Full wwPDB NMR Structure Validation Report (i)

Dec 20, 2023 – 09:35 AM EST

PDB ID	:	1B69
Title	:	THE SOLUTION STRUCTURE OF TN916 INTEGRASE N-TERMINAL DOMAIN/DNA COMPLEX
Authors	:	Clubb, R.T.; Wojciak, J.M.; Connolly, K.M.
Deposited on	:	1999-01-21

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

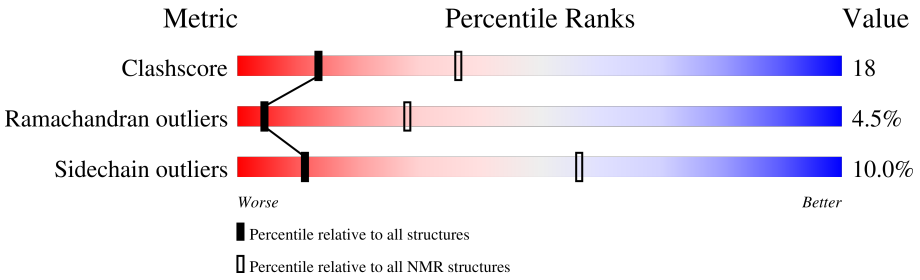
SOFTWARE-VERSIONS INFOmissingINFO

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 <=5%

Mol	Chain	Length	Quality of chain
1	B	13	8% 69% 23%
2	C	13	15% 38% 46%
3	A	69	67% 26% 7%

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1991 atoms, of which 888 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	B	13	Total	C	H	N	O	P	0
			416	129	149	51	75	12	

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

Mol	Chain	Residues	Atoms						Trace
2	C	13	Total	C	H	N	O	P	0
			410	127	150	44	77	12	

- Molecule 3 is a protein called PROTEIN (INTEGRASE).

Mol	Chain	Residues	Atoms					Trace
3	A	69	Total	C	H	N	O	0
			1165	362	589	109	105	

There is a discrepancy between the modelled and reference sequences:

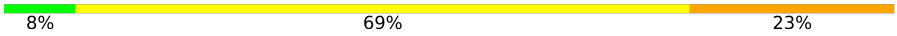
Chain	Residue	Modelled	Actual	Comment	Reference
A	57	ALA	CYS	engineered mutation	UNP P22886

4 Residue-property plots (i)

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.

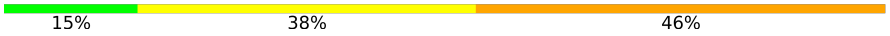
• Molecule 1: DNA (5'-D(*GP*AP*GP*TP*AP*GP*TP*AP*AP*AP*TP*TP*C)-3')

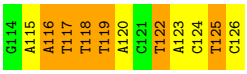
Chain B:



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

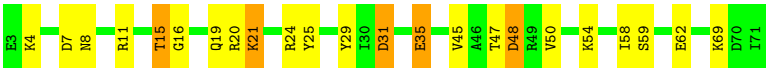
Chain C:





• Molecule 3: PROTEIN (INTEGRASE)

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *DISTANCE GEOMETRY AND SIMULATED ANNEALING*.

Of the 100 calculated structures, 1 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
X-PLOR	refinement	3.843
X-PLOR	structure solution	3.843

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	1.23	4/300 (1.3%)	1.92	6/462 (1.3%)
2	C	1.29	5/290 (1.7%)	1.93	5/445 (1.1%)
3	A	1.09	0/585 (0.0%)	0.88	0/780 (0.0%)
All	All	1.18	9/1175 (0.8%)	1.53	11/1687 (0.7%)

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	125	DT	C5-C7	6.07	1.53	1.50
1	B	112	DT	C5-C7	5.94	1.53	1.50
2	C	119	DT	C5-C7	5.93	1.53	1.50
2	C	118	DT	C5-C7	5.86	1.53	1.50
1	B	111	DT	C5-C7	5.74	1.53	1.50
1	B	104	DT	C5-C7	5.44	1.53	1.50
2	C	122	DT	C5-C7	5.16	1.53	1.50
1	B	107	DT	C5-C7	5.12	1.53	1.50
2	C	117	DT	C5-C7	5.07	1.53	1.50

All angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	122	DT	C6-C5-C7	-6.24	119.16	122.90
1	B	107	DT	C6-C5-C7	-5.80	119.42	122.90
2	C	117	DT	C6-C5-C7	-5.63	119.52	122.90
1	B	101	DG	O4'-C1'-N9	5.38	111.76	108.00
2	C	116	DA	O4'-C1'-N9	5.34	111.74	108.00
2	C	122	DT	C4-C5-C6	5.25	121.15	118.00
1	B	107	DT	C4-C5-C6	5.19	121.12	118.00
1	B	104	DT	C6-C5-C7	-5.17	119.80	122.90
2	C	117	DT	C4-C5-C6	5.13	121.08	118.00
1	B	106	DG	O4'-C1'-N9	5.09	111.56	108.00
1	B	103	DG	O4'-C1'-N9	5.09	111.56	108.00

There are no chirality outliers.

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	267	149	149	6
2	C	260	150	150	15
3	A	576	589	588	19
All	All	1103	888	887	36

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 18.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)
2:C:123:DA:H2"	2:C:124:DC:O5'	0.58	1.99
3:A:31:ASP:OD1	3:A:31:ASP:N	0.54	2.40
2:C:123:DA:C5	2:C:124:DC:C4	0.54	2.95
2:C:122:DT:C4	2:C:123:DA:N6	0.54	2.76
1:B:108:DA:H2"	1:B:109:DA:O5'	0.53	2.03
2:C:116:DA:C2	2:C:117:DT:C2	0.53	2.96
3:A:8:ASN:OD1	3:A:19:GLN:NE2	0.52	2.43
1:B:102:DA:H2"	1:B:103:DG:O5'	0.51	2.06
2:C:122:DT:H2"	2:C:123:DA:O5'	0.50	2.05
3:A:58:ILE:HD12	3:A:62:GLU:HB3	0.50	1.82
3:A:15:THR:HG22	3:A:16:GLY:N	0.50	2.21
1:B:104:DT:H2"	1:B:105:DA:O5'	0.49	2.06
3:A:11:ARG:NH2	3:A:25:TYR:OH	0.49	2.46
3:A:21:LYS:CB	3:A:21:LYS:NZ	0.48	2.76
2:C:125:DT:H2"	2:C:126:DC:O5'	0.48	2.09
1:B:109:DA:H2"	1:B:110:DA:O5'	0.47	2.09
3:A:45:VAL:O	3:A:47:THR:N	0.47	2.46
2:C:119:DT:H2"	2:C:120:DA:O5'	0.47	2.10
2:C:118:DT:H2"	2:C:119:DT:O5'	0.46	2.09
3:A:7:ASP:N	3:A:7:ASP:OD1	0.46	2.49
3:A:54:LYS:N	3:A:54:LYS:CD	0.45	2.79
3:A:58:ILE:HD12	3:A:62:GLU:CB	0.45	2.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)
3:A:15:THR:CG2	3:A:16:GLY:N	0.44	2.80
1:B:105:DA:C6	1:B:106:DG:C6	0.43	3.06
3:A:31:ASP:OD1	3:A:35:GLU:O	0.43	2.37
2:C:117:DT:H2'	2:C:118:DT:H72	0.42	1.90
3:A:50:VAL:HG23	3:A:50:VAL:O	0.42	2.14
2:C:118:DT:P	3:A:24:ARG:HH22	0.41	2.38
2:C:119:DT:C7	3:A:24:ARG:NH1	0.41	2.83
2:C:115:DA:H2"	2:C:116:DA:O5'	0.41	2.15
1:B:105:DA:H2"	1:B:106:DG:O5'	0.41	2.15
2:C:124:DC:H2"	2:C:125:DT:O5'	0.41	2.16
2:C:118:DT:P	3:A:24:ARG:NH2	0.40	2.94
3:A:29:TYR:O	3:A:29:TYR:CD1	0.40	2.74
3:A:59:SER:O	3:A:62:GLU:N	0.40	2.49
2:C:119:DT:H71	3:A:24:ARG:NH1	0.40	2.32

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	67/69 (97%)	49 (73%)	15 (22%)	3 (4%)	4	28
All	All	67/69 (97%)	49 (73%)	15 (22%)	3 (4%)	4	28

All 3 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
3	A	15	THR
3	A	48	ASP
3	A	69	LYS

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	60/60 (100%)	54 (90%)	6 (10%)	11 56
All	All	60/60 (100%)	54 (90%)	6 (10%)	11 56

All 6 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
3	A	4	LYS
3	A	20	ARG
3	A	21	LYS
3	A	31	ASP
3	A	35	GLU
3	A	48	ASP

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