

# wwPDB X-ray Structure Validation Summary Report (i)

#### Nov 3, 2022 - 03:34 am GMT

PDB ID : 7ZSG

Title : Structure of Orange Carotenoid Protein with canthaxanthin bound after 1

minute of illumination

Authors: Chukhutsina, V.U.; Baxter, J.M.; Fadini, A.; Morgan, R.M.; Pope, M.A.;

Maghlaoui, K.; Orr, C.; Wagner, A.; van Thor, J.J.

Deposited on : 2022-05-06

Resolution : 1.39 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 (i)) were used in the production of this report:

MolProbity: 4.02b-467

Mogul : 1.8.4, CSD as541be (2020)

Xtriage (Phenix) : 1.13

EDS : 2.31.2

buster-report : 1.1.7 (2018)

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

Refmac : 5.8.0267

CCP4 : 7.1.010 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

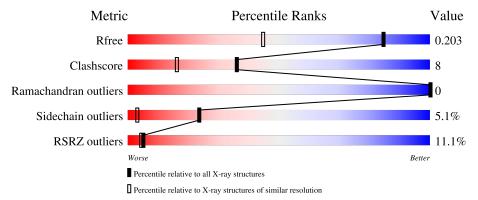
Validation Pipeline (wwPDB-VP) : 2.31.2

## 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 1.39 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},{\rm resolution\ range}(\mathring{\rm A})) \end{array}$
$R_{free}$	130704	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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	
			11%			
1	1	327		65%	25%	6% •



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5724 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 Orange carotenoid-binding protein.

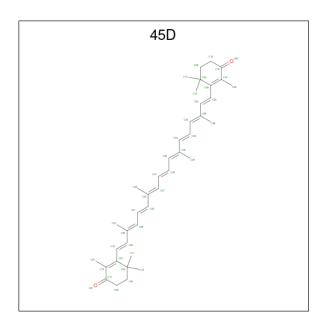
Mol	Chain	Residues		Atoms			ZeroOcc	AltConf	Trace	
1	1	314	Total 4818	C 3083	N 813	O 900	S 22	0	314	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
1	-9	MET	-	initiating methionine	UNP P74102
1	-8	GLY	-	expression tag	UNP P74102
1	-7	SER	-	expression tag	UNP P74102
1	-6	SER	-	expression tag	UNP P74102
1	-5	HIS	-	expression tag	UNP P74102
1	-4	HIS	-	expression tag	UNP P74102
1	-3	HIS	-	expression tag	UNP P74102
1	-2	HIS	-	expression tag	UNP P74102
1	-1	HIS	-	expression tag	UNP P74102
1	0	HIS	-	expression tag	UNP P74102

• Molecule 2 is beta, beta-carotene-4,4'-dione (three-letter code: 45D) (formula:  $C_{40}H_{52}O_2$ ) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
2	1	1	Total	С	О	0	1
	1	1	84	80	4		1

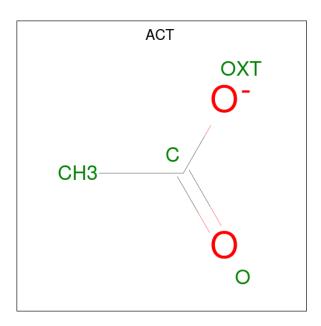
 $\bullet$  Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $\mathrm{C_3H_8O_3}).$ 



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

 $\bullet$  Molecule 4 is ACETATE ION (three-letter code: ACT) (formula:  $\mathrm{C_2H_3O_2}).$ 





Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	
4	1	1	Total 8	C 4	O 4	0	1

• Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	1	1	Total Cl 2 2	0	1

• Molecule 6 is water.

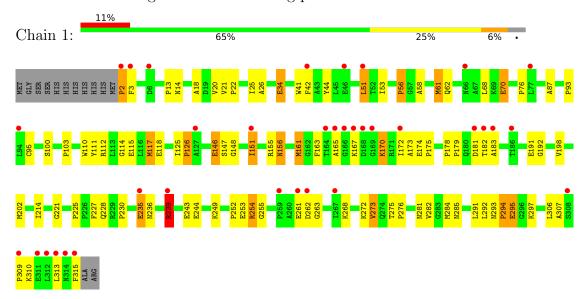
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	1	400	Total O 800 800	0	400



