wwPDB X-ray Structure Validation Summary Report (i)

Oct 17, 2021 – 07:28 AM EDT

PDB ID	:	1MS8
Title	:	Triclinic form of Trypanosoma cruzi trans-sialidase, in complex with 3-deoxy-2,3-dehydro-N-acetylneuraminic acid (DANA)
Authors	:	Buschiazzo, A.; Amaya, M.F.; Cremona, M.L.; Frasch, A.C.; Alzari, P.M.
Deposited on	:	2002-09-19
Resolution	:	2.00 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 following versions of software and data (see references (1)) were used in the production of this report:

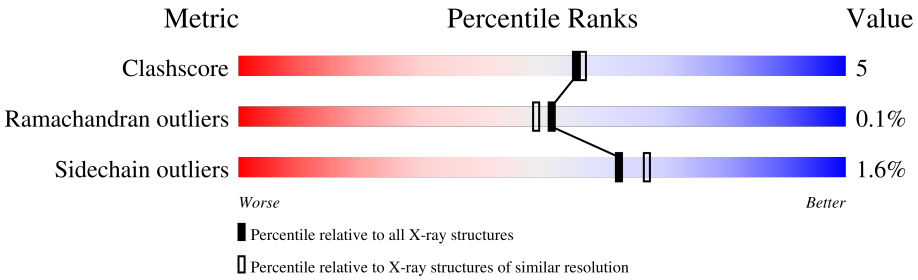
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.00 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	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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	648	85% 11%
1	B	648	85% 10%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	DAN	A	700	-	-	X	-
2	DAN	B	701	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10704 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 trans-sialidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	624	Total	C	N	O	S	0	11	0
			4931	3119	861	934	17			
1	B	623	Total	C	N	O	S	0	11	0
			4924	3114	858	935	17			

There are 50 discrepancies between the modelled and reference sequences:

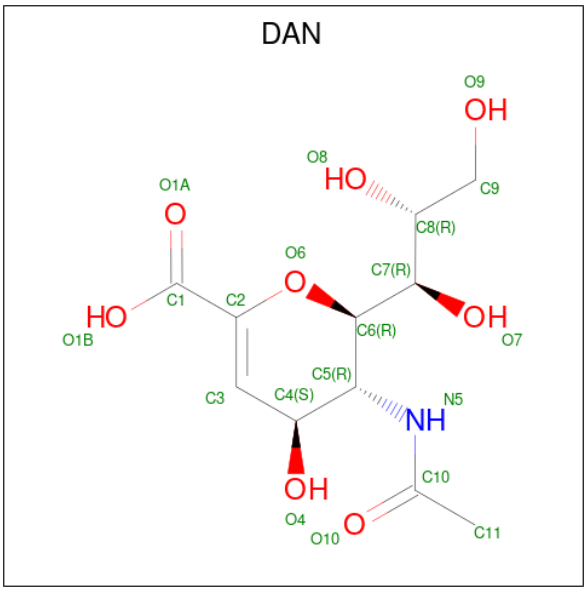
Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	expression tag	UNP Q26964
A	-12	GLY	-	expression tag	UNP Q26964
A	-11	GLY	-	expression tag	UNP Q26964
A	-10	SER	-	expression tag	UNP Q26964
A	-9	HIS	-	expression tag	UNP Q26964
A	-8	HIS	-	expression tag	UNP Q26964
A	-7	HIS	-	expression tag	UNP Q26964
A	-6	HIS	-	expression tag	UNP Q26964
A	-5	HIS	-	expression tag	UNP Q26964
A	-4	HIS	-	expression tag	UNP Q26964
A	-3	GLY	-	expression tag	UNP Q26964
A	-2	MET	-	expression tag	UNP Q26964
A	-1	ALA	-	expression tag	UNP Q26964
A	0	SER	-	expression tag	UNP Q26964
A	58	PHE	ASN	engineered mutation	UNP Q26964
A	262	THR	SER	SEE REMARK 999	UNP Q26964
A	476	HIS	ARG	SEE REMARK 999	UNP Q26964
A	484	LEU	VAL	SEE REMARK 999	UNP Q26964
A	495	LYS	SER	engineered mutation	UNP Q26964
A	496	GLY	VAL	engineered mutation	UNP Q26964
A	520	LYS	GLU	engineered mutation	UNP Q26964
A	558	VAL	GLU	SEE REMARK 999	UNP Q26964
A	593	GLY	ASP	engineered mutation	UNP Q26964
A	597	ASP	ILE	engineered mutation	UNP Q26964
A	599	ARG	HIS	engineered mutation	UNP Q26964

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-13	MET	-	expression tag	UNP Q26964
B	-12	GLY	-	expression tag	UNP Q26964
B	-11	GLY	-	expression tag	UNP Q26964
B	-10	SER	-	expression tag	UNP Q26964
B	-9	HIS	-	expression tag	UNP Q26964
B	-8	HIS	-	expression tag	UNP Q26964
B	-7	HIS	-	expression tag	UNP Q26964
B	-6	HIS	-	expression tag	UNP Q26964
B	-5	HIS	-	expression tag	UNP Q26964
B	-4	HIS	-	expression tag	UNP Q26964
B	-3	GLY	-	expression tag	UNP Q26964
B	-2	MET	-	expression tag	UNP Q26964
B	-1	ALA	-	expression tag	UNP Q26964
B	0	SER	-	expression tag	UNP Q26964
B	58	PHE	ASN	engineered mutation	UNP Q26964
B	262	THR	SER	SEE REMARK 999	UNP Q26964
B	476	HIS	ARG	SEE REMARK 999	UNP Q26964
B	484	LEU	VAL	SEE REMARK 999	UNP Q26964
B	495	LYS	SER	engineered mutation	UNP Q26964
B	496	GLY	VAL	engineered mutation	UNP Q26964
B	520	LYS	GLU	engineered mutation	UNP Q26964
B	558	VAL	GLU	SEE REMARK 999	UNP Q26964
B	593	GLY	ASP	engineered mutation	UNP Q26964
B	597	ASP	ILE	engineered mutation	UNP Q26964
B	599	ARG	HIS	engineered mutation	UNP Q26964

- Molecule 2 is 2-DEOXY-2,3-DEHYDRO-N-ACETYL-NEURAMINIC ACID (three-letter code: DAN) (formula: $C_{11}H_{17}NO_8$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C N O 20 11 1 8	0	0
2	B	1	Total C N O 20 11 1 8	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	377	Total O 377 377	0	0
3	B	432	Total O 432 432	0	0

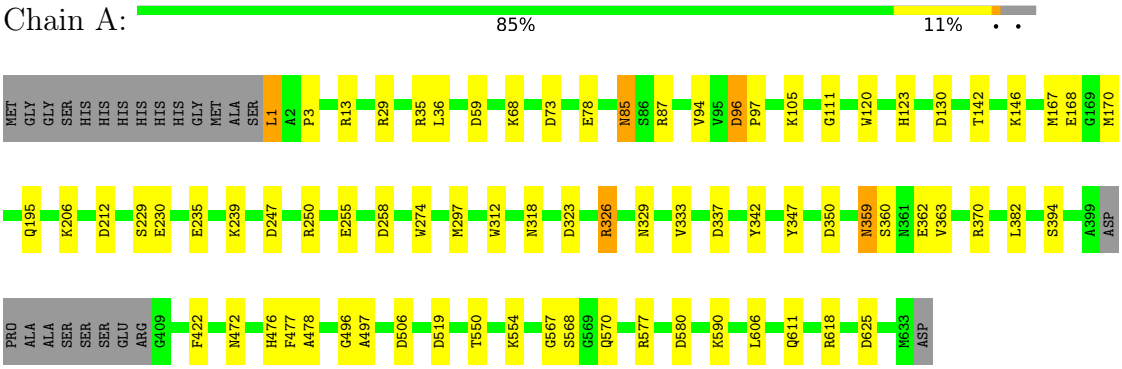
3 Residue-property plots (i)

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.

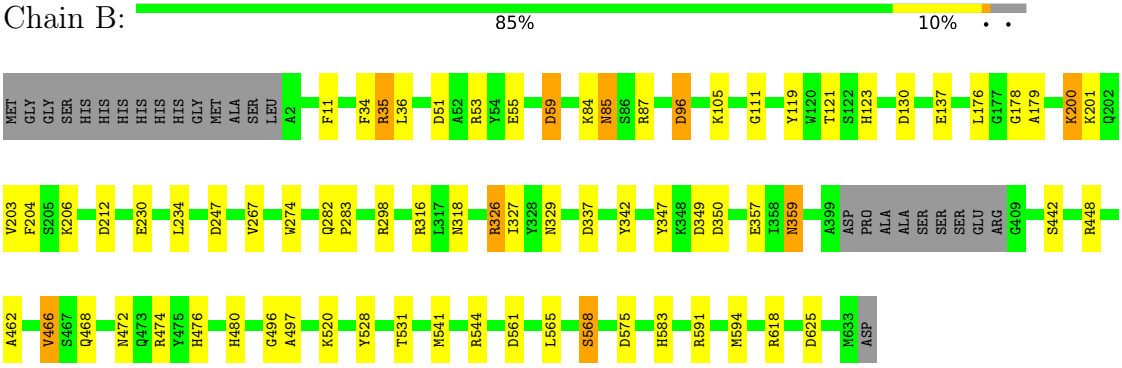
• Molecule 1: trans-sialidase

Chain A:



• Molecule 1: trans-sialidase

Chain B:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	51.50Å 74.18Å 87.57Å 86.10° 84.20° 88.20°	Depositor
Resolution (Å)	19.88 – 2.00	Depositor
% Data completeness (in resolution range)	86.0 (19.88-2.00)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.03	Depositor
Refinement program	REFMAC 5.0	Depositor
R, R_{free}	0.163 , 0.222	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	10704	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

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

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	1.08	7/5045 (0.1%)	1.02	20/6848 (0.3%)
1	B	1.08	1/5038 (0.0%)	1.04	18/6840 (0.3%)
All	All	1.08	8/10083 (0.1%)	1.03	38/13688 (0.3%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	235	GLU	CD-OE1	-6.16	1.18	1.25
1	A	394	SER	CB-OG	6.11	1.50	1.42
1	A	255	GLU	CD-OE1	6.00	1.32	1.25
1	B	200	LYS	CE-NZ	5.67	1.63	1.49
1	A	239	LYS	CD-CE	5.25	1.64	1.51

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	130	ASP	CB-CG-OD1	9.45	126.80	118.30
1	B	51	ASP	CB-CG-OD1	8.35	125.82	118.30
1	B	575	ASP	CB-CG-OD2	8.31	125.78	118.30
1	B	349	ASP	CB-CG-OD2	7.24	124.82	118.30
1	A	326	ARG	NE-CZ-NH2	-7.10	116.75	120.30

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	4931	0	4847	44	0
1	B	4924	0	4829	57	0
2	A	20	0	16	11	0
2	B	20	0	16	7	0
3	A	377	0	0	0	0
3	B	432	0	0	4	0
All	All	10704	0	9708	101	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 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:A:342[B]:TYR:OH	2:A:700:DAN:C1	1.73	1.36
1:B:342[B]:TYR:OH	2:B:701:DAN:C1	1.87	1.22
1:B:342[B]:TYR:HH	2:B:701:DAN:C1	1.50	1.21
1:A:342[B]:TYR:HH	2:A:700:DAN:C1	1.54	1.12
1:B:327:ILE:HB	1:B:474:ARG:HG2	1.35	1.05

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	631/648 (97%)	602 (95%)	29 (5%)	0	100	100
1	B	630/648 (97%)	603 (96%)	26 (4%)	1 (0%)	47	44
All	All	1261/1296 (97%)	1205 (96%)	55 (4%)	1 (0%)	51	49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	466	VAL

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	538/547 (98%)	529 (98%)	9 (2%)	60	65
1	B	537/547 (98%)	529 (98%)	8 (2%)	65	69
All	All	1075/1094 (98%)	1058 (98%)	17 (2%)	62	67

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	541	MET
1	B	618	ARG
1	A	590	LYS
1	A	618	ARG
1	B	84	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	604	ASN
1	B	603	ASN
1	B	85	ASN
1	B	472	ASN
1	A	611	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	$\#\|Z\| > 2$
2	DAN	B	701	-	17,20,20	3.49	4 (23%)	18,28,28	2.15	6 (33%)
2	DAN	A	700	-	17,20,20	4.05	3 (17%)	18,28,28	1.78	4 (22%)

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	DAN	B	701	-	-	1/14/34/34	0/1/1/1
2	DAN	A	700	-	-	6/14/34/34	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	700	DAN	C3-C2	15.71	1.51	1.32
2	B	701	DAN	C3-C2	13.49	1.48	1.32
2	A	700	DAN	O6-C2	4.16	1.44	1.37
2	B	701	DAN	O10-C10	2.39	1.28	1.23
2	B	701	DAN	C9-C8	2.26	1.58	1.52

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	701	DAN	O8-C8-C9	4.49	119.68	109.14
2	B	701	DAN	C8-C7-C6	-4.25	104.97	113.03
2	A	700	DAN	C6-C5-N5	-4.08	104.13	110.91
2	B	701	DAN	C4-C3-C2	-3.81	115.16	121.60
2	A	700	DAN	C4-C3-C2	-3.46	115.76	121.60

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	700	DAN	C5-C6-C7-C8
2	A	700	DAN	C5-C6-C7-O7
2	A	700	DAN	O6-C6-C7-O7
2	A	700	DAN	O8-C8-C9-O9
2	A	700	DAN	O6-C6-C7-C8

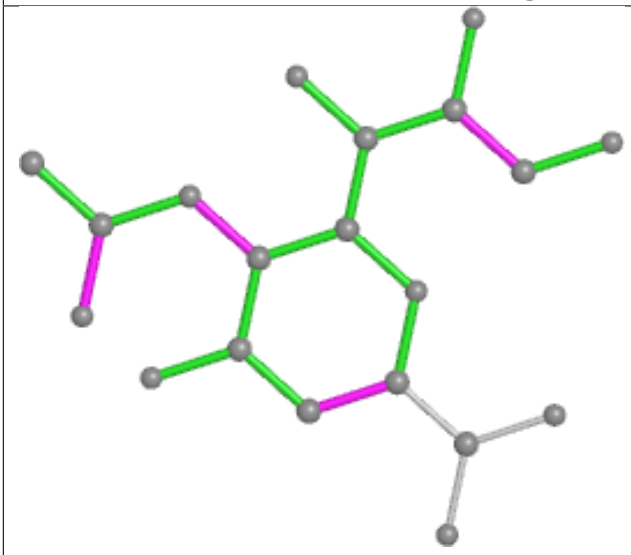
There are no ring outliers.

2 monomers are involved in 18 short contacts:

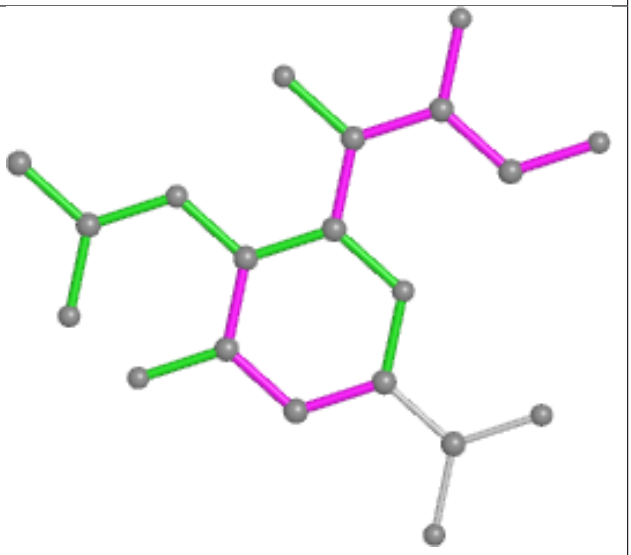
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	701	DAN	7	0
2	A	700	DAN	11	0

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.

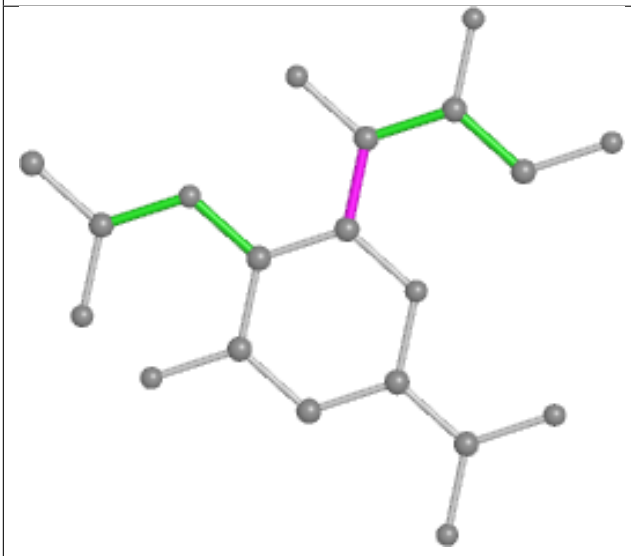
Ligand DAN B 701



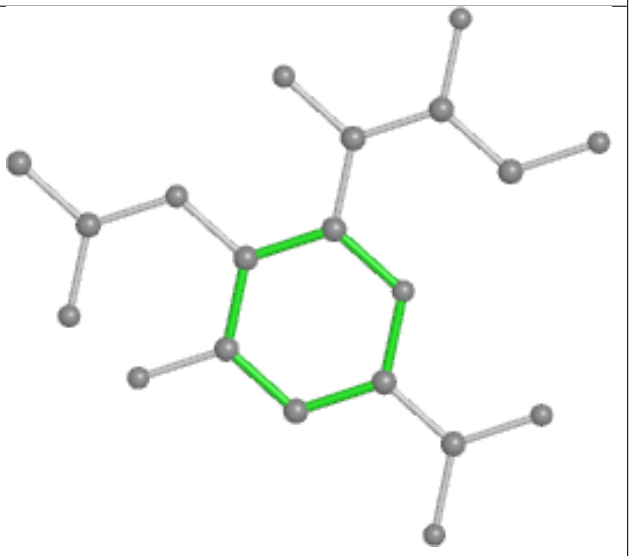
Bond lengths



Bond angles

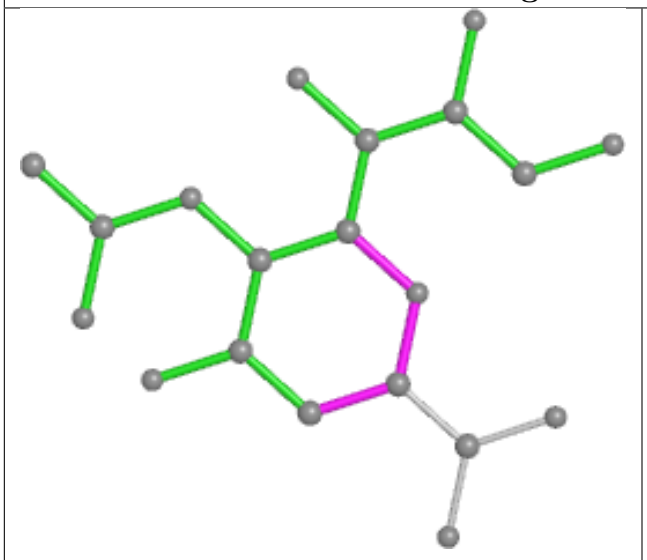


Torsions

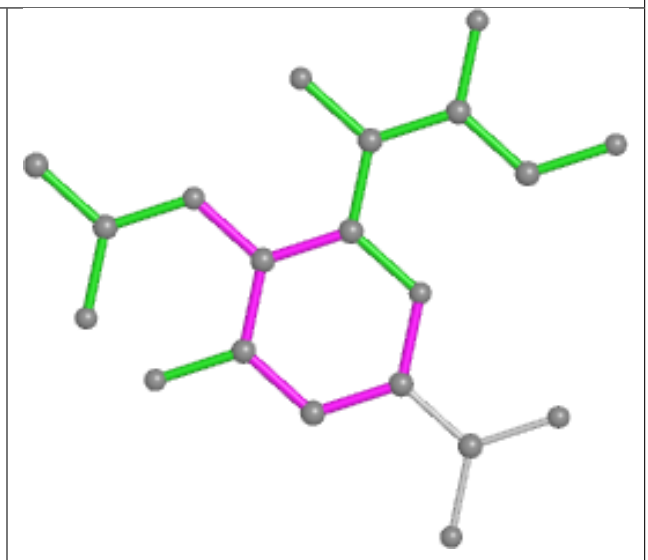


Rings

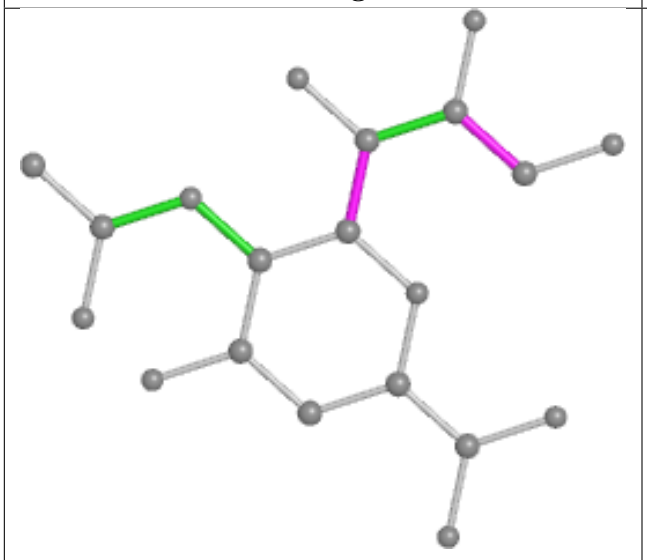
Ligand DAN A 700



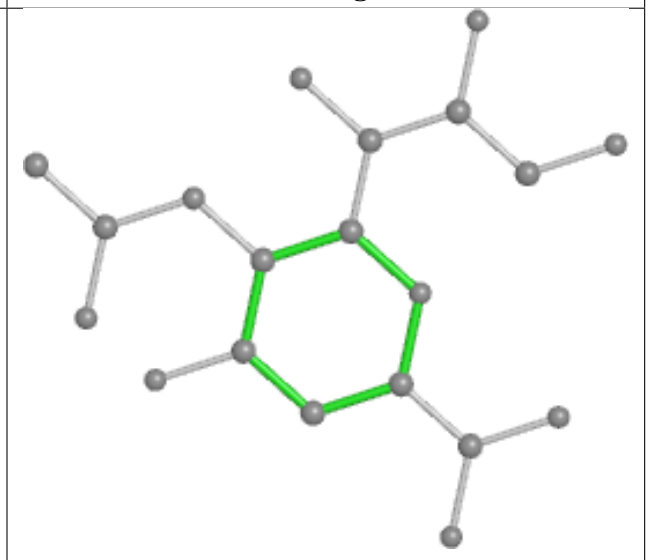
Bond lengths



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.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.