wwPDB NMR Structure Validation Summary Report

Sep 18, 2023 – 10:04 PM EDT

PDB ID	:	1WR1
Title	:	The complex structure of Dsk2p UBA with ubiquitin
Authors	:	Ohno, A.; Jee, J.G.; Fujiwara, K.; Tenno, T.; Goda, N.; Tochio, H.; Hiroaki, H.; kobayashi, H.; Shirakawa, M.
Deposited on	:	2004-10-08

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

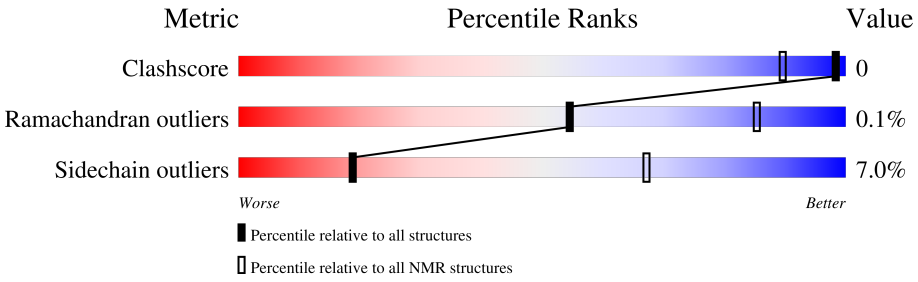
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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	76	86% 9% 5%
2	B	58	71% 9% 21%

2 Ensemble composition and analysis

This entry contains 20 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. The authors have identified model 16 as representative, based on the following criterion: *closest to the average*. No medoid model was calculated.

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:1-A:72, B:328-B:373 (118)	Not calculated	Not calculated

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: Inconsistent models in file

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2072 atoms, of which 1036 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms						Trace
1	A	76	Total	C	H	N	O	S	0
			1226	375	625	105	120	1	

- Molecule 2 is a protein called Ubiquitin-like protein DSK2.

Mol	Chain	Residues	Atoms						Trace
2	B	58	Total	C	H	N	O	S	0
			846	265	411	81	88	1	

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	316	PRO	-	cloning artifact	UNP P48510
B	317	GLY	-	cloning artifact	UNP P48510
B	318	ILE	-	cloning artifact	UNP P48510
B	319	SER	-	cloning artifact	UNP P48510
B	320	GLY	-	cloning artifact	UNP P48510
B	321	GLY	-	cloning artifact	UNP P48510
B	322	GLY	-	cloning artifact	UNP P48510
B	323	GLY	-	cloning artifact	UNP P48510
B	324	GLY	-	cloning artifact	UNP P48510
B	325	ILE	-	cloning artifact	UNP P48510
B	326	LEU	-	cloning artifact	UNP P48510
B	327	ASP	-	cloning artifact	UNP P48510

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

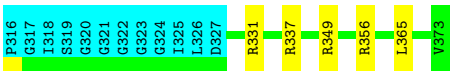
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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: Ubiquitin

Chain A:

• Molecule 2: Ubiquitin-like protein DSK2

Chain B:



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

The representative model is number 16. Colouring as in section 4.1 above.

• Molecule 1: Ubiquitin

Chain A:

• Molecule 2: Ubiquitin-like protein DSK2

Chain B:

5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 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
CYANA	structure solution	2.0.17
Amber	refinement	7

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.64±0.00	0±0/579 (0.0± 0.0%)	0.98±0.03	2±1/780 (0.2± 0.2%)
2	B	0.74±0.01	0±0/371 (0.0± 0.0%)	1.08±0.03	3±1/495 (0.5± 0.2%)
All	All	0.68	0/18999 (0.0%)	1.02	92/25498 (0.4%)

There are no bond-length outliers.

5 of 12 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	356	ARG	NE-CZ-NH1	7.45	124.02	120.30	9	11
1	A	42	ARG	NE-CZ-NH1	7.15	123.87	120.30	11	10
1	A	54	ARG	NE-CZ-NH1	7.13	123.86	120.30	14	12
2	B	337	ARG	NE-CZ-NH1	7.10	123.85	120.30	16	8
2	B	349	ARG	NE-CZ-NH1	6.98	123.79	120.30	6	19

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

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	573	595	595	0±1
2	B	366	342	342	0±0
All	All	18779	18740	18736	7

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:VAL:HG21	1:A:56:LEU:HD21	0.48	1.84	12	3
1:A:15:LEU:HD22	1:A:29:LYS:HB3	0.45	1.88	4	2
2:B:362:GLN:HE21	2:B:365:LEU:HD13	0.41	1.75	15	1
2:B:362:GLN:HE21	2:B:365:LEU:HD21	0.41	1.75	12	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	71/76 (93%)	69±1 (97±2%)	2±1 (3±2%)	0±0 (0±0%)	100	100
2	B	45/58 (78%)	43±1 (96±3%)	2±1 (4±3%)	0±0 (0±1%)	50	82
All	All	2320/2680 (87%)	2241 (97%)	77 (3%)	2 (0%)	54	85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
2	B	343	GLY	2

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	67/69 (97%)	61±1 (91±2%)	6±1 (9±2%)	13	60
2	B	38/44 (86%)	37±1 (96±3%)	1±1 (4±3%)	37	85
All	All	2099/2260 (93%)	1952 (93%)	147 (7%)	19	67

5 of 27 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	5	VAL	20
1	A	63	LYS	20
1	A	67	LEU	15
1	A	69	LEU	15
1	A	6	LYS	13

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

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

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided