WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 11:58 AM EDT

|		
PDB ID	:	6SU7
Title	:	Complex between a UDP-glucosyltransferase from Polygonum tinctorium capable of glucosylating indoxyl and 3,4-Dichloroaniline
Authors	:	Fredslund, F.; Teze, D.; Svensson, B.; Adams, P.D.; Welner, D.H.
Deposited on	:	2019-09-13
Resolution	:	2.75 Å(reported)

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 2.75 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.
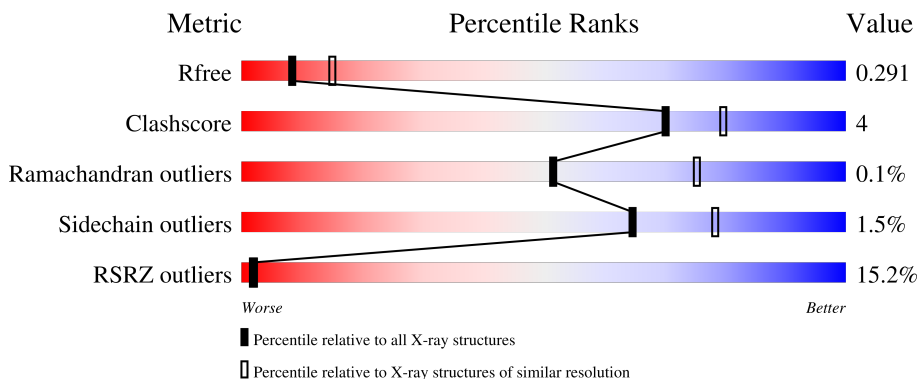
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	499	5% / 83% 8% 9%
1	B	499	11% / 84% 6% 9%
1	C	499	11% / 79% 10% 11%
1	D	499	27% / 78% 11% 11%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 27755 atoms, of which 13943 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 Glycosyltransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	452	Total 6981	C 2222	H 3508	N 597	O 640	S 14	0	1	0
1	B	452	Total 6980	C 2222	H 3507	N 597	O 640	S 14	0	0	0
1	C	446	Total 6910	C 2201	H 3473	N 590	O 633	S 13	0	0	0
1	D	444	Total 6875	C 2192	H 3455	N 585	O 630	S 13	0	0	0

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	initiating methionine	UNP A0A2R2JFJ4
A	-20	GLY	-	expression tag	UNP A0A2R2JFJ4
A	-19	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-18	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-17	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-16	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-15	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-14	HIS	-	expression tag	UNP A0A2R2JFJ4
A	-13	ASP	-	expression tag	UNP A0A2R2JFJ4
A	-12	TYR	-	expression tag	UNP A0A2R2JFJ4
A	-11	ASP	-	expression tag	UNP A0A2R2JFJ4
A	-10	ILE	-	expression tag	UNP A0A2R2JFJ4
A	-9	PRO	-	expression tag	UNP A0A2R2JFJ4
A	-8	THR	-	expression tag	UNP A0A2R2JFJ4
A	-7	THR	-	expression tag	UNP A0A2R2JFJ4
A	-6	GLU	-	expression tag	UNP A0A2R2JFJ4
A	-5	ASN	-	expression tag	UNP A0A2R2JFJ4
A	-4	LEU	-	expression tag	UNP A0A2R2JFJ4
A	-3	TYR	-	expression tag	UNP A0A2R2JFJ4
A	-2	PHE	-	expression tag	UNP A0A2R2JFJ4
A	-1	GLN	-	expression tag	UNP A0A2R2JFJ4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-21	MET	-	initiating methionine	UNP A0A2R2JFJ4
B	-20	GLY	-	expression tag	UNP A0A2R2JFJ4
B	-19	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-18	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-17	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-16	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-15	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-14	HIS	-	expression tag	UNP A0A2R2JFJ4
B	-13	ASP	-	expression tag	UNP A0A2R2JFJ4
B	-12	TYR	-	expression tag	UNP A0A2R2JFJ4
B	-11	ASP	-	expression tag	UNP A0A2R2JFJ4
B	-10	ILE	-	expression tag	UNP A0A2R2JFJ4
B	-9	PRO	-	expression tag	UNP A0A2R2JFJ4
B	-8	THR	-	expression tag	UNP A0A2R2JFJ4
B	-7	THR	-	expression tag	UNP A0A2R2JFJ4
B	-6	GLU	-	expression tag	UNP A0A2R2JFJ4
B	-5	ASN	-	expression tag	UNP A0A2R2JFJ4
B	-4	LEU	-	expression tag	UNP A0A2R2JFJ4
B	-3	TYR	-	expression tag	UNP A0A2R2JFJ4
B	-2	PHE	-	expression tag	UNP A0A2R2JFJ4
B	-1	GLN	-	expression tag	UNP A0A2R2JFJ4
C	-21	MET	-	initiating methionine	UNP A0A2R2JFJ4
C	-20	GLY	-	expression tag	UNP A0A2R2JFJ4
C	-19	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-18	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-17	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-16	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-15	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-14	HIS	-	expression tag	UNP A0A2R2JFJ4
C	-13	ASP	-	expression tag	UNP A0A2R2JFJ4
C	-12	TYR	-	expression tag	UNP A0A2R2JFJ4
C	-11	ASP	-	expression tag	UNP A0A2R2JFJ4
C	-10	ILE	-	expression tag	UNP A0A2R2JFJ4
C	-9	PRO	-	expression tag	UNP A0A2R2JFJ4
C	-8	THR	-	expression tag	UNP A0A2R2JFJ4
C	-7	THR	-	expression tag	UNP A0A2R2JFJ4
C	-6	GLU	-	expression tag	UNP A0A2R2JFJ4
C	-5	ASN	-	expression tag	UNP A0A2R2JFJ4
C	-4	LEU	-	expression tag	UNP A0A2R2JFJ4
C	-3	TYR	-	expression tag	UNP A0A2R2JFJ4
C	-2	PHE	-	expression tag	UNP A0A2R2JFJ4
C	-1	GLN	-	expression tag	UNP A0A2R2JFJ4

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-21	MET	-	initiating methionine	UNP A0A2R2JFJ4
D	-20	GLY	-	expression tag	UNP A0A2R2JFJ4
D	-19	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-18	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-17	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-16	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-15	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-14	HIS	-	expression tag	UNP A0A2R2JFJ4
D	-13	ASP	-	expression tag	UNP A0A2R2JFJ4
D	-12	TYR	-	expression tag	UNP A0A2R2JFJ4
D	-11	ASP	-	expression tag	UNP A0A2R2JFJ4
D	-10	ILE	-	expression tag	UNP A0A2R2JFJ4
D	-9	PRO	-	expression tag	UNP A0A2R2JFJ4
D	-8	THR	-	expression tag	UNP A0A2R2JFJ4
D	-7	THR	-	expression tag	UNP A0A2R2JFJ4
D	-6	GLU	-	expression tag	UNP A0A2R2JFJ4
D	-5	ASN	-	expression tag	UNP A0A2R2JFJ4
D	-4	LEU	-	expression tag	UNP A0A2R2JFJ4
D	-3	TYR	-	expression tag	UNP A0A2R2JFJ4
D	-2	PHE	-	expression tag	UNP A0A2R2JFJ4
D	-1	GLN	-	expression tag	UNP A0A2R2JFJ4

- Molecule 2 is 3,4-Dichloroaniline (three-letter code: LV5) (formula: $C_6H_5Cl_2N$) (labeled as "Ligand of Interest" by depositor).
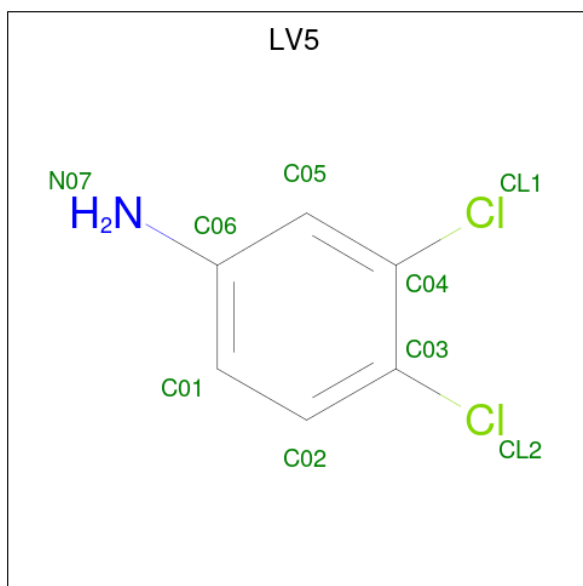
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	Cl	N		
2	A	1	9	6	2	1	0	0

