wwPDB NMR Structure Validation Summary Report (i)

Jan 11, 2017 – 06:32 AM EST

PDB ID	:	5T82
Title	:	HIV-1 reverse transcriptase thumb subdomain
Authors	:	Gronenborn, A.M.; Sharaf, N.G.; Byeon, I.-J.L.
Deposited on	:	2016-09-06

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:

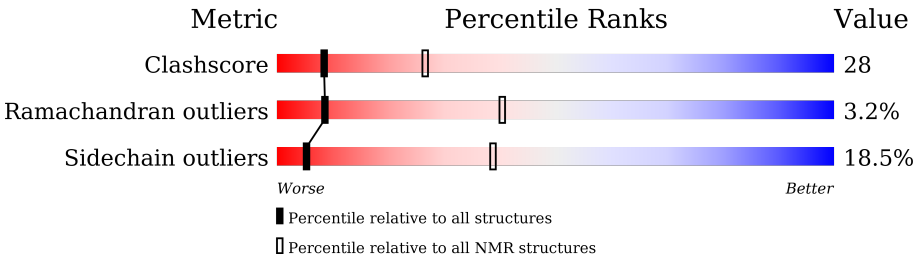
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20028442
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

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 80%.

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	A	90	32% 39% • 28%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 17 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:244-A:283, A:288-A:312 (65)	0.28	17

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. No single-model clusters were found.

Cluster number	Models
1	3, 6, 7, 9, 12, 13, 16, 17, 19, 20, 21, 22, 26, 27, 28, 30
2	4, 18, 23, 25, 29
3	2, 8, 11, 14, 24
4	1, 5, 10, 15

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1491 atoms, of which 751 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Reverse transcriptase.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	90	1491	475	751	131	134	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	280	SER	CYS	engineered mutation	UNP Q74596
A	319	LEU	-	expression tag	UNP Q74596
A	320	GLU	-	expression tag	UNP Q74596
A	321	HIS	-	expression tag	UNP Q74596
A	322	HIS	-	expression tag	UNP Q74596
A	323	HIS	-	expression tag	UNP Q74596
A	324	HIS	-	expression tag	UNP Q74596
A	325	HIS	-	expression tag	UNP Q74596
A	326	HIS	-	expression tag	UNP Q74596

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: Reverse transcriptase

Chain A: 32% 39% • 28%

D237 K238 W239 T240 V241 Q242 P243 L246 P247 E248 K249 W252 T253 V254 N255 D256 I257 Q258 K259 L260 V261 G262 K263 L264 N265 W266 A267 S268 Q269 I270 Y271 I274 K275 V276 K281 L282 L283 R284 G285 T286 K287 A288 E291 T296 E297 E298 A299 E300 L301 E302 L303 N306

I309 L310 K311 E312 P313 V314 H315 G316 V317 Y318 L319 E320 H321 H322 H323 H324 H325 H326

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

The representative model is number 17. Colouring as in section 4.1 above.

- Molecule 1: Reverse transcriptase

Chain A: 28% 32% 12% 28%

D237 K238 W239 T240 V241 Q242 P243 I244 V245 L246 K249 D250 S251 W252 T253 V254 N255 D256 I257 Q258 K259 L260 V261 G262 K263 L264 N265 W266 A267 S268 Q269 I270 Y271 I274 K275 V276 R277 Q278 K281 L282 L283 R284 G285 T286 K287 A288 E291 L295 T296 A299 E300 L301 E302 L303

N306 R307 E308 I309 L310 K311 E312 P313 V314 H315 G316 V317 Y318 L319 E320 H321 H322 H323 H324 H325 H326

5 Refinement protocol and experimental data overview (i)

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

Of the 256 calculated structures, 30 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
CNS	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	5t82_cs.cif
Number of chemical shift lists	1
Total number of shifts	939
Number of shifts mapped to atoms	939
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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)

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	523	552	552	30±4
All	All	15690	16560	16560	902

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 28.

