wwPDB X-ray Structure Validation Summary Report (i)

Jun 13, 2024 – 10:09 AM EDT

PDB ID	:	1LYA
Title	:	CRYSTAL STRUCTURES OF NATIVE AND INHIBITED FORMS OF HUMAN CATHEPSIN D: IMPLICATIONS FOR LYSOSOMAL TARGETING AND DRUG DESIGN
Authors	:	Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.
Deposited on	:	1993-04-22
Resolution	:	2.50 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:

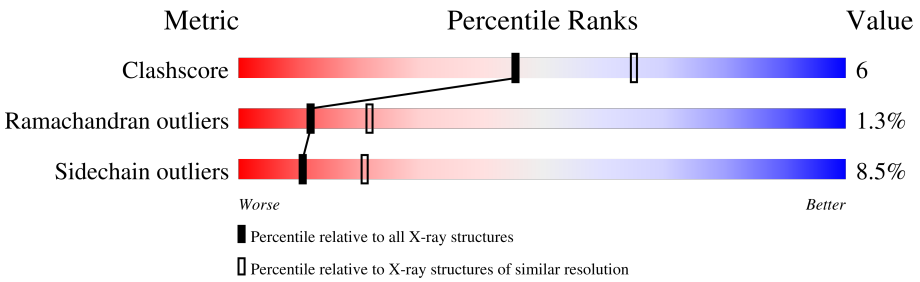
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (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:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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))
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	97	68% 30% •
1	C	97	77% 22% •
2	B	241	74% 23% •
2	D	241	75% 23% •
3	E	4	100%
3	F	4	100%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6662 atoms, of which 1302 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 CATHEPSIN D.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	97	904	478	156	119	146	5	0	0	0
1	C	97	904	478	156	119	146	5	0	0	0

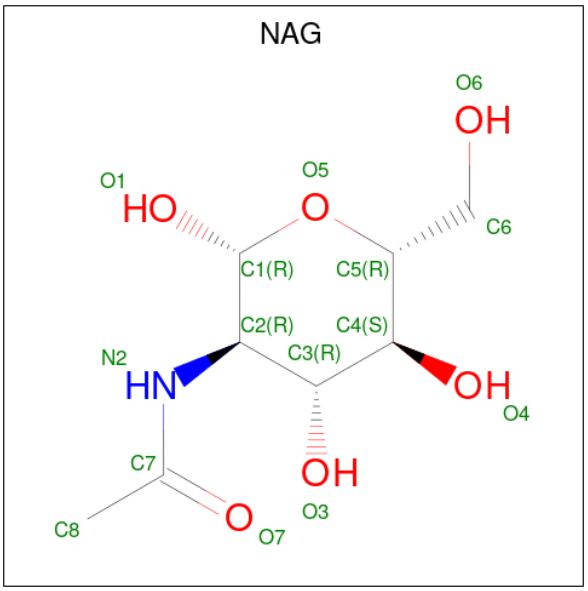
- Molecule 2 is a protein called CATHEPSIN D.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	241	2234	1184	389	302	348	11	0	0	0
2	D	241	2234	1184	389	302	348	11	0	0	0

- Molecule 3 is an oligosaccharide called alpha-D-mannopyranose-(1-6)-beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	H	N	O			
3	E	4	96	28	46	2	20	0	0	0
3	F	4	96	28	46	2	20	0	0	0

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	N	O		
4	B	1	28	8	14	1	5	0	0
			Total	C	H	N	O		
4	D	1	28	8	14	1	5	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	H	O		
5	A	11	33	22	11	0	0
			Total	H	O		
5	B	12	36	24	12	0	0
			Total	H	O		
5	C	8	24	16	8	0	0
			Total	H	O		
5	D	15	45	30	15	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.

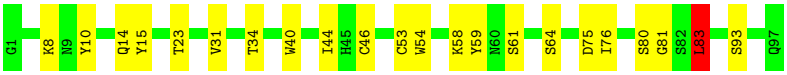
Note EDS was not executed.

• Molecule 1: CATHEPSIN D

Chain A:

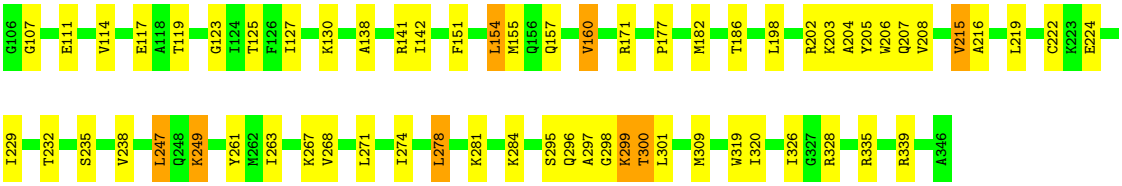
• Molecule 1: CATHEPSIN D

Chain C:



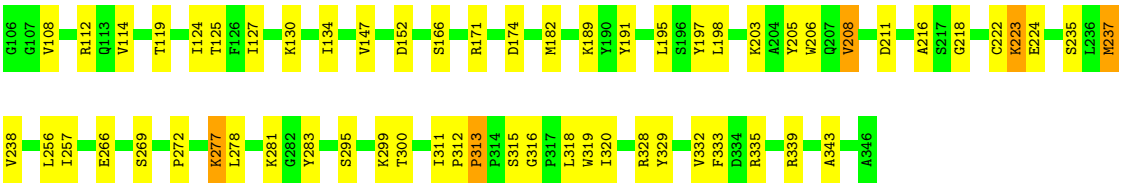
• Molecule 2: CATHEPSIN D

Chain B:



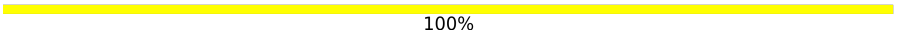
• Molecule 2: CATHEPSIN D

Chain D:



• Molecule 3: alpha-D-mannopyranose-(1-6)-beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E:



NAG1
NAG2
BMA3
MAN4

• Molecule 3: alpha-D-mannopyranose-(1-6)-beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain F: 100%

NAG1
NAG2
BMA3
MAN4

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	125.90Å 125.90Å 104.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	10.00 – 2.50	Depositor
% Data completeness (in resolution range)	(Not available) (10.00-2.50)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.0	Depositor
R, R_{free}	0.188 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	6662	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

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: MAN, BMA, NAG

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	0.81	0/771	1.48	9/1051 (0.9%)
1	C	0.82	0/771	1.44	11/1051 (1.0%)
2	B	0.73	0/1884	1.45	14/2551 (0.5%)
2	D	0.70	0/1884	1.40	11/2551 (0.4%)
All	All	0.75	0/5310	1.44	45/7204 (0.6%)

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
2	D	0	1

There are no bond length outliers.

The worst 5 of 45 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	40	TRP	CD1-CG-CD2	9.12	113.59	106.30
1	C	83	LEU	CA-CB-CG	8.71	135.32	115.30
1	A	54	TRP	CD1-CG-CD2	8.49	113.09	106.30
1	C	40	TRP	CD1-CG-CD2	8.35	112.98	106.30
2	D	319	TRP	CD1-CG-CD2	8.22	112.88	106.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	D	313	PRO	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	A	748	156	702	18	0
1	C	748	156	702	9	0
2	B	1845	389	1849	29	0
2	D	1845	389	1849	24	0
3	E	50	46	43	0	0
3	F	50	46	43	0	0
4	B	14	14	13	0	0
4	D	14	14	13	0	0
5	A	11	22	0	0	0
5	B	12	24	0	1	0
5	C	8	16	0	1	0
5	D	15	30	0	1	0
All	All	5360	1302	5214	67	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 6.

The worst 5 of 67 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:B:268:VAL:HA	2:B:271:LEU:HD22	1.71	0.72
2:D:114:VAL:HG13	2:D:147:VAL:HG13	1.72	0.71
2:D:316:GLY:HA2	2:D:318:LEU:HG	1.75	0.67
2:B:238:VAL:HG13	2:B:320:ILE:HB	1.81	0.62
2:D:311:ILE:HD11	2:D:320:ILE:HD11	1.83	0.61

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	95/97 (98%)	90 (95%)	3 (3%)	2 (2%)	7	11
1	C	95/97 (98%)	93 (98%)	1 (1%)	1 (1%)	14	26
2	B	239/241 (99%)	223 (93%)	12 (5%)	4 (2%)	9	16
2	D	239/241 (99%)	223 (93%)	14 (6%)	2 (1%)	19	35
All	All	668/676 (99%)	629 (94%)	30 (4%)	9 (1%)	12	21

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	SER
1	C	80	SER
2	B	224	GLU
2	B	300	THR
2	B	299	LYS

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	85/86 (99%)	77 (91%)	8 (9%)	8	17
1	C	85/86 (99%)	81 (95%)	4 (5%)	26	49
2	B	199/199 (100%)	181 (91%)	18 (9%)	9	19
2	D	199/199 (100%)	181 (91%)	18 (9%)	9	19
All	All	568/570 (100%)	520 (92%)	48 (8%)	10	21

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	83	LEU
2	D	198	LEU
1	C	93	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	152	ASP
2	D	208	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	60	ASN
2	D	207	GLN
2	D	338	ASN
2	B	207	GLN
2	B	121	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)

8 monosaccharides 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	NAG	E	1	1,3	14,14,15	0.76	0	17,19,21	1.92	3 (17%)
3	NAG	E	2	3	14,14,15	0.80	1 (7%)	17,19,21	1.17	1 (5%)
3	BMA	E	3	3	11,11,12	0.93	0	15,15,17	1.02	2 (13%)
3	MAN	E	4	3	11,11,12	0.93	0	15,15,17	1.32	1 (6%)
3	NAG	F	1	1,3	14,14,15	0.78	0	17,19,21	1.61	2 (11%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	NAG	F	2	3	14,14,15	0.71	0	17,19,21	2.18	7 (41%)
3	BMA	F	3	3	11,11,12	0.98	0	15,15,17	2.43	6 (40%)
3	MAN	F	4	3	11,11,12	0.96	0	15,15,17	1.14	1 (6%)

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	NAG	E	1	1,3	-	2/6/23/26	0/1/1/1
3	NAG	E	2	3	-	0/6/23/26	0/1/1/1
3	BMA	E	3	3	-	0/2/19/22	0/1/1/1
3	MAN	E	4	3	-	2/2/19/22	0/1/1/1
3	NAG	F	1	1,3	-	0/6/23/26	0/1/1/1
3	NAG	F	2	3	-	1/6/23/26	0/1/1/1
3	BMA	F	3	3	-	0/2/19/22	0/1/1/1
3	MAN	F	4	3	-	0/2/19/22	1/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	2	NAG	C1-C2	2.37	1.55	1.52

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	1	NAG	C1-O5-C5	5.63	119.82	112.19
3	F	3	BMA	C1-C2-C3	-5.06	103.45	109.67
3	F	1	NAG	C1-O5-C5	4.48	118.27	112.19
3	F	3	BMA	C6-C5-C4	4.35	123.18	113.00
3	F	2	NAG	O4-C4-C3	4.12	119.87	110.35

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	F	2	NAG	C1-C2-N2-C7
3	E	1	NAG	O5-C5-C6-O6
3	E	4	MAN	O5-C5-C6-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	E	1	NAG	C4-C5-C6-O6
3	E	4	MAN	C4-C5-C6-O6

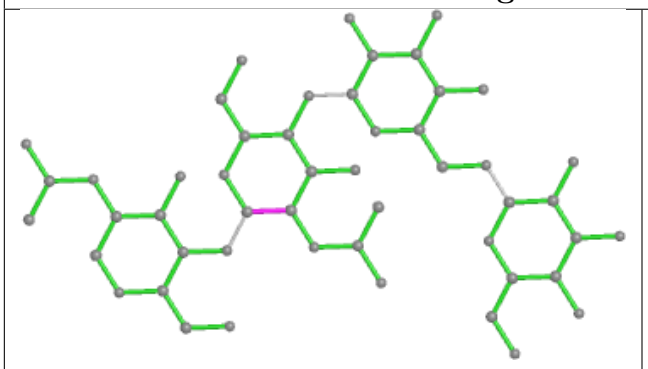
All (1) ring outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	F	4	MAN	C1-C2-C3-C4-C5-O5

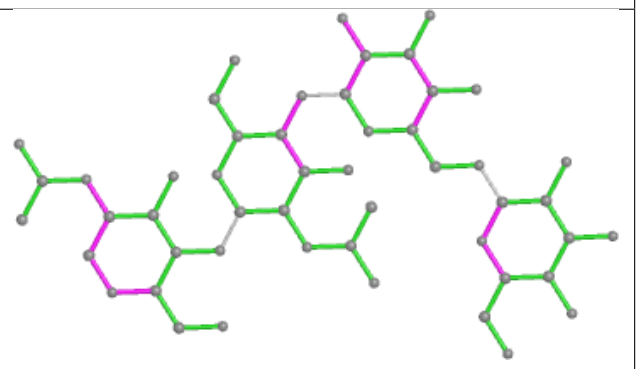
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

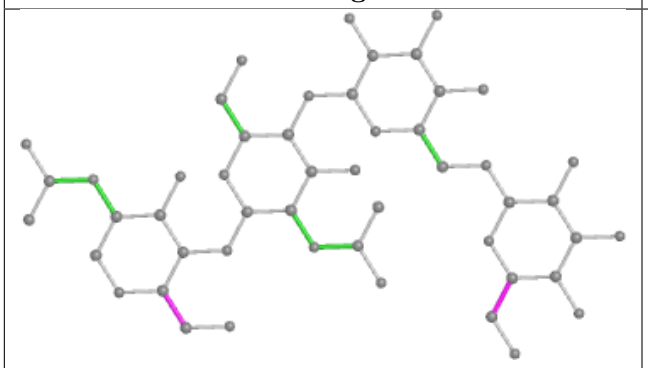
Oligosaccharide Chain E



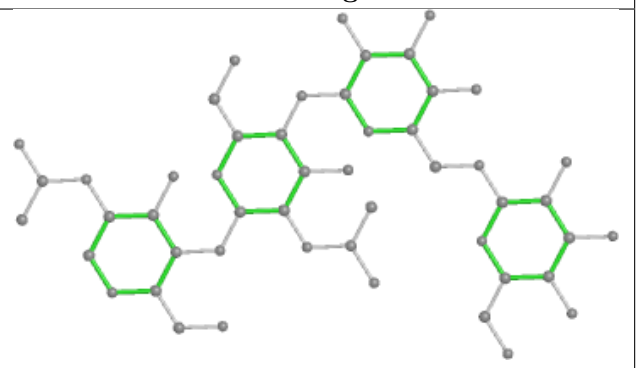
Bond lengths



Bond angles

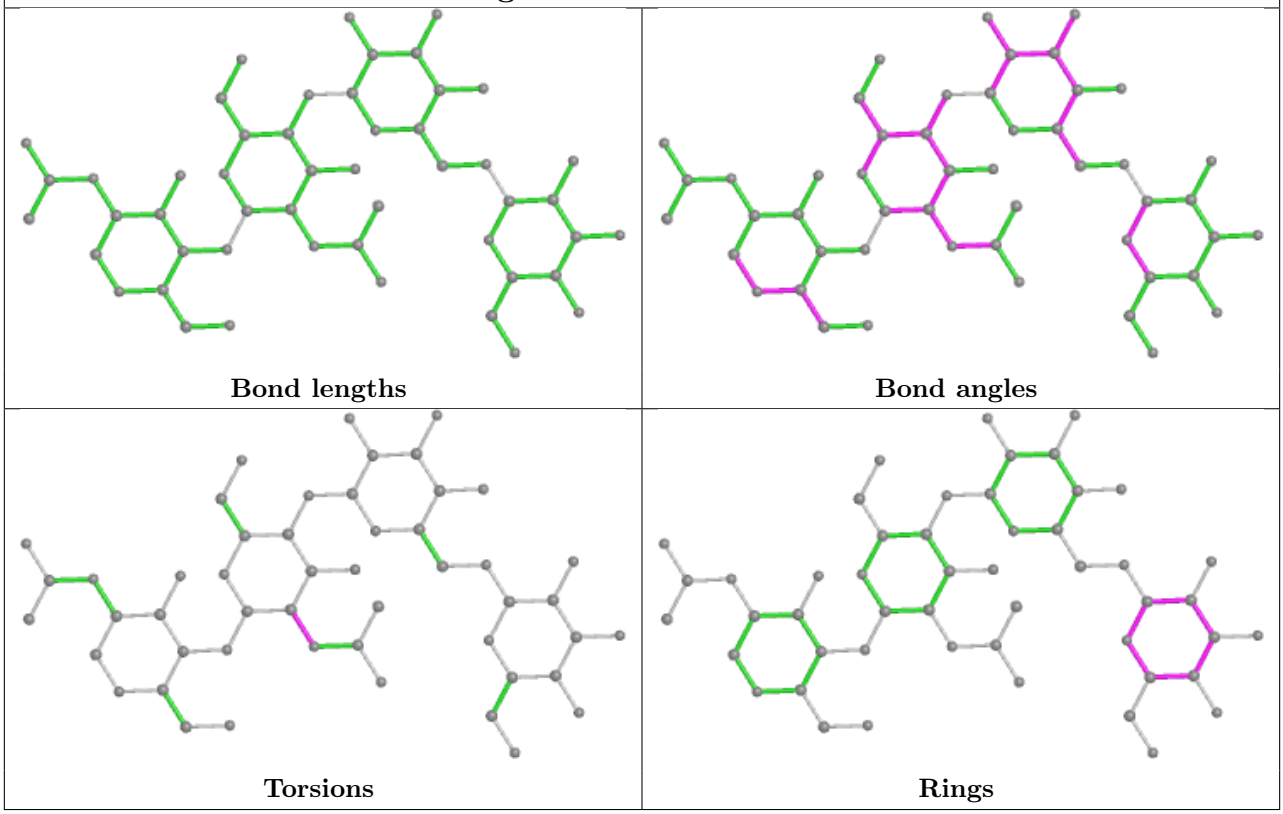


Torsions



Rings

Oligosaccharide Chain F



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	$\#\|Z\| > 2$
4	NAG	B	1	2	14,14,15	0.93	1 (7%)	17,19,21	1.58	3 (17%)
4	NAG	D	1	2	14,14,15	0.79	0	17,19,21	1.60	3 (17%)

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	1	2	-	1/6/23/26	0/1/1/1
4	NAG	D	1	2	-	2/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	1	NAG	C1-C2	2.74	1.56	1.52

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	1	NAG	C1-O5-C5	4.41	118.16	112.19
4	B	1	NAG	C3-C4-C5	-3.14	104.63	110.24
4	B	1	NAG	C6-C5-C4	2.94	119.90	113.00
4	B	1	NAG	O4-C4-C5	2.33	115.08	109.30
4	D	1	NAG	C4-C3-C2	-2.13	107.90	111.02

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	1	NAG	C4-C5-C6-O6
4	D	1	NAG	O5-C5-C6-O6
4	B	1	NAG	C3-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.