Full wwPDB NMR Structure Validation Report (i)

Feb 28, 2022 – 06:20 PM EST

PDB ID	:	2EPR
Title	:	Solution structure of the secound zinc finger domain of Zinc finger protein 278
Authors	:	Tanabe, W.; Suzuki, S.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-03-30

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:

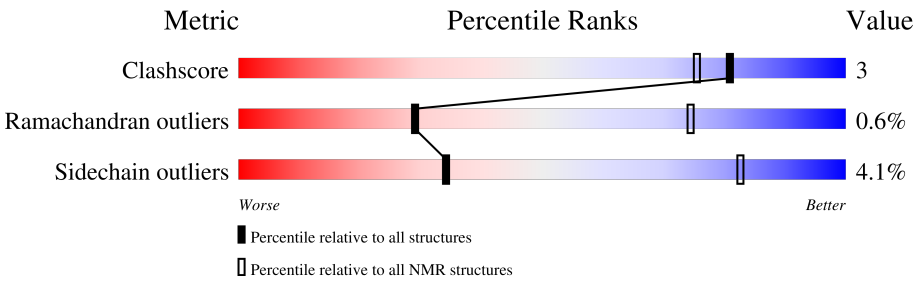
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	48	46% · 50%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 20 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:354-A:377 (24)	0.19	20

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 5 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 710 atoms, of which 349 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called POZ-, AT hook-, and zinc finger-containing protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	48	709	217	349	72	69	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	343	GLY	-	expression tag	UNP Q9HBE1
A	344	SER	-	expression tag	UNP Q9HBE1
A	345	SER	-	expression tag	UNP Q9HBE1
A	346	GLY	-	expression tag	UNP Q9HBE1
A	347	SER	-	expression tag	UNP Q9HBE1
A	348	SER	-	expression tag	UNP Q9HBE1
A	349	GLY	-	expression tag	UNP Q9HBE1
A	385	SER	-	expression tag	UNP Q9HBE1
A	386	GLY	-	expression tag	UNP Q9HBE1
A	387	PRO	-	expression tag	UNP Q9HBE1
A	388	SER	-	expression tag	UNP Q9HBE1
A	389	SER	-	expression tag	UNP Q9HBE1
A	390	GLY	-	expression tag	UNP Q9HBE1

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

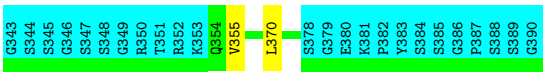
4 Residue-property plots (i)

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: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A:



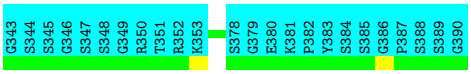
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

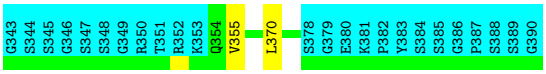
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

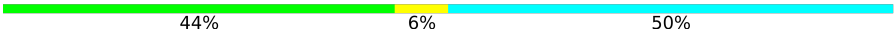
Chain A:

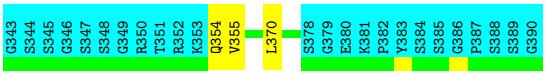


4.2.3 Score per residue for model 3

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A:

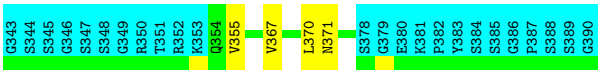




4.2.4 Score per residue for model 4

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A:



4.2.5 Score per residue for model 5

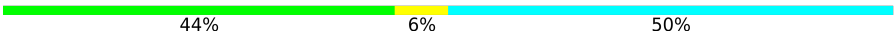
• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

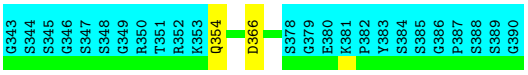
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 48% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 S376 H377 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.9 Score per residue for model 9

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 44% 6% 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 V355 K362 L370 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.10 Score per residue for model 10

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 48% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.11 Score per residue for model 11

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 48% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 K362 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.12 Score per residue for model 12

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 46% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 K362 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.13 Score per residue for model 13

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 48% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 C357 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.14 Score per residue for model 14

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 44% 6% 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 V355 E358 L370 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.15 Score per residue for model 15

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 46% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 V355 L370 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.16 Score per residue for model 16

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 40% 10% 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 Q354 V355 V367 L370 N371 R372 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.17 Score per residue for model 17

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

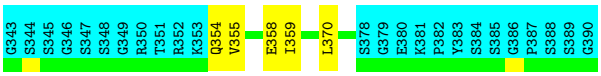
Chain A: 46% • 50%

G343 S344 S345 G346 S347 S348 G349 R350 T351 R352 K353 I359 H369 S378 G379 E380 K381 P382 Y383 S384 S385 G386 P387 S388 S389 G390

4.2.18 Score per residue for model 18

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 40% 10% 50%



4.2.19 Score per residue for model 19

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 40% 10% 50%

4.2.20 Score per residue for model 20 (medoid)

• Molecule 1: POZ-, AT hook-, and zinc finger-containing protein 1

Chain A: 40% 10% 50%

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, 20 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy, target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.1
CYANA	refinement	2.1

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	200	198	198	1±1
All	All	4020	3960	3960	21

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:355:VAL:HG11	1:A:370:LEU:HD22	0.72	1.61	2	10
1:A:372:ARG:HA	1:A:375:LEU:HD12	0.57	1.76	20	3
1:A:367:VAL:HG12	1:A:371:ASN:OD1	0.53	2.04	4	2
1:A:355:VAL:CG1	1:A:370:LEU:HD22	0.49	2.36	19	6

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	24/48 (50%)	22±1 (91±5%)	2±1 (9±4%)	0±0 (1±1%)	29	74
All	All	480/960 (50%)	436 (91%)	41 (9%)	3 (1%)	29	74

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	A	359	ILE	2
1	A	357	CYS	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	22/40 (55%)	21±1 (96±3%)	1±1 (4±3%)	34	82
All	All	440/800 (55%)	422 (96%)	18 (4%)	34	82

All 8 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	354	GLN	5
1	A	362	LYS	3
1	A	358	GLU	3
1	A	366	ASP	2
1	A	372	ARG	2
1	A	376	SER	1
1	A	369	HIS	1
1	A	374	LYS	1

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