



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 1, 2022 – 10:06 AM EDT

PDB ID : 8CUB  
Title : Crystal Structure of ABCG5/G8 in Complex with Cholesterol  
Authors : Rezaei, F.; Farhat, D.; Lee, J.Y.  
Deposited on : 2022-05-17  
Resolution : 4.05 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	:	2.29
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)	:	2.29

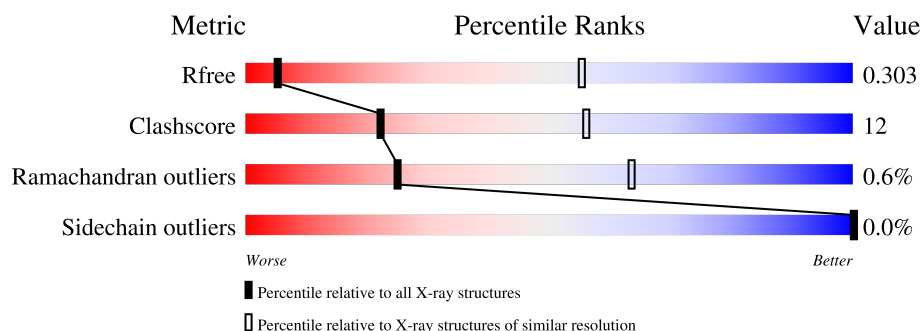
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.05 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	1127 (4.42-3.70)
Clashscore	141614	1033 (4.40-3.72)
Ramachandran outliers	138981	1145 (4.42-3.70)
Sidechain outliers	138945	1133 (4.42-3.70)

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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	664	
1	C	664	
2	B	687	
2	D	687	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 18498 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 ATP-binding cassette sub-family G member 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	588	Total	C	N	O	S	0	0	0
			4613	2987	777	820	29			
1	A	588	Total	C	N	O	S	0	0	0
			4619	2990	780	820	29			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	652	GLY	-	expression tag	UNP Q9H222
C	653	SER	-	expression tag	UNP Q9H222
C	654	HIS	-	expression tag	UNP Q9H222
C	655	HIS	-	expression tag	UNP Q9H222
C	656	HIS	-	expression tag	UNP Q9H222
C	657	HIS	-	expression tag	UNP Q9H222
C	658	HIS	-	expression tag	UNP Q9H222
C	659	HIS	-	expression tag	UNP Q9H222
C	660	GLY	-	expression tag	UNP Q9H222
C	661	HIS	-	expression tag	UNP Q9H222
C	662	HIS	-	expression tag	UNP Q9H222
C	663	HIS	-	expression tag	UNP Q9H222
C	664	HIS	-	expression tag	UNP Q9H222
C	665	HIS	-	expression tag	UNP Q9H222
C	666	HIS	-	expression tag	UNP Q9H222
A	652	GLY	-	expression tag	UNP Q9H222
A	653	SER	-	expression tag	UNP Q9H222
A	654	HIS	-	expression tag	UNP Q9H222
A	655	HIS	-	expression tag	UNP Q9H222
A	656	HIS	-	expression tag	UNP Q9H222
A	657	HIS	-	expression tag	UNP Q9H222
A	658	HIS	-	expression tag	UNP Q9H222
A	659	HIS	-	expression tag	UNP Q9H222
A	660	GLY	-	expression tag	UNP Q9H222
A	661	HIS	-	expression tag	UNP Q9H222

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	662	HIS	-	expression tag	UNP Q9H222
A	663	HIS	-	expression tag	UNP Q9H222
A	664	HIS	-	expression tag	UNP Q9H222
A	665	HIS	-	expression tag	UNP Q9H222
A	666	HIS	-	expression tag	UNP Q9H222

- Molecule 2 is a protein called ATP-binding cassette sub-family G member 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	579	Total	C	N	O	S	0	0	0
			4602	2985	785	804	28			
2	B	579	Total	C	N	O	S	0	0	0
			4608	2988	786	806	28			

There are 30 discrepancies between the modelled and reference sequences:

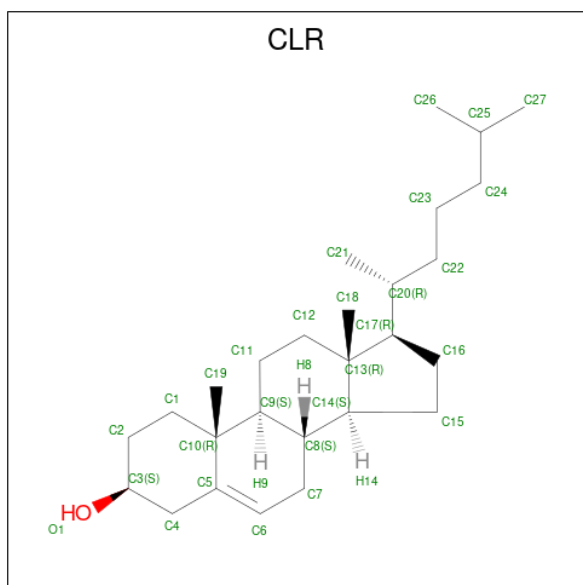
Chain	Residue	Modelled	Actual	Comment	Reference
D	-1	MET	-	expression tag	UNP Q9H221
D	0	GLY	-	expression tag	UNP Q9H221
D	1	SER	-	expression tag	UNP Q9H221
D	674	ALA	-	expression tag	UNP Q9H221
D	675	SER	-	expression tag	UNP Q9H221
D	676	ASN	-	expression tag	UNP Q9H221
D	677	SER	-	expression tag	UNP Q9H221
D	678	LEU	-	expression tag	UNP Q9H221
D	679	GLU	-	expression tag	UNP Q9H221
D	680	VAL	-	expression tag	UNP Q9H221
D	681	LEU	-	expression tag	UNP Q9H221
D	682	PHE	-	expression tag	UNP Q9H221
D	683	GLN	-	expression tag	UNP Q9H221
D	684	MET	-	expression tag	UNP Q9H221
D	685	GLU	-	expression tag	UNP Q9H221
B	-1	MET	-	expression tag	UNP Q9H221
B	0	GLY	-	expression tag	UNP Q9H221
B	1	SER	-	expression tag	UNP Q9H221
B	674	ALA	-	expression tag	UNP Q9H221
B	675	SER	-	expression tag	UNP Q9H221
B	676	ASN	-	expression tag	UNP Q9H221
B	677	SER	-	expression tag	UNP Q9H221
B	678	LEU	-	expression tag	UNP Q9H221
B	679	GLU	-	expression tag	UNP Q9H221
B	680	VAL	-	expression tag	UNP Q9H221

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	681	LEU	-	expression tag	UNP Q9H221
B	682	PHE	-	expression tag	UNP Q9H221
B	683	GLN	-	expression tag	UNP Q9H221
B	684	MET	-	expression tag	UNP Q9H221
B	685	GLU	-	expression tag	UNP Q9H221

- Molecule 3 is CHOLESTEROL (three-letter code: CLR) (formula:  $C_{27}H_{46}O$ ) (labeled as "Ligand of Interest" by depositor).

