wwPDB NMR Structure Validation Summary Report (i)

Feb 15, 2022 – 06:45 AM EST

PDB ID	:	1L6N
Title	:	STRUCTURE OF THE N-TERMINAL 283-RESIDUE FRAGMENT OF THE HIV-1 GAG POLYPROTEIN
Authors	:	Tang, C.; Ndassa, Y.; Summers, M.F.
Deposited on	:	2002-03-11

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 (i) symbol.

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

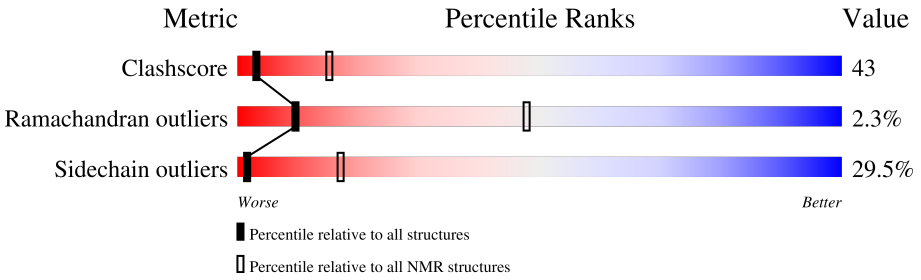
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	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	289	25% 42% 6% 27%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 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:7-A:20, A:29-A:105 (91)	0.21	17
2	A:147-A:215, A:230-A:279 (119)	0.42	2

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 5 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 4541 atoms, of which 2276 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Gag Polyprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	288	4541	1418	2276	414	423	10	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	284	HIS	-	expression tag	UNP Q72497
A	285	HIS	-	expression tag	UNP Q72497
A	286	HIS	-	expression tag	UNP Q72497
A	287	HIS	-	expression tag	UNP Q72497
A	288	HIS	-	expression tag	UNP Q72497
A	289	HIS	-	expression tag	UNP Q72497

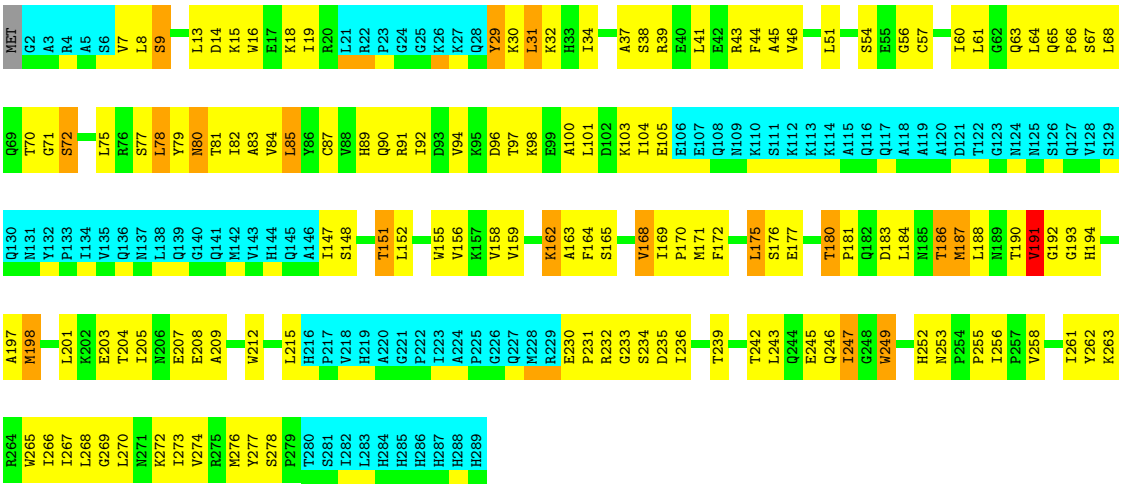
4 Residue-property plots (i)

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: Gag Polyprotein

Chain A: 25% 42% 6% 27%

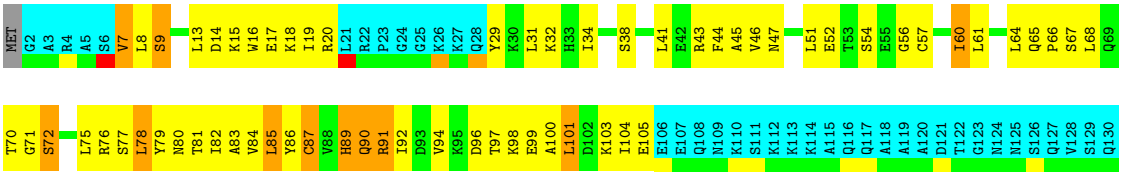


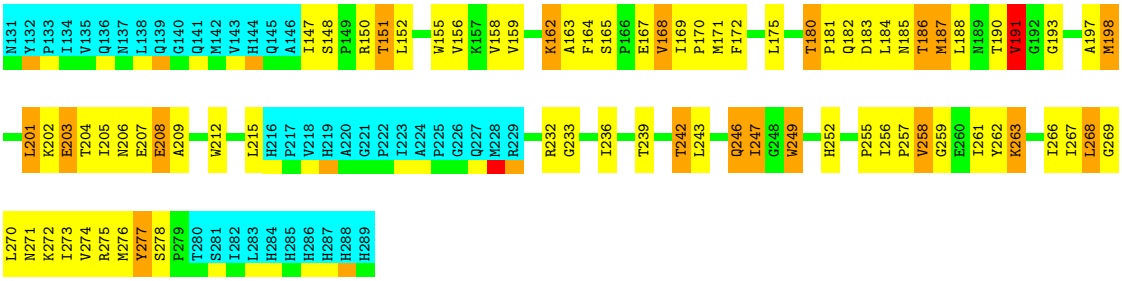
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

• Molecule 1: Gag Polyprotein

Chain A: 24% 38% 10% 27%




N131
Y132
P133
I134
V135
Q136
N137
L138
Q139
G140
Q141
M142
V143
H144
Q145
A146
I147
S148
P149
R150
T151
L152
W155
V156
K157
V158
V159
K162
A163
F164
S165
P166
E167
V168
I169
P170
M171
F172
L175
T180
P181
Q182
D183
L184
N185
T186
M187
L188
N189
T190
V191
G192
G193
A197
M198
L201
K202
E203
T204
I205
N206
E207
E208
A209
W212
L215
H216
P217
V218
H219
A220
G221
P222
I223
A224
P225
G226
Q227
M228
R229
R232
G233
I236
T239
T242
L243
Q246
I247
G248
W249
H252
P255
I256
P257
V258
G259
E260
I261
Y262
K263
I266
I267
L268
G269
L270
N271
K272
I273
V274
R275
M276
Y277
S278
P279
T280
S281
I282
L283
H284
H285
H286
H287
H288
H289

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, torsion angle dynamics.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
DYANA	structure solution	1.5
DYANA	refinement	1.5

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	A	1667	1684	1680	142±11
All	All	33340	33680	33600	2850

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 43.

5 of 664 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:7:VAL:HG11	1:A:34:ILE:HG21	1.04	1.28	12	20
1:A:83:ALA:HB1	1:A:104:ILE:HD13	1.03	1.31	3	18
1:A:83:ALA:HB1	1:A:104:ILE:CD1	0.98	1.87	13	20
1:A:147:ILE:HD12	1:A:183:ASP:OD1	0.93	1.64	1	6
1:A:162:LYS:O	1:A:168:VAL:HG21	0.92	1.64	3	7

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
1	A	210/289 (73%)	185±2 (88±1%)	20±2 (10±1%)	5±1 (2±1%)	9 48
All	All	4200/5780 (73%)	3704 (88%)	400 (10%)	96 (2%)	9 48

5 of 12 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	72	SER	20
1	A	191	VAL	20
1	A	177	GLU	14
1	A	7	VAL	9
1	A	194	HIS	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
1	A	183/246 (74%)	129±4 (70±2%)	54±4 (30±2%)	1 17
All	All	3660/4920 (74%)	2579 (70%)	1081 (30%)	1 17

5 of 131 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)
1	A	9	SER	20
1	A	148	SER	20
1	A	175	LEU	20
1	A	186	THR	20
1	A	204	THR	20

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 (i)

No chemical shift data were provided