



Full wwPDB NMR Structure Validation Report ⓘ

Oct 10, 2021 – 04:55 PM EDT

PDB ID : 2KDK
Title : Structure of human circadian clock protein BMAL2 C-terminal PAS domain
Authors : Wasielewski, E.; Correia, C.; Prendergast, F.G.; Mer, G.
Deposited on : 2009-01-10

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

<https://www.wwpdb.org/validation/2017/NMRValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : 2.23.2
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.2

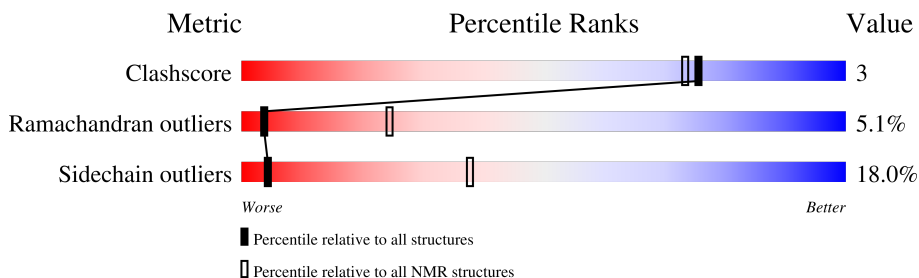
1 Overall quality at a glance

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	121	 61% 21% .. 12% .

2 Ensemble composition and analysis

This entry contains 10 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	A:317-A:420 (104)	0.40	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 1 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 9, 10
Single-model clusters	8

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1890 atoms, of which 932 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Aryl hydrocarbon receptor nuclear translocator-like protein 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1890	613	932	158	186	1	0

There are 4 discrepancies between the modelled and reference sequences:

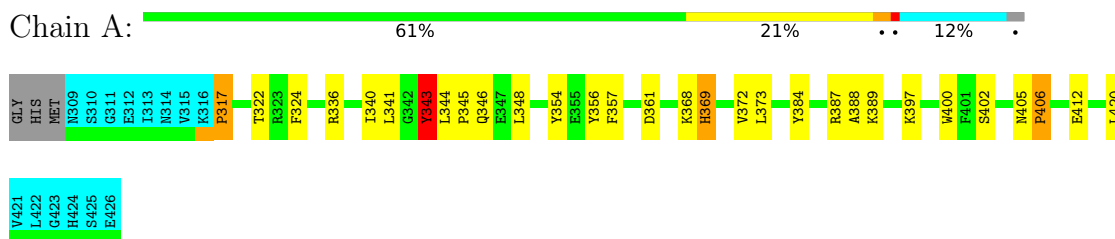
Chain	Residue	Modelled	Actual	Comment	Reference
A	306	GLY	-	expression tag	UNP Q8WYA1
A	307	HIS	-	expression tag	UNP Q8WYA1
A	308	MET	-	expression tag	UNP Q8WYA1
A	407	ASP	TRP	engineered mutation	UNP Q8WYA1

4 Residue-property plots i

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: Aryl hydrocarbon receptor nuclear translocator-like protein 2

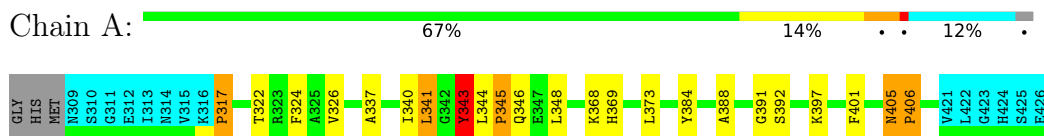


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

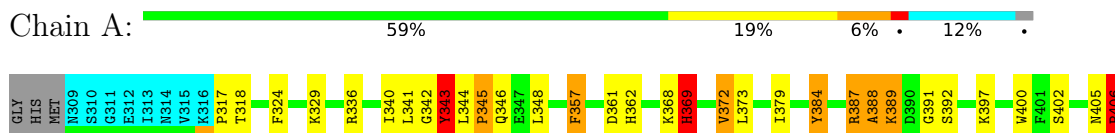
4.2.1 Score per residue for model 1

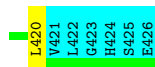
- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2



4.2.2 Score per residue for model 2

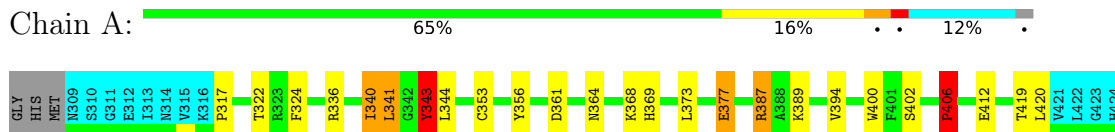
- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2





