



wwPDB NMR Structure Validation Summary Report ⓘ

Mar 3, 2022 – 07:35 AM EST

PDB ID : 2HDC
Title : STRUCTURE OF TRANSCRIPTION FACTOR GENESIS/DNA COM-
PLEX
Authors : Jin, C.; Marsden, I.; Chen, X.; Liao, X.
Deposited on : 1999-05-05

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.

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

MolProbity : 4.02b-467
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

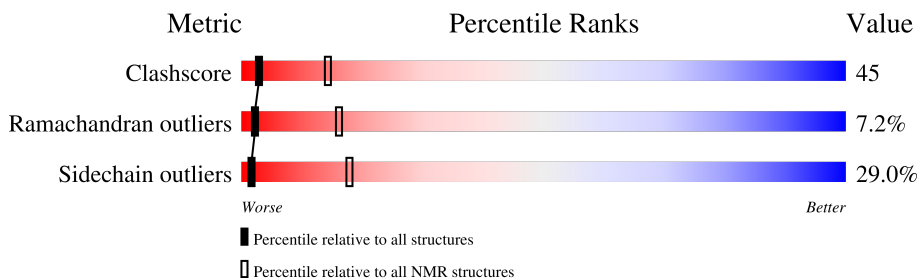
1 Overall quality at a glance

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
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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	B	17	29% (Yellow) 71% (Orange)
2	C	17	18% (Yellow) 82% (Orange)
3	A	97	19% (Green) 58% (Yellow) 7% (Orange) 16% (Cyan)

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:7-A:87 (81)	0.52	6

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20
2	1, 13

3 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2581 atoms, of which 1080 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	17	493	167	145	67	97	17	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	17	484	169	135	56	107	17	0

- Molecule 3 is a protein called PROTEIN (TRANSCRIPTION FACTOR).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	A	97	1604	517	800	143	140	4	0

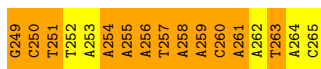
4 Residue-property plots [i](#)

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.

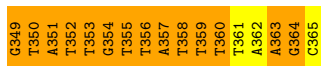
- Molecule 1: DNA (5'-D(P*GP*CP*TP*TP*AP*AP*AP*AP*TP*AP*AP*CP*AP*AP*TP*A P*C)-3')

Chain B:  29% 71%



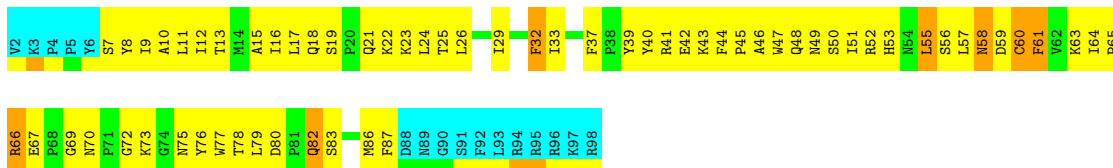
- Molecule 2: DNA (5'-D(P*GP*TP*AP*TP*TP*GP*TP*TP*AP*TP*TP*TP*TP*AP*AP*G P*C)-3')

Chain C:  18% 82%



- Molecule 3: PROTEIN (TRANSCRIPTION FACTOR)

Chain A:  19% 58% 7% 16%



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

The representative model is number 6. Colouring as in section 4.1 above.

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

Chain B:  35% 65%


G249
C250
T251
T252
A253
A254
A255
A256
T257
A258
A259
C260
A261
A262
T263
A264
C265

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

Chain C:  35% 65%

G349
T350
A351
T352
G353
G354
T355
T356
A357
T358
T359
T360
T361
A362
A363
G364
C365

- Molecule 3: PROTEIN (TRANSCRIPTION FACTOR)

Chain A:  22% 46% 13% 16%

V2
K3
P4
P5
Y6
S7
Y8
I9
A10
L11
I12
T13
M14
A15
Q18
S19
P20
Q21
K22
K23
L24
T25
L26
I29
F32
I33
S34
N35
R36
F37
P38
Y39
Y40
R41
E42
K43
F44
P45
A46
W47
Q48
M49
S50
I51
R52
H53
M54
L55
S56
L57
N58
D59
C60
F61
V62
K63
I64

P65
R66
E67
P68
G69
M70
P71
G72
M75
Y76
T77
T78
L79
D80
P81
Q82
D85
D88
M89
G90
S91
F92
L93
R94
R95
R96
K97
R98

5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry*.

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
DYANA	refinement	1.5
DYANA	structure solution	

No chemical shift data was provided.

6 Model quality i

6.1 Standard geometry i

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	4.98±2.25	13±1/391 (3.4± 0.3%)	5.51±0.64	49±1/600 (8.1± 0.2%)
2	C	3.59±0.38	12±2/389 (3.2± 0.4%)	4.76±0.32	62±2/599 (10.4± 0.3%)
3	A	0.72±0.00	0±0/685 (0.0± 0.0%)	0.92±0.00	0±0/930 (0.0± 0.0%)
All	All	3.42	513/29300 (1.8%)	3.93	2223/42580 (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	B	8.1±1.1	0.0±0.0
2	C	4.2±1.0	0.0±0.0
All	All	246	0

5 of 34 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	249	DG	C4'-O4'	277.59	4.22	1.45	20	15
2	C	359	DT	C4'-O4'	60.77	2.05	1.45	18	19
2	C	363	DA	C4'-O4'	44.67	1.89	1.45	9	7
1	B	260	DC	C4'-O4'	42.10	1.87	1.45	3	20
1	B	254	DA	C4'-O4'	-37.77	1.07	1.45	2	20

5 of 132 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
1	B	249	DG	O4'-C4'-C3'	-121.26	33.25	106.00	20	12
1	B	249	DG	C1'-O4'-C4'	-79.09	31.00	110.10	20	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
2	C	359	DT	O4'-C4'-C3'	-56.61	72.04	106.00	18	14
2	C	363	DA	O4'-C4'-C3'	-48.22	77.07	106.00	9	19
1	B	260	DC	O4'-C4'-C3'	-45.03	78.98	106.00	11	20

5 of 28 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	B	255	DA	C4'	20
1	B	260	DC	C4'	20
1	B	263	DT	C4'	18
2	C	359	DT	C4'	18
2	C	353	DT	C4'	16

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	B	348	145	192	26±6
2	C	349	135	197	22±5
3	A	664	652	652	72±13
All	All	27220	18640	20820	2141

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 45.

5 of 846 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:261:DA:O4'	1:B:261:DA:C4'	1.39	1.64	13	8
2:C:361:DT:C4'	2:C:361:DT:O4'	1.38	1.63	9	1
1:B:256:DA:O4'	1:B:256:DA:C4'	1.37	1.64	11	12
2:C:364:DG:O4'	2:C:364:DG:C4'	1.36	1.64	15	8
1:B:251:DT:O4'	1:B:251:DT:C4'	1.35	1.65	3	4

6.3 Torsion angles [i](#)

6.3.1 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 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	81/97 (84%)	62±2 (77±3%)	13±2 (16±2%)	6±1 (7±2%)	2	16
All	All	1620/1940 (84%)	1244 (77%)	260 (16%)	116 (7%)	2	16

5 of 16 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	58	ASN	20
3	A	82	GLN	17
3	A	46	ALA	14
3	A	66	ARG	14
3	A	60	CYS	13

6.3.2 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 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	74/89 (83%)	53±4 (71±5%)	21±4 (29±5%)	2	18
All	All	1480/1780 (83%)	1051 (71%)	429 (29%)	2	18

5 of 61 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	32	PHE	19
3	A	63	LYS	18
3	A	61	PHE	16
3	A	55	LEU	14
3	A	21	GLN	13

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](#)

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

No chemical shift data were provided