wwPDB NMR Structure Validation Summary Report (i)

Mar 6, 2022 – 11:19 PM EST

PDB ID	:	2YQF
Title	:	Solution structure of the death domain of Ankyrin-1
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 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 (i)) were used in the production of this report:

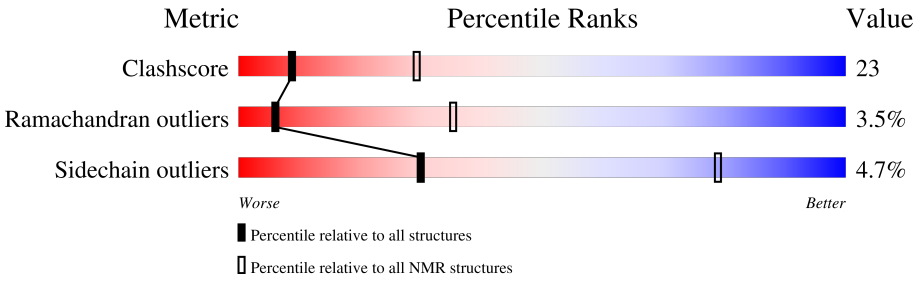
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	111	45% 25% • 28%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 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:1405-A:1484 (80)	0.18	4

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 4 clusters and 7 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1694 atoms, of which 841 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ankyrin-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	111	1694	520	841	156	173	4	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1387	GLY	-	expression tag	UNP P16157
A	1388	SER	-	expression tag	UNP P16157
A	1389	SER	-	expression tag	UNP P16157
A	1390	GLY	-	expression tag	UNP P16157
A	1391	SER	-	expression tag	UNP P16157
A	1392	SER	-	expression tag	UNP P16157
A	1393	GLY	-	expression tag	UNP P16157

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: Ankyrin-1

Chain A: 45% 25% • 28%

G1387 S1388 S1389 G1390 S1391 S1392 G1393 P1394 G1395 S1396 L1397 S1398 G1399 T1400 E1401 Q1402 A1403 E1404 M1405 K1406 V1409 H1413 L1414 G1415 L1416 S1417 W1418 A1419 E1424 L1425 I1432 N1433 R1434 N1439 P1440 L1443 Q1446 S1447 W1454 N1461 A1462 N1463 M1464 L1467 Y1468 L1471 I1474

M1482 L1483 E1484 G1485 S1486 G1487 R1488 Q1489 S1490 R1491 N1492 L1493 K1494 P1495 D1496 R1497

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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Ankyrin-1

Chain A: 50% 19% • 28%

G1387 S1388 S1389 G1390 S1391 S1392 G1393 P1394 G1395 S1396 L1397 S1398 G1399 T1400 E1401 Q1402 A1403 E1404 V1409 H1413 L1414 G1415 L1416 S1417 W1418 A1419 L1425 I1432 N1433 R1434 N1439 P1440 S1447 W1454 N1461 A1462 N1463 M1464 E1465 M1466 L1467 Y1468 L1471 I1474 G1485 S1486 G1487 R1488 Q1489

S1490 R1491 N1492 L1493 K1494 P1495 D1496 R1497

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	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	641	642	642	30±5
All	All	12820	12840	12840	602

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 23.

5 of 139 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1409:VAL:HG11	1:A:1483:LEU:HD21	0.81	1.51	13	11
1:A:1422:ALA:HB2	1:A:1450:LEU:HD11	0.75	1.59	14	1
1:A:1425:LEU:HD23	1:A:1454:TRP:CD1	0.73	2.18	18	2
1:A:1458:GLU:HB2	1:A:1462:ALA:HB2	0.71	1.62	13	7
1:A:1468:TYR:HA	1:A:1471:LEU:HD12	0.71	1.62	15	7

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	80/111 (72%)	64±3 (80±3%)	13±3 (17±3%)	3±1 (4±1%)	6 35
All	All	1600/2220 (72%)	1277 (80%)	267 (17%)	56 (4%)	6 35

5 of 6 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	1415	GLY	19
1	A	1417	SER	17
1	A	1432	ILE	8
1	A	1462	ALA	6
1	A	1474	ILE	5

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	71/94 (76%)	68±1 (95±1%)	3±1 (5±1%)	30 79
All	All	1420/1880 (76%)	1353 (95%)	67 (5%)	30 79

5 of 17 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	1433	ASN	14
1	A	1406	LYS	11
1	A	1424	GLU	11
1	A	1416	LEU	6
1	A	1461	ASN	5

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)

There are no ligands in this entry.

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