wwPDB NMR Structure Validation Summary Report (i)

Feb 23, 2022 – 10:59 AM EST

PDB ID	:	1XCZ
Title	:	Structure of DNA containing the beta-anomer of a carbocyclic abasic site
Authors	:	de los Santos, C.; El-khateeb, M.; Rege, P.; Tian, K.; Johnson, F.
Deposited on	:	2004-09-03

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 (1)) were used in the production of this report:

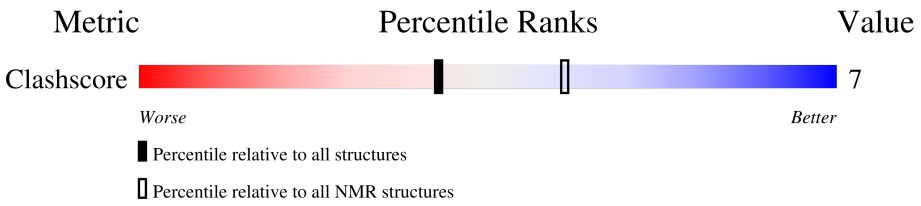
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	11	64% 36%
2	B	11	100%

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

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
1	A	DXD	6	9	-

2 Ensemble composition and analysis (i)

This entry contains 9 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 686 atoms, of which 249 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	11	335	102	124	36	63	10	0

- Molecule 2 is a DNA chain called 5'-D(*GP*CP*AP*TP*GP*AP*GP*TP*AP*CP*G)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	11	351	108	125	45	63	10	0

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: 5'-D(*CP*GP*TP*AP*CP*(DXD)P*CP*AP*TP*GP*C)-3'

Chain A: 64% 36%

C1 G2 T3 A4 C5 N6 C7 A8 T9 G10 C11

• Molecule 2: 5'-D(*GP*CP*AP*TP*GP*AP*GP*TP*AP*CP*G)-3'

Chain B: 100%

G12 C13 A14 T15 G16 A17 G18 T19 A20 C21 G22

4.2 Residue scores for the representative (author defined) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

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

Chain A: 27% 73%

C1 G2 T3 A4 C5 N6 C7 A8 T9 G10 C11

• Molecule 2: 5'-D(*GP*CP*AP*TP*GP*AP*GP*TP*AP*CP*G)-3'

Chain B: 100%

G12 C13 A14 T15 G16 A17 G18 T19 A20 C21 G22

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 20 calculated structures, 9 were deposited, based on the following criterion: *converging structures (<1.5A) and the least restraints violations*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR	refinement	3.85

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: DXD

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	2.53±0.01	11±0/221 (5.0± 0.0%)	2.22±0.02	16±1/336 (4.8± 0.3%)
2	B	2.56±0.00	15±0/254 (5.9± 0.0%)	2.34±0.01	15±1/391 (3.8± 0.3%)
All	All	2.55	234/4275 (5.5%)	2.29	280/6543 (4.3%)

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	1.0±0.0	0.0±0.0
All	All	9	0

5 of 26 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	B	12	DG	N9-C8	12.60	1.46	1.37	2	9
1	A	8	DA	C6-N6	12.22	1.43	1.33	4	9
2	B	14	DA	C6-N6	12.10	1.43	1.33	6	9
2	B	17	DA	C6-N6	12.09	1.43	1.33	4	9
1	A	4	DA	C6-N6	12.07	1.43	1.33	4	9

5 of 37 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(°)	Ideal(°)	Models	
								Worst	Total
2	B	12	DG	O4'-C1'-N9	14.54	118.17	108.00	3	9
2	B	17	DA	O4'-C1'-N9	14.19	117.93	108.00	9	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	17	DA	O4'-C4'-C3'	-11.47	99.12	106.00	3	9
1	A	5	DC	O4'-C1'-N1	10.96	115.67	108.00	1	8
2	B	12	DG	C8-N9-C1'	9.68	139.58	127.00	6	9

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	6	DXD	C1'	9

There are no planarity outliers.

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	211	124	124	5±1
2	B	226	125	125	0±1
All	All	3933	2241	2241	44

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 7.

5 of 12 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:6:DXD:H2'1	1:A:6:DXD:O5'	0.69	1.88	3	9
1:A:4:DA:C2	1:A:5:DC:C2	0.47	3.03	9	1
1:A:4:DA:C2	2:B:20:DA:C2	0.46	3.03	9	3
1:A:9:DT:C2	1:A:10:DG:C8	0.45	3.04	6	9
1:A:6:DXD:O5'	1:A:6:DXD:C2'	0.45	2.65	8	6

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)

1 non-standard protein/DNA/RNA residue 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
1	DXD	A	6	1	8,12,13	4.12±0.01	2±0 (25±0%)

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
1	DXD	A	6	1	10,16,19	2.34±0.03	5±0 (50±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
1	DXD	A	6	1	1±0,1,4,4	0±0,3,17,18	0±0,1,1,1

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	6	DXD	C6'-C1'	11.11	1.73	1.52	8	9
1	A	6	DXD	C2'-C3'	3.37	1.61	1.52	2	9

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
1	A	6	DXD	C5'-C4'-C3'	3.84	107.17	114.66	8	9
1	A	6	DXD	O6'-C1'-C2'	3.73	119.26	110.03	7	9
1	A	6	DXD	C2'-C1'-C6'	3.43	99.36	103.87	1	9
1	A	6	DXD	C3'-C2'-C1'	2.89	96.97	103.81	6	9
1	A	6	DXD	C2'-C3'-C4'	2.42	98.36	104.59	1	9

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	6	DXD	C1'	9

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

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