# 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.
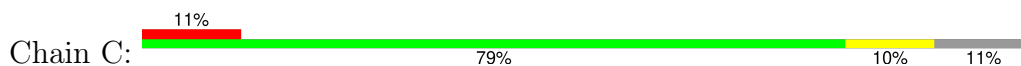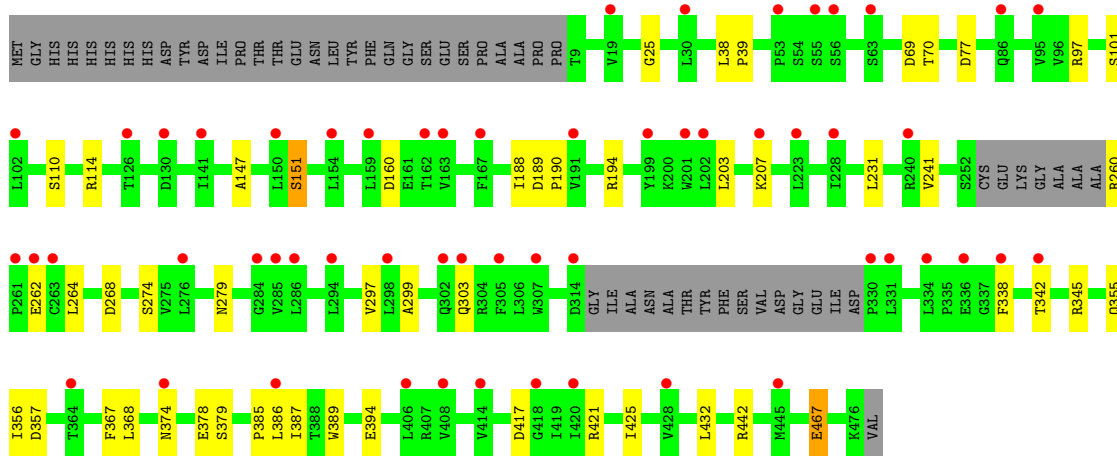
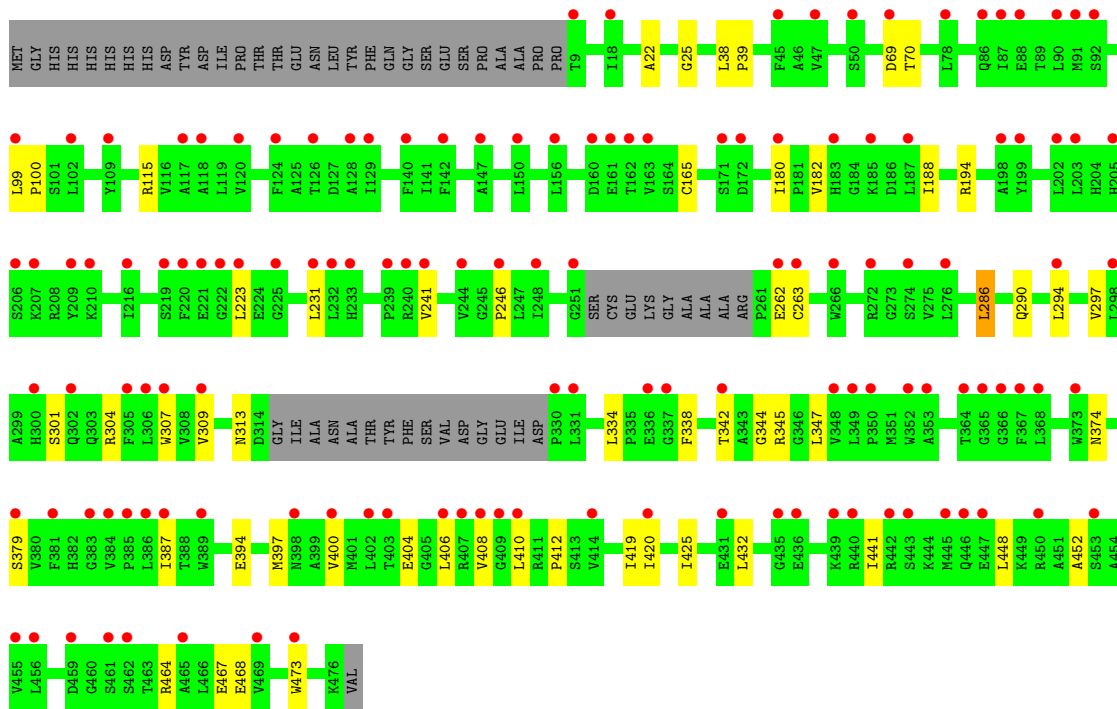- Molecule 1: Glycosyltransferase
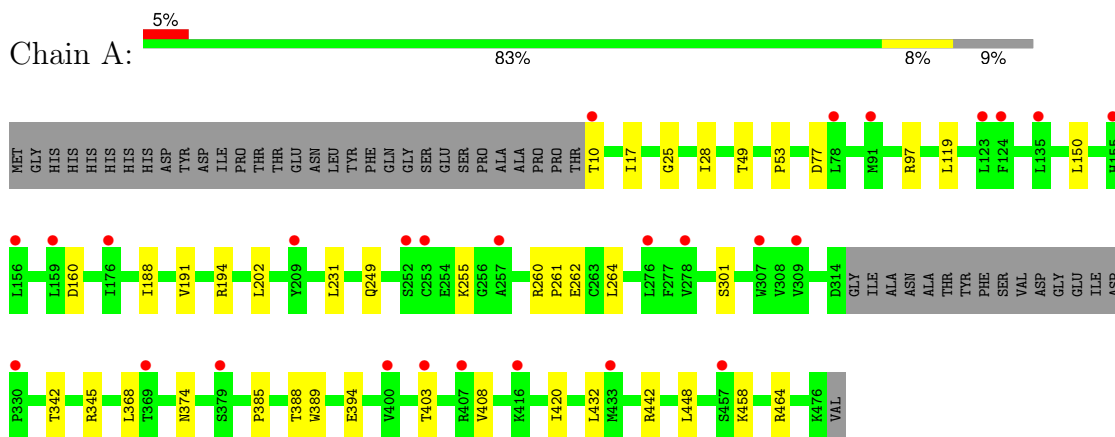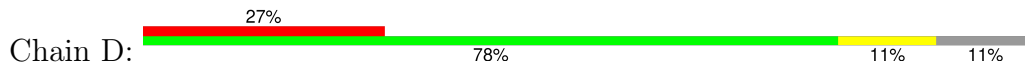
Chain A:

- Molecule 1: Glycosyltransferase

Chain B:

- Molecule 1: Glycosyltransferase

Chain C:
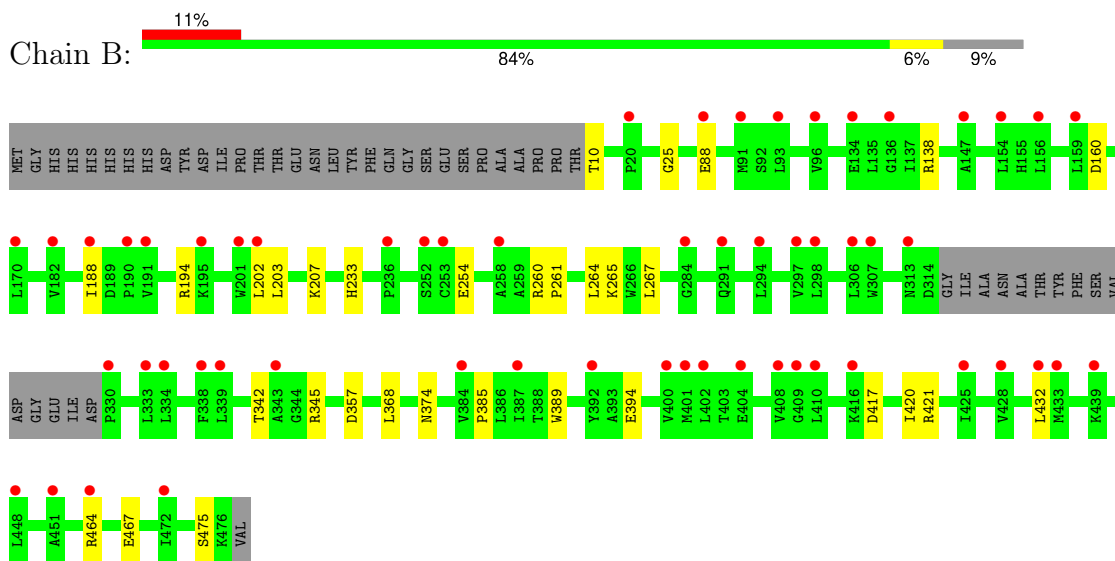
• Molecule 1: Glycosyltransferase

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	93.35Å 119.36Å 170.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.28 – 2.75 51.34 – 2.75	Depositor EDS
% Data completeness (in resolution range)	99.3 (50.28-2.75) 99.5 (51.34-2.75)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.70 (at 2.73Å)	Xtriage
Refinement program	PHENIX 1.16_3549, PHENIX 1.16_3549	Depositor
R, R_{free}	0.256 , 0.287 0.263 , 0.291	Depositor DCC
R_{free} test set	2475 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å^2)	56.1	Xtriage
Anisotropy	0.397	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 39.2	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	27755	wwPDB-VP
Average B, all atoms (Å^2)	99.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.05 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.6138e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: LV5

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	0.24	0/3548	0.42	0/4820
1	B	0.24	0/3557	0.42	0/4833
1	C	0.24	0/3520	0.42	0/4783
1	D	0.24	0/3503	0.43	0/4760
All	All	0.24	0/14128	0.42	0/19196

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	3473	3508	3508	21	1
1	B	3473	3507	3508	21	0
1	C	3437	3473	3472	28	1
1	D	3420	3455	3455	32	0
2	A	9	0	0	0	0
All	All	13812	13943	13943	101	1

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 4.

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

The worst 5 of 101 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:C:417:ASP:OD2	1:C:421:ARG:NH2	2.24	0.70
1:B:385:PRO:HB2	1:B:432:LEU:HD11	1.73	0.69
1:D:342:THR:HG22	1:D:345:ARG:HH11	1.57	0.68
1:A:385:PRO:HB2	1:A:432:LEU:HD11	1.77	0.66
1:C:342:THR:HG22	1:C:345:ARG:HH21	1.65	0.61

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:464:ARG:NH2	1:C:467:GLU:OE1[1_455]	2.19	0.01

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	448/499 (90%)	427 (95%)	21 (5%)	0	100	100
1	B	448/499 (90%)	423 (94%)	25 (6%)	0	100	100
1	C	440/499 (88%)	414 (94%)	26 (6%)	0	100	100
1	D	438/499 (88%)	413 (94%)	24 (6%)	1 (0%)	47	69
All	All	1774/1996 (89%)	1677 (94%)	96 (5%)	1 (0%)	51	75

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	22	ALA

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	379/419 (90%)	375 (99%)	4 (1%)	73	84
1	B	380/419 (91%)	378 (100%)	2 (0%)	88	92
1	C	378/419 (90%)	372 (98%)	6 (2%)	62	77
1	D	376/419 (90%)	366 (97%)	10 (3%)	44	65
All	All	1513/1676 (90%)	1491 (98%)	22 (2%)	65	78

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	286	LEU
1	D	410	LEU
1	D	408	VAL
1	D	419	ILE
1	C	101	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	269	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 (i)

1 ligand is 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	#$|Z| > 2$
2	LV5	A	1001	-	9,9,9	1.29	3 (33%)	12,12,12	0.56	0

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	LV5	A	1001	-	-	-	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1001	LV5	C04-CL1	2.26	1.79	1.73
2	A	1001	LV5	C03-CL2	2.22	1.78	1.73
2	A	1001	LV5	C06-N07	2.11	1.45	1.38

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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 LV5 A 1001

Bond lengths · Bond angles · Torsions · Rings
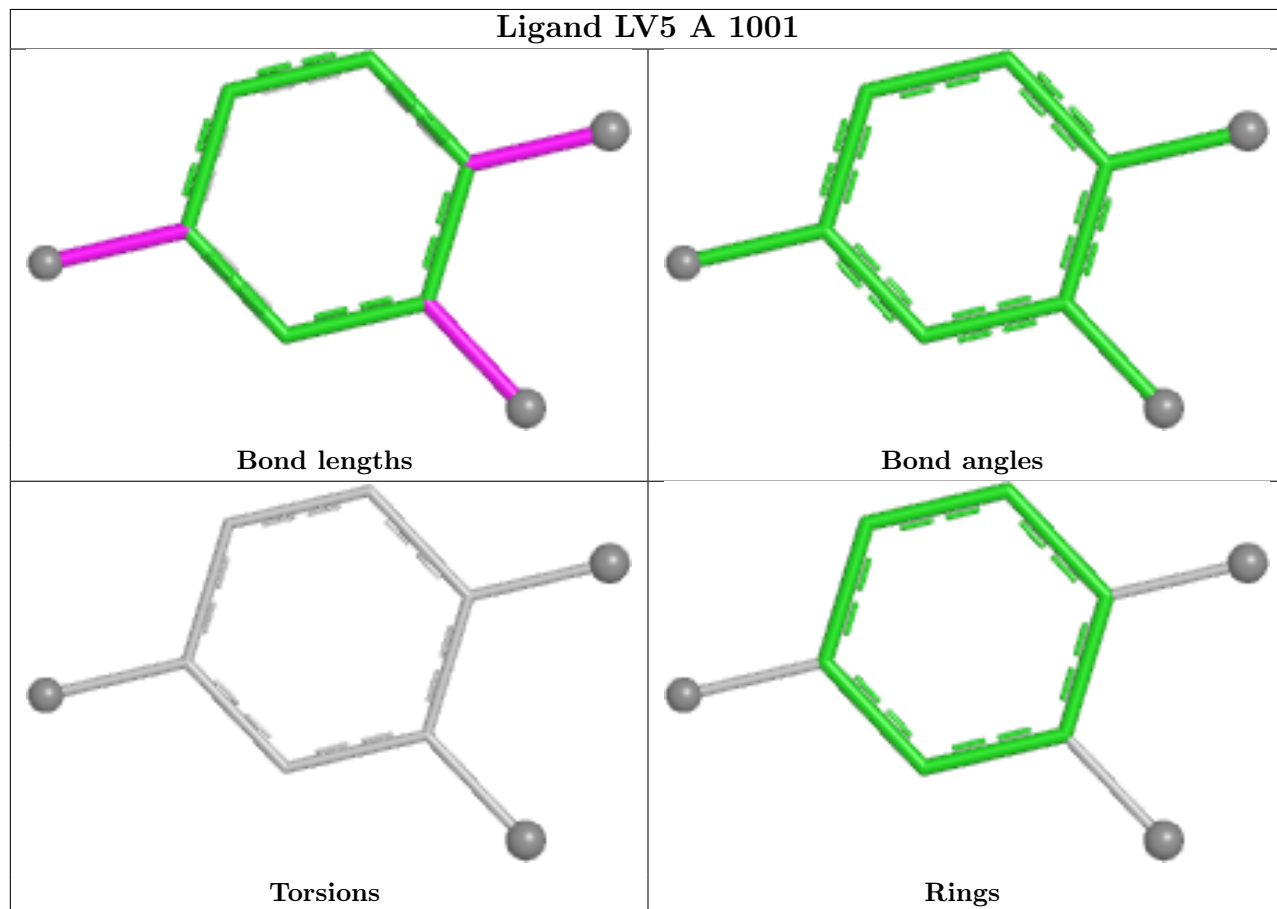
5.7 Other polymers ⓘ

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	452/499 (90%)	0.52	27 (5%)	21	26	27, 55, 104, 161	0
1	B	452/499 (90%)	0.88	57 (12%)	3	4	38, 75, 137, 211	0
1	C	446/499 (89%)	0.77	56 (12%)	3	4	42, 81, 136, 211	0
1	D	444/499 (88%)	1.60	133 (29%)	0	0	53, 136, 206, 269	0
All	All	1794/1996 (89%)	0.94	273 (15%)	2	2	27, 81, 174, 269	0

The worst 5 of 273 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	209	TYR	11.1
1	D	162	THR	9.9
1	D	436	GLU	9.5
1	D	223	LEU	9.2
1	D	336	GLU	8.9

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

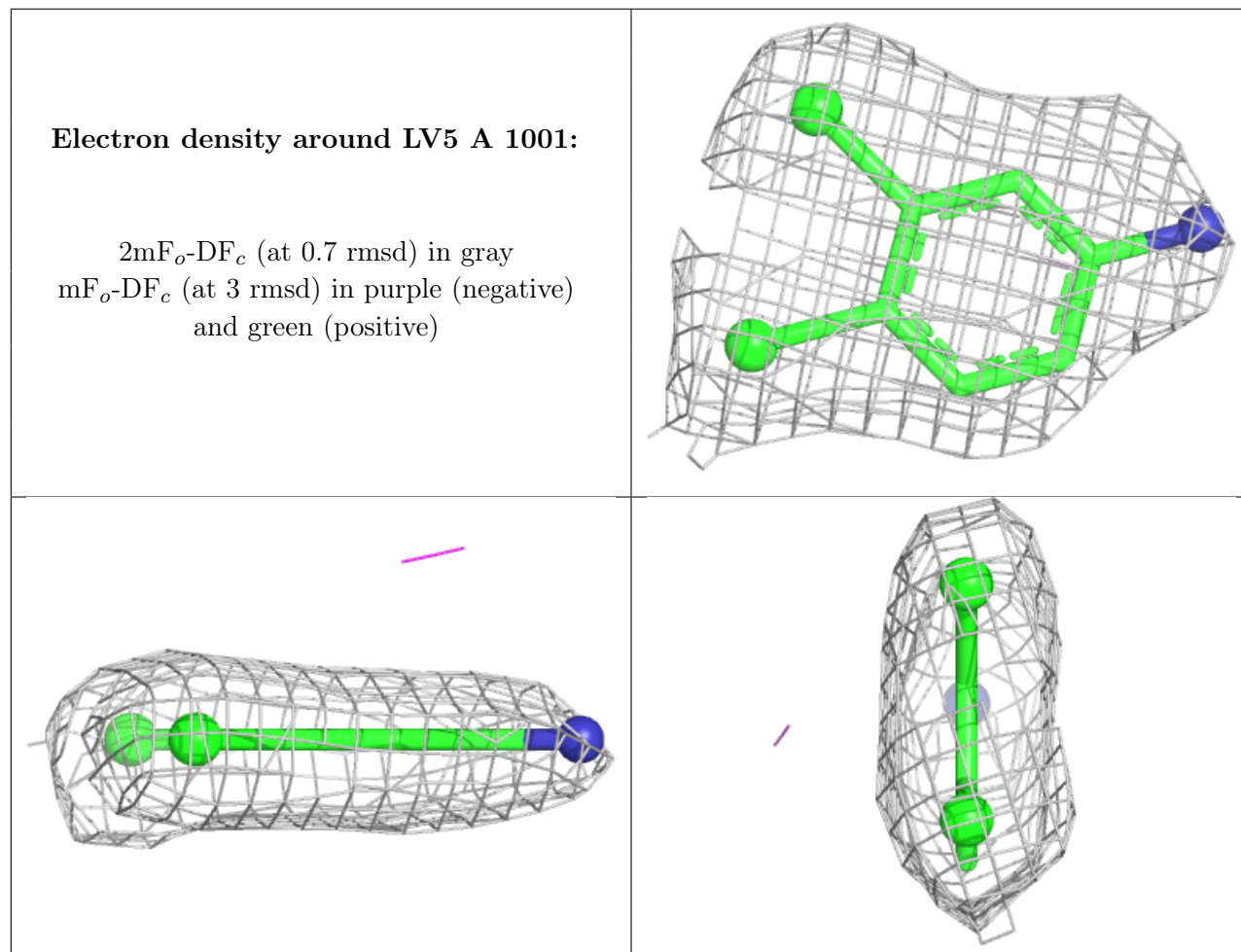There are no non-standard protein/DNA/RNA residues in this entry.

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
2	LV5	A	1001	9/9	0.91	0.22	48,51,61,62	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 LV5 A 1001:

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.