wwPDB NMR Structure Validation Summary Report

Apr 21, 2024 – 12:04 AM EDT

PDB ID	:	2LJY
BMRB ID	:	17968
Title	:	Haddock model structure of the N-terminal domain dimer of HPV16 E6
Authors	:	Zanier, K.; Muhamed Sidi, A.; Boulade-Ladame, C.; Rybin, V.; Chappelle, A.; Atkinson, A.; Kieffer, B.; Trave, G.
Deposited on	:	2011-09-30

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

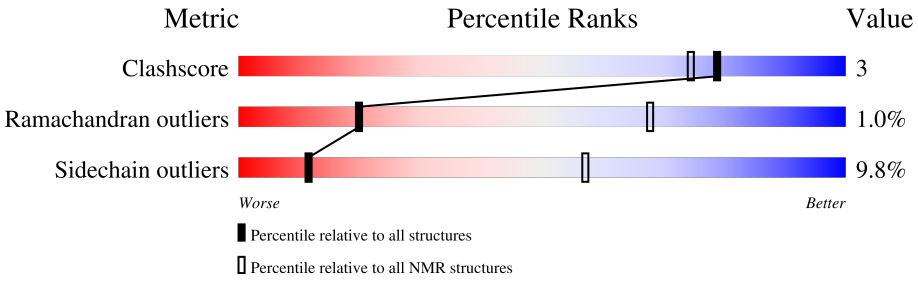
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 38%.

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	84	67% • 13% 17%
1	B	84	68% 7% 8% 17%

2 Ensemble composition and analysis

This entry contains 20 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:9-A:67 (59)	0.58	20
2	B:8-B:70 (63)	0.69	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 3 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2358 atoms, of which 1170 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein E6.

Mol	Chain	Residues	Atoms						Trace
1	A	70	Total	C	H	N	O	S	0
			1178	380	585	101	105	7	
1	B	70	Total	C	H	N	O	S	0
			1178	380	585	101	105	7	

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P03126
A	80	SER	CYS	engineered mutation	UNP P03126
B	-1	GLY	-	expression tag	UNP P03126
B	80	SER	CYS	engineered mutation	UNP P03126

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1
2	B	1	Total	Zn
			1	1

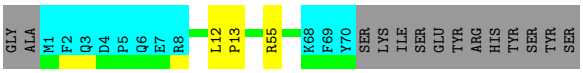
4 Residue-property plots (i)

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: Protein E6

Chain A:



• Molecule 1: Protein E6

Chain B:

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: Protein E6

Chain A:

• Molecule 1: Protein E6

Chain B:

5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 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
HADDOCK webserver	structure calculation	
HADDOCK webserver	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	891
Number of shifts mapped to atoms	794
Number of unparsed shifts	0
Number of shifts with mapping errors	97
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	38%

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	489	490	490	3±2
1	B	532	534	534	3±1
All	All	20460	20480	20495	113

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 3.

5 of 52 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:28:LEU:HD23	1:B:61:ALA:HB2	0.68	1.63	14	1
1:B:30:CYS:O	1:B:34:LYS:HA	0.63	1.94	20	5
1:B:15:LEU:HD22	1:B:52:ILE:HD13	0.59	1.75	17	1
1:A:16:CYS:SG	1:A:23:ILE:HG12	0.58	2.39	13	3
1:B:63:CYS:SG	1:B:65:LYS:HE3	0.57	2.39	13	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	
1	A	59/84 (70%)	56±2 (95±3%)	3±2 (4±3%)	1±1 (1±1%)	21	69
1	B	62/84 (74%)	58±2 (93±3%)	4±1 (6±2%)	1±1 (1±1%)	20	68
All	All	2420/3360 (72%)	2271 (94%)	126 (5%)	23 (1%)	20	68

5 of 8 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	57	GLY	7
1	B	57	GLY	5
1	B	8	ARG	4
1	A	9	PRO	2
1	B	56	ASP	2

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	56/79 (71%)	51±2 (90±4%)	5±2 (10±4%)	12	58
1	B	60/79 (76%)	54±2 (90±3%)	6±2 (10±3%)	11	57
All	All	2320/3160 (73%)	2093 (90%)	227 (10%)	11	57

5 of 63 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	B	17	THR	12
1	A	17	THR	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	B	55	ARG	10
1	A	53	VAL	10
1	B	53	VAL	8

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)

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

The completeness of assignment taking into account all chemical shift lists is 38% for the well-defined parts and 39% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *E6Ndimer*

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	891
Number of shifts mapped to atoms	794
Number of unparsed shifts	0
Number of shifts with mapping errors	97
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 97) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	71	SER	CA	60.13	0.30	1
1	A	71	SER	CB	63.63	0.30	1
1	A	71	SER	H	7.88	0.02	1
1	A	71	SER	HA	4.32	0.02	1
1	A	71	SER	HB2	3.94	0.02	1
1	A	71	SER	N	117.21	0.30	1
1	A	72	LYS	CA	56.93	0.30	1
1	A	72	LYS	CB	32.63	0.30	1
1	A	72	LYS	CG	24.93	0.30	1
1	A	72	LYS	CD	29.13	0.30	1
1	A	72	LYS	CE	42.33	0.30	1
1	A	72	LYS	H	7.8	0.02	1
1	A	72	LYS	HA	4.26	0.02	1
1	A	72	LYS	HB2	1.86	0.02	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	72	LYS	HB3	1.75	0.02	2
1	A	72	LYS	HG2	1.42	0.02	2
1	A	72	LYS	HG3	1.36	0.02	2
1	A	72	LYS	HD2	1.61	0.02	1
1	A	72	LYS	HE2	2.92	0.02	1
1	A	72	LYS	N	122.81	0.30	1
1	A	73	ILE	CA	61.93	0.30	1
1	A	73	ILE	CB	38.63	0.30	1
1	A	73	ILE	CG1	27.63	0.30	1
1	A	73	ILE	CG2	17.83	0.30	1
1	A	73	ILE	CD1	13.43	0.30	1
1	A	73	ILE	H	7.76	0.02	1
1	A	73	ILE	HA	4.1	0.02	1
1	A	73	ILE	HB	1.88	0.02	1
1	A	73	ILE	HG12	1.4	0.02	2
1	A	73	ILE	HG13	1.12	0.02	2
1	A	73	ILE	HG21	0.85	0.02	1
1	A	73	ILE	HG22	0.85	0.02	1
1	A	73	ILE	HG23	0.85	0.02	1
1	A	73	ILE	HD11	0.77	0.02	1
1	A	73	ILE	HD12	0.77	0.02	1
1	A	73	ILE	HD13	0.77	0.02	1
1	A	73	ILE	N	121.61	0.30	1
1	A	74	SER	CA	59.33	0.30	1
1	A	74	SER	CB	63.93	0.30	1
1	A	74	SER	H	8.16	0.02	1
1	A	74	SER	HA	4.3	0.02	1
1	A	74	SER	HB2	3.85	0.02	2
1	A	74	SER	HB3	3.77	0.02	2
1	A	74	SER	N	120.31	0.30	1
1	A	75	GLU	CA	57.33	0.30	1
1	A	75	GLU	CB	30.13	0.30	1
1	A	75	GLU	CG	36.43	0.30	1
1	A	75	GLU	H	8.25	0.02	1
1	A	75	GLU	HA	4.18	0.02	1
1	A	75	GLU	HB2	1.94	0.02	2
1	A	75	GLU	HB3	1.87	0.02	2
1	A	75	GLU	HG2	2.13	0.02	2
1	A	75	GLU	HG3	2.16	0.02	2
1	A	75	GLU	N	123.81	0.30	1
1	A	76	TYR	CA	58.13	0.30	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	76	TYR	CB	38.73	0.30	1
1	A	76	TYR	H	8.0	0.02	1
1	A	76	TYR	HA	4.46	0.02	1
1	A	76	TYR	HB2	2.95	0.02	2
1	A	76	TYR	HB3	2.92	0.02	2
1	A	76	TYR	HD1	7.0	0.02	1
1	A	76	TYR	HE1	6.74	0.02	1
1	A	76	TYR	N	122.31	0.30	1
1	A	77	ARG	CA	56.03	0.30	1
1	A	77	ARG	CB	31.03	0.30	1
1	A	77	ARG	CG	27.03	0.30	1
1	A	77	ARG	CD	43.43	0.30	1
1	A	77	ARG	H	7.9	0.02	1
1	A	77	ARG	HA	4.19	0.02	1
1	A	77	ARG	HB3	1.59	0.02	1
1	A	77	ARG	HG2	1.39	0.02	1
1	A	77	ARG	HD2	3.07	0.02	1
1	A	77	ARG	N	124.51	0.30	1
1	A	78	HIS	CA	56.33	0.30	1
1	A	78	HIS	CB	30.73	0.30	1
1	A	78	HIS	H	8.13	0.02	1
1	A	78	HIS	HA	4.51	0.02	1
1	A	78	HIS	HB2	3.01	0.02	2
1	A	78	HIS	HB3	2.94	0.02	2
1	A	78	HIS	HD2	6.95	0.02	1
1	A	78	HIS	HE1	7.91	0.02	1
1	A	78	HIS	N	122.11	0.30	1
1	A	79	TYR	CA	57.93	0.30	1
1	A	79	TYR	CB	39.13	0.30	1
1	A	79	TYR	H	8.01	0.02	1
1	A	79	TYR	HA	4.58	0.02	1
1	A	79	TYR	HB2	3.07	0.02	2
1	A	79	TYR	HB3	2.87	0.02	2
1	A	79	TYR	HD1	7.06	0.02	1
1	A	79	TYR	HE1	6.77	0.02	1
1	A	79	TYR	N	123.61	0.30	1
1	A	80	SER	CA	60.23	0.30	1
1	A	80	SER	CB	65.13	0.30	1
1	A	80	SER	H	7.84	0.02	1
1	A	80	SER	HA	4.19	0.02	1
1	A	80	SER	HB2	3.8	0.02	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	80	SER	N	124.41	0.30	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	80	-0.45 ± 0.34	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	79	0.05 ± 0.34	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	76	-1.28 ± 0.49	Should be applied

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 38%, i.e. 675 atoms were assigned a chemical shift out of a possible 1792. 0 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	231/600 (38%)	116/240 (48%)	59/244 (24%)	56/116 (48%)
Sidechain	428/1047 (41%)	286/677 (42%)	137/320 (43%)	5/50 (10%)
Aromatic	16/145 (11%)	16/69 (23%)	0/74 (0%)	0/2 (0%)
Overall	675/1792 (38%)	418/986 (42%)	196/638 (31%)	61/168 (36%)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	37	LEU	CG	32.63	21.37 – 32.19	5.4
1	A	61	ALA	HB1	0.12	0.14 – 2.58	-5.1
1	A	61	ALA	HB2	0.12	0.14 – 2.58	-5.1
1	A	61	ALA	HB3	0.12	0.14 – 2.58	-5.1

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: