WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

Mar 7, 2022 – 04:53 AM EST

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
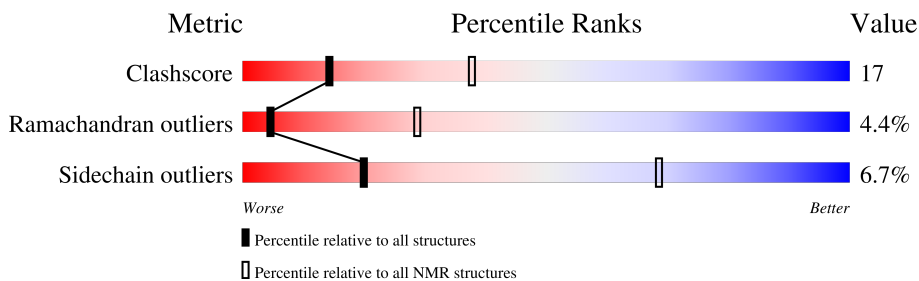with specific help available everywhere you see the ⓘ symbol.

1   Overall quality at a glance ⓘ

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	Percentile Ranks	Value
Clashscore		17
Ramachandran outliers		4.4%
Sidechain outliers		6.7%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	B	16	38% 56% 6%
2	C	16	31% 44% 25%
3	A	66	53% 32% 15%

2 Ensemble composition and analysis ⓘ

This entry contains 34 models. Model 22 is the overall representative, medoid model (most similar to other models).

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	Residue range (total)	Backbone RMSD (Å)	Medoid model
1	A:3-A:58 (56)	0.30	22

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 6 clusters and 3 single-model clusters were found.

Cluster number	Models
1	3, 7, 9, 13, 14, 15, 16, 20, 21, 22, 23, 24, 25, 28
2	1, 2, 4, 10, 11, 26, 29, 30
3	18, 19, 31
4	12, 32
5	8, 34
6	5, 27
Single-model clusters	6; 17; 33

3 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2088 atoms, of which 913 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*GP*TP*TP*GP*CP*AP*GP*AP*TP*AP*AP*AP*CP*AP*TP*T)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	16	510	158	183	61	93	15	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	16	507	157	184	56	95	15	0

- Molecule 3 is a protein called ERYTHROID TRANSCRIPTION FACTOR GATA-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	66	1070	316	546	110	92	6	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	1	1	1

4 Residue-property plots ⓘ

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*TP*TP*GP*CP*AP*GP*AP*TP*AP*AP*AP*CP*AP*TP*T)-3')
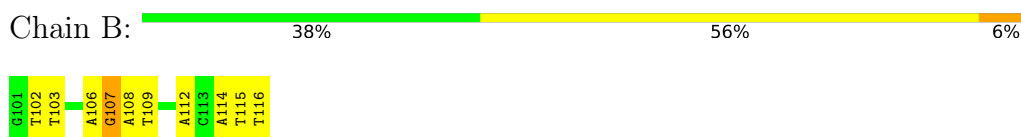
Chain B:

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

Chain C:

• Molecule 3: ERYTHROID TRANSCRIPTION FACTOR GATA-1

Chain A:

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

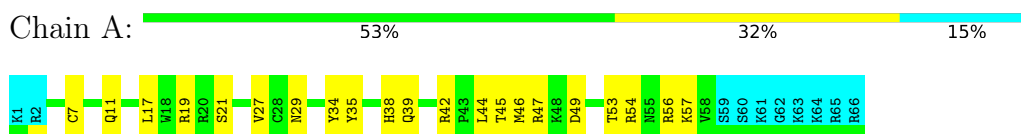The representative model is number 22. Colouring as in section 4.1 above.

• Molecule 1: DNA (5'-D(*GP*TP*TP*GP*CP*AP*GP*AP*TP*AP*AP*AP*CP*AP*TP*T)-3')
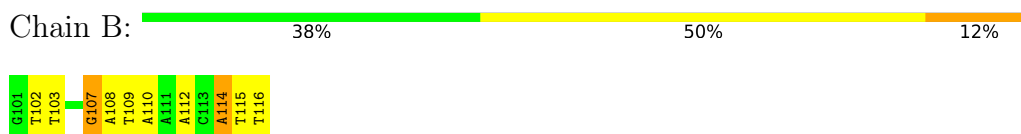
Chain B:

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

Chain C:

44%   31%   25%

A117 A118 T119 G120 T121 T122 T123 A124 T125 C126 T127 C132

• Molecule 3: ERYTHROID TRANSCRIPTION FACTOR GATA-1

Chain A:

47%   36%   • 15%

K1 R2 T5 V6 C7 S8 N9 C10 Q11 T12 S13 T14 T15 T16 L17 W18 R19 D25 P26 V27 C28 N29 Y34 Y35 H38 Q39 R42 P43 L44 T45 M46 D49 G50 I51 Q52 T53 K57 V58 S59 S60 K61 G62 K63 K64 R65 R66

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing.*

Of the 34 calculated structures, 34 were deposited, based on the following criterion: ?.

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

Software name	Classification	Version
X-PLOR	refinement	3.1
X-PLOR MODIFIED	structure solution	MODIFIED

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: 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 (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	B	1.38±0.02	5±0/367 (1.4± 0.0%)	1.90±0.01	5±2/565 (0.8± 0.3%)
2	C	1.39±0.03	5±1/361 (1.4± 0.2%)	1.94±0.01	10±2/555 (1.8± 0.4%)
3	A	1.06±0.01	0±0/446 (0.0± 0.0%)	0.96±0.01	0±0/604 (0.0± 0.0%)
All	All	1.27	347/39916 (0.9%)	1.65	495/58616 (0.8%)

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	B	0.0±0.0	0.0±0.2
All	All	0	1

5 of 11 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	B	102	DT	C5-C7	7.83	1.54	1.50	30	34
2	C	123	DT	C5-C7	7.22	1.54	1.50	30	34
2	C	121	DT	C5-C7	6.64	1.54	1.50	4	25
1	B	109	DT	C5-C7	6.63	1.54	1.50	15	34
2	C	122	DT	C5-C7	6.60	1.54	1.50	9	34

5 of 41 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	C	120	DG	O4'-C1'-N9	7.54	113.28	108.00	32	23

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models Worst	Models Total
2	C	125	DT	O4'-C1'-N1	7.31	113.12	108.00	25	33
1	B	107	DG	O4'-C1'-N9	7.17	113.02	108.00	9	34
2	C	121	DT	O4'-C1'-N1	6.94	112.86	108.00	26	32
1	B	112	DA	O4'-C1'-N9	6.60	112.62	108.00	32	33

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	B	108	DA	Sidechain	1

6.2 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 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	B	327	183	183	9±3
2	C	323	184	184	11±3
3	A	438	440	440	23±5
All	All	37026	27438	27438	1114

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 17.

5 of 242 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
3:A:38:HIS:O	3:A:40:VAL:N	0.97	1.96	33	1
3:A:17:LEU:HD23	3:A:19:ARG:NH1	0.86	1.85	11	30
2:C:117:DA:HO5'	2:C:117:DA:H8	0.85	1.12	25	23
3:A:10:CYS:O	3:A:12:THR:N	0.83	2.11	32	3
1:B:106:DA:N7	3:A:17:LEU:HD21	0.80	1.91	12	27

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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 NMR 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	
3	A	56/66 (85%)	50±1 (89±3%)	4±2 (6±3%)	2±1 (4±2%)	4	29
All	All	1904/2244 (85%)	1699 (89%)	121 (6%)	84 (4%)	4	29

5 of 9 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	49	ASP	26
3	A	57	LYS	16
3	A	44	LEU	10
3	A	39	GLN	10
3	A	45	THR	7

6.3.2 Protein sidechains ⓘ

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 NMR 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	
3	A	50/59 (85%)	47±1 (93±3%)	3±1 (7±3%)	20	68
All	All	1700/2006 (85%)	1586 (93%)	114 (7%)	20	68

5 of 18 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	53	THR	33
3	A	21	SER	24
3	A	42	ARG	12
3	A	47	ARG	9
3	A	54	ARG	7

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

No chemical shift data were provided