Full wwPDB X-ray Structure Validation Report (i)

Jan 3, 2024 – 02:36 pm GMT

PDB ID	:	5K8Y
Title	:	Structure of the Mus musclus Langerin carbohydrate recognition domain
Authors	:	Loll, B.; Aretz, J.; Rademacher, C.; Wahl, M.C.
Deposited on	:	2016-05-31
Resolution	:	2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

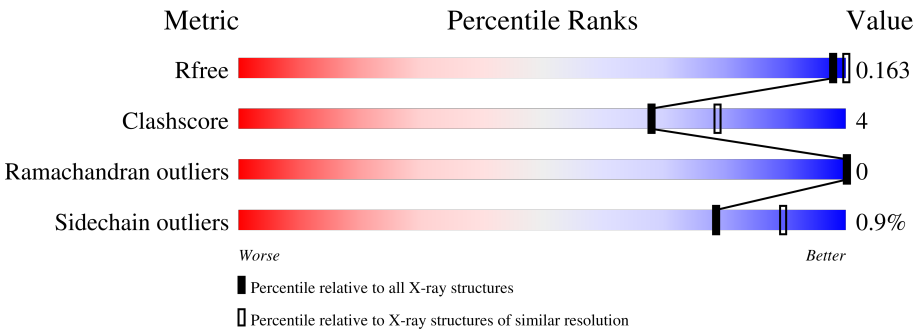
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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%

Mol	Chain	Length	Quality of chain
1	A	151	81% 9% 10%
1	B	151	77% 13% 10%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2366 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 C-type lectin domain family 4 member K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	136	1112	714	192	200	6	0	0	0
1	B	136	1130	724	195	205	6	0	2	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	191	MET	-	initiating methionine	UNP Q8VBX4
A	192	ALA	-	cloning artifact	UNP Q8VBX4
A	193	GLY	-	cloning artifact	UNP Q8VBX4
A	332	GLY	-	expression tag	UNP Q8VBX4
A	333	THR	-	expression tag	UNP Q8VBX4
A	334	TRP	-	expression tag	UNP Q8VBX4
A	335	SER	-	expression tag	UNP Q8VBX4
A	336	HIS	-	expression tag	UNP Q8VBX4
A	337	PRO	-	expression tag	UNP Q8VBX4
A	338	GLN	-	expression tag	UNP Q8VBX4
A	339	PHE	-	expression tag	UNP Q8VBX4
A	340	GLU	-	expression tag	UNP Q8VBX4
A	341	LYS	-	expression tag	UNP Q8VBX4
B	191	MET	-	initiating methionine	UNP Q8VBX4
B	192	ALA	-	cloning artifact	UNP Q8VBX4
B	193	GLY	-	cloning artifact	UNP Q8VBX4
B	332	GLY	-	expression tag	UNP Q8VBX4
B	333	THR	-	expression tag	UNP Q8VBX4
B	334	TRP	-	expression tag	UNP Q8VBX4
B	335	SER	-	expression tag	UNP Q8VBX4
B	336	HIS	-	expression tag	UNP Q8VBX4
B	337	PRO	-	expression tag	UNP Q8VBX4
B	338	GLN	-	expression tag	UNP Q8VBX4
B	339	PHE	-	expression tag	UNP Q8VBX4
B	340	GLU	-	expression tag	UNP Q8VBX4

Continued on next page...

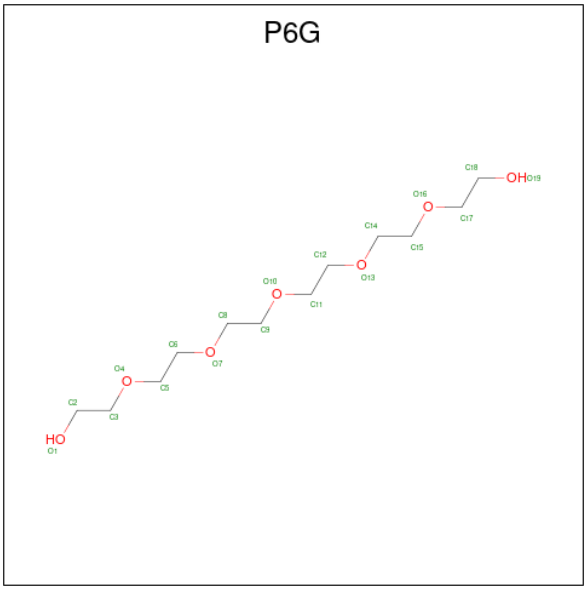
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	341	LYS	-	expression tag	UNP Q8VBX4

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

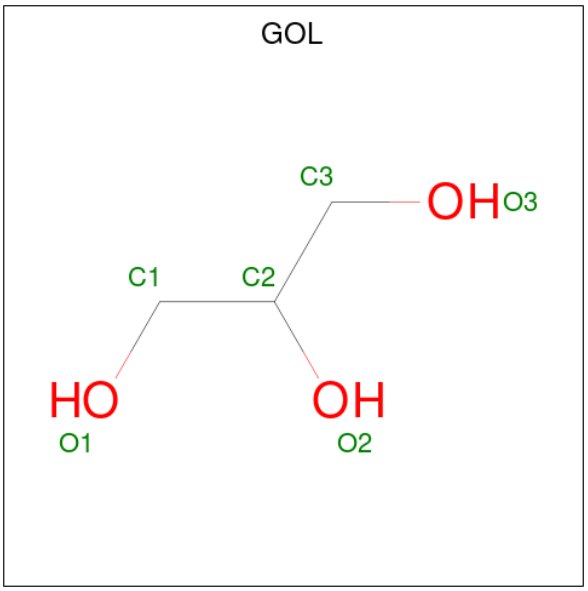
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Ca 1 1	0	0
2	B	1	Total Ca 1 1	0	0

- Molecule 3 is HEXAETHYLENE GLYCOL (three-letter code: P6G) (formula: $C_{12}H_{26}O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 10 6 4	0	0
3	A	1	Total C O 10 6 4	0	0
3	A	1	Total C O 16 10 6	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 5 is water.

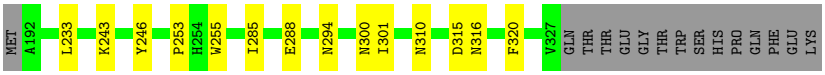
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	37	37	37	0	0
			Total	O		
5	B	43	43	43	0	0

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. 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.

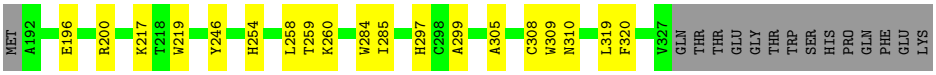
- Molecule 1: C-type lectin domain family 4 member K

Chain A:



- Molecule 1: C-type lectin domain family 4 member K

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α, β, γ	143.31Å 143.31Å 143.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.32 – 2.40 45.32 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.32-2.40) 99.9 (45.32-2.40)	Depositor EDS
R_{merge}	0.78	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.23 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.196 , 0.234 0.141 , 0.163	Depositor DCC
R_{free} test set	972 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	49.8	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 41.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.268 for -l,-k,-h	Xtriage
Reported twinning fraction	0.530 for H, K, L 0.470 for -H, L, K	Depositor
Outliers	0 of 19343 reflections	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2366	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, P6G, CA

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	0.46	0/1150	0.61	0/1559
1	B	0.46	0/1168	0.62	0/1583
All	All	0.46	0/2318	0.62	0/3142

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	1112	0	1038	8	0
1	B	1130	0	1050	9	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	36	0	47	2	0
4	B	6	0	8	0	0
5	A	37	0	0	1	0
5	B	43	0	0	0	0
All	All	2366	0	2143	17	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 4.

All (17) 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:B:254:HIS:HA	1:B:319:LEU:O	2.01	0.60
1:B:284:TRP:HA	1:B:309:TRP:HB2	1.87	0.57
1:B:258:LEU:HB3	1:B:299:ALA:HB3	1.88	0.55
1:B:196:GLU:O	1:B:200:ARG:HG3	2.08	0.54
1:B:260:LYS:HE3	1:B:297:HIS:HA	1.89	0.54
1:A:233:LEU:HD21	1:A:320:PHE:HB2	1.93	0.49
1:B:285:ILE:HD12	1:B:310:ASN:HB2	1.94	0.49
1:A:253:PRO:HA	1:A:301:ILE:O	2.13	0.48
1:B:217:LYS:HD2	1:B:320:PHE:CE1	2.50	0.47
1:A:285:ILE:O	1:A:288:GLU:HB2	2.17	0.44
1:A:316:ASN:HA	3:A:403:P6G:H52	2.00	0.44
1:A:315:ASP:HB3	3:A:402:P6G:H51	2.01	0.43
1:A:243:LYS:HA	1:A:246:TYR:CE2	2.54	0.42
1:A:310:ASN:ND2	5:A:502:HOH:O	2.53	0.41
1:B:219:TRP:CZ2	1:B:259:THR:HB	2.55	0.41
1:A:255:TRP:CE3	1:A:300:ASN:HB3	2.56	0.41
1:B:305:ALA:O	1:B:308:CYS:HB3	2.21	0.41

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	134/151 (89%)	128 (96%)	6 (4%)	0	100	100
1	B	136/151 (90%)	131 (96%)	5 (4%)	0	100	100
All	All	270/302 (89%)	259 (96%)	11 (4%)	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	Rotameric	Outliers	Percentiles	
1	A	116/130 (89%)	115 (99%)	1 (1%)	78	90
1	B	118/130 (91%)	117 (99%)	1 (1%)	81	91
All	All	234/260 (90%)	232 (99%)	2 (1%)	78	90

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	294	ASN
1	B	246	TYR

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 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	P6G	A	404	-	15,15,18	0.51	0	14,14,17	0.22	0
3	P6G	A	402	-	9,9,18	0.49	0	8,8,17	0.23	0
3	P6G	A	403	-	9,9,18	0.56	0	8,8,17	0.30	0
4	GOL	B	402	-	5,5,5	0.29	0	5,5,5	0.33	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	P6G	A	404	-	-	6/13/13/16	-
3	P6G	A	402	-	-	4/7/7/16	-
3	P6G	A	403	-	-	5/7/7/16	-
4	GOL	B	402	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (17) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	402	GOL	O1-C1-C2-C3
3	A	404	P6G	O4-C5-C6-O7
3	A	403	P6G	O1-C2-C3-O4
4	B	402	GOL	O1-C1-C2-O2
3	A	403	P6G	O4-C5-C6-O7
3	A	403	P6G	O7-C8-C9-O10
3	A	404	P6G	C9-C8-O7-C6
3	A	402	P6G	C9-C8-O7-C6
3	A	402	P6G	C6-C5-O4-C3
3	A	402	P6G	C2-C3-O4-C5
3	A	403	P6G	C9-C8-O7-C6
3	A	404	P6G	C15-C14-O13-C12
3	A	404	P6G	C11-C12-O13-C14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	404	P6G	O10-C11-C12-O13
3	A	403	P6G	C2-C3-O4-C5
3	A	404	P6G	C2-C3-O4-C5
3	A	402	P6G	O4-C5-C6-O7

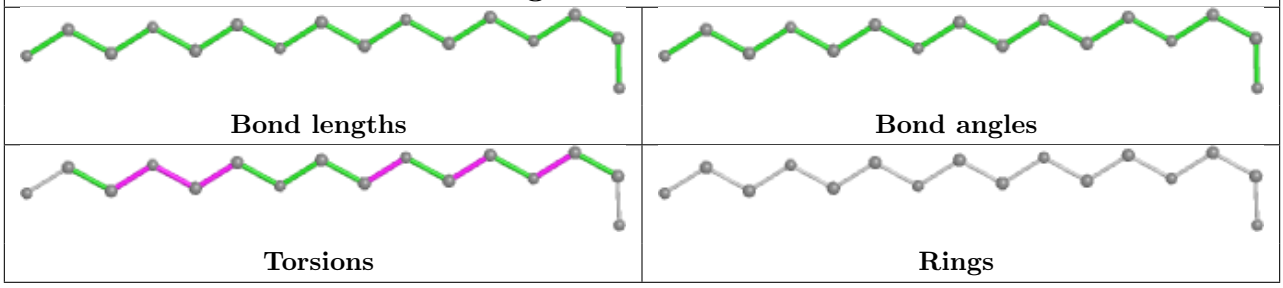
There are no ring outliers.

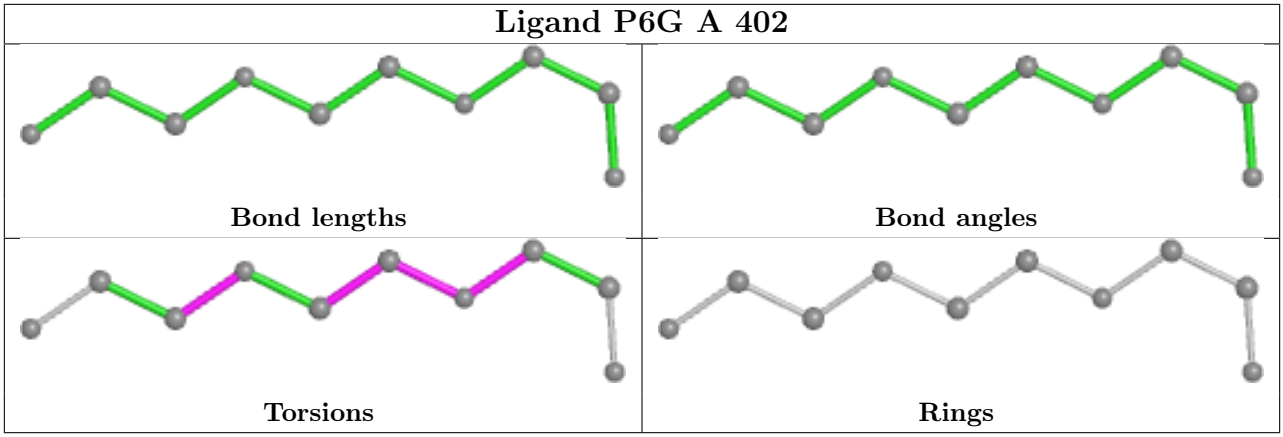
2 monomers are involved in 2 short contacts:

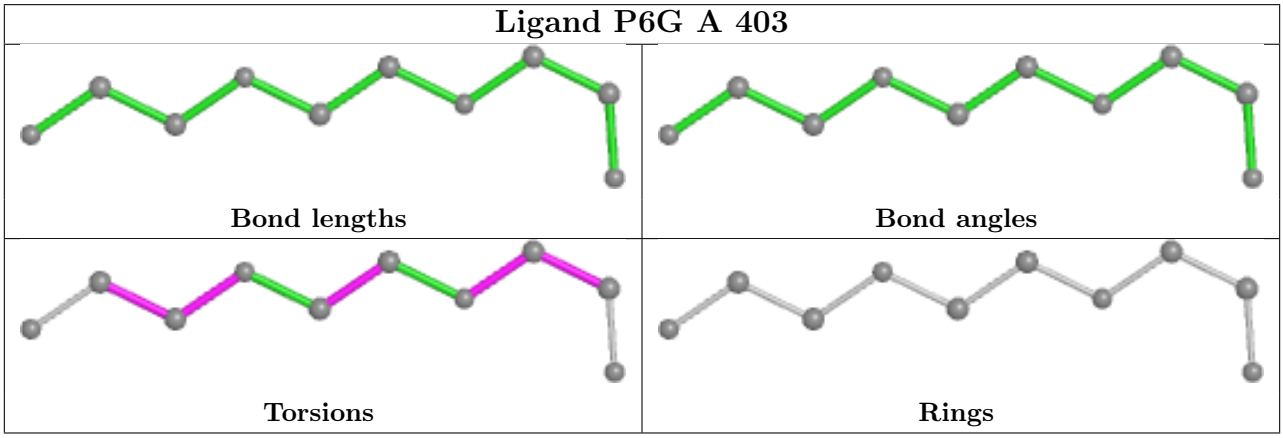
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	402	P6G	1	0
3	A	403	P6G	1	0

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 P6G A 404







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.

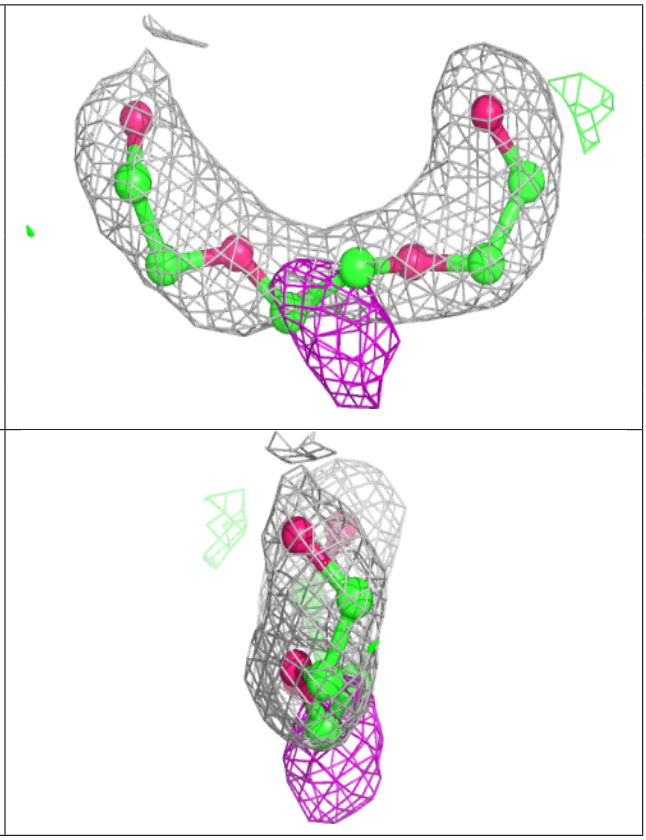
6.4 Ligands (i)

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.

