



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 20, 2023 – 11:27 AM EDT

PDB ID : 5KGJ  
Title : X-ray structure of a glucosamine N-Acetyltransferase from *Clostridium acetobutylicum* in complex with galactosamine  
Authors : Dopkins, B.J.; Thoden, J.B.; Tipton, P.A.; Holden, H.M.  
Deposited on : 2016-06-13  
Resolution : 1.90 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 types of validation reports are described at

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

---

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) : 1.13  
EDS : 2.35.1  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35.1

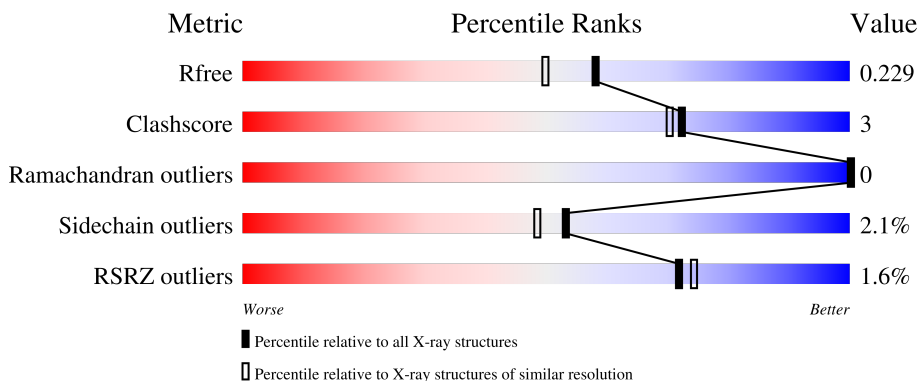
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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))
$R_{free}$	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	328	

## 2 Entry composition i

There are 7 unique types of molecules in this entry. The entry contains 2927 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.

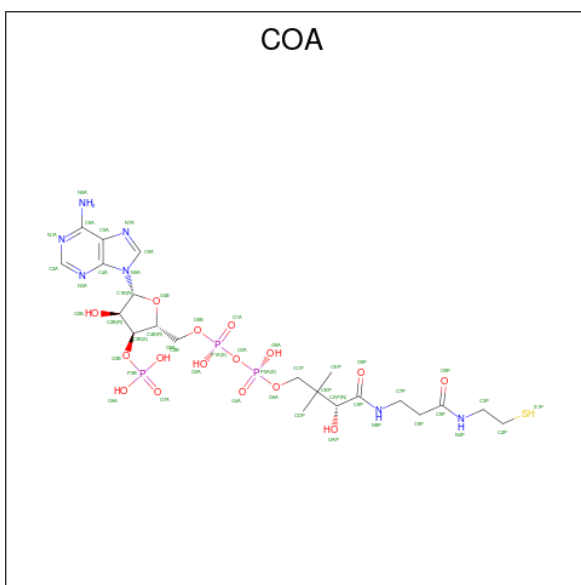
- Molecule 1 is a protein called Predicted acetyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	316	2636	1734	429	463	10	0	5	0

There are 8 discrepancies between the modelled and reference sequences:

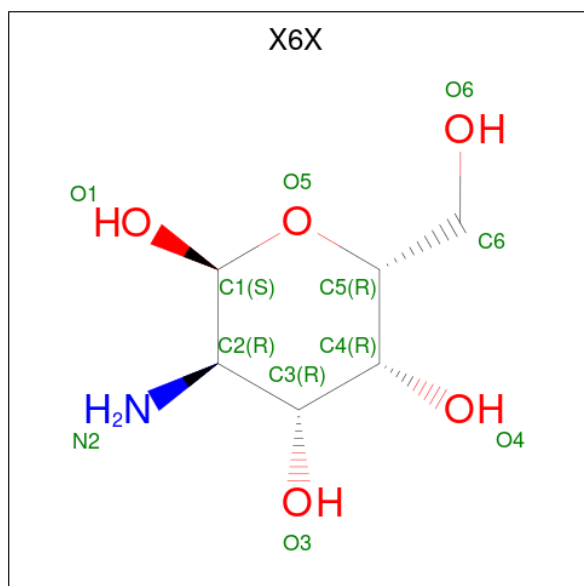
Chain	Residue	Modelled	Actual	Comment	Reference
A	321	GLY	-	expression tag	UNP Q97ML2
A	322	GLY	-	expression tag	UNP Q97ML2
A	323	HIS	-	expression tag	UNP Q97ML2
A	324	HIS	-	expression tag	UNP Q97ML2
A	325	HIS	-	expression tag	UNP Q97ML2
A	326	HIS	-	expression tag	UNP Q97ML2
A	327	HIS	-	expression tag	UNP Q97ML2
A	328	HIS	-	expression tag	UNP Q97ML2

- Molecule 2 is COENZYME A (three-letter code: COA) (formula:  $C_{21}H_{36}N_7O_{16}P_3S$ ).



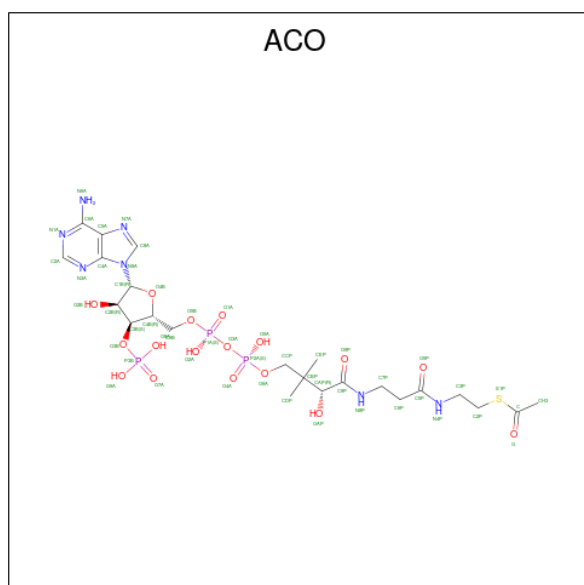
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	23	9	2	10	2	0	0

- Molecule 3 is 2-amino-2-deoxy-alpha-D-galactopyranose (three-letter code: X6X) (formula:  $C_6H_{13}NO_5$ ).



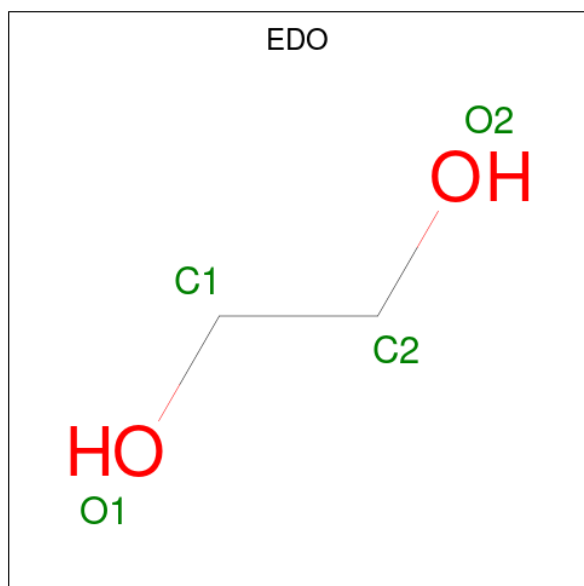
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	12	6	1	5	0	0

- Molecule 4 is ACETYL COENZYME \*A (three-letter code: ACO) (formula:  $C_{23}H_{38}N_7O_{17}P_3S$ ).



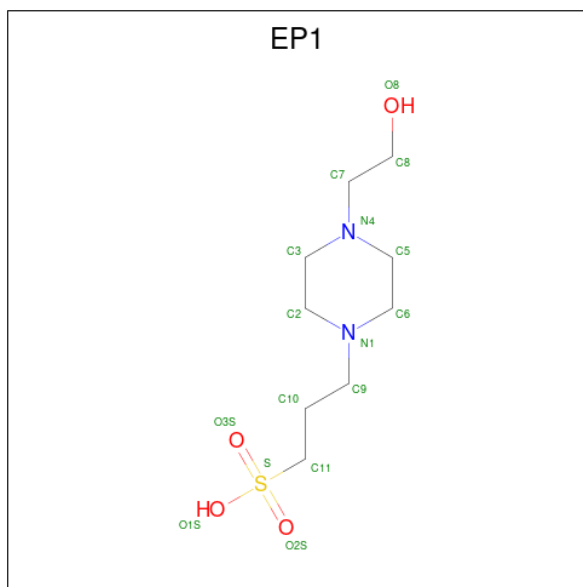
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	N	O	P			S
4	A	1	51	23	7	17	3	1	0	0

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	1	4	2	0	0
5	A	1	4	2	0	0
5	A	1	4	2	0	0

- Molecule 6 is 3-[4-(2-HYDROXYETHYL)PIPERAZIN-1-YL]PROPANE-1-SULFONIC ACID (three-letter code: EP1) (formula:  $C_9H_{20}N_2O_4S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
6	A	1	16	9	2	4	1	0	0

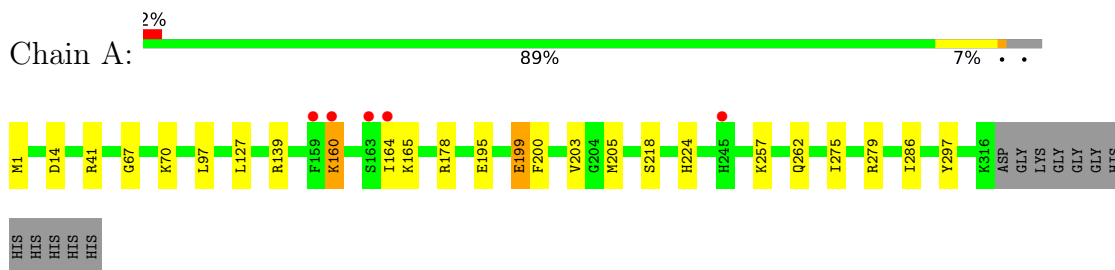
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	177	177	177	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 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: Predicted acetyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	119.08Å 43.84Å 74.17Å 90.00° 120.81° 90.00°	Depositor
Resolution (Å)	50.00 – 1.90 29.52 – 1.90	Depositor EDS
% Data completeness (in resolution range)	97.9 (50.00-1.90) 98.0 (29.52-1.90)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 1.91Å)	Xtrriage
Refinement program	REFMAC 5.8.0124	Depositor
R, $R_{free}$	0.167 , 0.215 0.179 , 0.229	Depositor DCC
$R_{free}$ test set	1275 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.6	Xtrriage
Anisotropy	0.066	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2927	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.72% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: ACO, EDO, EP1, X6X, COA

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	0.61	0/2720	0.85	3/3655 (0.1%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	178	ARG	NE-CZ-NH2	-6.64	116.98	120.30
1	A	178	ARG	NE-CZ-NH1	5.95	123.28	120.30
1	A	14	ASP	CB-CG-OD2	-5.22	113.60	118.30

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	2636	0	2671	15	0
2	A	23	0	15	0	0
3	A	12	0	11	1	0
4	A	51	0	34	1	0
5	A	12	0	18	0	0
6	A	16	0	19	3	0
7	A	177	0	0	3	0
All	All	2927	0	2768	19	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 3.

The worst 5 of 19 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:199:GLU:HG3	7:A:594:HOH:O	1.85	0.77
1:A:41:ARG:HD2	7:A:667:HOH:O	1.87	0.74
4:A:403:ACO:O9A	7:A:501:HOH:O	2.09	0.70
1:A:97:LEU:HD23	1:A:139:ARG:CD	2.23	0.68
1:A:97:LEU:HD23	1:A:139:ARG:HD3	1.79	0.63

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	319/328 (97%)	308 (97%)	11 (3%)	0	<b>100</b>   <b>100</b>

There are no Ramachandran outliers to report.

### 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	287/291 (99%)	281 (98%)	6 (2%)	53   48

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	199	GLU
1	A	218	SER
1	A	257	LYS
1	A	70	LYS
1	A	1	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	49	ASN
1	A	144	GLN
1	A	274	GLN
1	A	295	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

7 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	EDO	A	405	-	3,3,3	0.42	0	2,2,2	0.33	0
5	EDO	A	406	-	3,3,3	0.59	0	2,2,2	0.19	0
3	X6X	A	402	-	12,12,12	0.79	0	16,17,17	2.39	4 (25%)
5	EDO	A	404	-	3,3,3	0.35	0	2,2,2	0.56	0
6	EP1	A	407	-	16,16,16	2.57	2 (12%)	21,21,21	1.63	2 (9%)
4	ACO	A	403	-	45,53,53	0.88	1 (2%)	56,79,79	1.53	6 (10%)
2	COA	A	401	-	17,22,50	0.89	0	27,33,75	1.99	7 (25%)

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	EDO	A	405	-	-	1/1/1/1	-
5	EDO	A	406	-	-	0/1/1/1	-
3	X6X	A	402	-	-	0/2/22/22	0/1/1/1
5	EDO	A	404	-	-	0/1/1/1	-
6	EP1	A	407	-	-	8/10/20/20	0/1/1/1
4	ACO	A	403	-	-	3/47/67/67	0/3/3/3
2	COA	A	401	-	-	2/29/29/64	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	A	407	EP1	C11-S	-8.22	1.65	1.77
6	A	407	EP1	O2S-S	5.68	1.61	1.45
4	A	403	ACO	C5A-C4A	2.48	1.47	1.40

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	402	X6X	O5-C1-C2	6.48	117.07	109.51
2	A	401	COA	CEP-CBP-CCP	5.34	116.94	108.23
6	A	407	EP1	O1S-S-C11	5.16	114.12	105.77
4	A	403	ACO	N3A-C2A-N1A	-5.06	120.77	128.68
3	A	402	X6X	C1-O5-C5	4.36	121.88	113.66

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

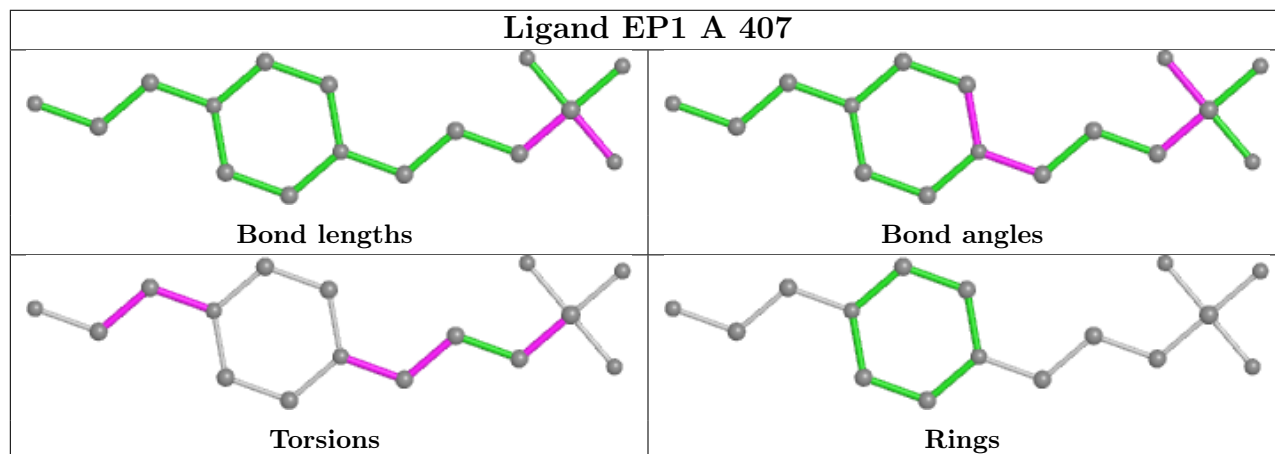
Mol	Chain	Res	Type	Atoms
4	A	403	ACO	C5B-O5B-P1A-O1A
4	A	403	ACO	C5B-O5B-P1A-O2A
4	A	403	ACO	C5B-O5B-P1A-O3A
6	A	407	EP1	C10-C11-S-O2S
6	A	407	EP1	C10-C11-S-O3S

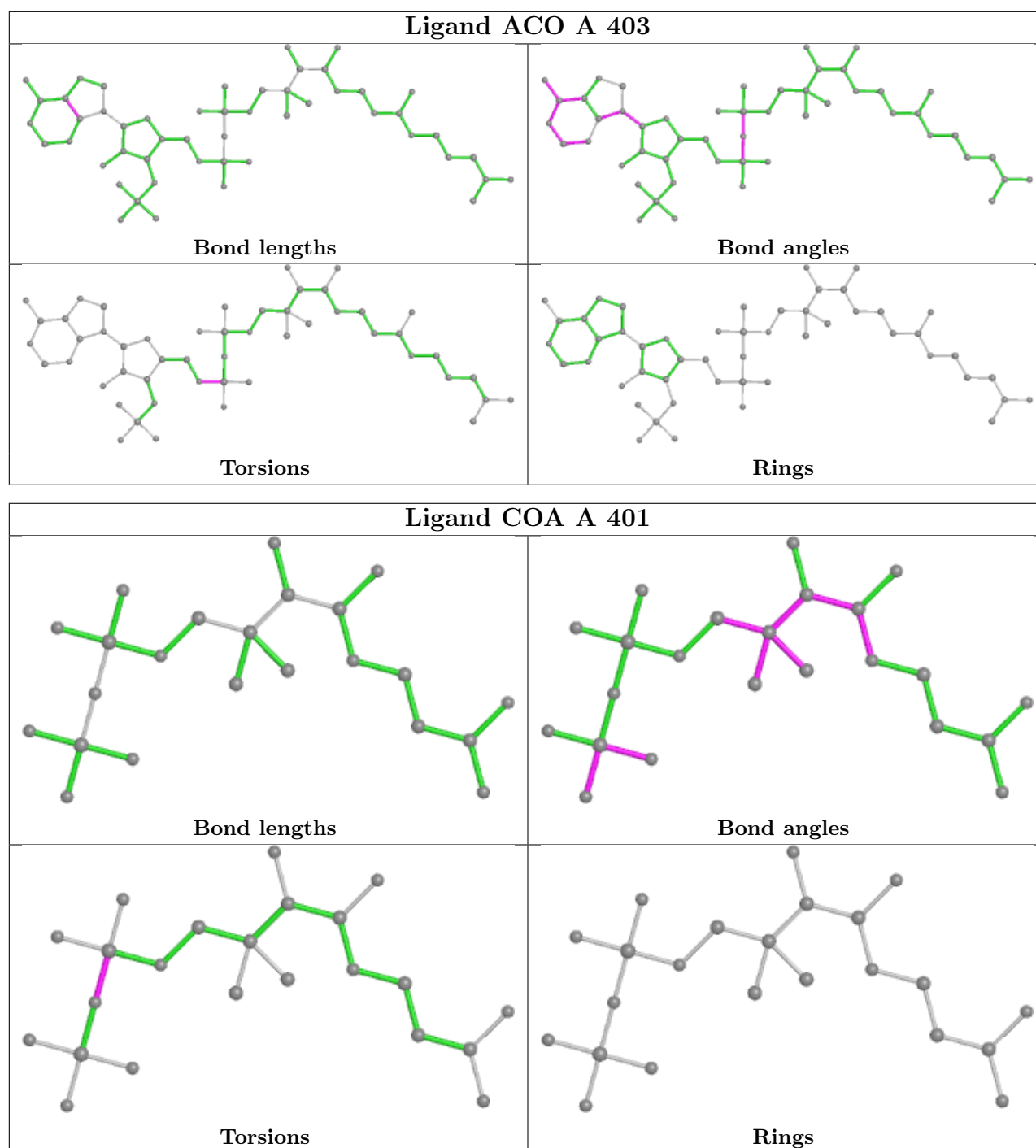
There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	402	X6X	1	0
6	A	407	EP1	3	0
4	A	403	ACO	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 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	316/328 (96%)	-0.18	5 (1%) 72 74	9, 20, 41, 85	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	164	ILE	8.1
1	A	245	HIS	3.2
1	A	160	LYS	3.1
1	A	163	SER	2.7
1	A	159[A]	PHE	2.2

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

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

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.4 Ligands [i](#)

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, 95<sup>th</sup> 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.

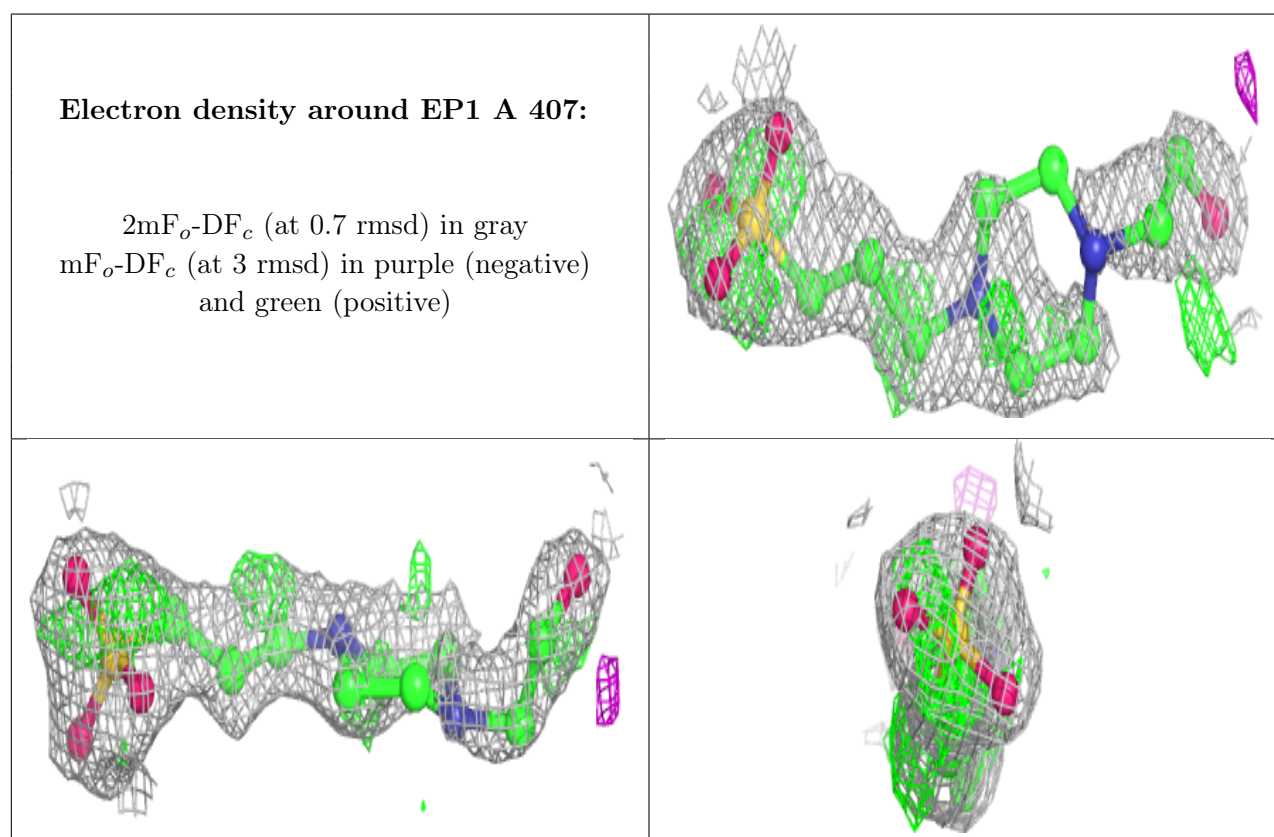
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
3	X6X	A	402	12/12	0.91	0.13	28,33,36,38	0
6	EP1	A	407	16/16	0.91	0.29	17,20,23,24	16

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
5	EDO	A	405	4/4	0.94	0.13	14,19,20,25	4
2	COA	A	401	23/48	0.96	0.08	23,25,31,37	0
5	EDO	A	406	4/4	0.96	0.08	17,21,24,25	0
5	EDO	A	404	4/4	0.96	0.08	26,27,28,31	0
4	ACO	A	403	51/51	0.97	0.08	9,14,24,31	0

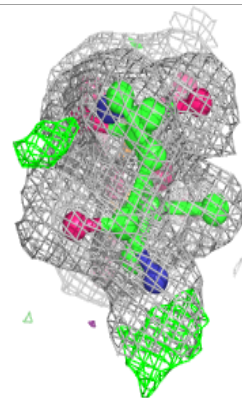
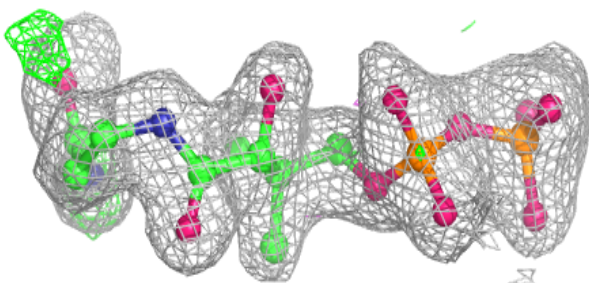
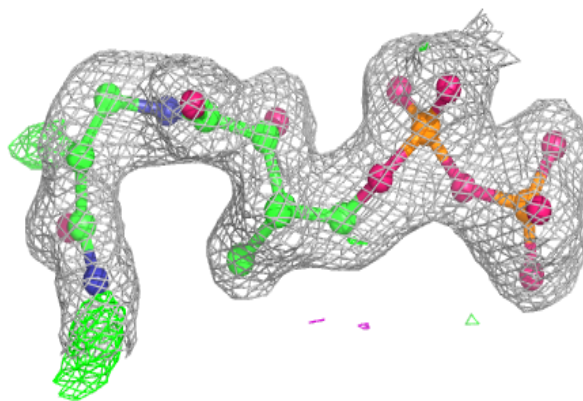
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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



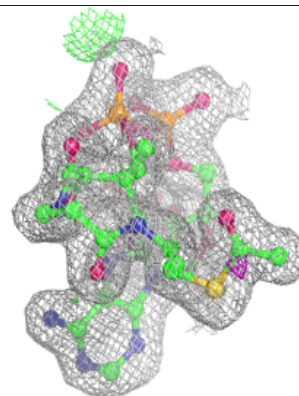
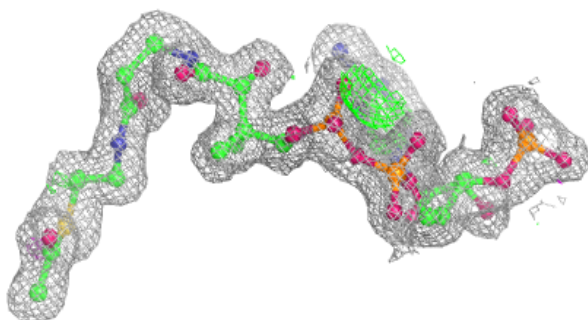
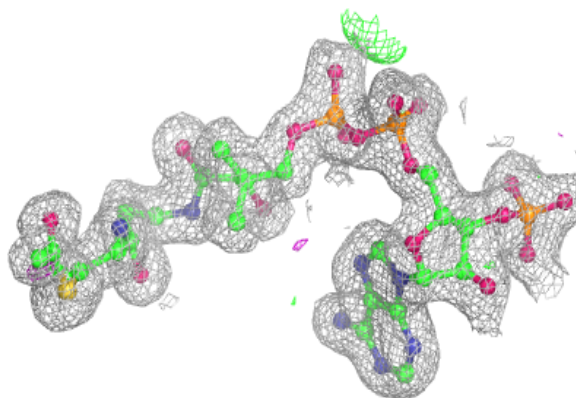


**Electron density around COA A 401:**

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

**Electron density around ACO A 403:**

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



## 6.5 Other polymers [i](#)

There are no such residues in this entry.