# 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: Orange carotenoid-binding protein





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants	82.58Å 82.58Å 87.22Å	Donositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $120.00^{\circ}$	Depositor
Resolution (Å)	55.36 - 1.39	Depositor
Resolution (A)	55.30 - 1.39	EDS
% Data completeness	99.9 (55.36-1.39)	Depositor
(in resolution range)	100.0 (55.30-1.39)	EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.73 (at 1.39Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
$R, R_{free}$	0.194 , 0.197	Depositor
10, 10 free	0.194 , $0.203$	DCC
$R_{free}$ test set	3548 reflections $(5.11%)$	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.5	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	(Not available), (Not available)	EDS
L-test for twinning <sup>2</sup>	$< L >=0.51, < L^2>=0.34$	Xtriage
Estimated twinning fraction	0.017 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	5724	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.50% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

## 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: GOL, 45D, CL, ACT

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	Bo	ond lengths	В	ond angles
MIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z >5
1	1	1.67	$112/4926 \ (2.3\%)$	1.29	$32/6706 \ (0.5\%)$

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a maintain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	1	0	4

The worst 5 of 112 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\operatorname{Observed}(\text{\AA})$	$\operatorname{Ideal}( ext{\AA})$
1	1	161[A]	MET	C-N	-19.09	0.98	1.33
1	1	161[B]	MET	C-N	-19.09	0.98	1.33
1	1	51[A]	LEU	C-N	-14.95	0.99	1.34
1	1	51[B]	LEU	C-N	-14.95	0.99	1.34
1	1	273[A]	VAL	C-N	-14.44	1.00	1.34

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	1	51[A]	LEU	C-N-CA	14.05	156.81	121.70
1	1	51[B]	LEU	C-N-CA	14.05	156.81	121.70
1	1	51[A]	LEU	O-C-N	-13.19	101.59	122.70
1	1	51[B]	LEU	O-C-N	-13.19	101.59	122.70
1	1	161[A]	MET	C-N-CA	11.90	147.28	122.30

There are no chirality outliers.

All (4) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	1	111[A]	TYR	Mainchain
1	1	172[A]	ILE	Mainchain
1	1	273[A]	VAL	Peptide
1	1	51[A]	LEU	Peptide

#### 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	1	4818	0	4791	69	0
2	1	84	0	104	11	0
3	1	12	0	16	0	0
4	1	8	0	6	1	0
5	1	2	0	0	0	0
6	1	800	0	0	33	0
All	All	5724	0	4917	79	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 8.

The worst 5 of 79 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	$\begin{array}{c} \text{Clash} \\ \text{overlap (Å)} \end{array}$	
1:1:309[A]:PRO:O	1:1:315[A]:PHE:CE1	2.04	1.10	
1:1:315[A]:PHE:HD1	6:1:768[A]:HOH:O	1.38	1.06	
1:1:235[B]:GLU:OE1	6:1:508[B]:HOH:O	1.74	1.05	
1:1:2[A]:PRO:HG3	1:1:291[A]:LEU:HD21	1.39	1.04	
1:1:309[A]:PRO:O	1:1:315[A]:PHE:CD1	2.11	1.03	

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	1	624/327 (191%)	620 (99%)	4 (1%)	0	100 100

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	Analysed Rotameric Outliers			
1	1	508/266 (191%)	482 (95%)	26 (5%)	24 3	

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	1	170[A]	LYS
1	1	181[B]	ASP
1	1	310[A]	LYS
1	1	181[A]	ASP
1	1	239[A]	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

#### 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)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 for Mogul analysis.

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).

Mal	Mol Type Cl		Res	Link	В	Bond lengths			Bond angles		
MIOI	Mol Type Chain	nes	Lilik	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2		
3	GOL	1	402[A]	-	5,5,5	0.21	0	5,5,5	0.23	0	
4	ACT	1	403[B]	-	3,3,3	0.89	0	3,3,3	0.96	0	
2	45D	1	401[A]	-	43,43,43	2.41	19 (44%)	54,60,60	2.91	17 (31%)	
4	ACT	1	403[A]	-	3,3,3	0.91	0	3,3,3	0.82	0	
3	GOL	1	402[B]	-	5,5,5	0.37	0	5,5,5	0.27	0	
2	45D	1	401[B]	-	43,43,43	1.89	13 (30%)	54,60,60	1.20	6 (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
2	45D	1	401[A]	-	-	8/29/69/69	0/2/2/2
3	GOL	1	402[A]	-	-	0/4/4/4	-
2	45D	1	401[B]	-	-	2/29/69/69	0/2/2/2
3	GOL	1	402[B]	_	-	0/4/4/4	-

The worst 5 of 32 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(A)	$\operatorname{Ideal}(\text{\AA})$
2	1	401[A]	45D	C27-C25	-6.96	1.36	1.50
2	1	401[A]	45D	C30-C26	5.44	1.43	1.35
2	1	401[A]	45D	C38-C36	4.60	1.41	1.35
2	1	401[A]	45D	C37-C35	4.30	1.41	1.35
2	1	401[A]	45D	C39-C35	-3.87	1.42	1.50



The worst	5	of 2	23	bond	angle	outliers	are	listed	below:
110 WOID	$\overline{}$	01.		OIIG	211510	Cathere	COL C	IID CCC	OCIO III.