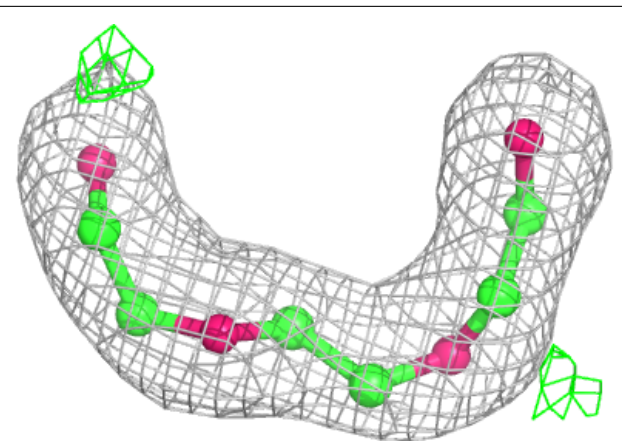
Electron density around P6G A 402:

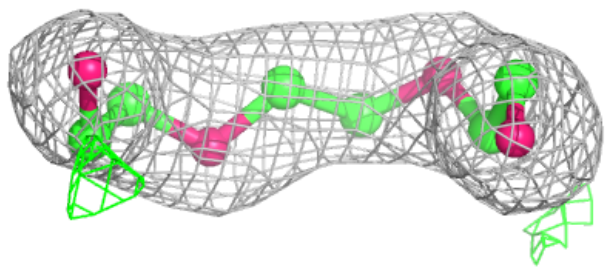
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

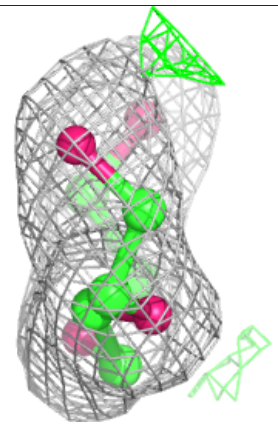


Electron density around P6G A 403:

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

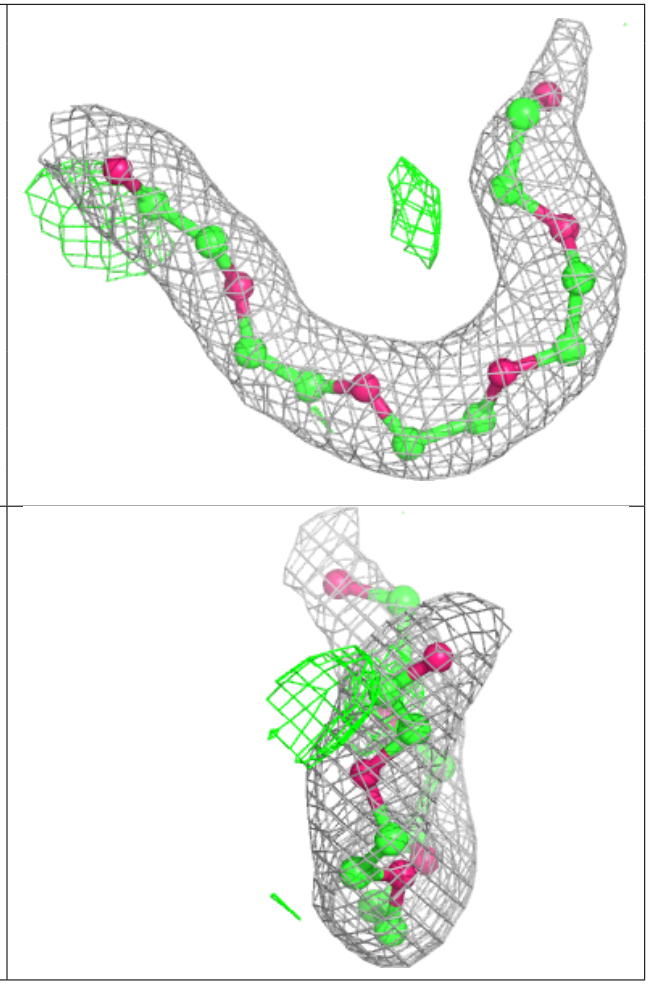






Electron density around P6G A 404:

$2mF_o$-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)

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