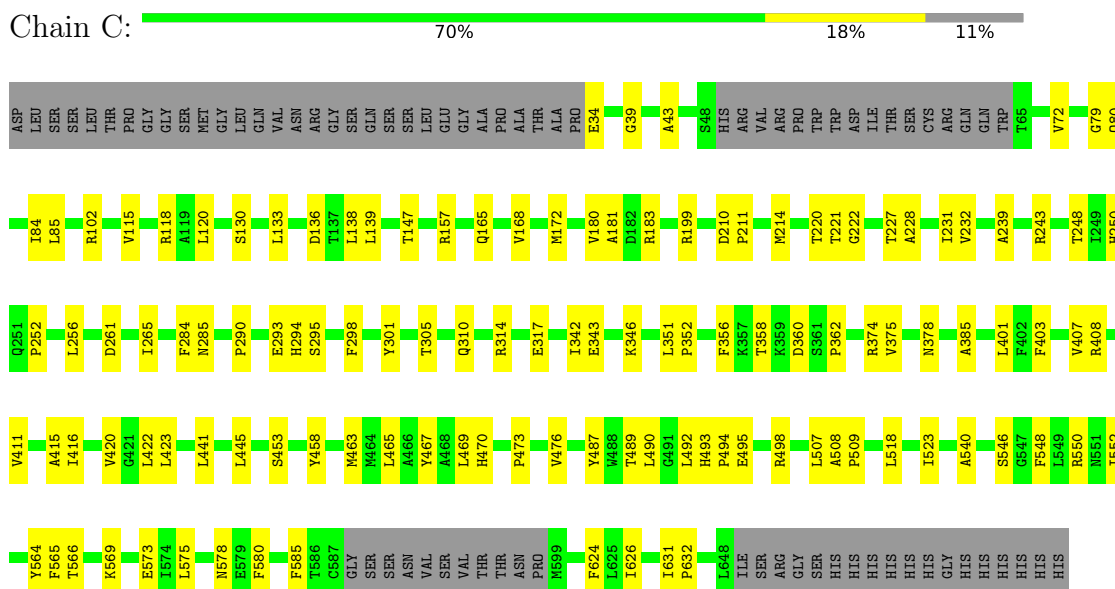


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	C	1	Total	C	O	0	0
			28	27	1		
3	A	1	Total	C	O	0	0
			28	27	1		

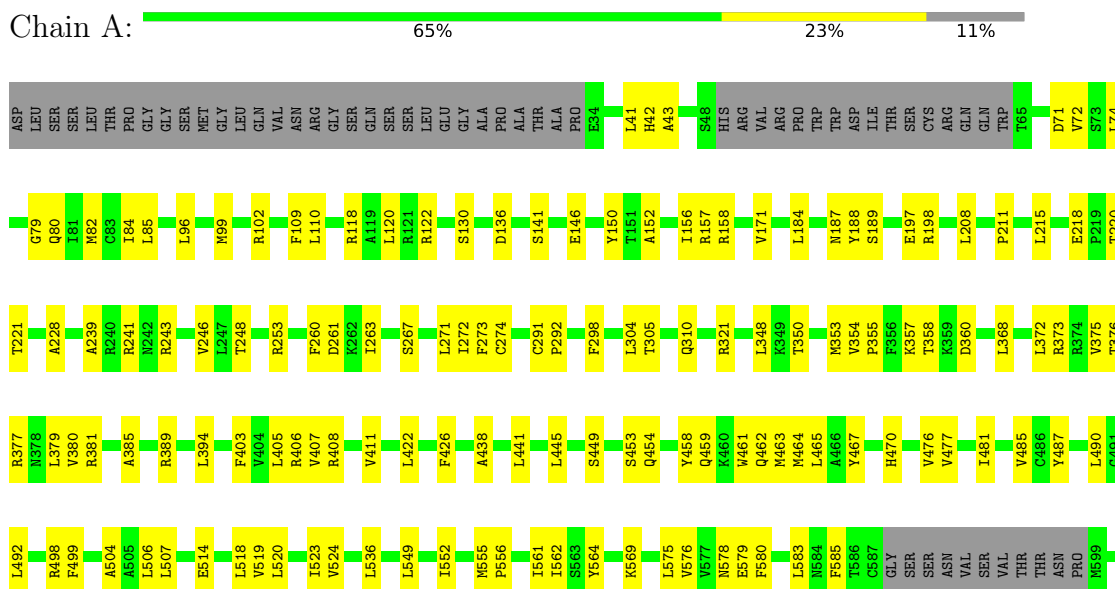
### 3 Residue-property plots

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. 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.

- Molecule 1: ATP-binding cassette sub-family G member 5



- Molecule 1: ATP-binding cassette sub-family G member 5





Y656	Y657	Y658	Q657	K668	P669	S670	GLN	ASP	TRP	ALA	SER	ASN	SER	LEU	GLU	VAL	LEU	PHE	GLN	MET	GLU
------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	173.27Å 230.11Å 249.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.91 – 4.05 30.91 – 4.05	Depositor EDS
% Data completeness (in resolution range)	65.0 (30.91-4.05) 76.6 (30.91-4.05)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.09 (at 4.12Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, $R_{free}$	0.246 , 0.303 0.246 , 0.303	Depositor DCC
$R_{free}$ test set	2017 reflections (5.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.1	Xtriage
Anisotropy	1.372	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.22 , -30.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.76	EDS
Total number of atoms	18498	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.55% 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: CLR

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	0.29	0/4710	0.59	0/6374
1	C	0.30	0/4704	0.56	0/6367
2	B	0.30	0/4715	0.59	0/6382
2	D	0.31	0/4709	0.59	0/6374
All	All	0.30	0/18838	0.58	0/25497

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 [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	4619	0	4752	129	0
1	C	4613	0	4741	93	0
2	B	4608	0	4674	126	0
2	D	4602	0	4663	115	0
3	A	28	0	46	3	0
3	C	28	0	46	4	0
All	All	18498	0	18922	438	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 12.

The worst 5 of 438 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:631:ILE:HG13	1:A:632:PRO:HD3	1.30	1.11
1:A:220:THR:HG23	1:A:228:ALA:HB1	1.52	0.91
2:B:115:LEU:HD13	2:B:268:SER:HB2	1.55	0.89
2:D:236:LEU:HD22	2:D:255:LEU:HD21	1.55	0.89
1:A:109:PHE:HB3	1:A:122:ARG:HH22	1.40	0.87

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	582/664 (88%)	546 (94%)	33 (6%)	3 (0%)	29	67
1	C	582/664 (88%)	545 (94%)	36 (6%)	1 (0%)	47	80
2	B	571/687 (83%)	530 (93%)	36 (6%)	5 (1%)	17	55
2	D	571/687 (83%)	536 (94%)	31 (5%)	4 (1%)	22	61
All	All	2306/2702 (85%)	2157 (94%)	136 (6%)	13 (1%)	25	63

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	629	SER
2	B	111	ARG
2	B	210	VAL
2	B	440	SER
2	D	210	VAL

### 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	513/580 (88%)	513 (100%)	0	100	100
1	C	512/580 (88%)	511 (100%)	1 (0%)	93	96
2	B	498/593 (84%)	498 (100%)	0	100	100
2	D	496/593 (84%)	496 (100%)	0	100	100
All	All	2019/2346 (86%)	2018 (100%)	1 (0%)	100	100

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	624	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	294	HIS
1	C	378	ASN
2	D	160	ASN
2	D	409	ASN
2	B	135	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](#)

2 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	$\# Z  > 2$
3	CLR	C	801	-	31,31,31	0.70	0	48,48,48	1.32	4 (8%)
3	CLR	A	801	-	31,31,31	0.70	0	48,48,48	1.48	9 (18%)

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
3	CLR	C	801	-	-	6/10/68/68	0/4/4/4
3	CLR	A	801	-	-	7/10/68/68	0/4/4/4

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	801	CLR	C4-C5-C10	4.99	123.04	116.42
3	A	801	CLR	C4-C5-C10	4.40	122.27	116.42
3	A	801	CLR	C13-C17-C20	-3.87	113.42	119.49
3	C	801	CLR	C4-C5-C6	-3.04	116.23	120.61
3	A	801	CLR	C17-C13-C14	2.82	103.41	100.07

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

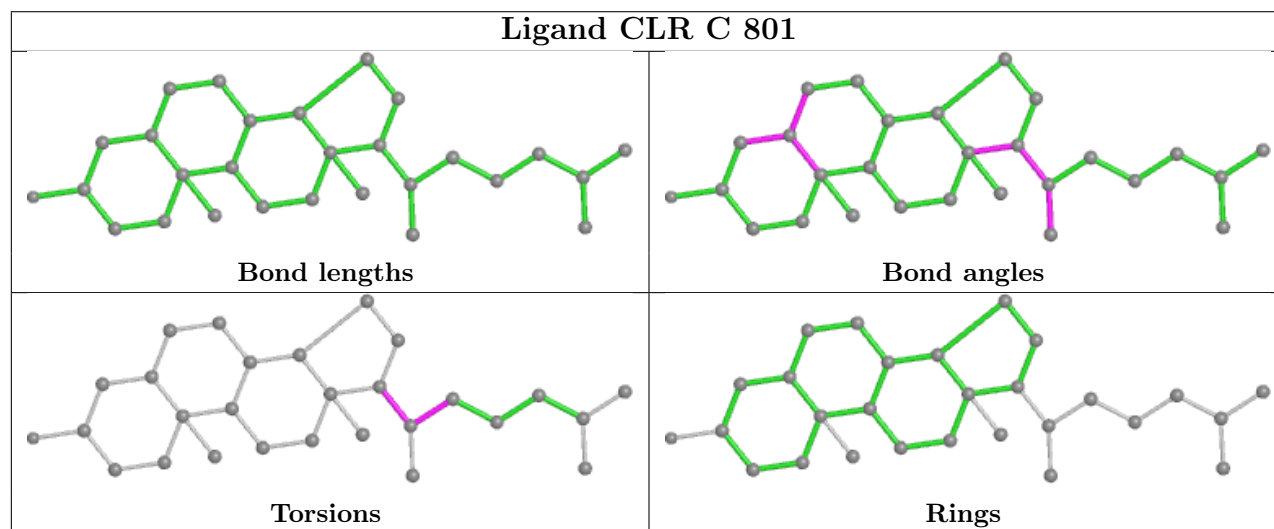
Mol	Chain	Res	Type	Atoms
3	C	801	CLR	C13-C17-C20-C21
3	A	801	CLR	C13-C17-C20-C21
3	C	801	CLR	C16-C17-C20-C21
3	A	801	CLR	C16-C17-C20-C21
3	C	801	CLR	C13-C17-C20-C22

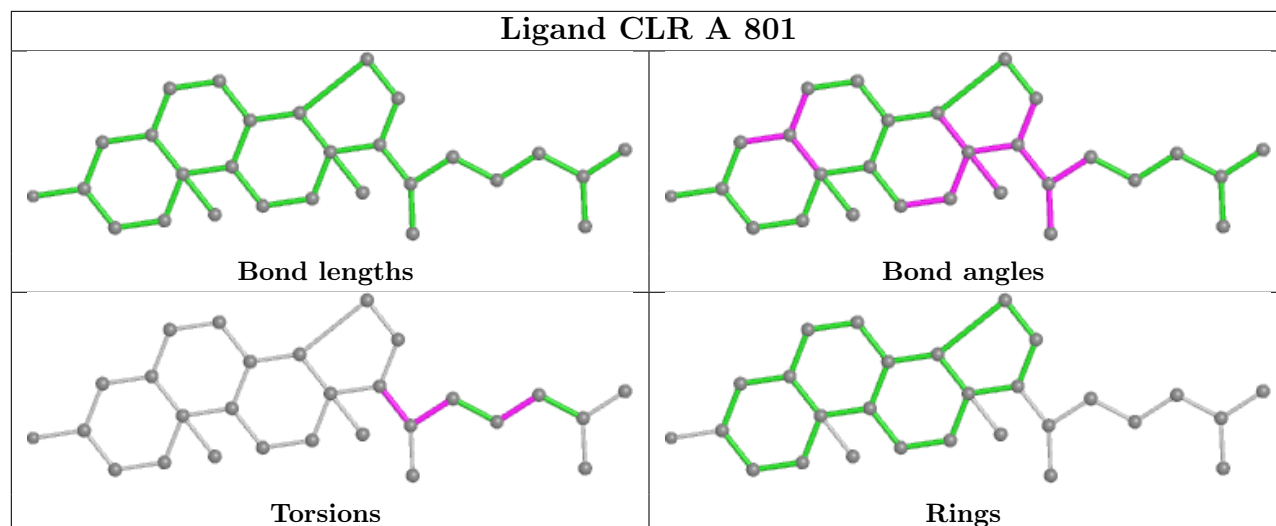
There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	801	CLR	4	0
3	A	801	CLR	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 [i](#)

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

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

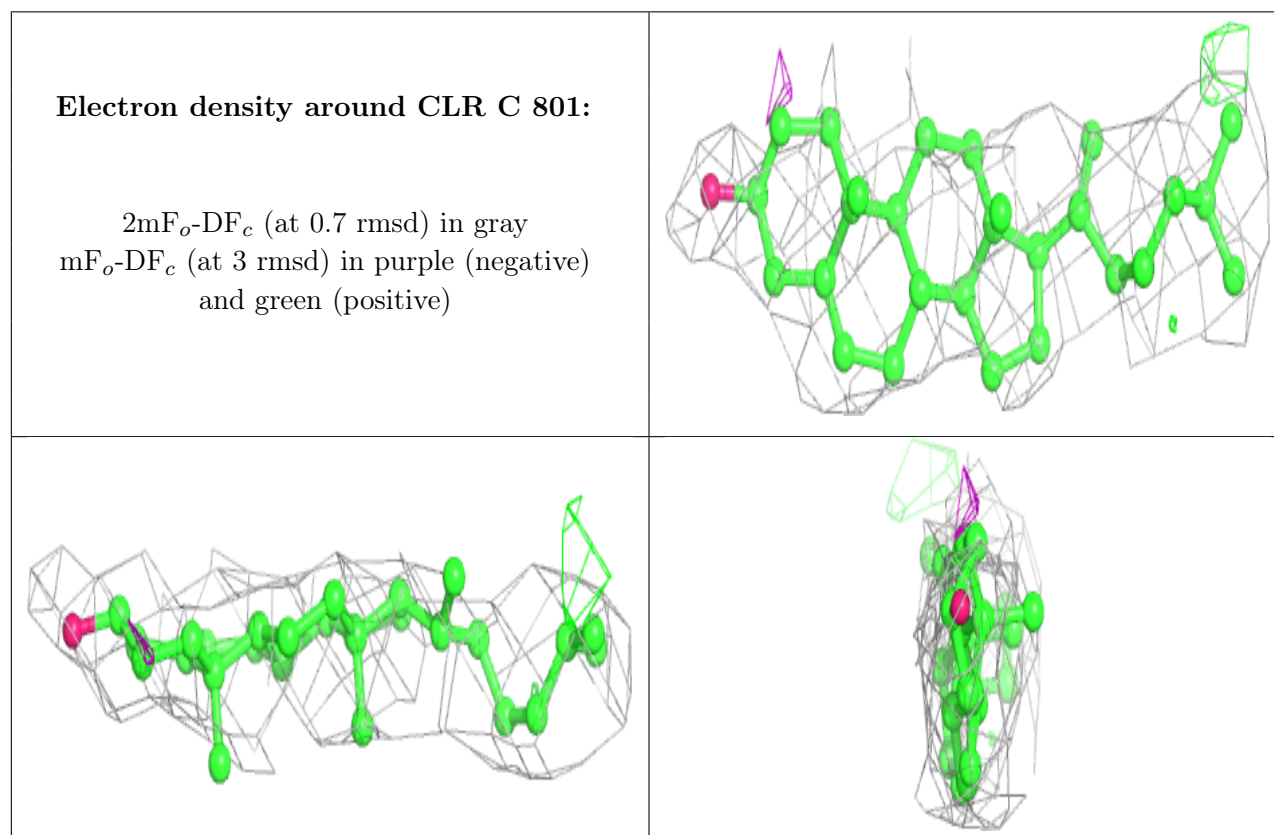
### 6.3 Carbohydrates [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