Mol	Chain	Res	Type	Atoms	Z	$Observed(^o)$	$\operatorname{Ideal}({}^o)$
2	1	401[A]	45D	C23-C25-C29	12.14	137.57	118.94
2	1	401[A]	45D	C27-C25-C23	-6.33	108.10	118.08
2	1	401[A]	45D	C27-C25-C29	-6.15	114.30	122.92
2	1	401[A]	45D	C24-C26-C30	5.90	127.99	118.94
2	1	401[A]	45D	C28-C26-C30	-5.34	115.45	122.92

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	1	401[A]	45D	C03-C07-C19-C23
2	1	401[A]	45D	C19-C23-C25-C27
2	1	401[A]	45D	C19-C23-C25-C29
2	1	401[A]	45D	C37-C41-C42-C38
2	1	401[A]	45D	C15-C07-C19-C23

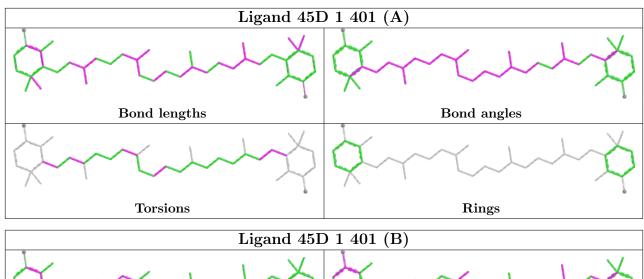
There are no ring outliers.

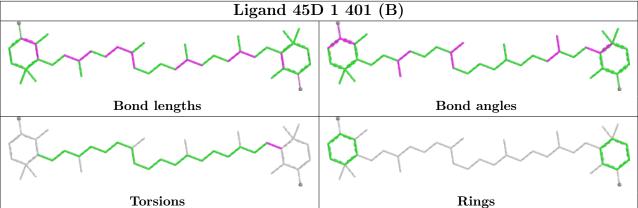
3 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	1	401[A]	45D	9	0
4	1	403[A]	ACT	1	0
2	1	401[B]	45D	2	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.







## 5.7 Other polymers (i)

There are no such residues in this entry.

### 5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	1	14

The worst 5 of 14 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	1	14[A]:ASN	С	15[A]:THR	N	1.19
1	1	42[A]:PHE	С	43[A]:ALA	N	1.19
1	1	191[A]:GLU	С	192[A]:GLY	N	1.19
1	1	239[A]:ARG	С	240[A]:PHE	N	1.19
1	1	214[A]:ILE	С	215[A]:GLU	N	1.16



## 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,  $95^{th}$  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 $>$	$\# \mathrm{RSRZ}{>}2$	$OWAB(Å^2)$	Q<0.9
1	1	314/327 (96%)	0.99	35 (11%) 5 4	13, 23, 44, 96	1 (0%)

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	1	168[A]	ASP	10.9
1	1	3[A]	PHE	10.0
1	1	166[A]	GLY	9.4
1	1	315[A]	PHE	9.2
1	1	2[A]	PRO	7.9

### 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,  $95^{th}$  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	$oxed{ \mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2) }$	Q<0.9
4	ACT	1	403[A]	4/4	0.82	0.15	63,65,68,69	4
4	ACT	1	403[B]	4/4	0.82	0.15	62,66,67,68	4

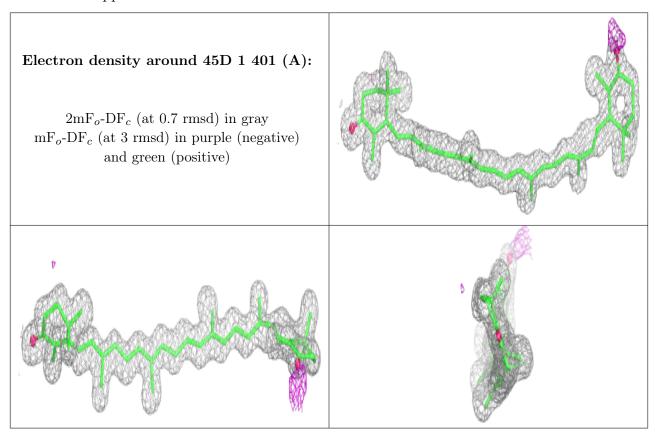
Continued on next page...



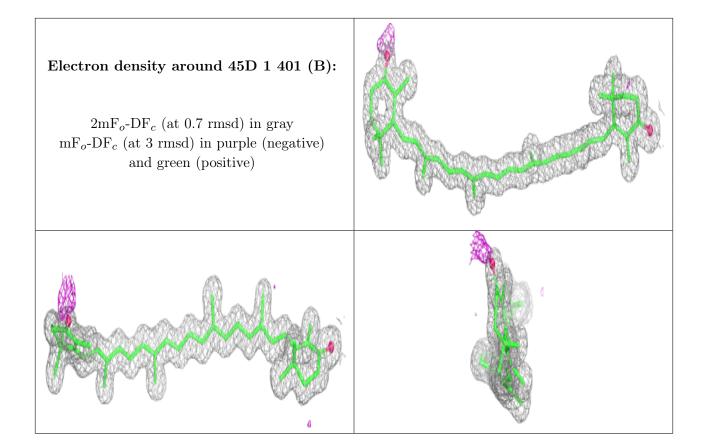
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
3	GOL	1	402[A]	6/6	0.91	0.07	26,27,30,30	6
3	GOL	1	402[B]	6/6	0.91	0.07	24,27,29,31	6
2	45D	1	401[A]	42/42	0.93	0.10	17,18,22,25	42
2	45D	1	401[B]	42/42	0.93	0.10	16,18,22,27	42
5	CL	1	404[A]	1/1	1.00	0.10	20,20,20,20	1
5	CL	1	404[B]	1/1	1.00	0.10	21,21,21,21	1

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.



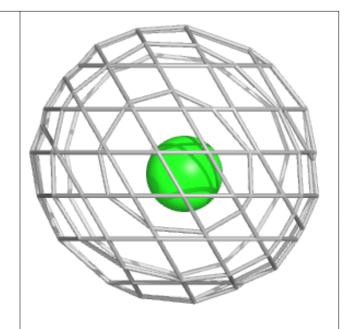


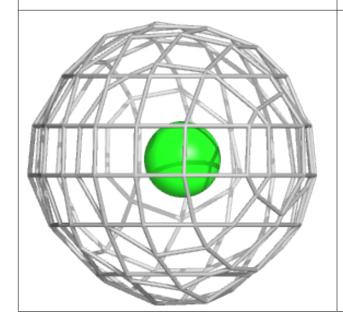


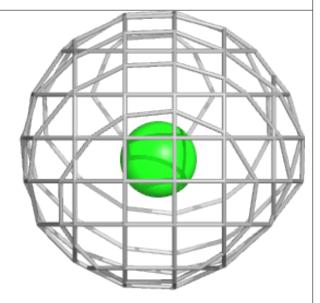


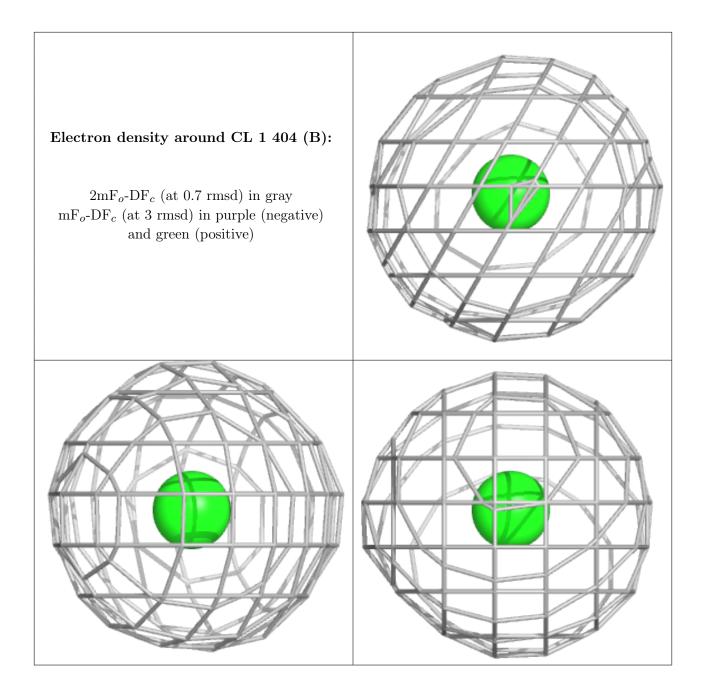
#### Electron density around CL 1 404 (A):

 $2 {
m mF}_o {
m -DF}_c$  (at 0.7 rmsd) in gray  ${
m mF}_o {
m -DF}_c$  (at 3 rmsd) in purple (negative) and green (positive)









## 6.5 Other polymers (i)

There are no such residues in this entry.

