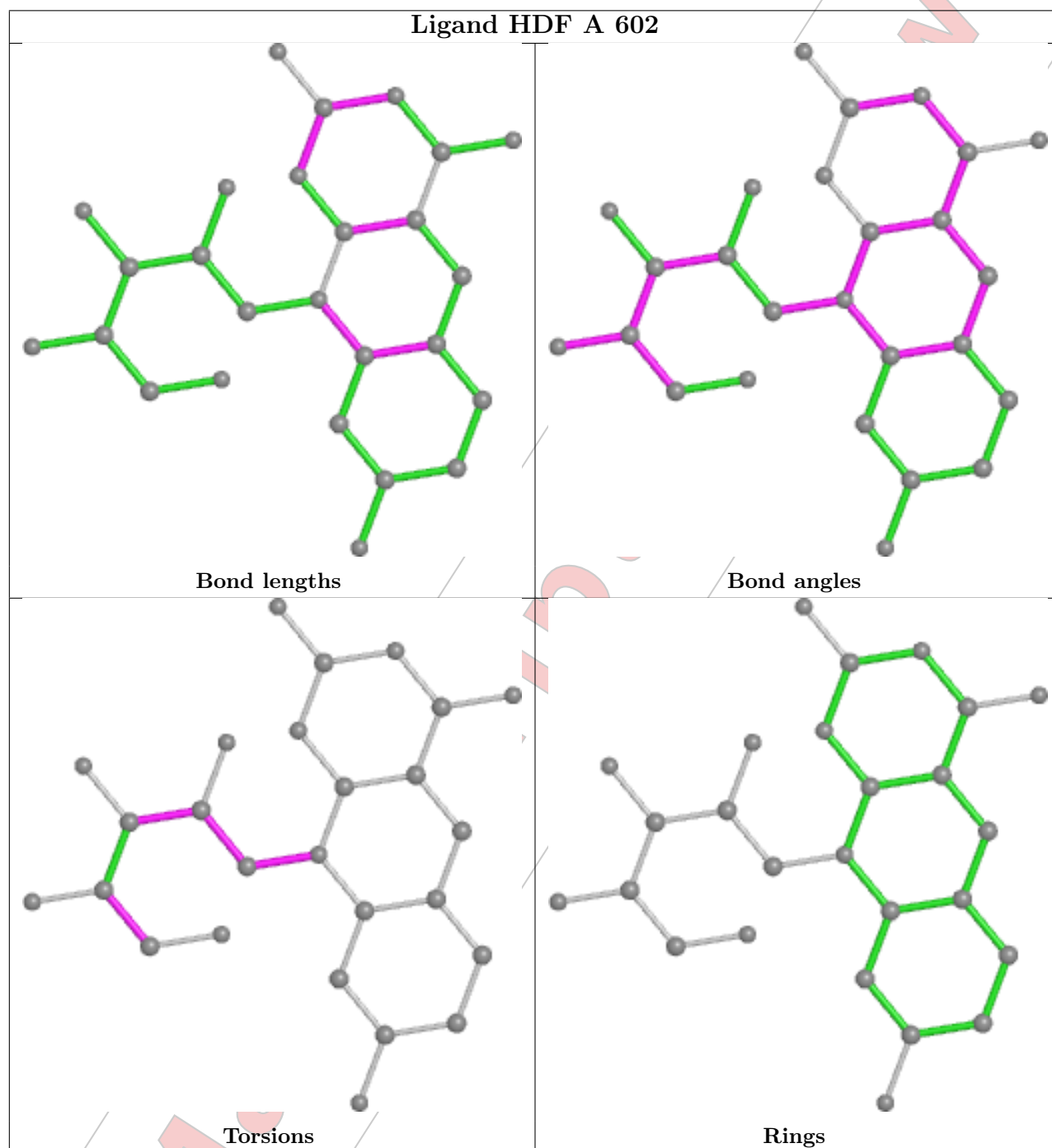
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	501/514 (97%)	-0.44	0	100	100	20, 30, 50, 73	0

There are no RSRZ outliers to report.

