wwPDB NMR Structure Validation Summary Report

Nov 9, 2021 – 10:14 PM EST

PDB ID	:	6XQJ
Title	:	Structure of HIV-1 Vpr in complex with the human nucleotide excision repair protein hHR23A
Authors	:	Byeon, I.-J.L.; Calero, G.; Wu, Y.; Byeon, C.H.; Gronenborn, A.M.
Deposited on	:	2020-07-09

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:

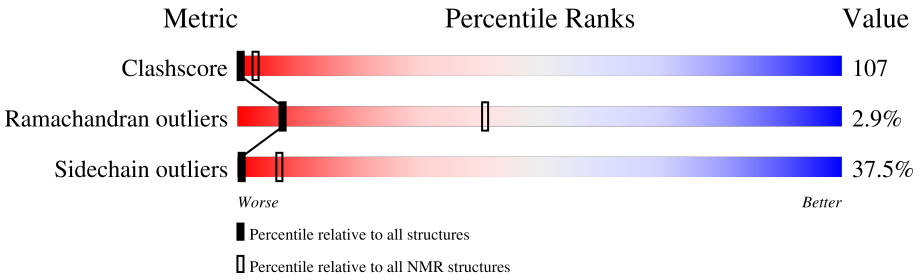
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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	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	234	5% 53% 15% 26%

2 Ensemble composition and analysis (i)

This entry contains 55 models. Model 43 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:14-A:79, A:230-A:288, A:314-A:360 (172)	0.52	43

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 13 clusters and 2 single-model clusters were found.

Cluster number	Models
1	7, 16, 21, 22, 26, 28, 36, 38, 48, 53
2	5, 19, 23, 25, 27, 32, 43, 45, 50
3	3, 4, 6, 14, 15, 34, 41
4	2, 8, 11, 12, 29, 47
5	35, 37, 46, 51
6	1, 9, 42
7	17, 54
8	24, 44
9	33, 39
10	10, 13
11	18, 49
12	31, 55
13	20, 52
Single-model clusters	30; 40

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2856 atoms, of which 1415 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein Vpr,UV excision repair protein RAD23 homolog A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	173	2855	925	1415	252	259	4	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	80	GLY	-	linker	UNP P12520
A	81	GLY	-	linker	UNP P12520
A	82	SER	-	linker	UNP P12520
A	83	GLY	-	linker	UNP P12520
A	84	GLY	-	linker	UNP P12520
A	85	SER	-	linker	UNP P12520
A	364	LEU	-	expression tag	UNP P54725
A	365	GLU	-	expression tag	UNP P54725
A	366	HIS	-	expression tag	UNP P54725
A	367	HIS	-	expression tag	UNP P54725
A	368	HIS	-	expression tag	UNP P54725
A	369	HIS	-	expression tag	UNP P54725
A	370	HIS	-	expression tag	UNP P54725
A	371	HIS	-	expression tag	UNP P54725

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

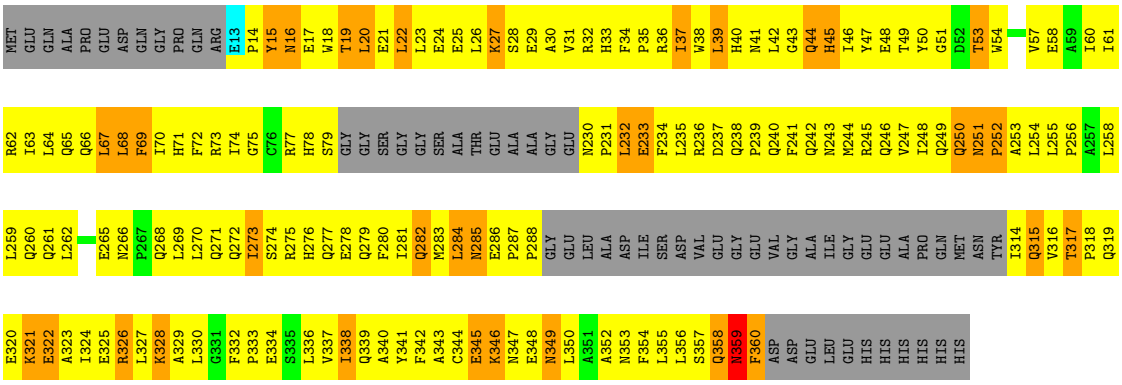
4 Residue-property plots

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: Protein Vpr,UV excision repair protein RAD23 homolog A

Chain A: 5% 53% 15% 26%

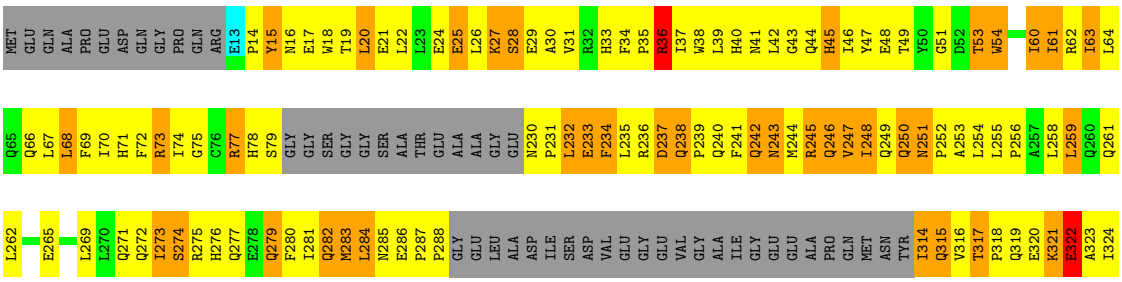


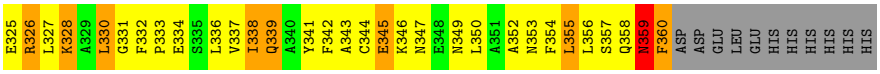
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 43. Colouring as in section 4.1 above.

• Molecule 1: Protein Vpr,UV excision repair protein RAD23 homolog A

Chain A: 11% 42% 20% 26%




E325
R326
L327
K328
A329
L330
G331
F332
P333
E334
S335
L336
V337
T338
Q339
A340
Y341
F342
A343
C344
E345
K346
N347
E348
N349
L350
A351
A352
N353
F354
L355
L356
S357
Q358
N359
F360
ASP
ASP
GLU
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

5 Refinement protocol and experimental data overview

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

Of the 960 calculated structures, 55 were deposited, based on the following criterion: *lowest energy structures*.

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

Software name	Classification	Version
X-PLOR NIH	structure calculation	2.48

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	2569
Number of shifts mapped to atoms	2569
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%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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

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	1431	1409	1410	305±23
All	All	78760	77495	77550	16754

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 107.

5 of 3759 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:31:VAL:HG23	1:A:39:LEU:HD22	1.12	1.18	54	14
1:A:316:VAL:HG13	1:A:342:PHE:CZ	1.09	1.81	52	38
1:A:31:VAL:HG22	1:A:39:LEU:HD21	1.07	1.15	55	10
1:A:258:LEU:HD13	1:A:259:LEU:N	1.07	1.63	26	1
1:A:248:ILE:HD11	1:A:255:LEU:HD13	1.07	1.19	51	10

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	167/234 (71%)	137±4 (82±2%)	25±4 (15±2%)	5±1 (3±1%)	7	41
All	All	9185/12870 (71%)	7533 (82%)	1385 (15%)	267 (3%)	7	41

5 of 26 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	359	ASN	51
1	A	15	TYR	37
1	A	252	PRO	37
1	A	315	GLN	26
1	A	322	GLU	24

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	155/200 (78%)	97±5 (62±3%)	58±5 (38±3%)	1	7
All	All	8525/11000 (78%)	5324 (62%)	3201 (38%)	1	7

5 of 133 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	45	HIS	55
1	A	321	LYS	55
1	A	284	LEU	54
1	A	317	THR	54
1	A	349	ASN	54

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

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

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *BMRB_PDBdep_temp.txt*

7.1.1 Bookkeeping

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	2569
Number of shifts mapped to atoms	2569
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	231	-0.49 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	209	0.48 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	206	0.55 ± 0.28	None needed (imprecise)

7.1.3 Completeness of resonance assignments

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. 1836 atoms were assigned a chemical shift out of a possible 2288. 0 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	655/838 (78%)	330/333 (99%)	172/344 (50%)	153/161 (95%)
Sidechain	1011/1253 (81%)	631/739 (85%)	346/450 (77%)	34/64 (53%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	170/197 (86%)	90/108 (83%)	77/83 (93%)	3/6 (50%)
Overall	1836/2288 (80%)	1051/1180 (89%)	595/877 (68%)	190/231 (82%)

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	346	LYS	HA	1.71	6.46 – 2.06	-5.8

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: