Full wwPDB NMR Structure Validation Report (i)

May 31, 2020 – 10:23 pm BST

PDB ID	:	6GVU
Title	:	NMR structure of the DNA-bound helix bundle domain from the functional pRN1 primase
Authors	:	Boudet, J.; Lipps, G.; Allain, F.H.-T.
Deposited on	:	2018-06-21

This is a Full wwPDB NMR Structure Validation 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:

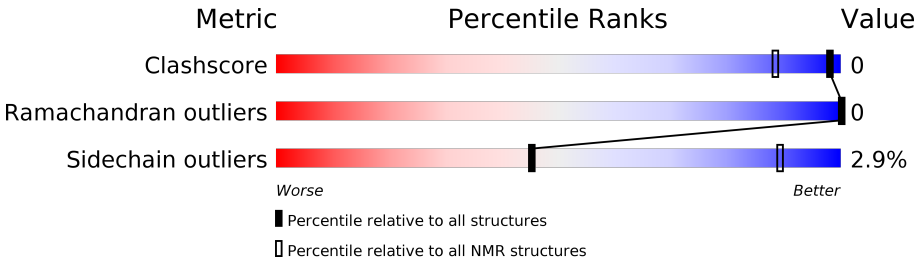
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 is 63%.

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	9	56% 44%
2	B	115	95% . .

2 Ensemble composition and analysis

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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	B:257-B:342, B:347-B:370 (110)	0.39	8

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2224 atoms, of which 1094 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	9	283	87	104	30	54	8	0

- Molecule 2 is a protein called functional pRN1 primase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	115	1941	611	990	160	177	3	0

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

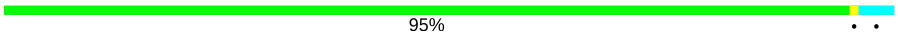
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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.

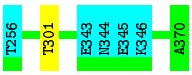
- Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:

- Molecule 2: functional pRN1 primase

Chain B:





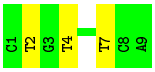
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

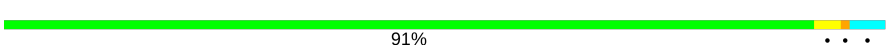
- Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

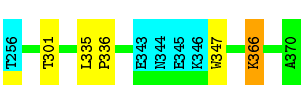
Chain A:



- Molecule 2: functional pRN1 primase

Chain B:





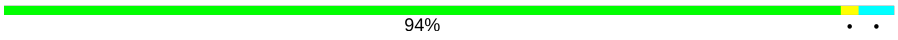
4.2.2 Score per residue for model 2

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



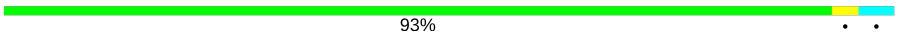
4.2.3 Score per residue for model 3

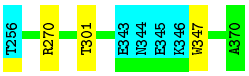
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:





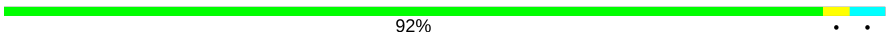
4.2.4 Score per residue for model 4

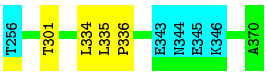
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:





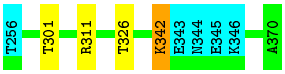
4.2.5 Score per residue for model 5

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

Chain A:

• Molecule 2: functional pRN1 primase

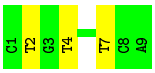
Chain B:



4.2.6 Score per residue for model 6

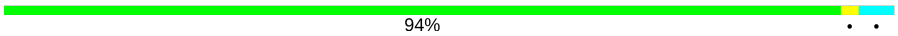
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:



• Molecule 2: functional pRN1 primase

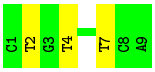
Chain B:



4.2.7 Score per residue for model 7

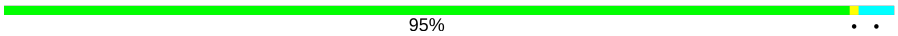
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

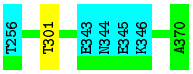
Chain A:



• Molecule 2: functional pRN1 primase

Chain B:





4.2.8 Score per residue for model 8 (medoid)

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

Chain A: 

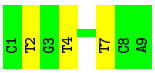
• Molecule 2: functional pRN1 primase

Chain B: 

4.2.9 Score per residue for model 9

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

Chain A: 



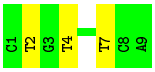
• Molecule 2: functional pRN1 primase

Chain B: 

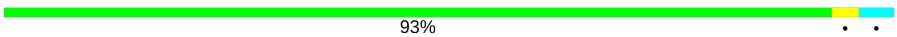
4.2.10 Score per residue for model 10

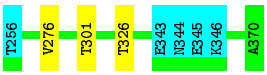
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A: 



• Molecule 2: functional pRN1 primase

Chain B: 



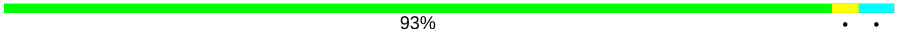
4.2.11 Score per residue for model 11

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



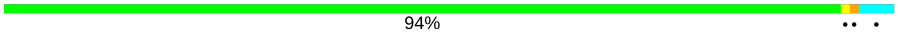
4.2.12 Score per residue for model 12

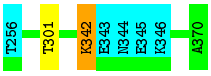
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:





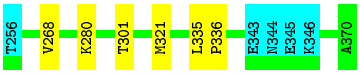
4.2.13 Score per residue for model 13

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



4.2.14 Score per residue for model 14

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:

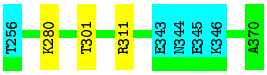
4.2.15 Score per residue for model 15

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



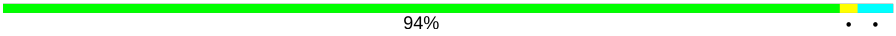
4.2.16 Score per residue for model 16

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



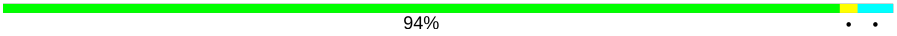
4.2.17 Score per residue for model 17

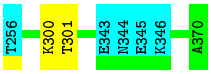
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:

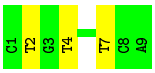




4.2.18 Score per residue for model 18

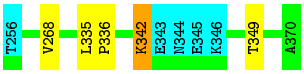
• Molecule 1: DNA (5'-D(*CP*TP*GP*TP*GP*CP*TP*CP*A)-3')

Chain A:



• Molecule 2: functional pRN1 primase

Chain B:



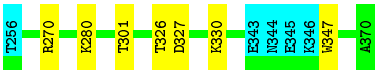
4.2.19 Score per residue for model 19

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

Chain A:

• Molecule 2: functional pRN1 primase

Chain B:



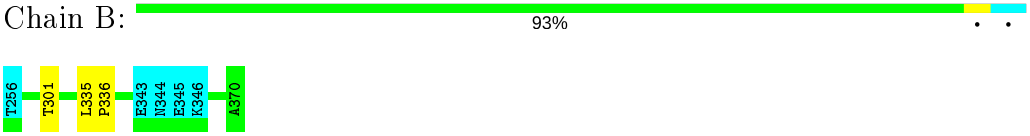
4.2.20 Score per residue for model 20

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

Chain A:

- Molecule 2: functional pRN1 primase

Chain B:



5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
CYANA	structure calculation	
Amber	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1186
Number of shifts mapped to atoms	1186
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	63%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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
2	B	909	950	949	1±1
All	All	21760	21080	21060	12

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:342:LYS:HA	2:B:342:LYS:HE2	0.49	1.84	5	1
2:B:342:LYS:HE3	2:B:342:LYS:HA	0.48	1.85	18	1
2:B:342:LYS:HE2	2:B:342:LYS:HA	0.48	1.84	12	1
2:B:366:LYS:HE2	2:B:366:LYS:HA	0.47	1.86	1	1
2:B:335:LEU:N	2:B:336:PRO:CD	0.43	2.81	20	4
2:B:335:LEU:N	2:B:336:PRO:HD3	0.43	2.28	4	2
2:B:291:LEU:HD22	2:B:291:LEU:C	0.40	2.36	9	1
2:B:327:ASP:HB3	2:B:330:LYS:HE2	0.40	1.93	19	1

5.2 Torsion angles (i)

5.2.1 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	
2	B	109/115 (95%)	107±1 (98±1%)	2±1 (2±1%)	0±0 (0±0%)	100	100
All	All	2180/2300 (95%)	2145 (98%)	35 (2%)	0 (0%)	100	100

There are no Ramachandran outliers.

5.2.2 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	
2	B	103/108 (95%)	100±1 (97±1%)	3±1 (3±1%)	45	89
All	All	2060/2160 (95%)	2001 (97%)	59 (3%)	45	89

All 17 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)
2	B	301	THR	19

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	270	ARG	6
2	B	326	THR	4
2	B	311	ARG	4
2	B	280	LYS	4
2	B	268	VAL	4
2	B	347	TRP	3
2	B	342	LYS	3
2	B	321	MET	3
2	B	349	THR	2
2	B	276	VAL	1
2	B	300	LYS	1
2	B	309	ARG	1
2	B	366	LYS	1
2	B	334	LEU	1
2	B	291	LEU	1
2	B	263	LEU	1

5.2.3 RNA

There are no RNA molecules in this entry.

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

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 63% for the well-defined parts and 62% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *starch_output*

6.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1186
Number of shifts mapped to atoms	1186
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	66	-0.97 ± 0.32	Should be applied
$^{13}C_{\beta}$	90	0.25 ± 0.21	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	101	1.29 ± 0.26	Should be applied

6.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 63%, i.e. 1023 atoms were assigned a chemical shift out of a possible 1633. 13 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	350/540 (65%)	188/215 (87%)	64/220 (29%)	98/105 (93%)
Sidechain	534/816 (65%)	334/479 (70%)	200/301 (66%)	0/36 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	66/101 (65%)	38/52 (73%)	25/43 (58%)	3/6 (50%)
Overall	1023/1633 (63%)	633/850 (74%)	289/625 (46%)	101/158 (64%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 62%, i.e. 1049 atoms were assigned a chemical shift out of a possible 1696. 13 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	361/565 (64%)	194/225 (86%)	66/230 (29%)	101/110 (92%)
Sidechain	549/854 (64%)	344/501 (69%)	205/315 (65%)	0/38 (0%)
Aromatic	66/101 (65%)	38/52 (73%)	25/43 (58%)	3/6 (50%)
Overall	1049/1696 (62%)	649/882 (74%)	296/649 (46%)	104/165 (63%)

6.1.4 Statistically unusual chemical shifts (i)

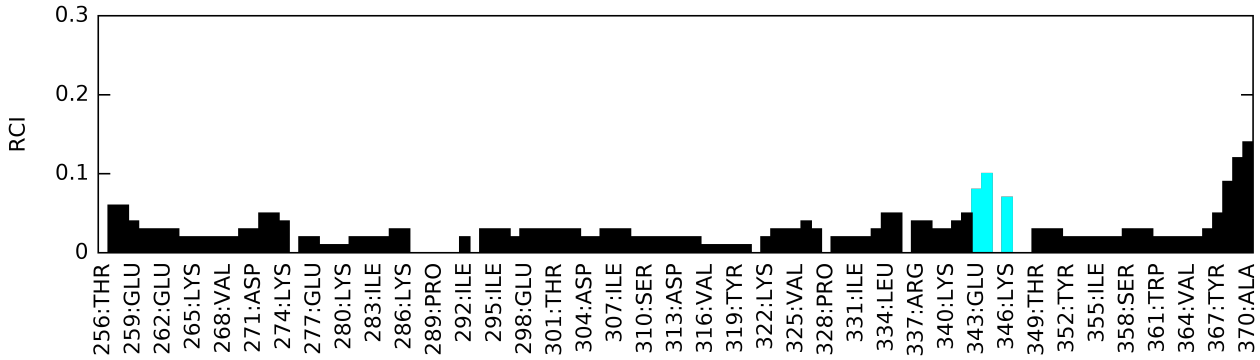
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
2	B	320	LEU	CD2	11.15	32.60 – 15.60	-7.6
2	B	320	LEU	HD12	2.35	2.16 – -0.64	5.7
2	B	320	LEU	HD13	2.35	2.16 – -0.64	5.7
2	B	320	LEU	HD11	2.35	2.16 – -0.64	5.7

6.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain B:


RCI
0.3
0.2
0.1
0
256:THR
259:GLU
262:GLU
265:LYS
268:VAL
271:ASP
274:LYS
277:GLU
280:LYS
283:ILE
286:LYS
289:PRO
292:ILE
295:ILE
298:GLU
301:THR
304:ASP
307:ILE
310:SER
313:ASP
316:VAL
319:TYR
322:LYS
325:VAL
328:PRO
331:ILE
334:LEU
337:ARG
340:LYS
343:GLU
346:LYS
349:THR
352:TYR
355:ILE
358:SER
361:TRP
364:VAL
367:TYR
370:ALA