wwPDB NMR Structure Validation Summary Report (i)

Nov 23, 2016 – 01:55 PM EST

PDB ID	:	4AAI
Title	:	THERMOSTABLE PROTEIN FROM HYPERTHERMOPHILIC VIRUS SSV-RH
Authors	:	Schlenker, C.; Goel, A.; Tripet, B.P.; Menon, S.; Lawrence, C.M.; Copie, V.
Deposited on	:	2011-12-02

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:

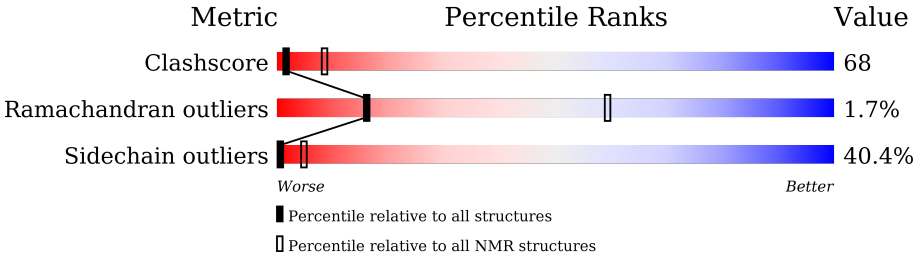
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20028320
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

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 23%.

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	73	14% 44% 22% 21%
1	B	73	12% 44% 23% 21%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:66, B:9-B:66 (116)	0.32	1

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 and 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2502 atoms, of which 1290 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ORF E73.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	73	1251	394	645	98	111	3	0
1	B	73	1251	394	645	98	111	3	0

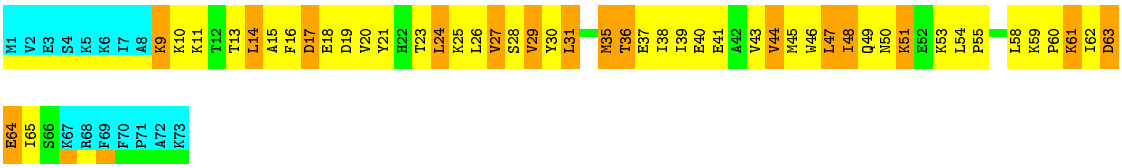
4 Residue-property plots (i)

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.

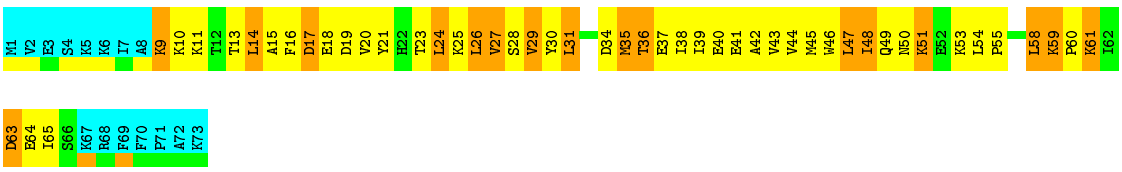
- Molecule 1: ORF E73

Chain A: 14% 44% 22% 21%



- Molecule 1: ORF E73

Chain B: 12% 44% 23% 21%

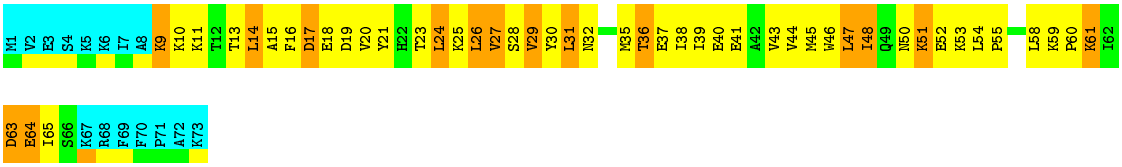


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

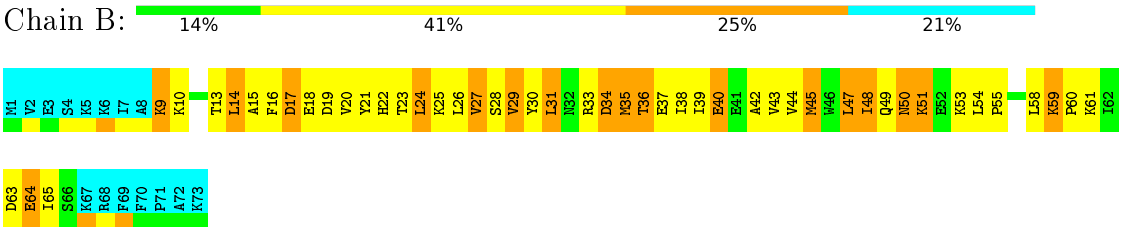
- Molecule 1: ORF E73

Chain A: 14% 45% 21% 21%



- Molecule 1: ORF E73

Chain B:


14%
41%
25%
21%
M1 Y2 E3 S4 K5 K6 I7 A8 K9 K10 T13 L14 A15 F16 D17 E18 D19 V20 Y21 H22 T23 L24 K25 L26 V27 S28 V29 Y30 L31 N32 R33 D34 M35 T36 E37 I38 I39 E40 E41 A42 V43 V44 M45 W46 L47 I48 Q49 N50 K51 E52 K53 L54 P55 L58 K59 P60 K61 I62
D63 E64 I65 S66 K67 R68 F69 F70 P71 A72 K73

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *ENERGY MINIMIZATION*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
CNS	refinement	
ARIA	structure solution	
CNS	structure solution	

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)	4aai_cs.cif
Number of chemical shift lists	1
Total number of shifts	575
Number of shifts mapped to atoms	501
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	74
Assignment completeness (well-defined parts)	23%

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	481	503	502	88±6
1	B	481	503	502	86±5
All	All	19240	20120	20080	2655

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 68.

5 of 507 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:ILE:HD13	1:B:14:LEU:HD13	1.11	1.15	8	12
1:A:14:LEU:HD22	1:B:36:THR:HG23	1.09	1.24	2	2
1:A:36:THR:HG23	1:B:14:LEU:HD22	1.07	1.25	12	8
1:A:14:LEU:HD11	1:B:14:LEU:HD11	1.03	1.29	1	7
1:A:14:LEU:HD23	1:B:36:THR:HG23	0.93	1.40	8	11

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	58/73 (79%)	51±1 (88±2%)	6±1 (10±2%)	1±0 (2±0%)	16 59
1	B	58/73 (79%)	52±1 (89±2%)	5±1 (9±2%)	1±0 (2±0%)	16 59
All	All	2320/2920 (79%)	2058 (89%)	222 (10%)	40 (2%)	16 59

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	17	ASP	20
1	A	17	ASP	20

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	56/69 (81%)	35±2 (62±4%)	21±2 (38±4%)	1 6
1	B	56/69 (81%)	32±3 (58±5%)	24±3 (42±5%)	0 3
All	All	2240/2760 (81%)	1336 (60%)	904 (40%)	0 5

5 of 90 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	61	LYS	20
1	A	29	VAL	20
1	A	31	LEU	20
1	A	13	THR	20
1	B	29	VAL	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 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 23% for the well-defined parts and 22% for the entire structure.

7.1 Chemical shift list 1

File name: 4aai_cs.cif

Chemical shift list name: *chemical_shift.str.csh*

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	575
Number of shifts mapped to atoms	501
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	74
Number of shift outliers (ShiftChecker)	49

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

- No matching atoms found in structure. First 5 (of 74) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	31	LEU	HN	8.417	-1.0	1
A	6	LYS	HD1	1.624	-1.0	1
A	33	ARG	HD1	2.769	-1.0	1
A	53	LYS	HN	0.12	-1.0	1
A	22	HIS	HN	8.851	-1.0	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
$^{13}C_{\alpha}$	47	-1.06 ± 0.21	Should be applied
$^{13}C_{\beta}$	45	-0.30 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	26	0.00 ± 0.00	None needed (< 0.5 ppm)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	69	-0.60 ± 0.24	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 23%, i.e. 351 atoms were assigned a chemical shift out of a possible 1524. 2 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	154/572 (27%)	40/228 (18%)	58/232 (25%)	56/112 (50%)
Sidechain	197/862 (23%)	122/500 (24%)	75/332 (23%)	0/30 (0%)
Aromatic	0/90 (0%)	0/46 (0%)	0/38 (0%)	0/6 (0%)
Overall	351/1524 (23%)	162/774 (21%)	133/602 (22%)	56/148 (38%)

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	60	PRO	C	0.05	184.42 – 169.02	-114.7
1	A	55	PRO	C	0.09	184.42 – 169.02	-114.7
1	A	71	PRO	C	0.18	184.42 – 169.02	-114.6
1	A	63	ASP	C	0.11	185.29 – 167.59	-99.6
1	A	66	SER	C	0.25	183.48 – 165.88	-99.1
1	A	56	ASN	C	0.11	184.41 – 166.21	-96.3
1	A	50	ASN	C	0.12	184.41 – 166.21	-96.3
1	A	62	ILE	C	0.09	185.63 – 166.13	-90.2
1	A	65	ILE	C	0.10	185.63 – 166.13	-90.1
1	A	48	ILE	C	0.12	185.63 – 166.13	-90.1
1	A	49	GLN	C	0.07	186.20 – 166.50	-89.5
1	A	64	GLU	C	0.07	186.83 – 167.03	-89.3
1	A	57	GLU	C	0.09	186.83 – 167.03	-89.3
1	A	52	GLU	C	0.09	186.83 – 167.03	-89.3
1	A	61	LYS	C	0.11	186.58 – 166.78	-89.2
1	A	53	LYS	C	0.19	186.58 – 166.78	-89.1
1	A	51	LYS	C	0.20	186.58 – 166.78	-89.1
1	A	59	LYS	C	0.22	186.58 – 166.78	-89.1
1	A	67	LYS	C	0.26	186.58 – 166.78	-89.1

Continued on next page...

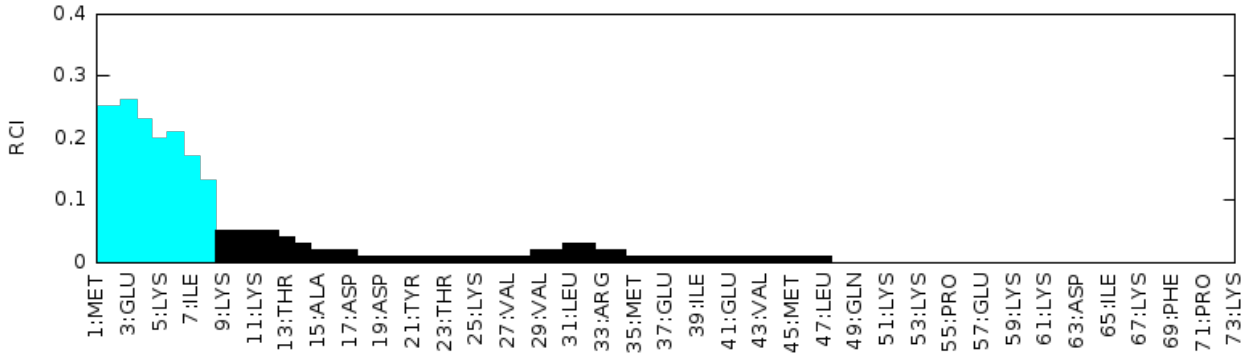
Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	73	LYS	C	0.30	186.58 – 166.78	-89.1
1	A	70	PHE	C	0.09	185.43 – 165.53	-88.1
1	A	69	PHE	C	0.12	185.43 – 165.53	-88.1
1	A	58	LEU	C	0.07	187.06 – 166.96	-88.0
1	A	54	LEU	C	0.08	187.06 – 166.96	-88.0
1	A	68	ARG	C	0.17	186.68 – 166.18	-86.0
1	A	72	ALA	C	0.04	188.57 – 166.97	-82.3
1	A	72	ALA	N	0.05	141.07 – 105.37	-34.5
1	A	57	GLU	N	0.04	138.24 – 103.14	-34.4
1	A	64	GLU	N	0.07	138.24 – 103.14	-34.4
1	A	52	GLU	N	0.15	138.24 – 103.14	-34.3
1	A	49	GLN	N	0.08	138.01 – 101.71	-33.0
1	A	68	ARG	N	0.12	139.40 – 102.20	-32.4
1	A	66	SER	N	0.03	134.24 – 98.34	-32.4
1	A	73	LYS	N	0.04	140.01 – 102.11	-31.9
1	A	59	LYS	N	0.05	140.01 – 102.11	-31.9
1	A	53	LYS	N	0.05	140.01 – 102.11	-31.9
1	A	51	LYS	N	0.06	140.01 – 102.11	-31.9
1	A	67	LYS	N	0.06	140.01 – 102.11	-31.9
1	A	61	LYS	N	0.07	140.01 – 102.11	-31.9
1	A	63	ASP	N	0.06	140.24 – 101.14	-30.9
1	A	54	LEU	N	0.04	141.70 – 102.00	-30.7
1	A	58	LEU	N	0.05	141.70 – 102.00	-30.7
1	A	50	ASN	N	0.04	139.12 – 98.82	-29.5
1	A	56	ASN	N	0.12	139.12 – 98.82	-29.5
1	A	69	PHE	N	0.07	141.27 – 99.77	-29.0
1	A	70	PHE	N	0.12	141.27 – 99.77	-29.0
1	A	62	ILE	N	0.05	143.11 – 99.91	-28.1
1	A	48	ILE	N	0.10	143.11 – 99.91	-28.1
1	A	65	ILE	N	0.10	143.11 – 99.91	-28.1

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.

Random coil index (RCI) for chain A:


RCI
0.4
0.3
0.2
0.1
0
1:MET
3:GLU
5:LYS
7:ILE
9:LYS
11:LYS
13:THR
15:ALA
17:ASP
19:ASP
21:TYR
23:THR
25:LYS
27:VAL
29:VAL
31:LEU
33:ARG
35:MET
37:GLU
39:ILE
41:GLU
43:VAL
45:MET
47:LEU
49:GLN
51:LYS
53:LYS
55:PRO
57:GLU
59:LYS
61:LYS
63:ASP
65:ILE
67:LYS
69:PHE
71:PRO
73:LYS