wwPDB NMR Structure Validation Summary Report (i)

Feb 26, 2022 – 02:31 PM EST

PDB ID	:	203D
Title	:	THE SOLUTION STRUCTURES OF PSORALEN MONOADDUCTED AND CROSSLINKED DNA OLIGOMERS BY NMR SPECTROSCOPY AND RESTRAINED MOLECULAR DYNAMICS
Authors	:	Spielmann, H.P.; Dwyer, T.J.; Hearst, J.E.; Wemmer, D.E.
Deposited on	:	1995-04-06

This is a wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

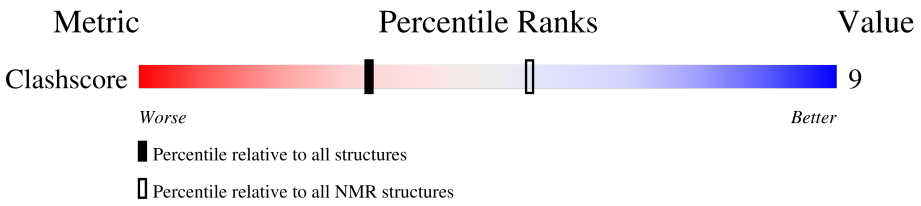
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

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)	NMR archive (#Entries)
Clashscore	158937	12864

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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	8	38% 12% 38% 12%
1	B	8	25% 12% 50% 12%

2 Ensemble composition and analysis (i)

This entry contains 40 models. This entry does not contain polypeptide chains, therefore identification of well-defined residues and clustering analysis are not possible. All residues are included in the validation scores.

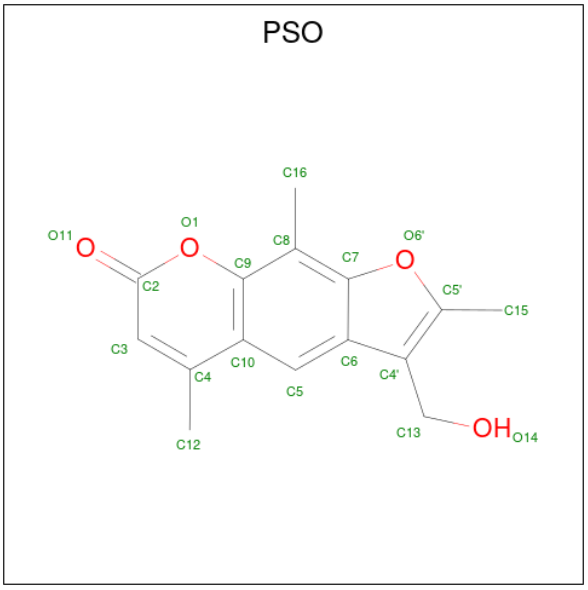
3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 537 atoms, of which 196 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*GP*CP*GP*TP*AP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	8	252	77	91	31	46	7	0
1	B	8	252	77	91	31	46	7	0

- Molecule 2 is 4'-HYDROXYMETHYL-4,5',8-TRIMETHYLPSORALEN (three-letter code: PSO) (formula: $C_{15}H_{14}O_4$).



Mol	Chain	Residues	Atoms			
			Total	C	H	O
2	A	1	33	15	14	4

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

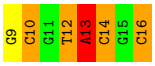
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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 outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: DNA (5'-D(*GP*CP*GP*TP*AP*CP*GP*C)-3')

Chain A: 38% 12% 38% 12%

• Molecule 1: DNA (5'-D(*GP*CP*GP*TP*AP*CP*GP*C)-3')

Chain B: 25% 12% 50% 12%



4.2 Residue scores for the first model from the NMR ensemble

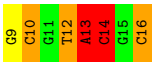
No representative models were identified. Colouring as in section 4.1 above.

• Molecule 1: DNA (5'-D(*GP*CP*GP*TP*AP*CP*GP*C)-3')

Chain A: 25% 25% 25% 25%

• Molecule 1: DNA (5'-D(*GP*CP*GP*TP*AP*CP*GP*C)-3')

Chain B: 25% 12% 38% 25%



5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 40 were deposited, based on the following criterion: *?*.

The authors did not provide any information on software used for structure solution, optimization or refinement.

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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

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 (average) 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	2.73±0.03	4±0/180 (2.1± 0.2%)	2.81±0.05	13±2/276 (4.9± 0.8%)
1	B	1.34±0.01	1±0/180 (0.6± 0.0%)	2.59±0.03	15±1/276 (5.6± 0.3%)
All	All	2.15	189/14400 (1.3%)	2.70	1154/22080 (5.2%)

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	Planarity
1	A	0.0±0.0	4.1±0.6
1	B	0.0±0.0	5.0±0.2
All	All	0	364

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	4	DT	C5-C6	27.84	1.53	1.34	19	40
1	A	4	DT	N1-C6	16.78	1.50	1.38	11	40
1	A	4	DT	C4-C5	9.27	1.53	1.45	5	40
1	B	12	DT	N1-C6	-6.15	1.33	1.38	32	40
1	A	4	DT	C5-C7	5.58	1.53	1.50	33	29

5 of 40 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	4	DT	C6-C5-C7	-24.63	108.12	122.90	38	40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	12	DT	C6-C5-C7	-16.41	113.06	122.90	11	40
1	A	6	DC	O4'-C1'-N1	15.05	118.53	108.00	3	32
1	A	5	DA	O4'-C1'-N9	12.89	117.02	108.00	15	40
1	B	16	DC	O4'-C1'-N1	12.04	116.42	108.00	33	40

There are no chirality outliers.

