



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 08:16 PM BST

PDB ID : 2EXF  
Title : Solution structure of the HIV-1 nucleocapsid (NCp7(12-55)) complexed with the DNA (-) Primer Binding Site  
Authors : Bourbigot, S.; Bouaziz, S.; Morellet, N.  
Deposited on : 2005-11-08

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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:

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

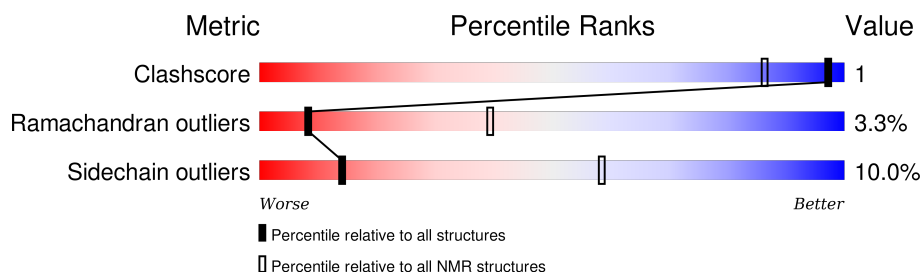
# 1 Overall quality at a glance

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  $\geq 3$ , 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	14	
2	A	44	

## 2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 4 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	A:13-A:51 (39)	0.48	4

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 3 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 4, 6, 7
2	2, 3, 5
3	8, 9, 10

### 3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 1122 atoms, of which 493 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
1	B	14	Total	C	H	N	O	P	0
			442	135	160	48	86	13	

- Molecule 2 is a protein called Nucleocapsid protein\* (NC\*).

Mol	Chain	Residues	Atoms						Trace
2	A	44	Total	C	H	N	O	S	0
			678	203	333	74	61	7	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ASN	-	CLONING ARTIFACT	UNP P03368

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
3	A	2	Total	Zn
			2	2

## 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: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B: 



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A: 

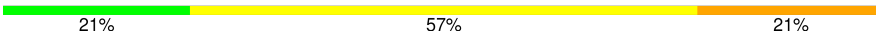


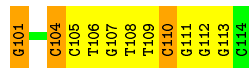
### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.


#### 4.2.1 Score per residue for model 1

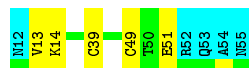
- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B: 



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A: 




### 4.2.2 Score per residue for model 2

- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B:  36% 50% 14%



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A:  80% 9% 11%



### 4.2.3 Score per residue for model 3

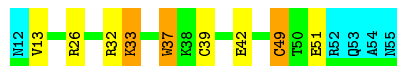
- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B:  36% 43% 21%



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A:  68% 14% 7% 11%




### 4.2.4 Score per residue for model 4 (medoid)

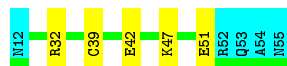
- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B:  14% 71% 14%



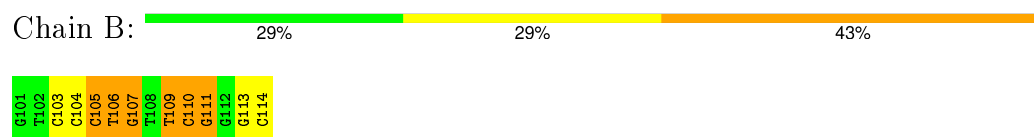
- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A:  77% 11% 11%

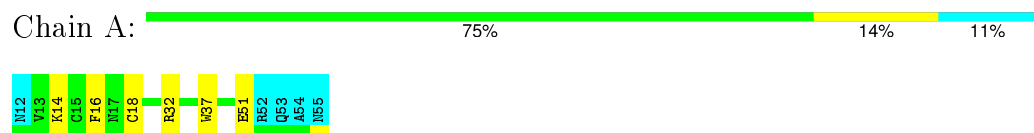


### 4.2.5 Score per residue for model 5

- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

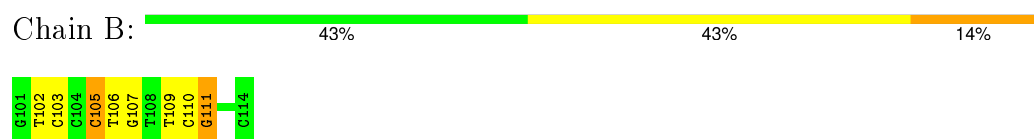


- Molecule 2: Nucleocapsid protein\* (NC\*)

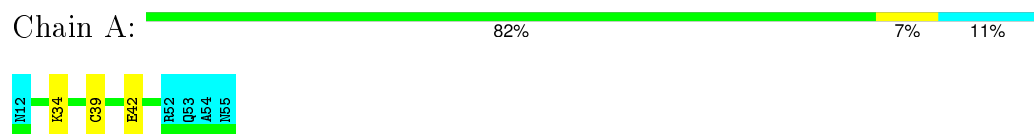


### 4.2.6 Score per residue for model 6

- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'



- Molecule 2: Nucleocapsid protein\* (NC\*)

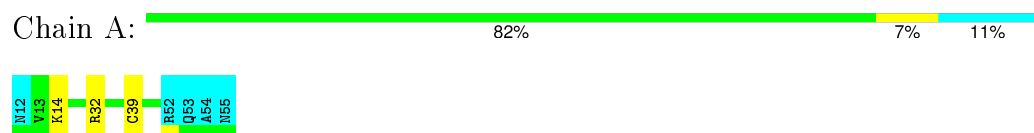


### 4.2.7 Score per residue for model 7

- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

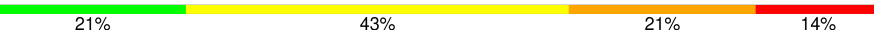


- Molecule 2: Nucleocapsid protein\* (NC\*)



### 4.2.8 Score per residue for model 8

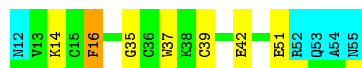
- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B: 



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A: 




### 4.2.9 Score per residue for model 9

- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B: 



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A: 




### 4.2.10 Score per residue for model 10

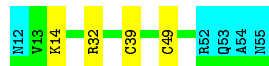
- Molecule 1: 5'-D(\*GP\*TP\*CP\*CP\*CP\*TP\*GP\*TP\*TP\*CP\*GP\*GP\*GP\*C)-3'

Chain B: 



- Molecule 2: Nucleocapsid protein\* (NC\*)

Chain A: 





## 5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *restrained molecular dynamics restrained simulated annealing energy minimization*.

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *structures with acceptable covalent geometry, structures with favorable non-bond energy, structures with the least restraint violations, structures with the lowest energy*.

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

Software name	Classification	Version
DISCOVER	refinement	2.98

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)

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	1.31±0.02	0±0/314 (0.0±0.1%)	2.22±0.06	16±2/483 (3.3±0.4%)
2	A	0.55±0.01	0±0/309 (0.0±0.0%)	0.94±0.04	0±0/408 (0.0±0.1%)
All	All	1.01	1/6230 (0.0%)	1.75	159/8910 (1.8%)

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	4.1±1.9
2	A	0.0±0.0	0.2±0.4
All	All	0	43

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	B	107	DG	C4'-O4'	-5.69	1.39	1.45	9	1

All 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	110	DC	O4'-C1'-N1	16.75	119.72	108.00	9	10
1	B	109	DT	O4'-C1'-N1	11.05	115.74	108.00	4	10
1	B	111	DG	O4'-C1'-N9	10.66	115.47	108.00	10	9
1	B	103	DC	O4'-C1'-N1	10.00	115.00	108.00	9	8
1	B	109	DT	C6-C5-C7	-9.25	117.35	122.90	4	10
1	B	113	DG	O4'-C1'-N9	9.01	114.31	108.00	4	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	B	101	DG	O4'-C1'-N9	8.98	114.28	108.00	8	4
1	B	107	DG	O4'-C1'-N9	8.75	114.13	108.00	7	6
1	B	101	DG	C4'-C3'-C2'	-8.40	95.54	103.10	2	2
1	B	106	DT	C6-C5-C7	-8.02	118.09	122.90	1	10
1	B	109	DT	O4'-C4'-C3'	-8.01	101.19	106.00	4	1
1	B	102	DT	O4'-C1'-N1	7.97	113.58	108.00	7	3
1	B	110	DC	O4'-C4'-C3'	-7.94	101.24	106.00	9	2
1	B	112	DG	O4'-C1'-N9	7.30	113.11	108.00	4	3
1	B	105	DC	O4'-C1'-N1	7.20	113.04	108.00	8	7
1	B	109	DT	C1'-O4'-C4'	-7.05	103.05	110.10	4	1
1	B	104	DC	O4'-C1'-N1	6.96	112.87	108.00	1	1
1	B	110	DC	N1-C2-O2	6.87	123.02	118.90	8	8
1	B	106	DT	N3-C2-O2	-6.73	118.26	122.30	9	10
1	B	103	DC	C1'-O4'-C4'	-6.61	103.49	110.10	9	5
1	B	108	DT	C6-C5-C7	-6.52	118.99	122.90	9	3
1	B	113	DG	P-O3'-C3'	6.39	127.36	119.70	5	1
1	B	102	DT	C6-C5-C7	-6.37	119.08	122.90	6	1
1	B	107	DG	C4'-C3'-C2'	-6.35	97.38	103.10	9	7
1	B	102	DT	O4'-C1'-C2'	-5.89	101.19	105.90	3	1
1	B	110	DC	C1'-O4'-C4'	-5.86	104.25	110.10	10	3
1	B	109	DT	C5'-C4'-C3'	5.84	124.62	114.10	4	1
1	B	103	DC	N1-C2-O2	5.62	122.27	118.90	8	6
1	B	101	DG	O4'-C1'-C2'	-5.58	101.44	105.90	1	2
1	B	109	DT	C4'-C3'-C2'	-5.58	98.08	103.10	6	3
1	B	107	DG	O4'-C1'-C2'	5.51	110.31	105.90	9	1
1	B	109	DT	P-O3'-C3'	5.46	126.25	119.70	6	2
1	B	102	DT	C4'-C3'-C2'	-5.45	98.19	103.10	2	1
1	B	111	DG	O4'-C1'-C2'	-5.44	101.55	105.90	1	3
1	B	107	DG	C1'-O4'-C4'	-5.41	104.69	110.10	9	1
1	B	112	DG	P-O3'-C3'	5.32	126.09	119.70	1	1
1	B	111	DG	C1'-O4'-C4'	-5.31	104.79	110.10	10	3
1	B	106	DT	C4-C5-C6	5.29	121.17	118.00	10	4
2	A	49	CYS	CA-CB-SG	5.25	123.45	114.00	3	1
1	B	108	DT	O4'-C1'-N1	5.10	111.57	108.00	1	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	B	104	DC	Sidechain	7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group	Models (Total)
1	B	109	DT	Sidechain	6
1	B	105	DC	Sidechain	6
1	B	110	DC	Sidechain	5
1	B	107	DG	Sidechain	5
1	B	111	DG	Sidechain	3
1	B	106	DT	Sidechain	3
2	A	16	PHE	Sidechain	2
1	B	114	DC	Sidechain	2
1	B	112	DG	Sidechain	2
1	B	108	DT	Sidechain	1
1	B	101	DG	Sidechain	1

## 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	282	160	160	1±1
2	A	303	293	293	0±1
All	All	5870	4530	4530	9

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:108:DT:C6	1:B:109:DT:H73	0.52	2.39	8	1
1:B:107:DG:C8	2:A:37:TRP:CH2	0.46	3.03	8	1
2:A:33:LYS:HG2	2:A:37:TRP:CD1	0.46	2.46	3	1
2:A:16:PHE:CZ	2:A:37:TRP:CZ3	0.42	3.07	8	1
1:B:107:DG:H4'	1:B:108:DT:H72	0.41	1.91	9	1
1:B:108:DT:C5	1:B:109:DT:H73	0.41	2.50	8	1
1:B:108:DT:C5	1:B:109:DT:C4	0.41	3.09	10	1
2:A:16:PHE:CD2	2:A:37:TRP:CZ3	0.40	3.09	5	1
1:B:107:DG:O4'	1:B:108:DT:H72	0.40	2.16	10	1

## 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	
2	A	39/44 (89%)	32±1 (83±3%)	5±1 (14±3%)	1±1 (3±2%)	8	39
All	All	390/440 (89%)	323 (83%)	54 (14%)	13 (3%)	8	39

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

Mol	Chain	Res	Type	Models (Total)
2	A	51	GLU	6
2	A	49	CYS	3
2	A	13	VAL	2
2	A	18	CYS	1
2	A	35	GLY	1

### 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	
2	A	32/36 (89%)	29±1 (90±4%)	3±1 (10±4%)	14	58
All	All	320/360 (89%)	288 (90%)	32 (10%)	14	58

All 10 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	A	39	CYS	7
2	A	14	LYS	6
2	A	32	ARG	6
2	A	42	GLU	6
2	A	33	LYS	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
2	A	49	CYS	1
2	A	26	ARG	1
2	A	47	LYS	1
2	A	37	TRP	1
2	A	34	LYS	1

### 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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

## 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