wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 09:10 PM BST

PDB ID	:	2HM8
Title	:	Solution Structure of the C-terminal MA-3 domain of Pdcd4
Authors	:	Waters, L.C.; Veverka, V.; Bohm, M.; Muskett, F.W.; Choong, P.T.; Klempnauer, K.H.; Carr, M.D.
Deposited on	:	2006-07-11

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
http://wwpdb.org/validation/2016/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:

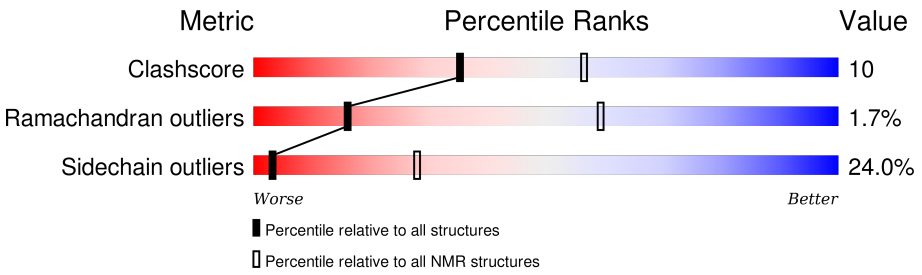
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	136	57% 27% • 13%

2 Ensemble composition and analysis (i)

This entry contains 49 models. Model 33 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:327-A:415, A:421-A:449 (118)	0.40	33

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 8 clusters and 5 single-model clusters were found.

Cluster number	Models
1	3, 4, 6, 7, 18, 19, 29, 35, 40, 45, 48
2	1, 10, 14, 22, 27, 32, 33, 34, 37, 43
3	2, 5, 16, 21, 25, 30, 39, 42
4	11, 17, 20, 26, 49
5	13, 28, 31
6	9, 24, 38
7	12, 15
8	8, 36
Single-model clusters	23; 41; 44; 46; 47

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2197 atoms, of which 1111 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Pdcd4 C-terminal MA-3 domain.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	136	2197	694	1111	177	208	7	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	314	GLY	-	INSERTION	UNP Q61823
A	315	PRO	-	INSERTION	UNP Q61823
A	316	LEU	-	INSERTION	UNP Q61823
A	317	GLY	-	INSERTION	UNP Q61823
A	318	SER	-	INSERTION	UNP Q61823

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: Pdcd4 C-terminal MA-3 domain

Chain A: 57% 27% • 13%

G314 P315 L316 G317 S318 G319 G320 Q321 Q322 P323 V324 N325 H326 L327 I334 L339 L340 S341 S345 H349 C350 L351 L354 E355 V356 F359 L363 A367 M370 V371 L372 E377 F380 I383 L384 L387 K388 S389 L390 W391 S394 T395 I396 T397 M401

K402 I412 P413 N416 L417 D418 V419 P420 H421 S422 Y423 S424 V425 L426 F429 V430 C433 I438 L443 R444 S449

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

The representative model is number 33. Colouring as in section 4.1 above.

- Molecule 1: Pdcd4 C-terminal MA-3 domain

Chain A: 56% 29% • 13%

G314 P315 L316 G317 S318 G319 G320 Q321 Q322 P323 V324 N325 H326 L327 I331 D332 E337 S341 E346 C350 L351 L354 E355 V356 H360 L363 A367 M370 V371 L372 E373 S374 F380 I383 L384 L387 W391 K392 S393 S394 T395 M401 K402 E406

R407 I408 I412 P413 D414 I415 N416 L417 D418 V419 P420 H421 S422 Y423 S424 V425 L426 F429 V430 C433 I438 K441 R444 D445 L446 S449

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 49 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
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.

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	961	992	966	20±5
All	All	47089	48608	47334	982

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 10.

5 of 239 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:331:ILE:HG23	1:A:363:LEU:HD12	0.80	1.53	25	10
1:A:351:LEU:HD12	1:A:363:LEU:HD13	0.78	1.55	43	4
1:A:367:ALA:HB1	1:A:383:ILE:HG23	0.78	1.55	25	49
1:A:351:LEU:HD13	1:A:363:LEU:HD13	0.72	1.62	23	6
1:A:364:VAL:HG12	1:A:368:ILE:HD11	0.72	1.61	23	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	117/136 (86%)	108±2 (92±2%)	7±2 (6±2%)	2±1 (2±1%)	16 59
All	All	5733/6664 (86%)	5270 (92%)	363 (6%)	100 (2%)	16 59

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	359	PHE	29
1	A	413	PRO	25
1	A	374	SER	17
1	A	421	HIS	13
1	A	377	GLU	11

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	110/124 (89%)	84±3 (76±3%)	26±3 (24±3%)	3 28
All	All	5390/6076 (89%)	4097 (76%)	1293 (24%)	3 28

5 of 79 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	412	ILE	47
1	A	384	LEU	46
1	A	327	LEU	41
1	A	354	LEU	41
1	A	370	MET	38

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 carbohydrates 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