Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 10:01 PM BST

PDB ID	:	2K1J
Title	:	Plan homeodomain finger of tumour supressor ING4
Authors	:	Palacios, A.; Garcia, P.; Padro, D.; Lopez-Hernandez, E.; Blanco, F.J.
Deposited on	:	2008-03-05

This is a Full wwPDB NMR Structure Validation 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:

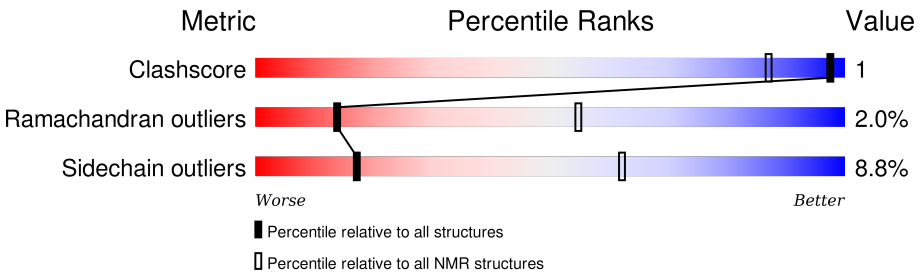
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	63	71% 8% 21%

2 Ensemble composition and analysis (i)

This entry contains 25 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: *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:196-A:245 (50)	0.18	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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 24, 25
2	5, 16, 23

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 988 atoms, of which 479 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Inhibitor of growth protein 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	63	986	318	479	87	92	10	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	187	MET	-	INITIATING METHIONINE	UNP Q9UNL4

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

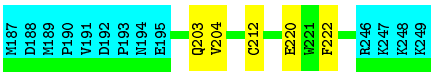
4 Residue-property plots (i)

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: Inhibitor of growth protein 4

Chain A:



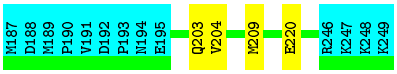
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Inhibitor of growth protein 4

Chain A:



4.2.2 Score per residue for model 2

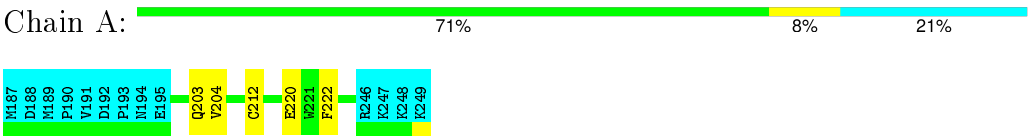
- Molecule 1: Inhibitor of growth protein 4

Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Inhibitor of growth protein 4

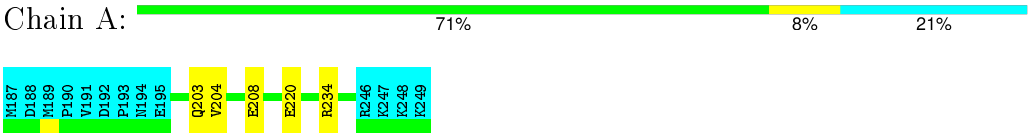
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Inhibitor of growth protein 4

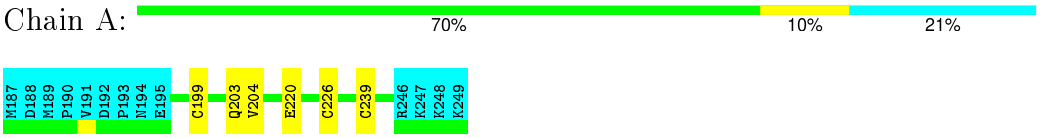
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Inhibitor of growth protein 4

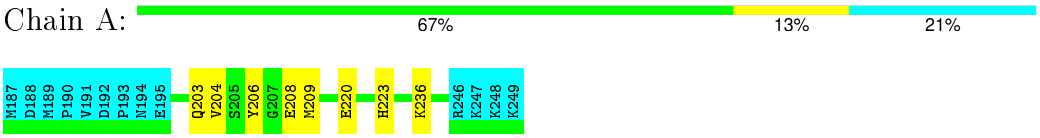
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Inhibitor of growth protein 4

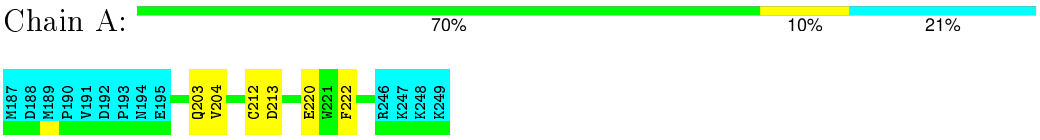
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Inhibitor of growth protein 4

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Inhibitor of growth protein 4

Chain A: 71% 8% 21%

M187 D188 M189 P190 V191 D192 P193 N194 E195 Q203 V204 E220 K236 C239 R246 K247 K248 K249

4.2.9 Score per residue for model 9 (medoid)

- Molecule 1: Inhibitor of growth protein 4

Chain A: 71% 8% 21%

M187 D188 M189 P190 V191 D192 P193 N194 E195 Q203 V204 C212 E220 W221 F222 R246 K247 K248 K249

4.2.10 Score per residue for model 10

- Molecule 1: Inhibitor of growth protein 4

Chain A: 70% 10% 21%

M187 D188 M189 P190 V191 D192 P193 N194 E195 Q203 V204 M209 E220 K236 C239 R246 K247 K248 K249

4.2.11 Score per residue for model 11

- Molecule 1: Inhibitor of growth protein 4

Chain A: 68% 11% 21%

M187 D188 M189 P190 V191 D192 P193 N194 E195 Q203 V204 M209 C212 E220 W221 F222 K236 R246 K247 K248 K249

4.2.12 Score per residue for model 12

- Molecule 1: Inhibitor of growth protein 4

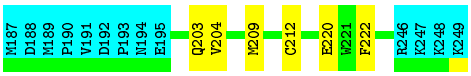
Chain A: 68% 11% 21%

M187 D188 M189 P190 V191 D192 P193 N194 E195 Q203 V204 M209 C212 E220 W221 F222 Q244 E245 R246 K247 K248 K249

4.2.13 Score per residue for model 13

- Molecule 1: Inhibitor of growth protein 4

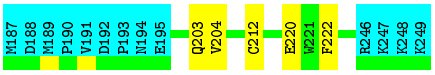
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Inhibitor of growth protein 4

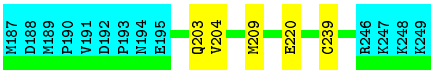
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Inhibitor of growth protein 4

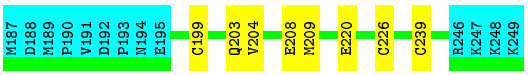
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Inhibitor of growth protein 4

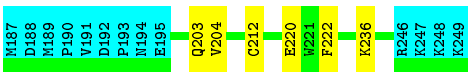
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Inhibitor of growth protein 4

Chain A:



4.2.18 Score per residue for model 18

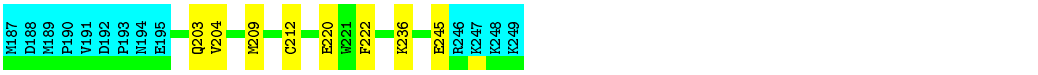
- Molecule 1: Inhibitor of growth protein 4

Chain A: 70% 10% 21%

4.2.19 Score per residue for model 19

- Molecule 1: Inhibitor of growth protein 4

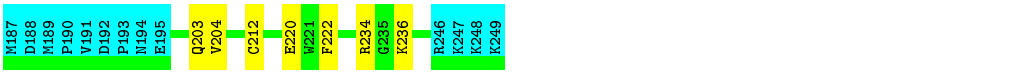
Chain A: 67% 13% 21%



4.2.20 Score per residue for model 20

- Molecule 1: Inhibitor of growth protein 4

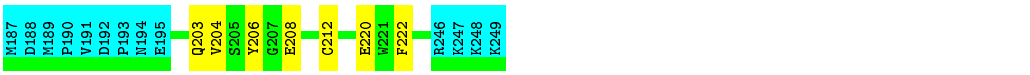
Chain A: 68% 11% 21%



4.2.21 Score per residue for model 21

- Molecule 1: Inhibitor of growth protein 4

Chain A: 68% 11% 21%



4.2.22 Score per residue for model 22

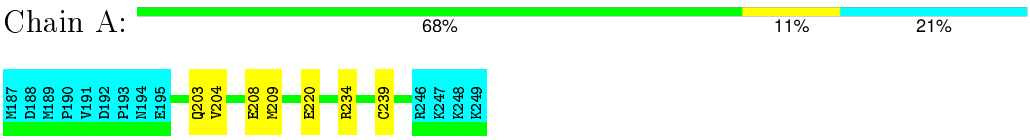
- Molecule 1: Inhibitor of growth protein 4

Chain A: 67% 13% 21%

4.2.23 Score per residue for model 23

- Molecule 1: Inhibitor of growth protein 4

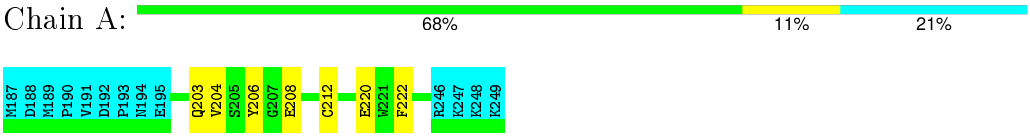
Chain A:



4.2.24 Score per residue for model 24

- Molecule 1: Inhibitor of growth protein 4

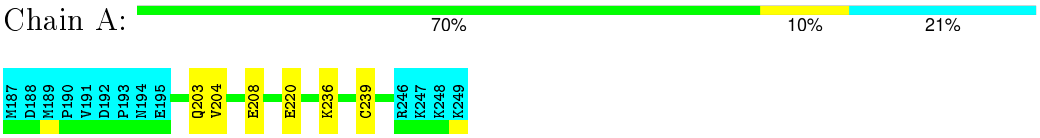
Chain A:



4.2.25 Score per residue for model 25

- Molecule 1: Inhibitor of growth protein 4

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 250 calculated structures, 25 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	
AMBER	refinement	7.0

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)

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 (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	398	364	364	1±0
All	All	10000	9100	9127	16

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:212:CYS:SG	1:A:222:PHE:CE1	0.61	2.94	20	14
1:A:206:TYR:CE2	1:A:223:HIS:CD2	0.40	3.09	18	2

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	50/63 (79%)	41±1 (82±3%)	8±1 (16±3%)	1±0 (2±0%)	14 55
All	All	1250/1575 (79%)	1028 (82%)	197 (16%)	25 (2%)	14 55

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	220	GLU	25

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	45/58 (78%)	41±1 (91±3%)	4±1 (9±3%)	17 62
All	All	1125/1450 (78%)	1026 (91%)	99 (9%)	17 62

All 14 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	203	GLN	25
1	A	204	VAL	25
1	A	209	MET	12
1	A	236	LYS	9
1	A	239	CYS	7
1	A	208	GLU	7
1	A	234	ARG	3
1	A	199	CYS	2
1	A	244	GLN	2
1	A	206	TYR	2
1	A	226	CYS	2
1	A	245	GLU	1
1	A	213	ASP	1
1	A	200	LEU	1

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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