wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 05:21 PM BST

PDB ID	:	1TN9
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 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
http://wwpdb.org/validation/2016/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

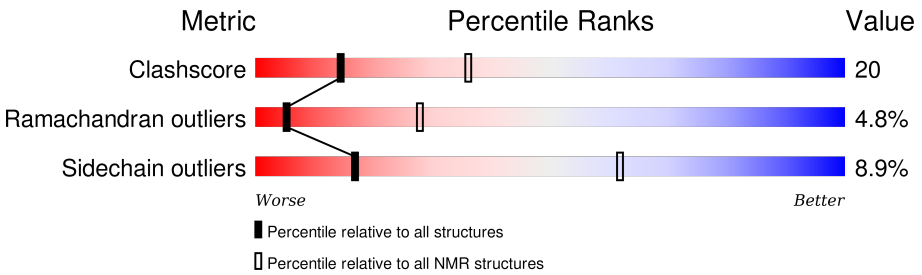
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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% 62% 31%
2	C	13	8% 38% 54%
3	A	69	55% 36% 9%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 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:6-A:68 (63)	0.36	3

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 7 clusters and 3 single-model clusters were found.

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

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
			Total	C	H	N	O	P	
1	B	13	416	129	149	51	75	12	0

- 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
			Total	C	H	N	O	P	
2	C	13	410	127	150	44	77	12	0

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

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

There is a discrepancy between the modelled and reference sequences:

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

4 Residue-property plots (i)

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(*GP*AP*GP*TP*AP*GP*TP*AP*AP*AP*TP*TP*C)-3')

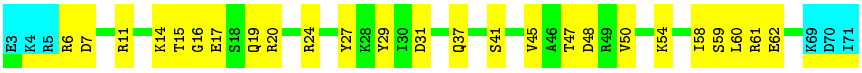
Chain B: 8% 62% 31%

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

Chain C: 8% 38% 54%

- Molecule 3: PROTEIN (INTEGRASE)

Chain A: 55% 36% 9%



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

The representative model is number 3. Colouring as in section 4.1 above.

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

Chain B: 15% 62% 23%

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

Chain C: 23% 31% 46%

G114 A115 A116 T117 T118 T119 A120 C121 T122 A123 C124 T125 C126

- Molecule 3: PROTEIN (INTEGRASE)

Chain A: 54% 28% 10% 9%

E3 K4 R5 G10 L13 K14 E17 S18 Q19 R20 K21 D22 G23 R24 Y27 K28 Y29 I30 D31 S32 F33 S41 W42 K43 L44 V45 A46 T47 D48 R49 V50 K54 I58 S59 E62 K63 K69 D70 I71

5 Refinement protocol and experimental data overview (i)

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

Of the 100 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
X-PLOR	refinement	3.843
XPLOR	structure solution	

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

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	1.24±0.01	3±0/300 (1.1±0.2%)	1.92±0.01	7±1/462 (1.5±0.2%)
2	C	1.28±0.01	4±1/290 (1.4±0.3%)	1.94±0.01	8±1/445 (1.8±0.3%)
3	A	1.09±0.00	0±0/532 (0.0±0.0%)	0.88±0.00	0±0/713 (0.0±0.0%)
All	All	1.18	150/22440 (0.7%)	1.56	302/32400 (0.9%)

5 of 9 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	112	DT	C5-C7	6.32	1.53	1.50	4	20
2	C	118	DT	C5-C7	6.30	1.53	1.50	17	20
2	C	119	DT	C5-C7	6.25	1.53	1.50	18	20
2	C	125	DT	C5-C7	6.25	1.53	1.50	16	20
1	B	111	DT	C5-C7	6.14	1.53	1.50	16	18

5 of 27 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	B	101	DG	O4'-C1'-N9	6.61	112.63	108.00	14	19
2	C	122	DT	C6-C5-C7	-6.61	118.94	122.90	3	20
2	C	117	DT	O4'-C1'-N1	6.48	112.53	108.00	11	17
1	B	107	DT	C6-C5-C7	-6.13	119.22	122.90	1	20
2	C	117	DT	C6-C5-C7	-5.89	119.37	122.90	3	19

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	9±2
2	C	260	150	150	12±2
3	A	522	529	529	20±6
All	All	20980	16560	16560	755

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 20.

5 of 255 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:C:116:DA:H2"	2:C:117:DT:O5'	0.66	1.90	8	10
3:A:13:LEU:HD12	3:A:13:LEU:H	0.64	1.52	14	1
3:A:28:LYS:NZ	3:A:38:PHE:CZ	0.63	2.66	5	2
3:A:37:GLN:NE2	3:A:38:PHE:N	0.63	2.45	20	1
3:A:27:TYR:CE2	3:A:29:TYR:CD1	0.63	2.87	7	8

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	63/69 (91%)	48±2 (76±3%)	12±2 (19±3%)	3±1 (5±2%)	5 28
All	All	1260/1380 (91%)	960 (76%)	240 (19%)	60 (5%)	5 28

5 of 15 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	48	ASP	15

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
3	A	36	PRO	8
3	A	58	ILE	8
3	A	10	GLY	7
3	A	20	ARG	6

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	54/60 (90%)	49±1 (91±3%)	5±1 (9±3%)	17	62
All	All	1080/1200 (90%)	984 (91%)	96 (9%)	17	62

5 of 29 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	41	SER	10
3	A	61	ARG	8
3	A	6	ARG	7
3	A	60	LEU	7
3	A	59	SER	6

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 carbohydrates 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