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.
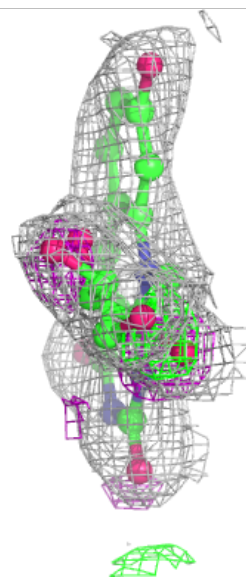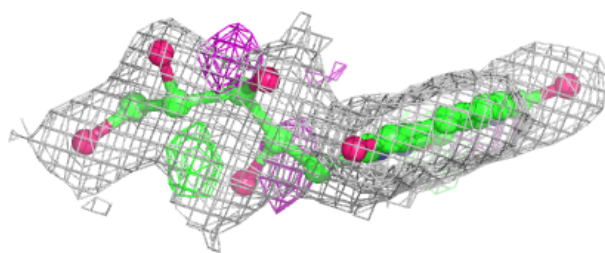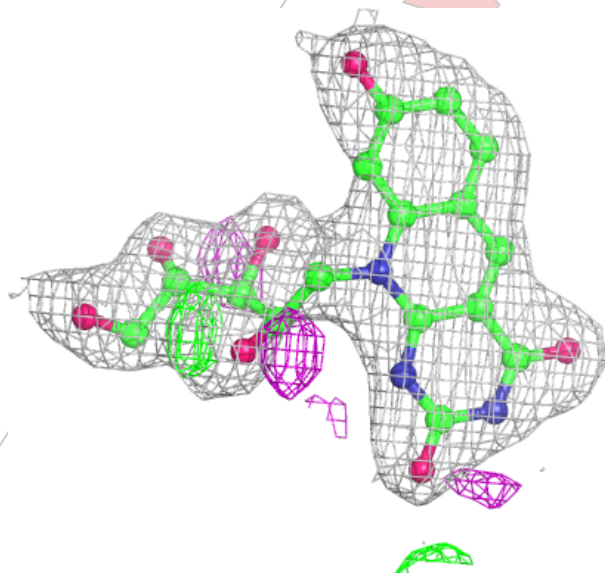
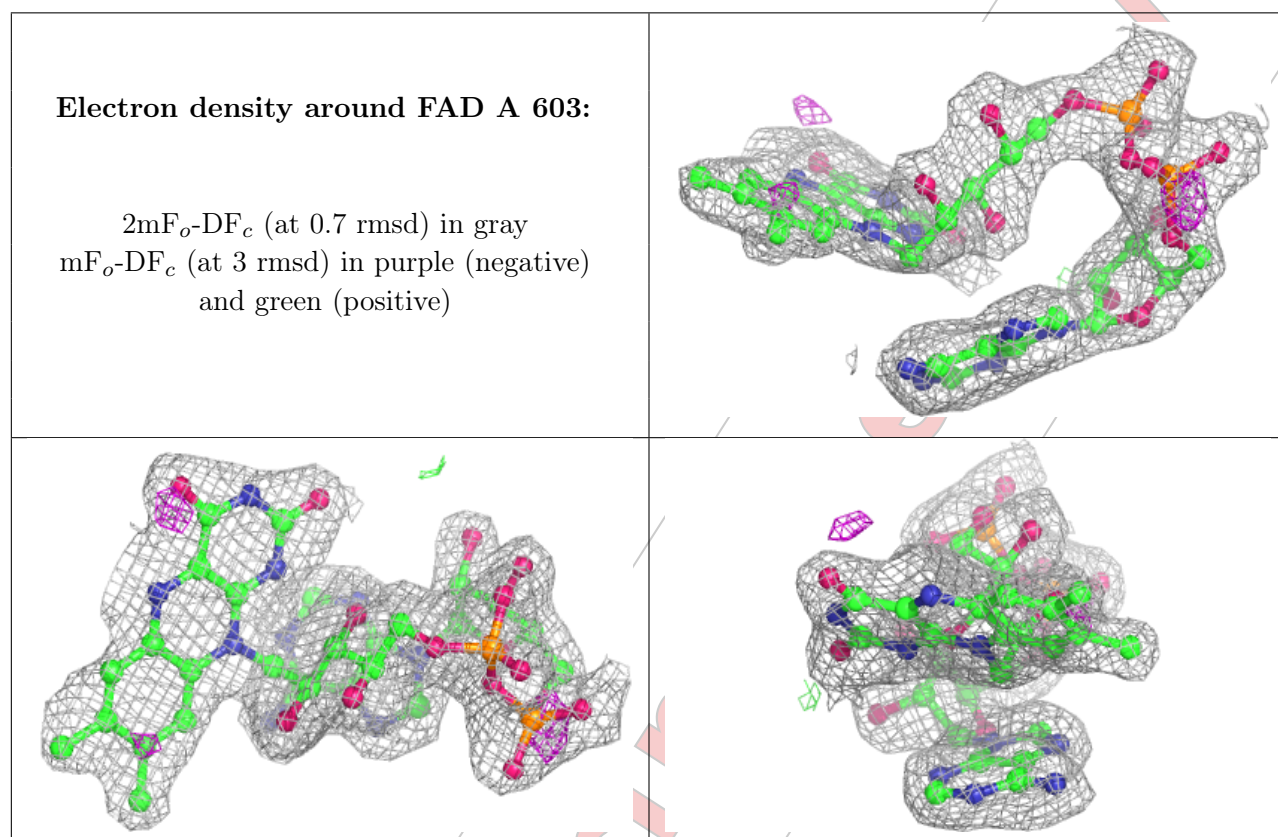| Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	GOL	A	605	6/6	0.80	0.18	44,52,54,56	0
3	HDF	A	602	26/26	0.90	0.15	25,31,34,39	0
6	TRS	A	606	8/8	0.93	0.14	39,46,47,47	0
5	GOL	A	604	6/6	0.94	0.13	39,41,46,49	0
4	FAD	A	603	53/53	0.95	0.11	25,30,33,36	0
2	SF4	A	601	8/8	0.99	0.08	23,25,26,26	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different

orientation to approximate a three-dimensional view.

Electron density around HDF A 602:

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

Electron density around FAD A 603:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.