wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:02 PM BST

PDB ID	:	1BU9
Title	:	SOLUTION STRUCTURE OF P18-INK4C, 21 STRUCTURES
Authors	:	Byeon, I.-J.L.; Li, J.; Tsai, M.-D.
Deposited on	:	1998-09-15

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:

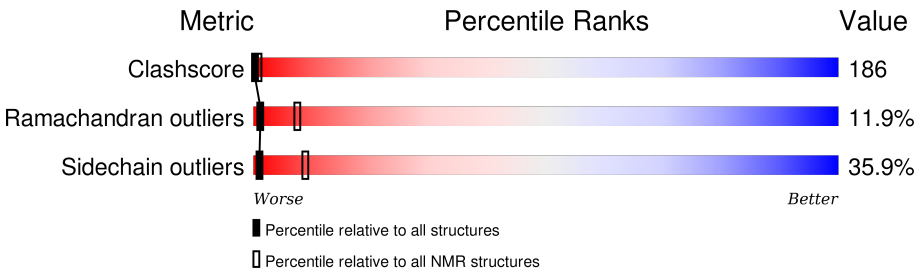
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal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

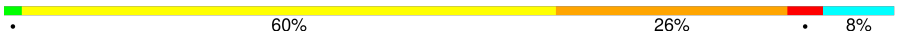
The overall completeness of chemical shifts assignment was not calculated.

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	168	

2 Ensemble composition and analysis (i)

This entry contains 21 models. Model 18 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:8-A:162 (155)	0.34	18

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 3 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2537 atoms, of which 1262 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN (CYCLIN-DEPENDENT KINASE 6 INHIBITOR).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	168	2537	784	1262	243	244	4	0

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: PROTEIN (CYCLIN-DEPENDENT KINASE 6 INHIBITOR)

Chain A:

M1 A2 E3 P4 W5 G6 N7 E8 L9 A10 S11 A12 A13 A14 R15 G16 D17 L18 E19 Q20 L21 T22 S23 L24 L25 Q26 N27 N28 V29 N30 V31 N32 A33 Q34 N35 G36 F37 G38 R39 T40 A41 L42 Q43 V44 M45 K46 L47 G48 N49 P50 E51 I52 A53 R54 R55 L56 L57 L58 R59 G60

A61 N62 P63 D64 L65 K66 D67 R68 T69 G70 F71 A72 V73 I74 H75 D76 A77 A78 R79 A80 G81 F82 L83 D84 T85 L86 Q87 T88 L89 L90 E91 F92 Q93 A94 D95 V96 N97 I98 E99 D100 N101 E102 G103 N104 L105 P106 L107 H108 L109 A110 A111 K112 E113 G114 H115 L116 R117 V118 V119 E120

F121 L122 V123 K124 H125 T126 A127 S128 N129 V130 G131 H132 R133 N134 H135 K136 G137 D138 T139 A140 C141 D142 L143 A144 R145 L146 Y147 G148 R149 N150 E151 V152 V153 S154 L155 M156 Q157 A158 N159 G160 A161 G162 G163 A164 T165 N166 L167 Q168

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

The representative model is number 18. Colouring as in section 4.1 above.

- Molecule 1: PROTEIN (CYCLIN-DEPENDENT KINASE 6 INHIBITOR)

Chain A:

M1 A2 E3 P4 W5 G6 N7 E8 L9 A10 S11 A12 A13 A14 R15 G16 D17 L18 E19 Q20 L21 T22 S23 L24 L25 Q26 N27 N28 V29 N30 V31 N32 A33 Q34 N35 G36 F37 G38 R39 T40 A41 L42 Q43 V44 M45 K46 L47 G48 N49 P50 E51 I52 A53 R54 R55 L56 L57 L58 R59 G60

A61 N62 P63 D64 L65 K66 D67 R68 T69 G70 F71 A72 V73 I74 H75 D76 A77 A78 R79 A80 G81 F82 L83 D84 T85 L86 Q87 T88 L89 L90 E91 F92 Q93 A94 D95 V96 N97 I98 E99 D100 N101 E102 G103 N104 L105 P106 L107 H108 L109 A110 A111 K112 E113 G114 H115 L116 R117 V118 V119 E120

F121 L122 V123 K124 H125 T126 A127 S128 N129 V130 G131 H132 R133 N134 H135 K136 G137 D138 T139 A140 C141 D142 L143 A144 R145 L146 Y147 G148 R149 N150 E151 V152 V153 S154 L155 A158 N159 G160 A161 G162 G163 A164 T165 N166 L167 Q168

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING*.

Of the 50 calculated structures, 21 were deposited, based on the following criterion: *CLOSEST TO MEAN STRUCTURE WHICH SHOWS GOOD AGREEMENT WITH THE CONSTRAINTS*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR	refinement	
X-PLOR	structure solution	

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)

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	1.04±0.00	0±0/1193 (0.0±0.0%)	1.29±0.01	1±1/1614 (0.0±0.0%)
All	All	1.04	0/25053 (0.0%)	1.29	16/33894 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	10.2±0.9
All	All	0	214

There are no bond-length outliers.

All 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
1	A	132	HIS	N-CA-CB	-5.67	100.39	110.60	4	1
1	A	92	PHE	N-CA-CB	-5.61	100.50	110.60	16	9
1	A	115	HIS	N-CA-CB	-5.38	100.92	110.60	16	6

There are no chirality outliers.

5 of 11 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	15	ARG	Sidechain	21
1	A	59	ARG	Sidechain	20
1	A	79	ARG	Sidechain	20
1	A	39	ARG	Sidechain	20
1	A	68	ARG	Sidechain	20

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	1178	1174	1174	436±16
All	All	24738	24654	24654	9163

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 186.

5 of 2175 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:105:LEU:HD21	1:A:132:HIS:CD2	1.19	1.71	18	16
1:A:105:LEU:HD13	1:A:108:HIS:CE1	1.17	1.72	2	1
1:A:111:ALA:HB1	1:A:144:ALA:HB2	1.16	1.16	19	20
1:A:96:VAL:HG23	1:A:122:LEU:HD23	1.13	1.14	16	8
1:A:72:ALA:HB2	1:A:98:ILE:HD12	1.12	1.13	9	21

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	155/168 (92%)	95±4 (61±2%)	42±4 (27±3%)	18±3 (12±2%)	1 7
All	All	3255/3528 (92%)	1995 (61%)	874 (27%)	386 (12%)	1 7

5 of 43 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	115	HIS	21
1	A	134	ASN	21
1	A	33	ALA	21

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	116	LEU	21
1	A	65	LEU	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	A	121/130 (93%)	78±4 (64±3%)	43±4 (36±3%)	1 9
All	All	2541/2730 (93%)	1628 (64%)	913 (36%)	1 9

5 of 97 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	136	LYS	21
1	A	88	THR	21
1	A	138	ASP	21
1	A	139	THR	21
1	A	59	ARG	21

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