W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 25, 2024 – 02:09 PM EDT

PDB ID	:	8SO5
Title	:	Crystal structure of the engineered quorum quenching acylase MacQ variant M1 - acylated form
Authors	:	Sompiyachoke, K.; Elias, M.
Deposited on	:	2023-04-28
Resolution	:	2.35 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

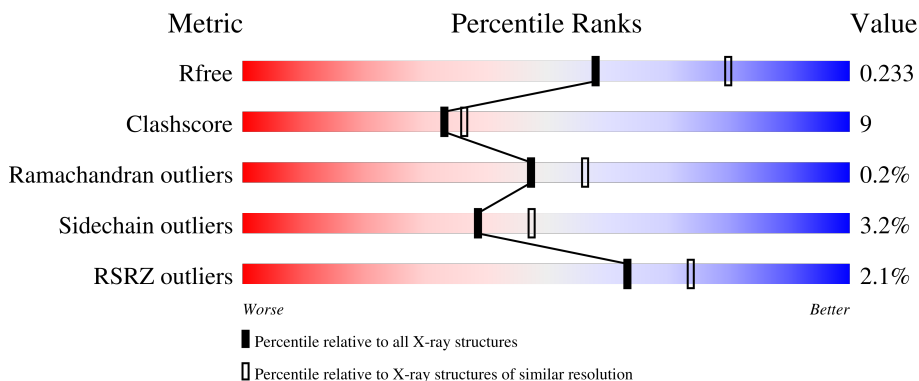|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.1

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 2.35 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		0.233
Clashscore		9
Ramachandran outliers		0.2%
Sidechain outliers		3.2%
RSRZ outliers		2.1%

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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	G	209	2% / 70% / 11% / 19%
1	J	209	67% / 13% / 19%
2	I	575	4% / 83% / 15% / •
2	L	575	% / 81% / 17% / •

WORLDWIDE PDB
PROTEIN DATA BANK

2   Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 11941 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 Protein related to penicillin acylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	169	Total 1304	C 810	N 246	O 242	S 6	0	0	0
1	J	169	Total 1304	C 810	N 246	O 242	S 6	0	0	0

- Molecule 2 is a protein called Protein related to penicillin acylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	574	Total 4365	C 2757	N 763	O 827	S 18	0	0	0
2	L	575	Total 4374	C 2762	N 764	O 830	S 18	0	0	0

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	17	LEU	VAL	conflict	UNP A0A0A1VBK6
I	85	PRO	GLN	conflict	UNP A0A0A1VBK6
I	87	ASP	GLU	conflict	UNP A0A0A1VBK6
I	93	VAL	ARG	conflict	UNP A0A0A1VBK6
I	146	GLN	GLY	conflict	UNP A0A0A1VBK6
I	186	GLY	SER	conflict	UNP A0A0A1VBK6
I	199	ASP	SER	conflict	UNP A0A0A1VBK6
I	209	VAL	ALA	conflict	UNP A0A0A1VBK6
I	218	LEU	THR	conflict	UNP A0A0A1VBK6
I	220	ARG	ALA	conflict	UNP A0A0A1VBK6
I	227	LYS	MET	conflict	UNP A0A0A1VBK6
I	229	PHE	VAL	conflict	UNP A0A0A1VBK6
I	230	ALA	GLY	conflict	UNP A0A0A1VBK6
I	257	PRO	LYS	conflict	UNP A0A0A1VBK6
I	260	ILE	VAL	conflict	UNP A0A0A1VBK6
I	316	VAL	ALA	conflict	UNP A0A0A1VBK6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
I	327	PRO	ALA	conflict	UNP A0A0A1VBK6
I	333	PRO	SER	conflict	UNP A0A0A1VBK6
I	343	ASN	SER	conflict	UNP A0A0A1VBK6
I	395	ALA	SER	conflict	UNP A0A0A1VBK6
I	443	VAL	ILE	conflict	UNP A0A0A1VBK6
I	463	VAL	ALA	conflict	UNP A0A0A1VBK6
I	527	TYR	PHE	conflict	UNP A0A0A1VBK6
I	574	LEU	-	expression tag	UNP A0A0A1VBK6
I	575	GLU	-	expression tag	UNP A0A0A1VBK6
L	17	LEU	VAL	conflict	UNP A0A0A1VBK6
L	85	PRO	GLN	conflict	UNP A0A0A1VBK6
L	87	ASP	GLU	conflict	UNP A0A0A1VBK6
L	93	VAL	ARG	conflict	UNP A0A0A1VBK6
L	146	GLN	GLY	conflict	UNP A0A0A1VBK6
L	186	GLY	SER	conflict	UNP A0A0A1VBK6
L	199	ASP	SER	conflict	UNP A0A0A1VBK6
L	209	VAL	ALA	conflict	UNP A0A0A1VBK6
L	218	LEU	THR	conflict	UNP A0A0A1VBK6
L	220	ARG	ALA	conflict	UNP A0A0A1VBK6
L	227	LYS	MET	conflict	UNP A0A0A1VBK6
L	229	PHE	VAL	conflict	UNP A0A0A1VBK6
L	230	ALA	GLY	conflict	UNP A0A0A1VBK6
L	257	PRO	LYS	conflict	UNP A0A0A1VBK6
L	260	ILE	VAL	conflict	UNP A0A0A1VBK6
L	316	VAL	ALA	conflict	UNP A0A0A1VBK6
L	327	PRO	ALA	conflict	UNP A0A0A1VBK6
L	333	PRO	SER	conflict	UNP A0A0A1VBK6
L	343	ASN	SER	conflict	UNP A0A0A1VBK6
L	395	ALA	SER	conflict	UNP A0A0A1VBK6
L	443	VAL	ILE	conflict	UNP A0A0A1VBK6
L	463	VAL	ALA	conflict	UNP A0A0A1VBK6
L	527	TYR	PHE	conflict	UNP A0A0A1VBK6
L	574	LEU	-	expression tag	UNP A0A0A1VBK6
L	575	GLU	-	expression tag	UNP A0A0A1VBK6

- Molecule 3 is DECANOIC ACID (three-letter code: DKA) (formula: $C_{10}H_{20}O_2$) (labeled as "Ligand of Interest" by depositor).

DKA
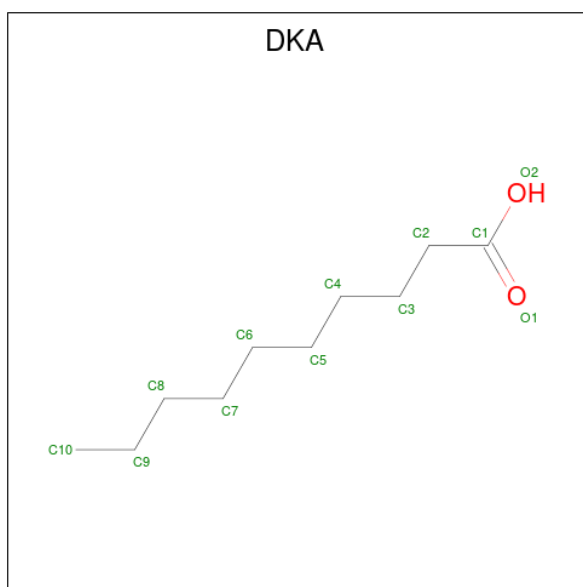
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	I	1	Total 11	C 10	O 1	0	0
3	L	1	Total 11	C 10	O 1	0	0

• Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	G	57	Total 57	O 57	0	0
4	I	172	Total 172	O 172	0	0
4	J	90	Total 90	O 90	0	0
4	L	253	Total 253	O 253	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.
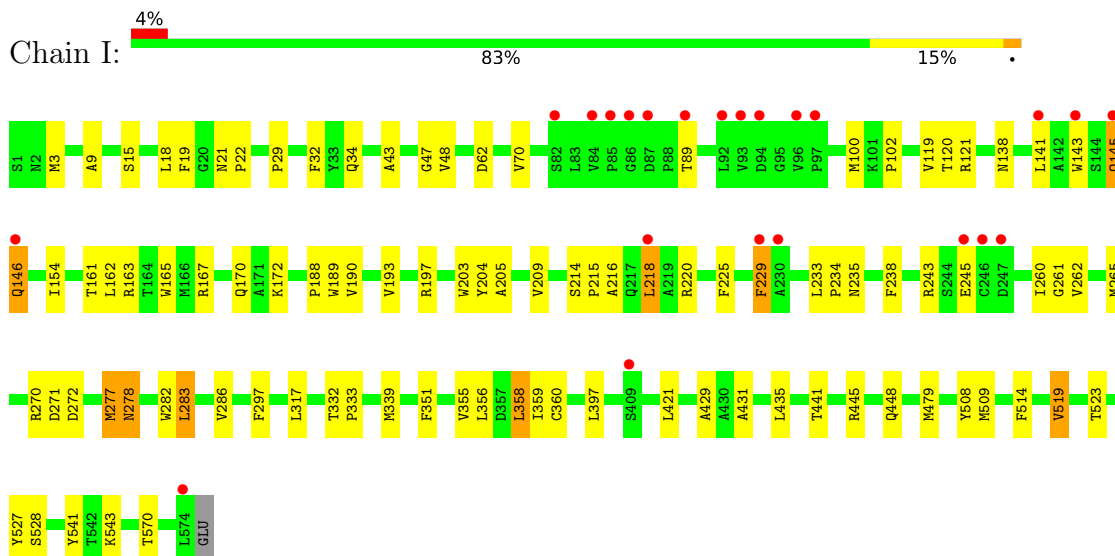
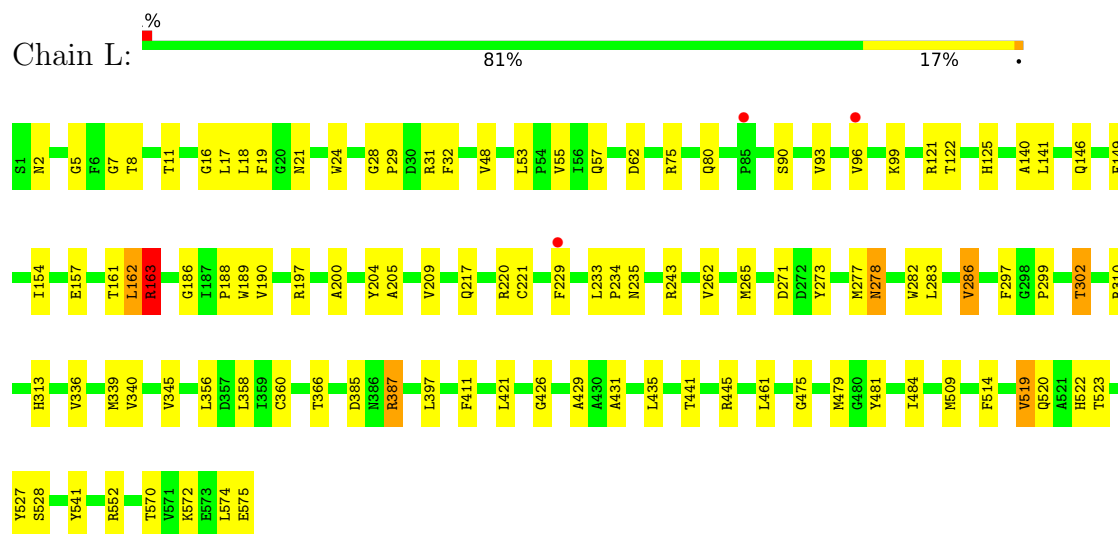• Molecule 1: Protein related to penicillin acylase
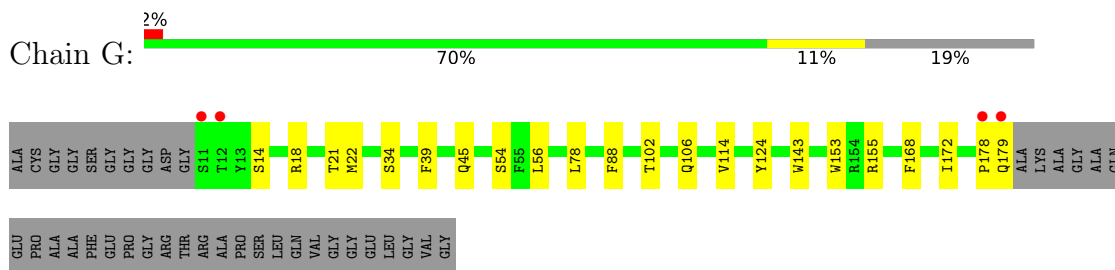
Chain G:

• Molecule 1: Protein related to penicillin acylase

Chain J:

• Molecule 2: Protein related to penicillin acylase

Chain I:

• Molecule 2: Protein related to penicillin acylase

Chain L:
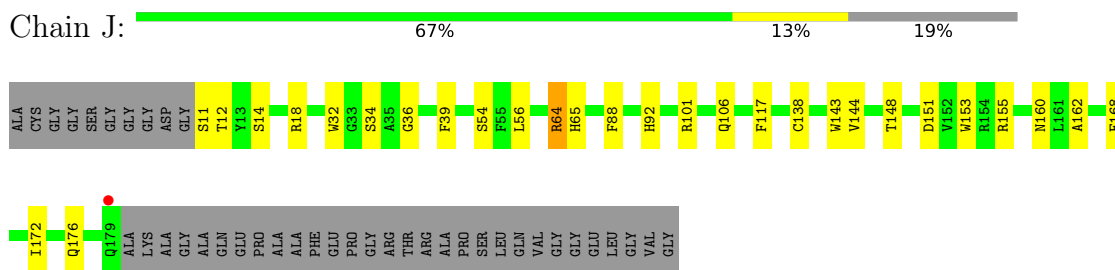
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	97.07Å 80.65Å 101.60Å 90.00° 90.81° 90.00°	Depositor
Resolution (Å)	53.16 – 2.35 53.16 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.8 (53.16-2.35) 98.9 (53.16-2.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.85 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0405	Depositor
R, R_{free}	0.185 , 0.228 0.192 , 0.233	Depositor DCC
R_{free} test set	3243 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	27.6	Xtriage
Anisotropy	0.093	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 30.6	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.005 for l,k,-h 0.036 for h,-k,-l 0.021 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11941	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% 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: DKA

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	G	0.26	0/1335	0.40	0/1808
1	J	0.26	0/1335	0.40	0/1808
2	I	0.26	0/4484	0.43	0/6125
2	L	0.26	0/4493	0.43	0/6137
All	All	0.26	0/11647	0.42	0/15878

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 mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	J	0	1
2	L	0	2
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	J	64	ARG	Sidechain
2	L	163	ARG	Sidechain
2	L	387	ARG	Sidechain

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	G	1304	0	1241	15	0
1	J	1304	0	1241	25	0
2	I	4365	0	4180	75	0
2	L	4374	0	4186	93	0
3	I	11	0	19	1	0
3	L	11	0	19	1	0
4	G	57	0	0	0	0
4	I	172	0	0	6	0
4	J	90	0	0	5	0
4	L	253	0	0	8	0
All	All	11941	0	10886	191	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 9.

The worst 5 of 191 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)
2:L:141:LEU:HD21	2:L:229:PHE:CZ	2.08	0.88
2:L:93:VAL:O	2:L:96:VAL:HG12	1.81	0.80
1:J:101:ARG:NH1	2:L:157:GLU:OE2	2.15	0.79
2:I:358:LEU:HD21	2:I:429:ALA:HA	1.63	0.78
2:I:32:PHE:CD1	2:I:48:VAL:CG1	2.66	0.77

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	G	167/209 (80%)	162 (97%)	5 (3%)	0	100	100
1	J	167/209 (80%)	161 (96%)	6 (4%)	0	100	100
2	I	572/575 (100%)	540 (94%)	30 (5%)	2 (0%)	41	47
2	L	573/575 (100%)	549 (96%)	23 (4%)	1 (0%)	47	56
All	All	1479/1568 (94%)	1412 (96%)	64 (4%)	3 (0%)	47	56

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	I	278	ASN
2	L	278	ASN
2	I	188	PRO

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	G	127/148 (86%)	124 (98%)	3 (2%)	49	59
1	J	127/148 (86%)	124 (98%)	3 (2%)	49	59
2	I	449/450 (100%)	431 (96%)	18 (4%)	31	39
2	L	450/450 (100%)	437 (97%)	13 (3%)	42	52
All	All	1153/1196 (96%)	1116 (97%)	37 (3%)	39	47

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	L	286	VAL
2	L	570	THR
2	L	302	THR
2	L	479	MET
2	I	283	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	L	125	HIS
2	L	145	GLN
2	L	538	HIS
2	L	448	GLN
2	L	493	GLN

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	DKA	L	601	2	10,10,11	0.23	0	9,9,11	0.13	0
3	DKA	I	601	2	10,10,11	0.26	0	9,9,11	0.16	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
3	DKA	L	601	2	-	6/7/8/9	-
3	DKA	I	601	2	-	4/7/8/9	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	I	601	DKA	C4-C5-C6-C7
3	L	601	DKA	C5-C6-C7-C8
3	L	601	DKA	C2-C3-C4-C5
3	I	601	DKA	C7-C8-C9-C10
3	L	601	DKA	C7-C8-C9-C10

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	L	601	DKA	1	0
3	I	601	DKA	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 DKA L 601

Ligand DKA I 601
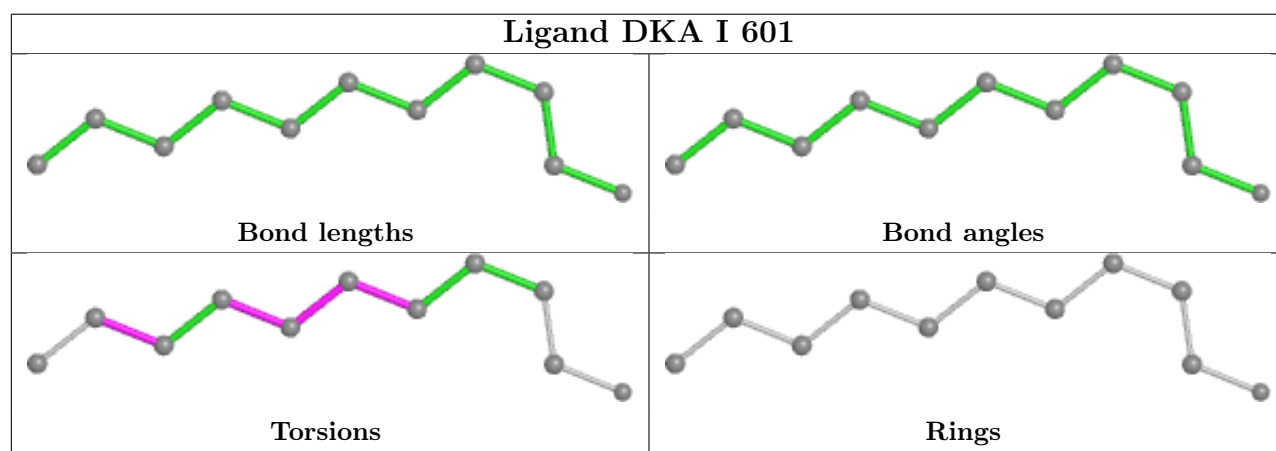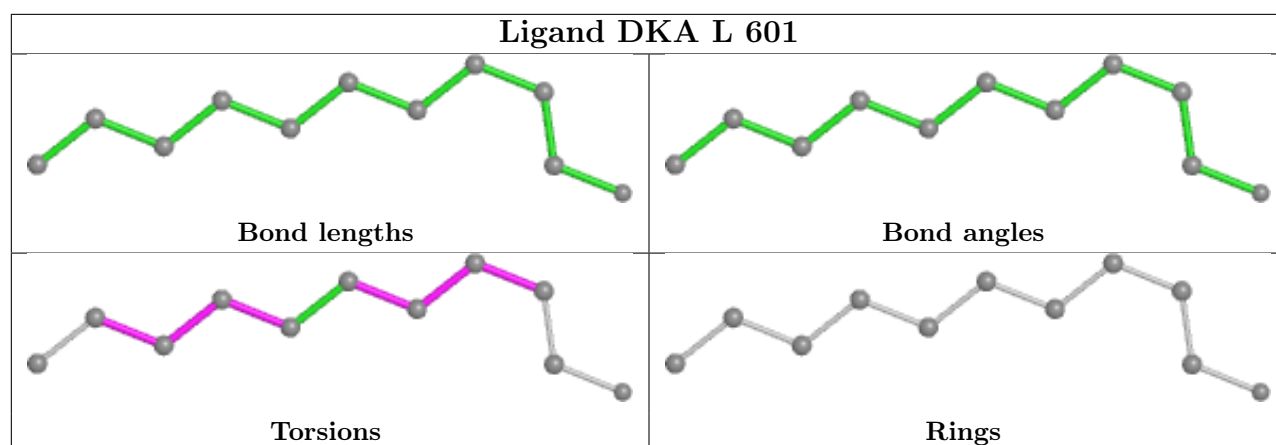
5.7 Other polymers ⓘ

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	G	169/209 (80%)	-0.02	4 (2%)	59	68	19, 30, 50, 97	0
1	J	169/209 (80%)	-0.15	1 (0%)	89	93	15, 23, 38, 68	0
2	I	574/575 (99%)	0.25	23 (4%)	38	51	18, 34, 65, 113	0
2	L	575/575 (100%)	-0.04	3 (0%)	91	95	14, 24, 43, 95	0
All	All	1487/1568 (94%)	0.06	31 (2%)	63	74	14, 27, 55, 113	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	11	SER	5.4
2	I	96	VAL	4.7
2	I	84	VAL	4.6
2	I	97	PRO	4.6
2	I	229	PHE	4.5

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

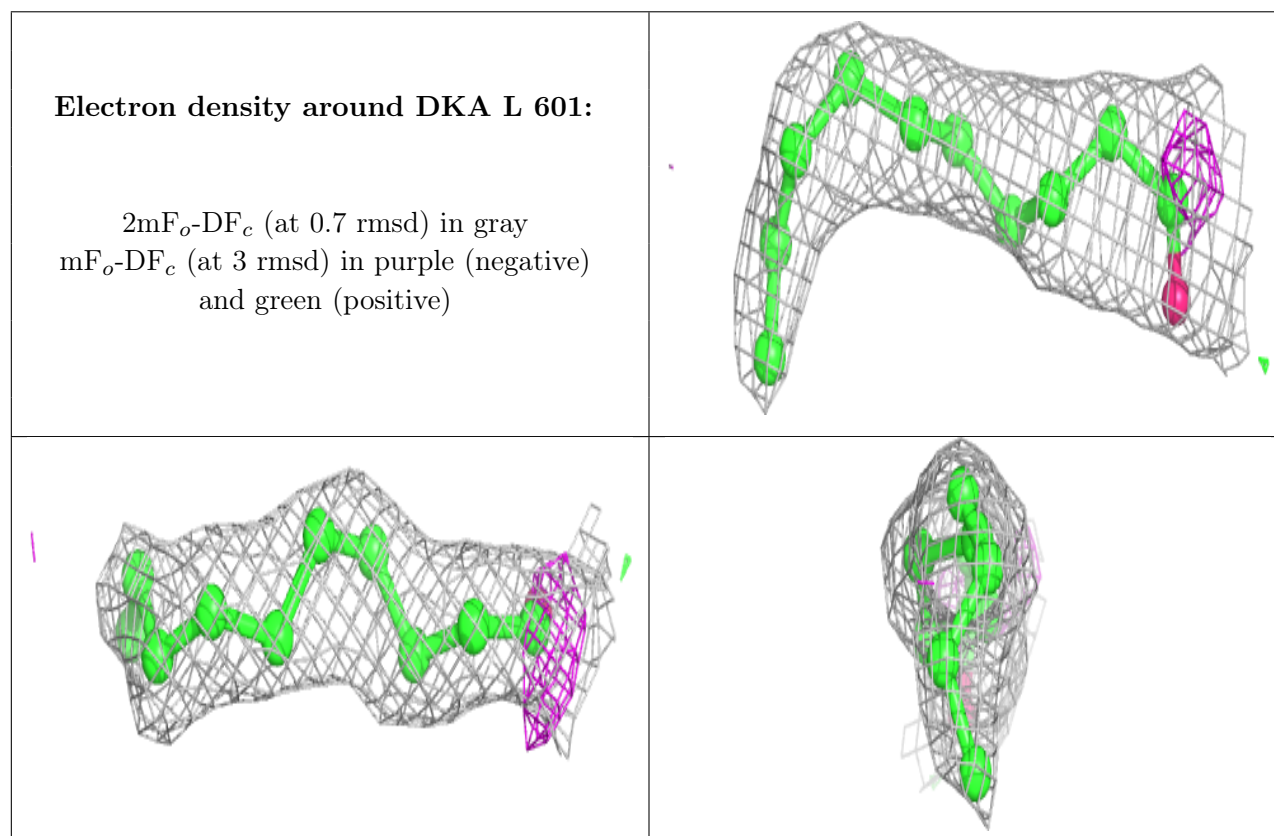There are no non-standard protein/DNA/RNA residues in this entry.

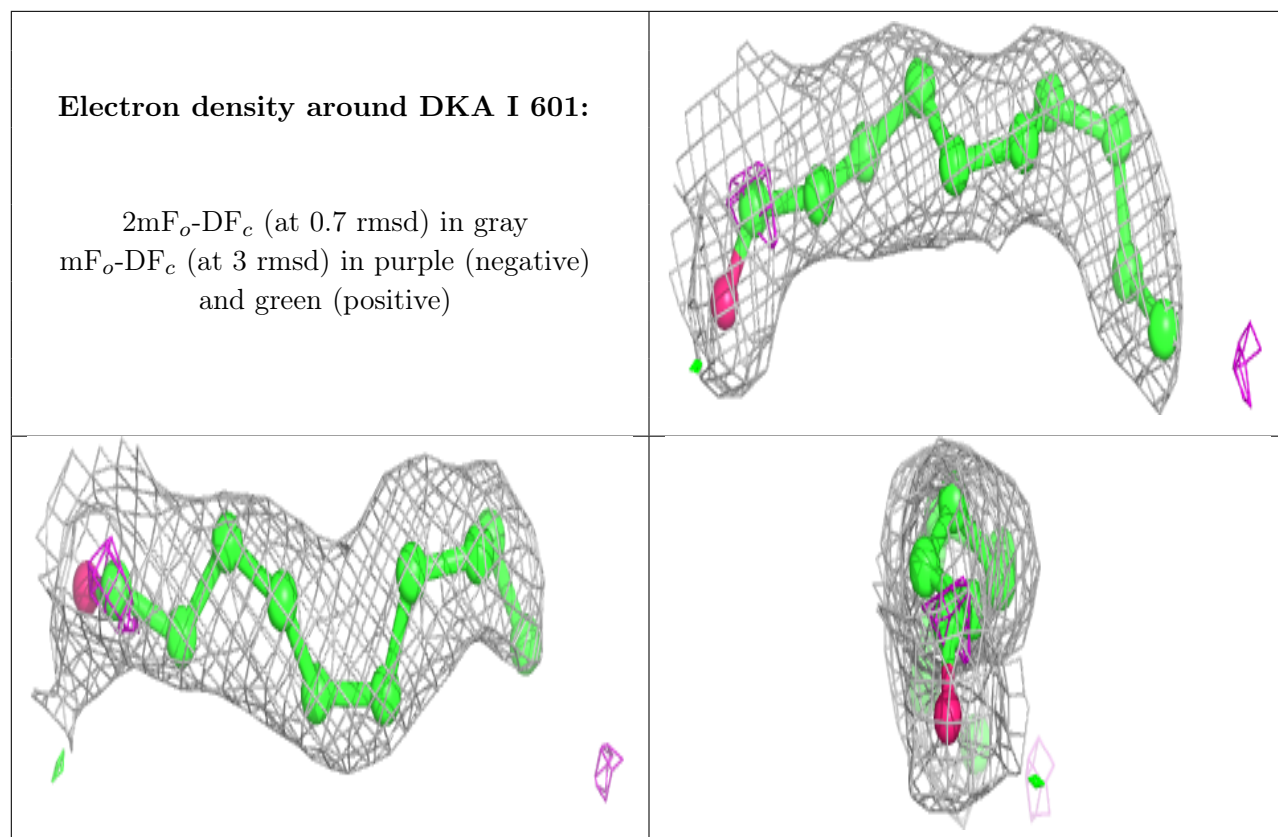## 6.3 Carbohydrates ⓘ

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	DKA	L	601	11/12	0.87	0.24	34,36,40,42	0
3	DKA	I	601	11/12	0.88	0.23	38,40,46,47	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 DKA L 601:

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

Electron density around DKA I 601:

$2mF_o$-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.