5 of 203 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:246:LEU:H	1:A:246:LEU:HD22	0.94	1.23	9	2
1:A:246:LEU:HD12	1:A:246:LEU:H	0.79	1.38	29	7
1:A:255:ASN:OD1	1:A:289:LEU:HD22	0.78	1.78	5	2
1:A:246:LEU:H	1:A:246:LEU:HD12	0.78	1.38	10	4
1:A:246:LEU:HD13	1:A:246:LEU:N	0.77	1.95	25	2

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	65/90 (72%)	53±2 (81±3%)	10±2 (16±3%)	2±1 (3±2%)	8 40
All	All	1950/2700 (72%)	1577 (81%)	311 (16%)	62 (3%)	8 40

5 of 9 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	288	ALA	17
1	A	247	PRO	15
1	A	310	LEU	11
1	A	276	VAL	7
1	A	248	GLU	4

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	59/82 (72%)	48±2 (82±4%)	11±2 (18±4%)	5 39
All	All	1770/2460 (72%)	1443 (82%)	327 (18%)	5 39

5 of 40 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	246	LEU	25
1	A	253	THR	24
1	A	254	VAL	23
1	A	300	GLU	22
1	A	275	LYS	19

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)

The completeness of assignment taking into account all chemical shift lists is 80% for the well-defined parts and 70% for the entire structure.

7.1 Chemical shift list 1

File name: 5t82_cs.cif

Chemical shift list name: *160902_2_chemical_shift.txt*

7.1.1 Bookkeeping (i)

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	939
Number of shifts mapped to atoms	939
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	82	-0.27 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	78	0.06 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	76	0.14 ± 0.27	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 80%, i.e. 669 atoms were assigned a chemical shift out of a possible 835. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	250/319 (78%)	123/127 (97%)	65/130 (50%)	62/62 (100%)
Sidechain	403/484 (83%)	239/280 (85%)	158/186 (85%)	6/18 (33%)

Continued on next page...

Continued from previous page...

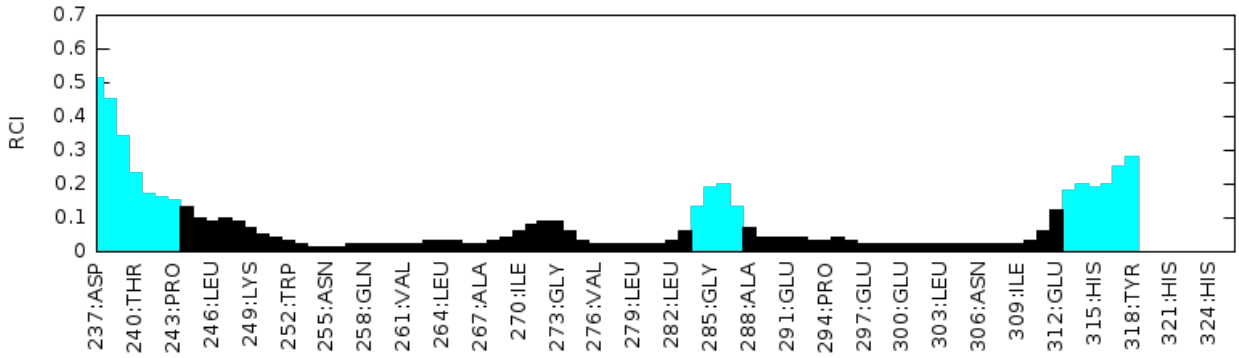
	Total	^{1}H	^{13}C	^{15}N
Aromatic	16/32 (50%)	14/16 (88%)	0/14 (0%)	2/2 (100%)
Overall	669/835 (80%)	376/423 (89%)	223/330 (68%)	70/82 (85%)

7.1.4 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
1	A	249	LYS	HB3	0.07	3.10 – 0.40	-6.2
1	A	278	GLN	H	11.48	11.17 – 5.27	5.5

7.1.5 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 A: