Full wwPDB NMR Structure Validation Report (i)

Feb 22, 2022 – 02:52 PM EST

PDB ID	:	1VA1
Title	:	Solution Structure of Transcription Factor Sp1 DNA Binding Domain (Zinc Finger 1)
Authors	:	Oka, S.; Shiraishi, Y.; Yoshida, T.; Ohkubo, T.; Sugiura, Y.; Kobayashi, Y.
Deposited on	:	2004-02-07

This is a Full wwPDB NMR Structure Validation 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:

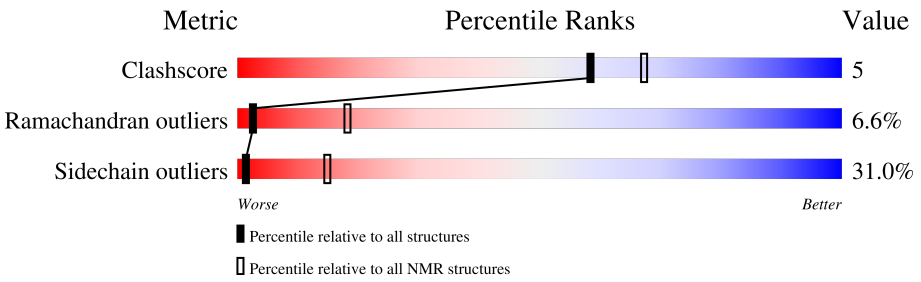
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	37	49% 24% 27%

2 Ensemble composition and analysis (i)

This entry contains 31 models. Model 1 is the overall representative, medoid model (most similar to other models). The authors have identified model 31 as representative, based on the following criterion: *minimized average structure*.

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:536-A:562 (27)	0.34	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 3 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 6, 7, 8, 10, 11, 14, 18, 21, 26, 27, 28, 31
2	9, 13, 15, 16, 22, 29, 30
3	12, 17, 19, 23
Single-model clusters	5; 20; 24; 25

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 604 atoms, of which 302 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription factor Sp1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	37	603	186	302	64	48	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P08047

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	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: Transcription factor Sp1

Chain A: 49% 24% 27%

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

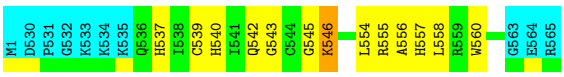
• Molecule 1: Transcription factor Sp1

Chain A: 43% 30% 27%

4.2.2 Score per residue for model 2

• Molecule 1: Transcription factor Sp1

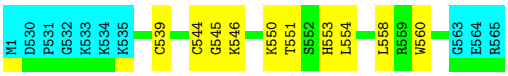
Chain A: 38% 32% • 27%



4.2.3 Score per residue for model 3

- Molecule 1: Transcription factor Sp1

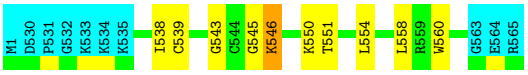
Chain A: 46% 27% 27%



4.2.4 Score per residue for model 4

- Molecule 1: Transcription factor Sp1

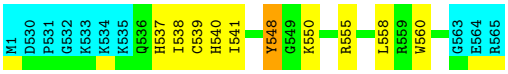
Chain A: 46% 24% • 27%



4.2.5 Score per residue for model 5

- Molecule 1: Transcription factor Sp1

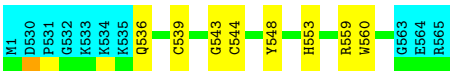
Chain A: 46% 24% • 27%



4.2.6 Score per residue for model 6

- Molecule 1: Transcription factor Sp1

Chain A: 51% 22% 27%



4.2.7 Score per residue for model 7

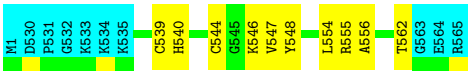
- Molecule 1: Transcription factor Sp1

Chain A: 49% 22% • 27%

4.2.8 Score per residue for model 8

- Molecule 1: Transcription factor Sp1

Chain A: 46% 27% 27%



4.2.9 Score per residue for model 9

- Molecule 1: Transcription factor Sp1

Chain A: 51% 19% • 27%

4.2.10 Score per residue for model 10

- Molecule 1: Transcription factor Sp1

Chain A: 43% 30% 27%

4.2.11 Score per residue for model 11

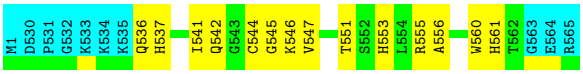
- Molecule 1: Transcription factor Sp1

Chain A: 38% 32% • 27%

4.2.12 Score per residue for model 12

- Molecule 1: Transcription factor Sp1

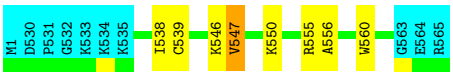
Chain A: 35% 38% 27%



4.2.13 Score per residue for model 13

- Molecule 1: Transcription factor Sp1

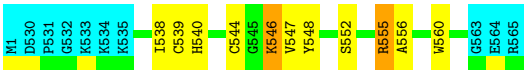
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Transcription factor Sp1

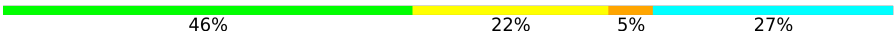
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Transcription factor Sp1

Chain A:



4.2.16 Score per residue for model 16

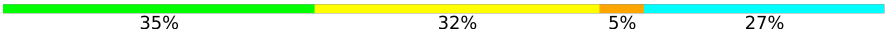
- Molecule 1: Transcription factor Sp1

Chain A:

4.2.17 Score per residue for model 17

- Molecule 1: Transcription factor Sp1

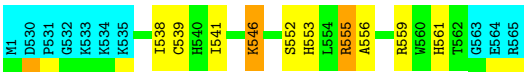
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Transcription factor Sp1

Chain A: 46% 22% 5% 27%



4.2.19 Score per residue for model 19

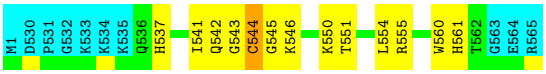
- Molecule 1: Transcription factor Sp1

Chain A: 32% 38% • 27%

4.2.20 Score per residue for model 20

- Molecule 1: Transcription factor Sp1

Chain A: 38% 32% • 27%



4.2.21 Score per residue for model 21

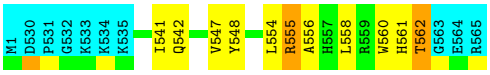
- Molecule 1: Transcription factor Sp1

Chain A: 38% 30% 5% 27%

4.2.22 Score per residue for model 22

- Molecule 1: Transcription factor Sp1

Chain A: 43% 24% 5% 27%



4.2.23 Score per residue for model 23

- Molecule 1: Transcription factor Sp1

Chain A: 49% 24% 27%

4.2.24 Score per residue for model 24

- Molecule 1: Transcription factor Sp1

Chain A: 38% 27% 8% 27%

4.2.25 Score per residue for model 25

- Molecule 1: Transcription factor Sp1

Chain A: 43% 27% • 27%

4.2.26 Score per residue for model 26

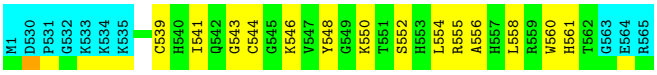
- Molecule 1: Transcription factor Sp1

Chain A: 49% 22% • 27%

4.2.27 Score per residue for model 27

- Molecule 1: Transcription factor Sp1

Chain A: 35% 38% 27%



4.2.28 Score per residue for model 28

- Molecule 1: Transcription factor Sp1

Chain A: 43% 24% 5% 27%

4.2.29 Score per residue for model 29

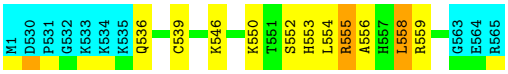
- Molecule 1: Transcription factor Sp1

Chain A: 38% 22% 14% 27%

4.2.30 Score per residue for model 30

- Molecule 1: Transcription factor Sp1

Chain A: 43% 24% 5% 27%



4.2.31 Score per residue for model 31

- Molecule 1: Transcription factor Sp1

Chain A: 32% 32% 8% 27%

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 31 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	structure solution	1.1
CNS	refinement	1.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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 (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	222	216	216	2±2
All	All	6913	6696	6696	67

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 5.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:537:HIS:NE2	1:A:551:THR:HG22	0.79	1.93	12	1
1:A:538:ILE:HG23	1:A:546:LYS:O	0.70	1.86	4	3
1:A:538:ILE:CD1	1:A:547:VAL:HG22	0.69	2.17	17	1
1:A:537:HIS:CE1	1:A:554:LEU:HD22	0.66	2.25	25	1
1:A:538:ILE:HD13	1:A:547:VAL:HG22	0.63	1.71	17	1
1:A:541:ILE:HD11	1:A:561:HIS:CE1	0.60	2.31	18	3
1:A:561:HIS:O	1:A:562:THR:HG22	0.60	1.97	29	1
1:A:541:ILE:HD13	1:A:561:HIS:CG	0.59	2.32	12	1
1:A:541:ILE:HG21	1:A:561:HIS:HB3	0.56	1.75	20	1
1:A:540:HIS:CE1	1:A:541:ILE:HD11	0.52	2.40	17	1
1:A:541:ILE:HD11	1:A:561:HIS:HE1	0.51	1.66	18	1
1:A:538:ILE:HG23	1:A:547:VAL:HG22	0.51	1.83	13	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:554:LEU:HD23	1:A:558:LEU:HD23	0.51	1.81	30	1
1:A:562:THR:O	1:A:562:THR:HG23	0.50	2.07	17	2
1:A:555:ARG:HG3	1:A:556:ALA:N	0.48	2.23	16	8
1:A:541:ILE:HD11	1:A:558:LEU:HD11	0.48	1.85	5	1
1:A:554:LEU:CD2	1:A:558:LEU:HD23	0.48	2.39	30	1
1:A:541:ILE:CD1	1:A:561:HIS:CE1	0.46	2.99	27	1
1:A:548:TYR:N	1:A:548:TYR:CD1	0.45	2.84	27	4
1:A:538:ILE:CG1	1:A:547:VAL:HG22	0.45	2.42	13	1
1:A:540:HIS:CE1	1:A:541:ILE:CD1	0.45	3.00	17	1
1:A:561:HIS:O	1:A:562:THR:CG2	0.45	2.64	29	1
1:A:537:HIS:CD2	1:A:554:LEU:CD1	0.44	2.99	31	1
1:A:548:TYR:CD1	1:A:548:TYR:N	0.44	2.85	29	1
1:A:541:ILE:HG22	1:A:542:GLN:N	0.43	2.27	22	2
1:A:554:LEU:HD23	1:A:558:LEU:CD1	0.43	2.42	27	1
1:A:555:ARG:CG	1:A:556:ALA:N	0.43	2.82	7	11
1:A:547:VAL:HG12	1:A:548:TYR:N	0.42	2.29	8	3
1:A:541:ILE:HG21	1:A:561:HIS:CB	0.42	2.44	20	1
1:A:554:LEU:HD23	1:A:558:LEU:HG	0.42	1.92	27	1
1:A:546:LYS:CE	1:A:557:HIS:NE2	0.41	2.82	2	2
1:A:547:VAL:O	1:A:548:TYR:CD1	0.41	2.73	26	1
1:A:538:ILE:HG22	1:A:539:CYS:N	0.41	2.31	5	1
1:A:556:ALA:O	1:A:559:ARG:CG	0.41	2.69	24	2
1:A:536:GLN:HB3	1:A:547:VAL:HG13	0.41	1.92	17	1
1:A:552:SER:O	1:A:555:ARG:HG2	0.40	2.16	27	1
1:A:537:HIS:O	1:A:548:TYR:CE1	0.40	2.75	29	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	27/37 (73%)	22±1 (83±5%)	3±2 (10±6%)	2±1 (7±5%)	2 18
All	All	837/1147 (73%)	697 (83%)	85 (10%)	55 (7%)	2 18

All 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	545	GLY	16
1	A	543	GLY	12
1	A	542	GLN	7
1	A	547	VAL	6
1	A	562	THR	5
1	A	544	CYS	4
1	A	536	GLN	2
1	A	539	CYS	2
1	A	546	LYS	1

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	23/31 (74%)	16±2 (69±7%)	7±2 (31±7%)	1 15
All	All	713/961 (74%)	492 (69%)	221 (31%)	1 15

All 18 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	546	LYS	27
1	A	560	TRP	25
1	A	539	CYS	22
1	A	544	CYS	18
1	A	550	LYS	15
1	A	553	HIS	15
1	A	555	ARG	13
1	A	554	LEU	12
1	A	558	LEU	11
1	A	559	ARG	11
1	A	540	HIS	11
1	A	542	GLN	8
1	A	536	GLN	7
1	A	551	THR	7
1	A	537	HIS	6
1	A	548	TYR	5
1	A	552	SER	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	562	THR	3

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 1 ligands modelled in this entry, 1 is 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 (i)

No chemical shift data were provided