5 of 12 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	3	DG	Sidechain	40
1	A	4	DT	Sidechain	40
1	A	8	DC	Sidechain	40
1	B	9	DG	Sidechain	40
1	B	10	DC	Sidechain	40

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	161	91	91	2±0
1	B	161	91	91	3±0
2	A	19	14	14	2±1
All	All	13640	7840	7840	198

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.

5 of 8 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:12:DT:H2"	1:B:13:DA:C5'	0.74	2.13	3	40
1:B:12:DT:H2"	1:B:13:DA:C4'	0.54	2.33	12	40
1:A:5:DA:C4	2:A:9:PSO:H163	0.52	2.40	20	38
1:A:4:DT:H2'	2:A:9:PSO:C5'	0.49	2.37	22	35
2:A:9:PSO:H123	1:B:12:DT:H72	0.48	1.86	3	19

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

There are no protein molecules in this entry.

6.3.2 Protein sidechains (i)

There are no protein molecules in this entry.

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	PSO	A	9	-	13,21,21	2.01±0.10	4±0 (27±3%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard

deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	PSO	A	9	-	18,32,32	2.26±0.08	5±0 (27±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
2	PSO	A	9	-	-	0±0,0,2,2	0±0,3,3,3

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	9	PSO	C15-C5'	4.68	1.54	1.48	8	21
2	A	9	PSO	O1-C9	4.09	1.29	1.36	39	40
2	A	9	PSO	C6-C7	3.76	1.35	1.43	22	40
2	A	9	PSO	C13-C4'	3.23	1.55	1.51	39	40

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	9	PSO	C15-C5'-C4'	5.74	107.05	126.51	11	40
2	A	9	PSO	C5-C6-C4'	4.95	125.63	134.17	12	40
2	A	9	PSO	C5-C10-C9	4.09	121.00	116.39	34	40
2	A	9	PSO	O1-C2-C3	3.39	123.59	119.27	15	40
2	A	9	PSO	O1-C9-C10	3.19	124.29	121.20	11	40

There are no chirality outliers.

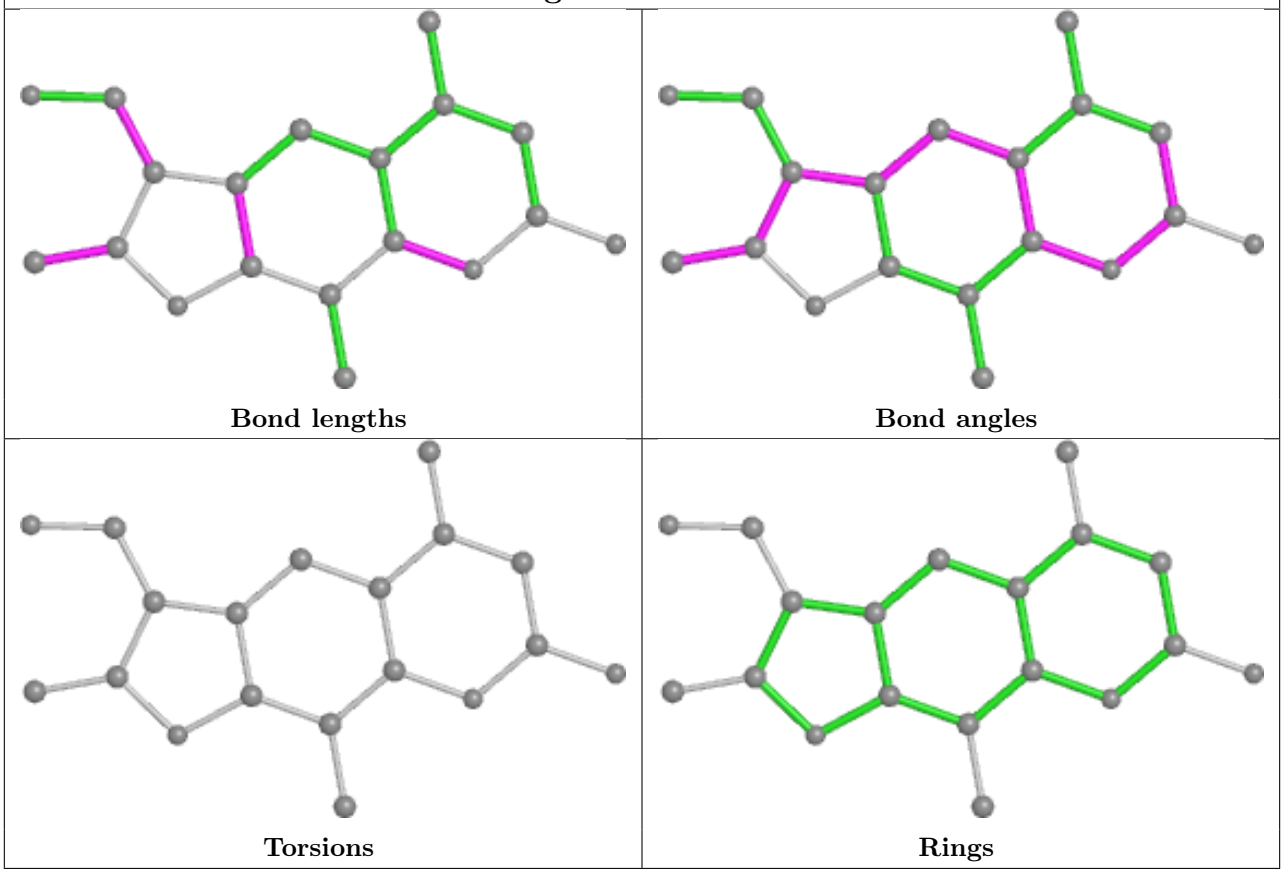
There are no torsion outliers.

There are no ring outliers.

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 PSO A 9



6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided