wwPDB NMR Structure Validation Summary Report

May 28, 2020 – 09:03 pm BST

PDB ID	:	1SV1
Title	:	NMR structure of the ThKaiA180C-CIIABD complex (25-structure ensemble)
Authors	:	Vakonakis, I.; LiWang, A.C.
Deposited on	:	2004-03-26

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 (1)) were used in the production of this report:

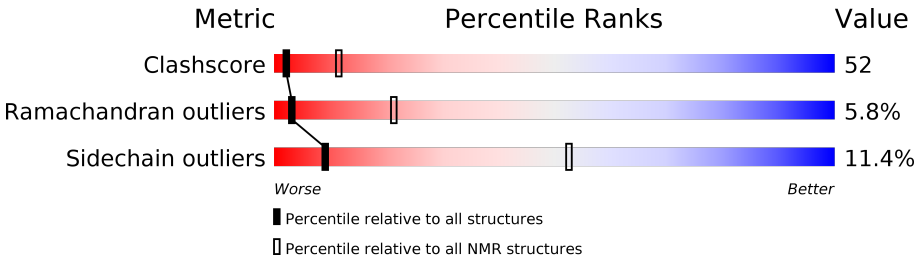
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance

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	107	41% 44% 6% 9%
1	B	107	43% 42% 6% 9%
2	C	34	. 44% 18% 35%
2	D	34	. 44% 18% 35%

2 Ensemble composition and analysis

This entry contains 25 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 11 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:8-A:104, B:208-B:304, C:409-C:430, D:509-D:530 (238)	0.33	10

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 12 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4596 atoms, of which 2328 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Circadian clock protein KaiA.

Mol	Chain	Residues	Atoms						Trace
1	A	107	Total	C	H	N	O	S	0
			1788	561	905	151	166	5	
1	B	107	Total	C	H	N	O	S	0
			1788	561	905	151	166	5	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	-	CLONING ARTIFACT	UNP Q79V62
A	2	MET	-	CLONING ARTIFACT	UNP Q79V62
A	3	ALA	-	CLONING ARTIFACT	UNP Q79V62
B	201	ALA	-	CLONING ARTIFACT	UNP Q79V62
B	202	MET	-	CLONING ARTIFACT	UNP Q79V62
B	203	ALA	-	CLONING ARTIFACT	UNP Q79V62

- Molecule 2 is a protein called Circadian clock protein KaiC.

Mol	Chain	Residues	Atoms						Trace
2	C	34	Total	C	H	N	O	S	0
			510	152	259	43	54	2	
2	D	34	Total	C	H	N	O	S	0
			510	152	259	43	54	2	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	401	ALA	-	CLONING ARTIFACT	UNP Q8RR33
C	402	MET	-	CLONING ARTIFACT	UNP Q8RR33
C	403	ALA	-	CLONING ARTIFACT	UNP Q8RR33
D	501	ALA	-	CLONING ARTIFACT	UNP Q8RR33
D	502	MET	-	CLONING ARTIFACT	UNP Q8RR33
D	503	ALA	-	CLONING ARTIFACT	UNP Q8RR33

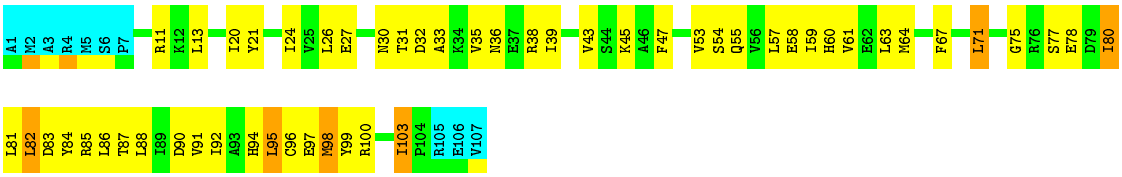
4 Residue-property plots

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.

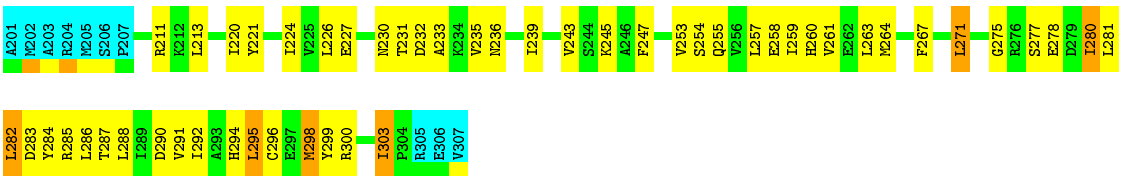
• Molecule 1: Circadian clock protein KaiA

Chain A:



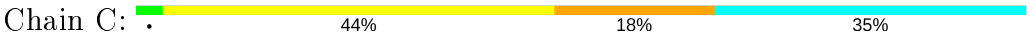
• Molecule 1: Circadian clock protein KaiA

Chain B:



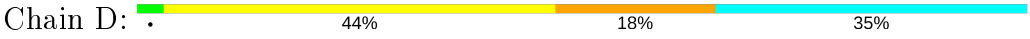
• Molecule 2: Circadian clock protein KaiC

Chain C:



• Molecule 2: Circadian clock protein KaiC

Chain D:

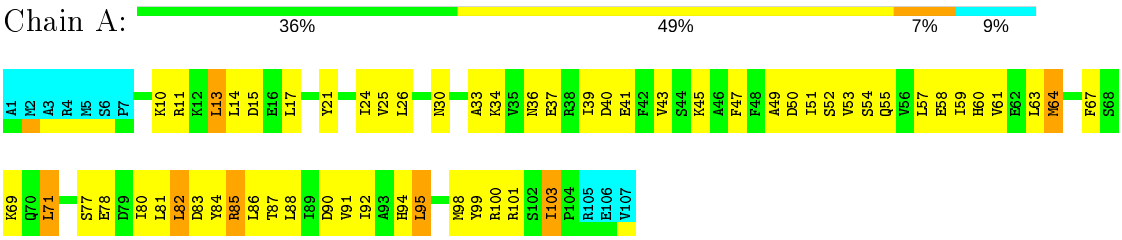


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 10. Colouring as in section 4.1 above.

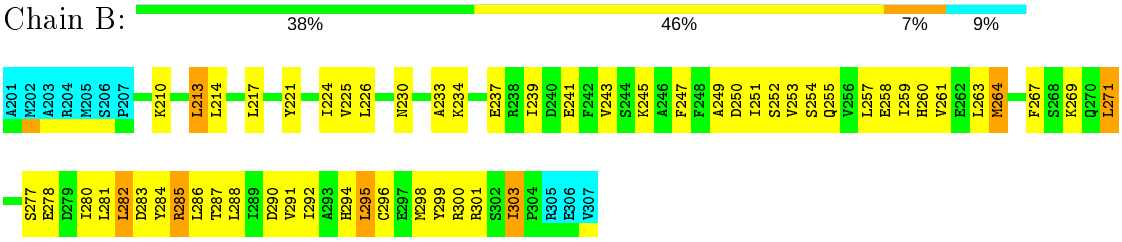
- Molecule 1: Circadian clock protein KaiA

Chain A:



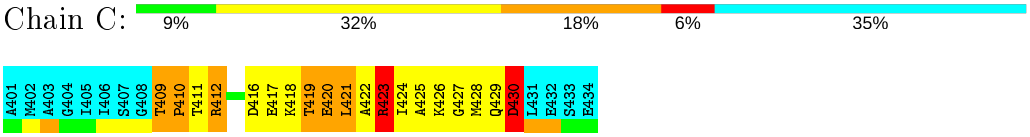
- Molecule 1: Circadian clock protein KaiA

Chain B:



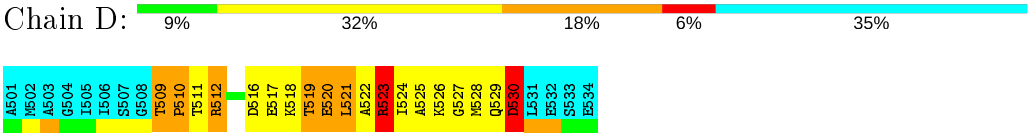
- Molecule 2: Circadian clock protein KaiC

Chain C:



- Molecule 2: Circadian clock protein KaiC

Chain D:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *Distance geometry, Simulated annealing.*

Of the 50 calculated structures, 25 were deposited, based on the following criterion: *lowest energy structures that satisfy all experimental restraints.*

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

Software name	Classification	Version
XPLOR-NIH	structure solution	2.9.1
XPLOR-NIH	refinement	2.9.1

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

6.1 Standard geometry

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	A	805	822	820	80±7
1	B	805	822	820	79±8
2	C	170	177	177	40±6
2	D	170	177	177	41±6
All	All	48750	49950	49850	5116

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 52.

5 of 1468 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:82:LEU:HD23	1:A:83:ASP:N	1.06	1.64	25	20
1:B:282:LEU:HD23	1:B:283:ASP:N	1.05	1.65	24	20
1:A:58:GLU:OE2	2:C:425:ALA:HB2	0.94	1.62	18	13
1:B:285:ARG:HH12	2:D:521:LEU:HD11	0.93	1.23	10	1
2:C:417:GLU:O	2:C:419:THR:HG22	0.93	1.63	16	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	97/107 (91%)	92±2 (95±2%)	4±1 (4±2%)	1±1 (1±1%)	24 71
1	B	97/107 (91%)	92±2 (95±2%)	4±2 (4±2%)	1±1 (1±1%)	24 71
2	C	22/34 (65%)	11±2 (49±9%)	5±2 (23±8%)	6±2 (28±8%)	0 1
2	D	22/34 (65%)	11±2 (49±9%)	5±2 (23±8%)	6±2 (28±8%)	0 1
All	All	5950/7050 (84%)	5154 (87%)	449 (8%)	347 (6%)	3 21

5 of 46 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	C	419	THR	21
2	D	519	THR	21
2	C	420	GLU	20
2	D	520	GLU	20
2	C	423	ARG	18

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	91/99 (92%)	83±1 (91±1%)	8±1 (9±1%)	13 59
1	B	91/99 (92%)	83±1 (91±1%)	8±1 (9±1%)	13 59
2	C	19/27 (70%)	15±1 (78±7%)	4±1 (22±7%)	3 29
2	D	19/27 (70%)	15±1 (78±7%)	4±1 (22±7%)	3 29
All	All	5500/6300 (87%)	4875 (89%)	625 (11%)	9 52

5 of 82 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	B	303	ILE	25
1	A	13	LEU	25
1	A	95	LEU	25
1	B	213	LEU	25
1	A	103	ILE	25

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 carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

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

No chemical shift data were provided