Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:24 pm BST

PDB ID	:	2KAK
Title	:	Solution structure of the beta-E-domain of wheat Ec-1 metallothionein
Authors	:	Peroza, E.A.; Schmucki, R.; Guntert, P.; Freisinger, E.; Zerbe, O.
Deposited on	:	2008-11-06

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 (1)) were used in the production of this report:

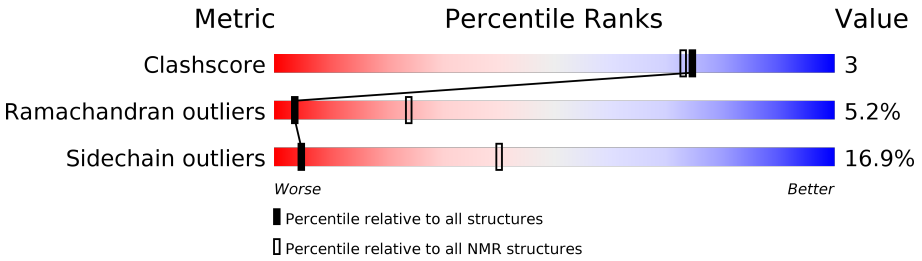
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 54%.

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	53	74% 11% • 13%

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:32-A:77 (46)	0.43	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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 637 atoms, of which 291 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called EC protein I/II.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	53	633	187	291	73	71	11	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	82	PRO	-	expression tag	UNP P30569
A	83	GLY	-	expression tag	UNP P30569

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	4	4	4

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: EC protein I/II

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 (medoid)

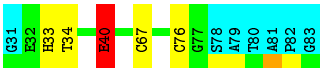
- Molecule 1: EC protein I/II

Chain A:

4.2.2 Score per residue for model 2

- Molecule 1: EC protein I/II

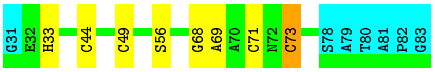
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: EC protein I/II

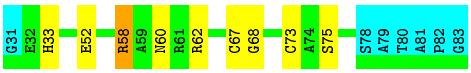
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: EC protein I/II

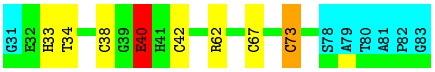
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: EC protein I/II

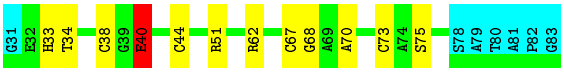
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: EC protein I/II

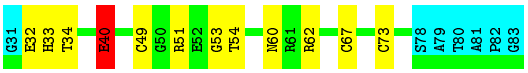
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: EC protein I/II

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: EC protein I/II

Chain A: 66% 17% • • 13%

G31 E32 H33 E40 H41 C42 T54 R62 C67 G68 A69 A70 G73 A74 S75 S78 A79 T80 A81 P82 G83

4.2.9 Score per residue for model 9

- Molecule 1: EC protein I/II

Chain A: 68% 19% 13%

G31 E32 H33 T34 G37 C42 G43 C44 R61 R62 C71 N72 C73 C76 G77 S78 A79 T80 A81 P82 G83

4.2.10 Score per residue for model 10

- Molecule 1: EC protein I/II

Chain A: 72% 13% • 13%

G31 E32 H33 C38 G39 E40 R51 N60 R61 C67 C73 S78 A79 T80 A81 P82 G83

4.2.11 Score per residue for model 11

- Molecule 1: EC protein I/II

Chain A: 66% 17% • • 13%

G31 E32 H33 T34 E40 C49 R61 C67 G68 A69 A70 S75 C76 G77 S78 A79 T80 A81 P82 G83

4.2.12 Score per residue for model 12

- Molecule 1: EC protein I/II

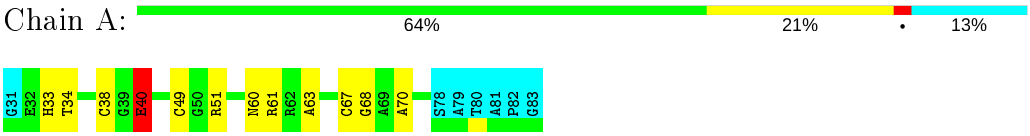
Chain A: 60% 23% • • 13%

G31 E32 H33 T34 T35 C36 E40 H41 C42 R51 E52 G53 G57 A63 C67 C73 A74 S75 C76 G77 S78 A79 T80 A81 P82 G83

4.2.13 Score per residue for model 13

- Molecule 1: EC protein I/II

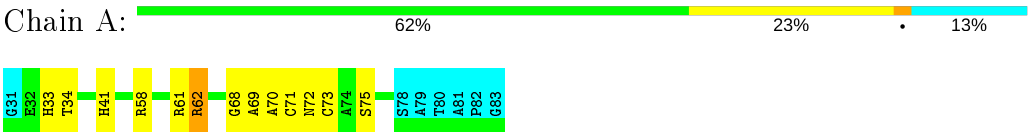
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: EC protein I/II

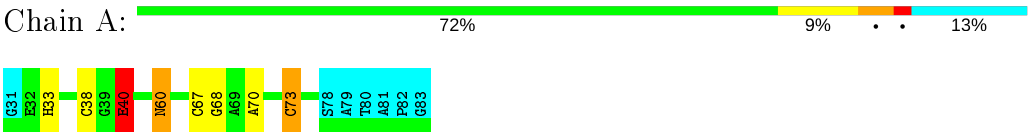
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: EC protein I/II

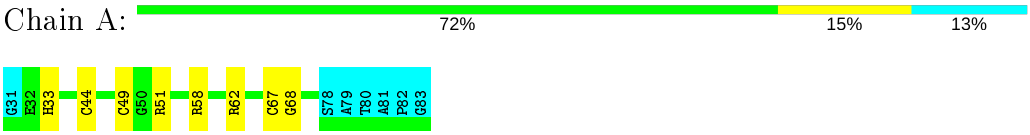
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: EC protein I/II

Chain A:



4.2.17 Score per residue for model 17

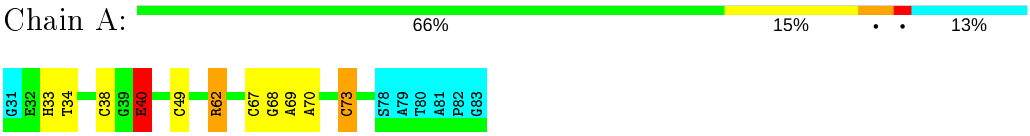
- Molecule 1: EC protein I/II

Chain A:

4.2.18 Score per residue for model 18

- Molecule 1: EC protein I/II

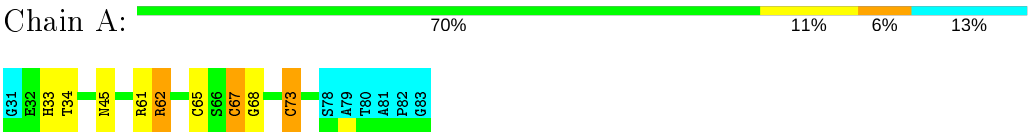
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: EC protein I/II

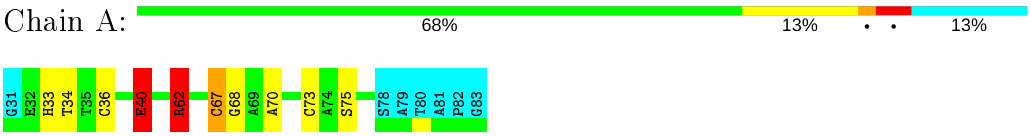
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: EC protein I/II

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CYANA	refinement	3.0
CYANA	structure solution	3.0
OPALp	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	311
Number of shifts mapped to atoms	311
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	54%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	304	256	256	2±1
All	All	6160	5120	5120	36

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:40:GLU:OE2	1:A:67:CYS:SG	0.56	2.64	8	11
1:A:40:GLU:OE1	1:A:67:CYS:SG	0.52	2.67	18	11
1:A:71:CYS:SG	1:A:76:CYS:SG	0.45	3.15	9	1
1:A:38:CYS:SG	1:A:40:GLU:OE1	0.44	2.76	5	6
1:A:34:THR:HG22	1:A:45:ASN:HD22	0.44	1.72	1	2
1:A:40:GLU:OE1	1:A:42:CYS:HB2	0.44	2.12	8	1
1:A:42:CYS:SG	1:A:44:CYS:SG	0.44	3.16	9	1
1:A:40:GLU:OE2	1:A:42:CYS:HB2	0.43	2.14	5	1
1:A:71:CYS:SG	1:A:73:CYS:HB3	0.42	2.54	3	1
1:A:65:CYS:SG	1:A:67:CYS:O	0.42	2.78	19	1

5.2 Torsion angles (i)

5.2.1 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	46/53 (87%)	37±2 (81±5%)	6±2 (14±4%)	2±1 (5±3%)	4 24
All	All	920/1060 (87%)	746 (81%)	126 (14%)	48 (5%)	4 24

All 14 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	68	GLY	13
1	A	70	ALA	10
1	A	60	ASN	5
1	A	62	ARG	4
1	A	53	GLY	2
1	A	67	CYS	2
1	A	58	ARG	2
1	A	54	THR	2
1	A	69	ALA	2
1	A	63	ALA	2
1	A	57	GLY	1
1	A	61	ARG	1
1	A	37	GLY	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	32	GLU	1

5.2.2 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	32/35 (91%)	27±2 (83±5%)	5±2 (17±5%)	5 40
All	All	640/700 (91%)	532 (83%)	108 (17%)	5 40

All 20 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	33	HIS	20
1	A	40	GLU	12
1	A	62	ARG	11
1	A	34	THR	11
1	A	73	CYS	10
1	A	75	SER	8
1	A	49	CYS	7
1	A	51	ARG	5
1	A	41	HIS	4
1	A	44	CYS	3
1	A	61	ARG	3
1	A	67	CYS	3
1	A	60	ASN	2
1	A	52	GLU	2
1	A	32	GLU	2
1	A	42	CYS	1
1	A	54	THR	1
1	A	56	SER	1
1	A	58	ARG	1
1	A	72	ASN	1

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 54% for the well-defined parts and 54% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	311
Number of shifts mapped to atoms	311
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	50	-1.81 ± 0.75	Should be applied

6.1.3 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 54%, i.e. 243 atoms were assigned a chemical shift out of a possible 451. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	134/226 (59%)	90/90 (100%)	0/92 (0%)	44/44 (100%)
Sidechain	105/211 (50%)	105/128 (82%)	0/67 (0%)	0/16 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	4/14 (29%)	4/8 (50%)	0/4 (0%)	0/2 (0%)
Overall	243/451 (54%)	199/226 (88%)	0/163 (0%)	44/62 (71%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 54%, i.e. 273 atoms were assigned a chemical shift out of a possible 504. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	153/259 (59%)	103/103 (100%)	0/106 (0%)	50/50 (100%)
Sidechain	116/231 (50%)	116/140 (83%)	0/75 (0%)	0/16 (0%)
Aromatic	4/14 (29%)	4/8 (50%)	0/4 (0%)	0/2 (0%)
Overall	273/504 (54%)	223/251 (89%)	0/185 (0%)	50/68 (74%)

6.1.4 Statistically unusual chemical shifts (i)

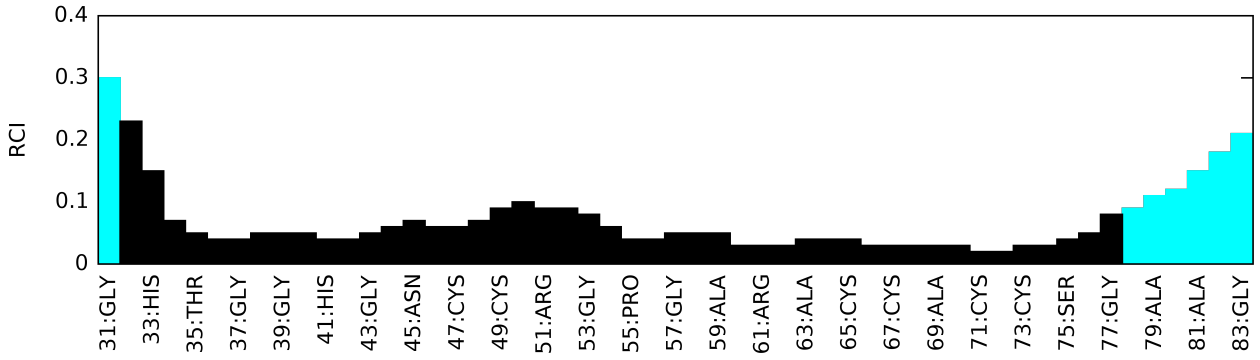
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	62	ARG	HG2	3.51	2.92 – 0.22	7.2
1	A	62	ARG	HG3	3.51	3.00 – 0.10	6.8
1	A	82	PRO	HG3	3.69	3.56 – 0.26	5.4
1	A	82	PRO	HG2	3.58	3.48 – 0.38	5.3

6.1.5 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
31:GLY
33:HIS
35:THR
37:GLY
39:GLY
41:HIS
43:GLY
45:ASN
47:CYS
49:CYS
51:ARG
53:GLY
55:PRO
57:GLY
59:ALA
61:ARG
63:ALA
65:CYS
67:CYS
69:ALA
71:CYS
73:CYS
75:SER
77:GLY
79:ALA
81:ALA
83:GLY