wwPDB EM Validation Summary Report

May 23, 2024 – 10:07 PM JST

PDB ID	:	8WLR
EMDB ID	:	EMD-37629
Title	:	Cryo-EM structure of SARS-CoV-2 prototype spike protein receptor-binding domain in complex with hippopotamus ACE2
Authors	:	Han, P.; Yang, R.R.; Li, S.H.
Deposited on	:	2023-09-30
Resolution	:	3.12 Å(reported)

This is a wwPDB EM 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/EMValidationReportHelp
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:

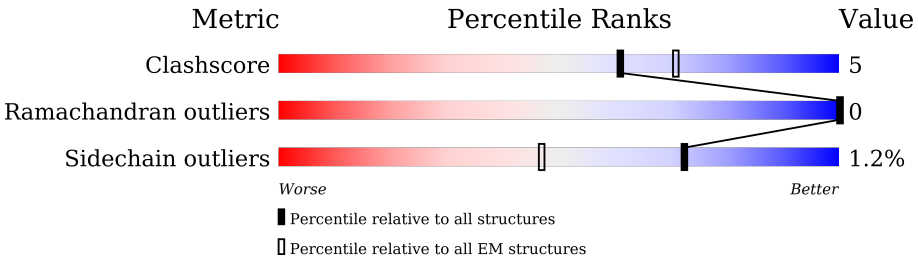
EMDB validation analysis	:	FAILED
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.12 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)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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	597	87% 13%
2	B	1217	14% · 84%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6431 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Angiotensin-converting enzyme.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	597	Total	C	N	O	S	0	0
			4873	3110	805	929	29		

- Molecule 2 is a protein called Spike glycoprotein.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	195	Total	C	N	O	S	0	0
			1543	989	257	289	8		

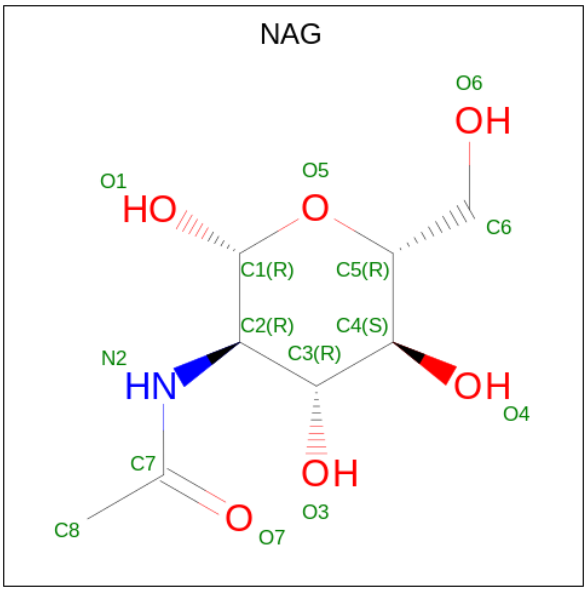
There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	682	GLY	ARG	engineered mutation	UNP P0DTC2
B	683	SER	ARG	engineered mutation	UNP P0DTC2
B	685	SER	ARG	engineered mutation	UNP P0DTC2
B	817	PRO	PHE	engineered mutation	UNP P0DTC2
B	892	PRO	ALA	engineered mutation	UNP P0DTC2
B	899	PRO	ALA	engineered mutation	UNP P0DTC2
B	942	PRO	ALA	engineered mutation	UNP P0DTC2
B	986	PRO	LYS	engineered mutation	UNP P0DTC2
B	987	PRO	VAL	engineered mutation	UNP P0DTC2

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		AltConf
3	A	1	Total	Zn	0
			1	1	

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$) (labeled as "Ligand of Interest" by depositor).



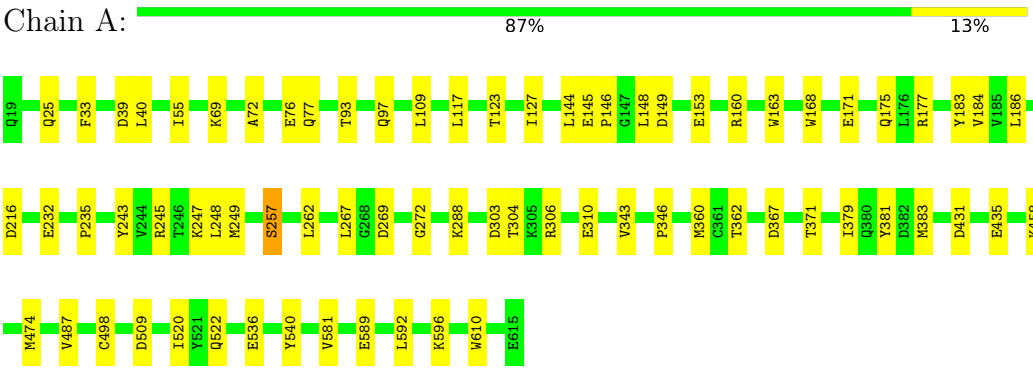
Mol	Chain	Residues	Atoms				AltConf
			Total	C	N	O	
4	B	1	14	8	1	5	0

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.

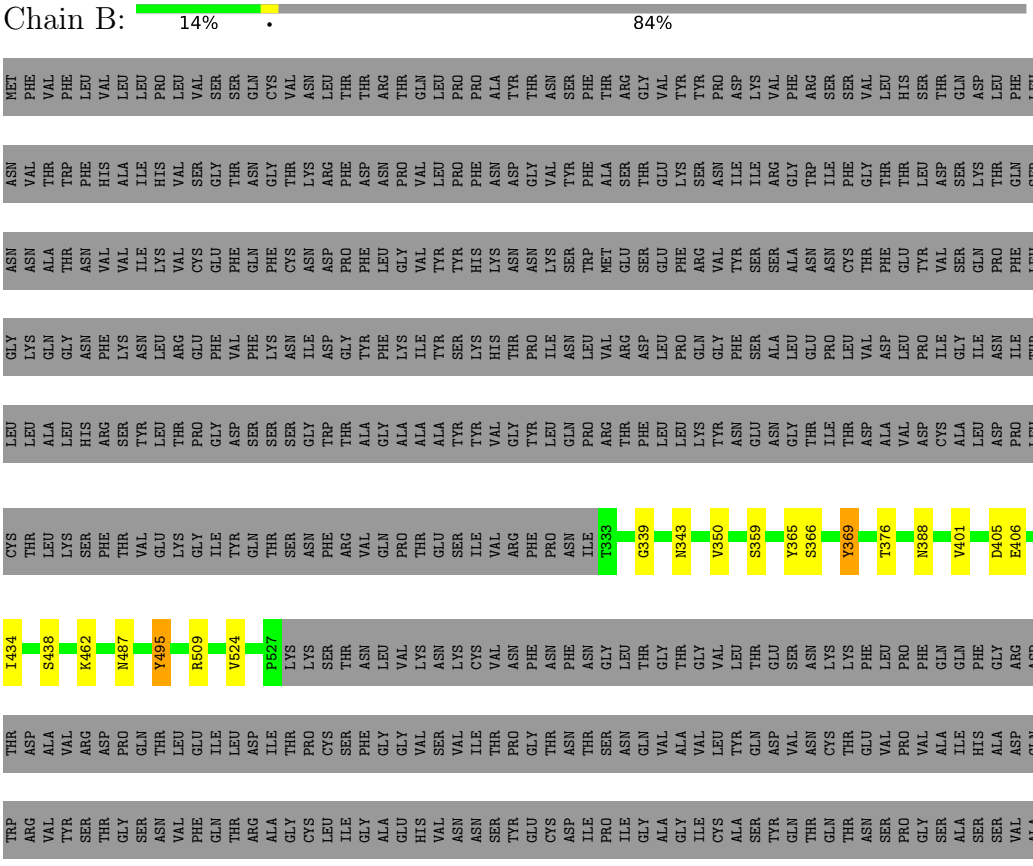
- Molecule 1: Angiotensin-converting enzyme

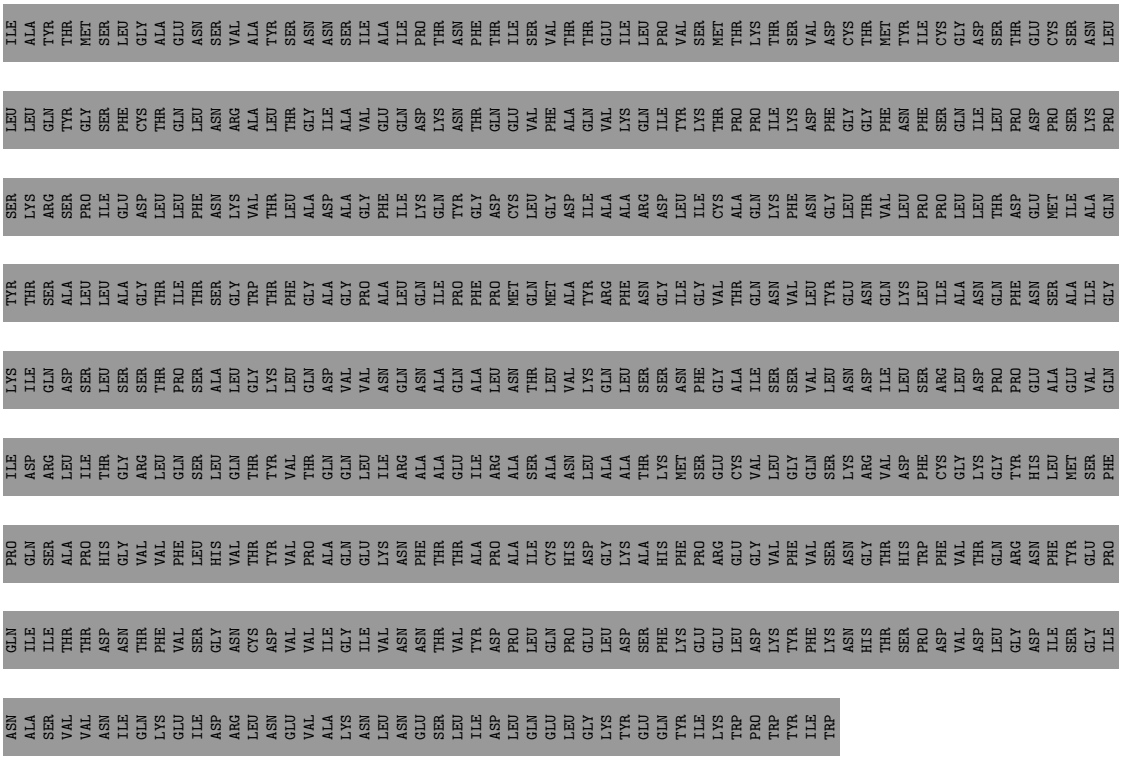
Chain A:



- Molecule 2: Spike glycoprotein

Chain B:




ILE ALA TYR THR MET SER LEU GLY ALA GLU ASN SER VAL ALA TYR SER ASN ASN SER ILE ALA ILE PRO THR ASN PHE THR ILE SER VAL THR THR GLU ILE LEU PRO VAL SER MET THR LYS THR SER VAL ASP CYS THR MET TYR ILE CYS GLY ASP SER THR GLU CYS SER LEU
LEU LEU GLN TYR GLY SER PHE CYS THR GLN LEU ASN ARG ALA LEU THR GLY ILE ALA VAL GLU GLN ASP LYS ASN THR GLN GLU VAL PHE ALA GLN VAL LYS GLN ILE TYR LYS THR PRO PRO ILE LYS ASP PHE GLY GLY PHE ASN PHE SER GLN ILE LEU PRO ASP PRO SER LYS PRO
SER LYS ARG SER PRO ILE GLU ASP LEU LEU PHE ASN LYS VAL THR LEU ALA ASP ALA GLY PHE ILE LYS GLN TYR GLY ASP CYS LEU GLY ASP ILE ALA ALA ARG ASP LEU ILE CYS ALA GLN LYS PHE ASN GLY LEU THR VAL LEU PRO PRO LEU LEU THR ASP GLU MET ILE ALA GLN
TYR THR SER ALA LEU LEU ALA GLY THR ILE THR SER GLY TRP THR PHE GLY ALA GLY PRO ALA LEU GLN ILE PRO PHE PRO MET GLN MET ALA TYR ARG PHE ASN GLY ILE GLY VAL THR GLN ASN VAL LEU TYR GLU ASN GLN LYS LEU ILE ALA ASN GLN PHE ASN SER ALA ILE GLY
LYS ILE GLN ASP SER LEU SER SER THR PRO SER ALA LEU GLY LYS LEU GLN ASP VAL VAL ASN GLN ASN ALA GLN ALA LEU ASN THR LEU VAL LYS GLN LEU SER SER ASN PHE GLY ALA ILE SER SER VAL LEU ASN ASP ILE LEU SER ARG LEU ASP PRO PRO GLU ALA GLU VAL GLN
ILE ASP ARG LEU ILE THR GLY ARG LEU GLN SER LEU GLN THR TYR VAL THR GLN GLN LEU ILE ARG ALA SER ALA ASN LEU ALA ALA THR LYS MET SER GLU CYS VAL LEU GLY GLN SER LYS ARG VAL ASP PHE CYS GLY LYS GLY TYR HIS LEU MET SER PHE
PRO GLN SER ALA PRO HIS GLY VAL VAL PHE LEU HIS VAL THR TYR VAL PRO ALA GLN GLU LYS ASN PHE THR THR ALA PRO ALA ILE CYS HIS ASP GLY LYS ALA HIS PHE PRO ARG GLU GLY VAL PHE VAL SER ASN GLY THR HIS TRP PHE VAL THR GLN ARG ASN PHE TYR GLU PRO
GLN ILE ILE THR THR ASP ASN THR PHE VAL SER GLY ASN CYS ASP VAL VAL ILE GLY ILE VAL ASN ASN THR VAL TYR ASP PRO LEU GLN PRO GLU LEU ASP SER PHE LYS GLU GLU LEU ASP LYS TYR PHE LYS ASN HIS THR SER PRO ASP VAL ASP LEU GLY ASP ILE SER GLY ILE
ASN ALA SER VAL VAL ASN ILE GLN LYS GLU ILE ASP ARG LEU ASN GLU VAL ALA LYS ASN LEU ASN GLU SER LEU ILE ASP LEU GLN GLU LEU GLY LYS TYR GLU GLN TYR ILE LYS TRP PRO TRP TYR ILE TRP

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	227185	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	60	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NAG, ZN

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	0.26	0/5007	0.47	0/6793
2	B	0.28	0/1587	0.51	0/2161
All	All	0.27	0/6594	0.48	0/8954

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	4873	0	4641	44	0
2	B	1543	0	1459	15	0
3	A	1	0	0	0	0
4	B	14	0	13	1	0
All	All	6431	0	6113	58	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 5.

The worst 5 of 58 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:160:ARG:HA	1:A:163:TRP:HD1	1.60	0.66
1:A:520:ILE:HD12	1:A:581:VAL:HG12	1.78	0.66
2:B:343:ASN:OD1	4:B:1301:NAG:N2	2.30	0.65
1:A:25:GLN:HE21	2:B:487:ASN:HD21	1.42	0.64
1:A:160:ARG:HA	1:A:163:TRP:CD1	2.35	0.60

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 PDB entries followed by that with respect to all EM entries.

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	595/597 (100%)	587 (99%)	8 (1%)	0	100	100
2	B	193/1217 (16%)	186 (96%)	7 (4%)	0	100	100
All	All	788/1814 (43%)	773 (98%)	15 (2%)	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 PDB entries followed by that with respect to all EM entries.

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	525/526 (100%)	519 (99%)	6 (1%)	73	88
2	B	168/1065 (16%)	166 (99%)	2 (1%)	71	87
All	All	693/1591 (44%)	685 (99%)	8 (1%)	72	87

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	495	TYR
2	B	369	TYR
1	A	522	GLN
1	A	381	TYR
1	A	589	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	137	ASN
2	B	474	GLN
1	A	374	HIS
2	B	501	ASN
2	B	414	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)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	$\#\|Z\|>2$
4	NAG	B	1301	2	14,14,15	0.37	0	17,19,21	0.75	1 (5%)

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
4	NAG	B	1301	2	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	1301	NAG	C1-O5-C5	2.58	115.69	112.19

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	1301	NAG	C4-C5-C6-O6
4	B	1301	NAG	O5-C5-C6-O6

There are no ring outliers.

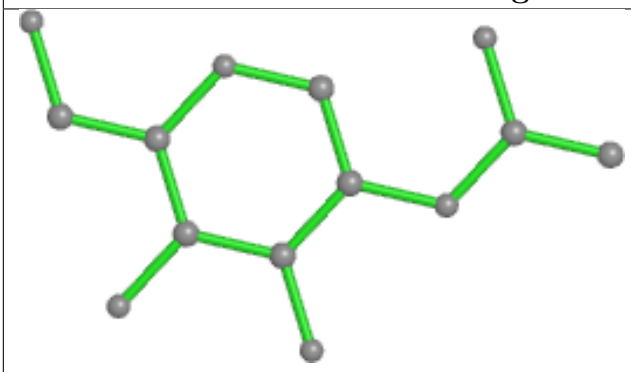
1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	1301	NAG	1	0

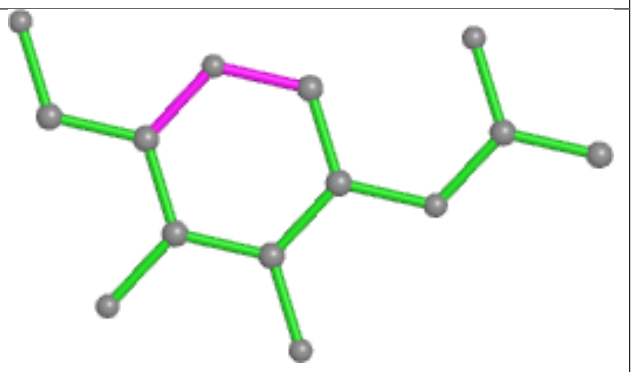
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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.

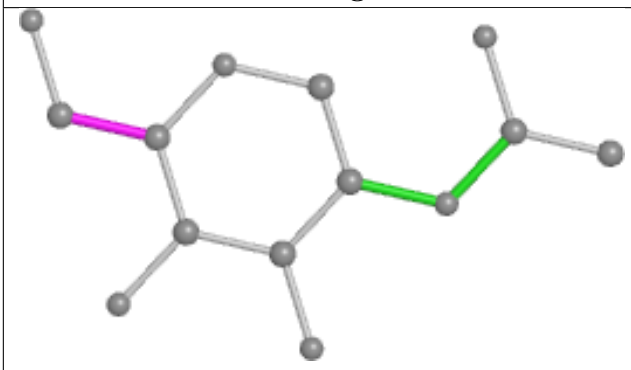
Ligand NAG B 1301



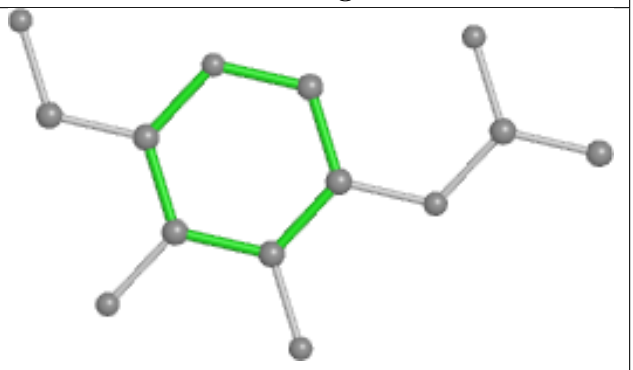
Bond lengths



Bond angles



Torsions



Rings

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.