wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 10:42 pm BST

PDB ID	:	2KV1
Title	:	Insights into Function, Catalytic Mechanism and Fold Evolution of Mouse Selenoprotein Methionine Sulfoxide Reductase B1 through Structural Analysis
Authors	:	Aachmann, F.L.; Sal, L.S.; Kim, H.Y.; Gladyshev, V.N.; Dikiy, A.
Deposited on	:	2010-03-04

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 following versions of software and data (see references (1)) were used in the production of this report:

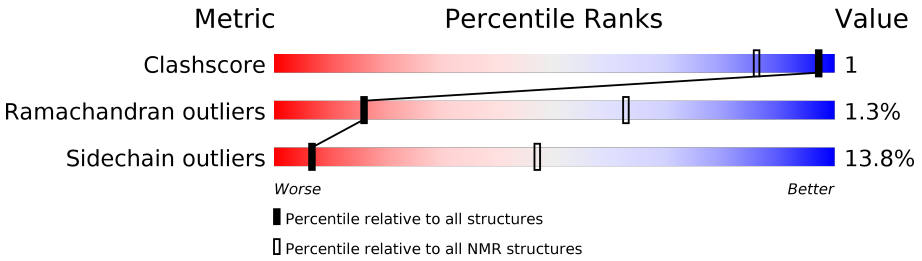
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 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	124	63% 9% 28%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:18-A:106 (89)	1.07	9

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 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1877 atoms, of which 903 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Methionine-R-sulfoxide reductase B1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	124	1876	615	903	175	175	8	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	CYS	SEC	ENGINEERED MUTATION	UNP Q9JLC3
A	117	LEU	-	EXPRESSION TAG	UNP Q9JLC3
A	118	GLU	-	EXPRESSION TAG	UNP Q9JLC3
A	119	HIS	-	EXPRESSION TAG	UNP Q9JLC3
A	120	HIS	-	EXPRESSION TAG	UNP Q9JLC3
A	121	HIS	-	EXPRESSION TAG	UNP Q9JLC3
A	122	HIS	-	EXPRESSION TAG	UNP Q9JLC3
A	123	HIS	-	EXPRESSION TAG	UNP Q9JLC3
A	124	HIS	-	EXPRESSION TAG	UNP Q9JLC3

- 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 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: Methionine-R-sulfoxide reductase B1

Chain A: 63% 9% 28%

M1 S2 P3 C4 S5 P6 F7 G8 G9 E10 V11 F12 Q13 N14 H15 F16 E17 Y21 S35 I50 K57 K61 M62 R63 G77 R89 R93 L101 K106 G107 K108 E109 A110 A111 A112 S113 Q114 G115 H116 L117 E118 H119 H120 H121 H122 H123 H124

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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: Methionine-R-sulfoxide reductase B1

Chain A: 56% 15% • 28%

M1 S2 P3 C4 S5 P6 F7 G8 G9 E10 V11 F12 Q13 N14 H15 F16 E17 V20 Y21 L30 F31 S35 K36 H39 T47 E48 T49 I50 K57 M62 R63 G77 R89 R93 F94 K102 K106 G107 K108 E109 A110 A111 A112 S113 Q114 G115 H116 L117

E118 H119 H120 H121 H122 H123 H124

5 Refinement protocol and experimental data overview (i)

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

Of the 96 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
AMBER	geometry optimization	10
CYANA	refinement	2.1

No chemical shift data was provided. 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	688	665	665	1±1
All	All	13780	13300	13300	15

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 1.

5 of 10 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:50:ILE:HD12	1:A:50:ILE:H	0.60	1.57	7	2
1:A:51:HIS:CD2	1:A:55:VAL:HG21	0.58	2.33	10	1
1:A:50:ILE:H	1:A:50:ILE:HD12	0.55	1.61	3	4
1:A:20:VAL:HG22	1:A:21:TYR:H	0.52	1.64	9	2
1:A:52:PRO:O	1:A:55:VAL:HG22	0.48	2.09	10	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	89/124 (72%)	78±3 (88±3%)	10±3 (11±3%)	1±1 (1±1%)	16 63
All	All	1780/2480 (72%)	1558 (88%)	199 (11%)	23 (1%)	16 63

5 of 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	77	GLY	13
1	A	92	SER	3
1	A	86	GLY	1
1	A	88	LYS	1
1	A	89	ARG	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	78/106 (74%)	67±3 (86±4%)	11±3 (14±4%)	7 47
All	All	1560/2120 (74%)	1344 (86%)	216 (14%)	7 47

5 of 44 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	57	LYS	20
1	A	101	LEU	14
1	A	35	SER	12
1	A	61	LYS	11
1	A	106	LYS	11

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 1 ligands modelled in this entry, 1 is 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)

No chemical shift data were provided