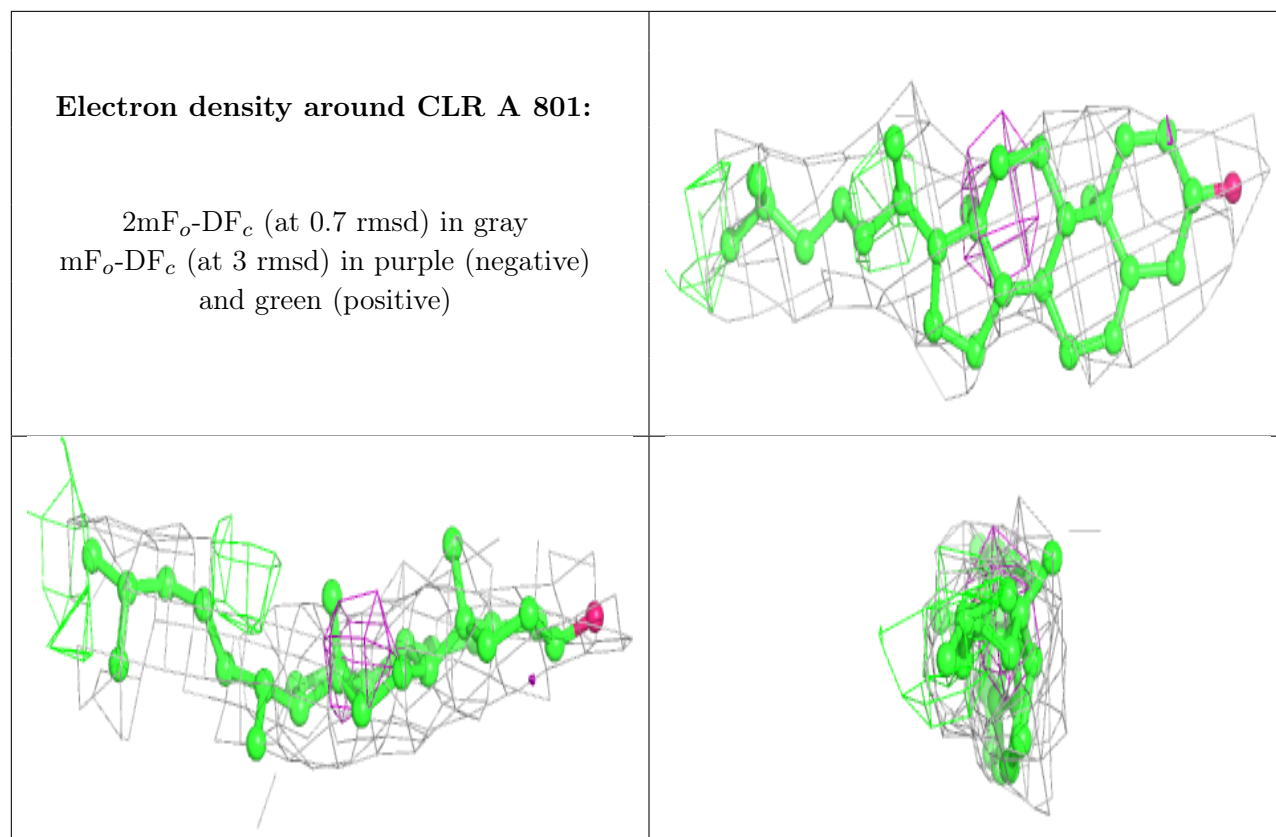
### 6.4 Ligands [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

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.







## 6.5 Other polymers [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.