wwPDB NMR Structure Validation Summary Report

Apr 20, 2024 – 09:53 PM EDT

PDB ID	:	2KWK
Title	:	Solution structures of the double PHD fingers of human transcriptional protein DPF3b bound to a H3 peptide wild type
Authors	:	Zeng, L.; Zhang, Q.; Li, S.; Plotnikov, A.N.; Walsh, M.J.; Zhou, M.
Deposited on	:	2010-04-12

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:

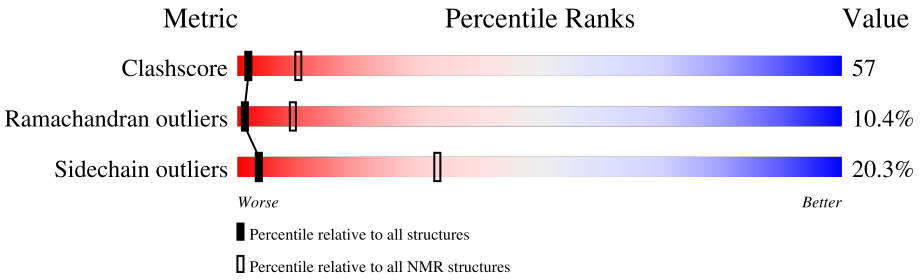
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	B	20	10% 10% 80%
2	A	114	20% 56% 11% • 11%

2 Ensemble composition and analysis

This entry contains 20 models. Model 7 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	B:1-B:4, A:260-A:274, A:278-A:363 (105)	0.45	7

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2021 atoms, of which 984 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone peptide.

Mol	Chain	Residues	Atoms					Trace
1	B	20	Total	C	H	N	O	0
			327	91	174	35	27	

- Molecule 2 is a protein called Zinc finger protein DPF3.

Mol	Chain	Residues	Atoms						Trace
2	A	114	Total	C	H	N	O	S	0
			1690	542	810	147	174	17	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	259	GLY	-	expression tag	UNP Q92784
A	260	SER	-	expression tag	UNP Q92784

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
3	A	4	Total	Zn
			4	4

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

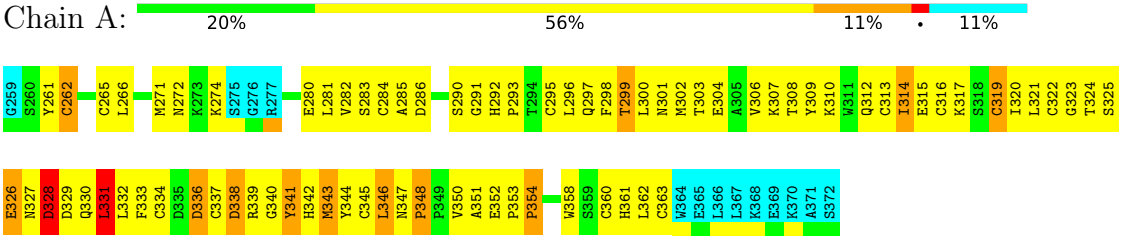
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for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

Chain B:

• Molecule 2: Zinc finger protein DPF3

Chain A:



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

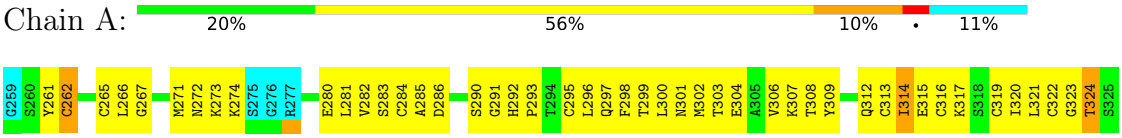
The representative model is number 7. Colouring as in section 4.1 above.

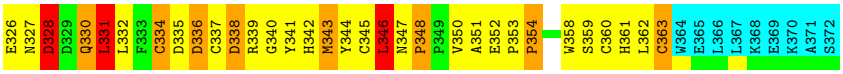
• Molecule 1: Histone peptide

Chain B:

• Molecule 2: Zinc finger protein DPF3

Chain A:




E326
N327
D328
D329
Q330
L331
L332
F333
C334
D335
D336
C337
D338
R339
G340
Y341
H342
M343
Y344
C345
L346
N347
P348
P349
V350
A351
E352
P353
P354
W358
S359
C360
H361
L362
C363
W364
E365
L366
L367
K368
E369
K370
A371
S372

5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing, torsion angle dynamics.*

Of the 200 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
ARIA	refinement	2.2
CNS	structure solution	1.2

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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

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	B	32	40	40	8±3
2	A	777	704	703	89±8
All	All	16260	14880	14860	1788

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 57.

5 of 329 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:346:LEU:HD12	2:A:350:VAL:HG23	0.82	1.48	2	19
2:A:322:CYS:SG	2:A:324:THR:HG22	0.79	2.18	8	1
2:A:298:PHE:CE2	2:A:306:VAL:HG11	0.78	2.13	14	20
2:A:284:CYS:HB2	2:A:313:CYS:N	0.77	1.94	4	2
2:A:314:ILE:HG22	2:A:315:GLU:N	0.77	1.94	9	18

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	B	3/20 (15%)	2±0 (72±16%)	0±1 (10±19%)	1±0 (18±17%)	0 3
2	A	101/114 (89%)	69±2 (69±2%)	22±3 (21±3%)	10±1 (10±1%)	1 9
All	All	2080/2680 (78%)	1428 (69%)	436 (21%)	216 (10%)	1 9

5 of 23 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	314	ILE	20
2	A	326	GLU	20
2	A	331	LEU	20
2	A	343	MET	20
2	A	348	PRO	20

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	B	3/15 (20%)	1±1 (45±24%)	2±1 (55±24%)	0 1
2	A	91/101 (90%)	74±3 (81±3%)	17±3 (19±3%)	4 35
All	All	1880/2320 (81%)	1498 (80%)	382 (20%)	3 33

5 of 52 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)
2	A	262	CYS	20
2	A	299	THR	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	A	309	TYR	20
2	A	336	ASP	20
2	A	346	LEU	20

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

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

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