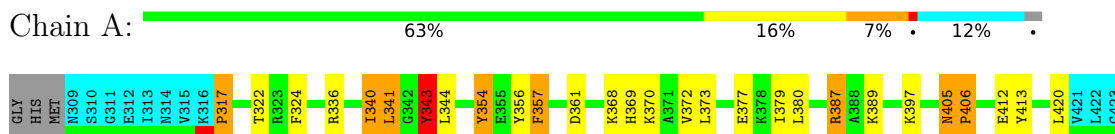
4.2.3 Score per residue for model 3

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2



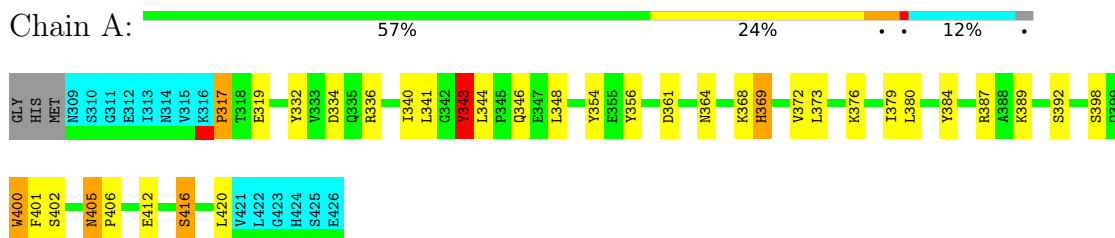
4.2.4 Score per residue for model 4

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2



4.2.5 Score per residue for model 5

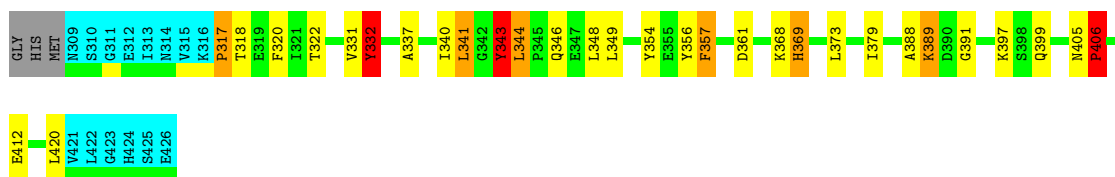
- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2



4.2.6 Score per residue for model 6

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2

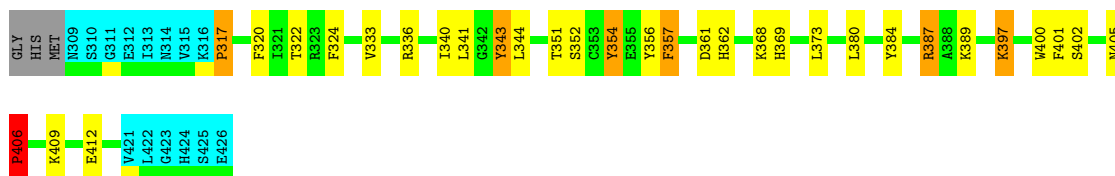




4.2.7 Score per residue for model 7 (medoid)

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2

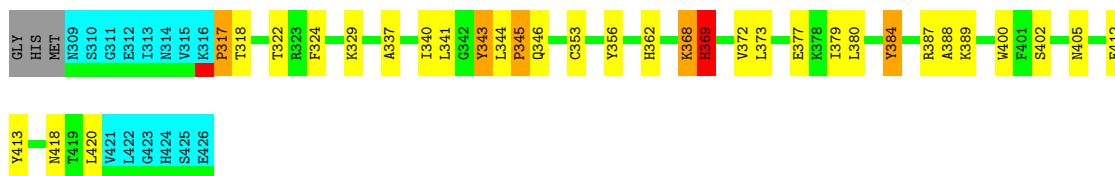
Chain A: 60% 21% 5% 12%



4.2.8 Score per residue for model 8

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2

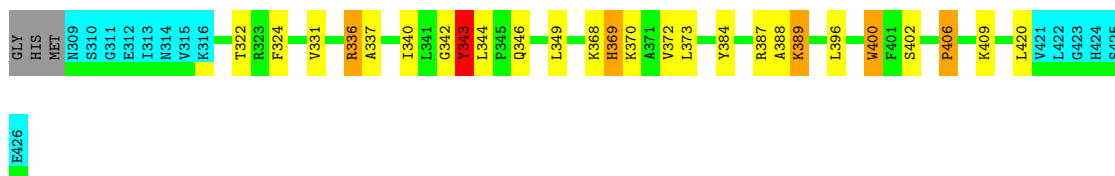
Chain A: 59% 22% 12%



4.2.9 Score per residue for model 9

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2

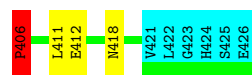
Chain A: 64% 17% 12%



4.2.10 Score per residue for model 10

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator-like protein 2

Chain A: 58% 19% 7% 12%



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 10 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	8.0

No chemical shift data was provided.

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	0.74±0.01	0±0/876 (0.0± 0.0%)	1.36±0.04	6±1/1187 (0.5± 0.1%)
All	All	0.74	0/8760 (0.0%)	1.36	60/11870 (0.5%)

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	6.5±1.4
All	All	0	65

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	343	TYR	CB-CG-CD2	-11.45	114.13	121.00	6	10
1	A	356	TYR	CB-CG-CD2	-9.06	115.56	121.00	5	3
1	A	387	ARG	CB-CA-C	8.72	127.85	110.40	5	4
1	A	343	TYR	CA-CB-CG	8.07	128.74	113.40	7	9
1	A	343	TYR	CB-CG-CD1	7.90	125.74	121.00	1	7
1	A	317	PRO	CA-N-CD	-6.86	101.90	111.50	1	5
1	A	356	TYR	CB-CG-CD1	-6.27	117.24	121.00	6	2
1	A	354	TYR	CB-CG-CD2	-6.05	117.37	121.00	7	2
1	A	387	ARG	C-N-CA	6.04	136.80	121.70	8	3
1	A	322	THR	CA-CB-CG2	6.01	120.81	112.40	3	3
1	A	372	VAL	CB-CA-C	6.00	122.80	111.40	2	1
1	A	372	VAL	CA-CB-CG1	5.72	119.49	110.90	2	1
1	A	332	TYR	CB-CG-CD2	-5.51	117.69	121.00	6	1
1	A	357	PHE	CB-CG-CD2	-5.43	117.00	120.80	2	1
1	A	397	LYS	N-CA-C	-5.41	96.39	111.00	2	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	400	TRP	CA-CB-CG	5.33	123.82	113.70	5	1
1	A	370	LYS	N-CA-CB	5.29	120.13	110.60	9	2
1	A	384	TYR	CB-CG-CD2	-5.06	117.96	121.00	1	1
1	A	388	ALA	N-CA-CB	-5.05	103.03	110.10	2	1

There are no chirality outliers.

All 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	344	LEU	Peptide,Mainchain	10
1	A	369	HIS	Sidechain	10
1	A	406	PRO	Peptide	9
1	A	405	ASN	Peptide	7
1	A	343	TYR	Sidechain	5
1	A	384	TYR	Sidechain	5
1	A	332	TYR	Sidechain	3
1	A	317	PRO	Peptide	2
1	A	320	PHE	Sidechain	2
1	A	387	ARG	Sidechain	1
1	A	356	TYR	Sidechain	1

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	854	832	832	5±2
All	All	8540	8320	8320	52

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:340:ILE:HG23	1:A:341:LEU:H	0.66	1.50	3	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:341:LEU:HD11	1:A:343:TYR:CE1	0.57	2.34	1	1
1:A:341:LEU:HD22	1:A:388:ALA:HB2	0.56	1.76	10	2
1:A:336:ARG:O	1:A:340:ILE:HG22	0.53	2.04	3	5
1:A:337:ALA:HA	1:A:340:ILE:HG22	0.52	1.79	10	5
1:A:344:LEU:HD13	1:A:346:GLN:HE22	0.49	1.68	6	1
1:A:340:ILE:HG23	1:A:341:LEU:HG	0.48	1.85	7	4
1:A:379:ILE:HD12	1:A:400:TRP:O	0.48	2.08	2	3
1:A:372:VAL:HG21	1:A:400:TRP:HB2	0.48	1.85	5	3
1:A:379:ILE:C	1:A:380:LEU:HD12	0.47	2.30	4	3
1:A:343:TYR:CE1	1:A:348:LEU:HD21	0.46	2.46	5	1
1:A:343:TYR:CE1	1:A:348:LEU:HG	0.45	2.47	2	3
1:A:369:HIS:HA	1:A:372:VAL:HG12	0.44	1.89	2	1
1:A:344:LEU:HD12	1:A:344:LEU:N	0.44	2.27	6	2
1:A:340:ILE:HG23	1:A:341:LEU:HD12	0.44	1.90	3	1
1:A:372:VAL:HA	1:A:380:LEU:HD13	0.42	1.91	5	3
1:A:343:TYR:CD2	1:A:388:ALA:HA	0.42	2.49	1	1
1:A:400:TRP:CE3	1:A:416:SER:HB2	0.42	2.49	5	1
1:A:353:CYS:HA	1:A:356:TYR:CE1	0.42	2.49	3	2
1:A:379:ILE:HD11	1:A:399:GLN:HB2	0.42	1.91	6	1
1:A:341:LEU:HB2	1:A:343:TYR:CE1	0.41	2.51	2	1
1:A:368:LYS:HA	1:A:368:LYS:HE3	0.41	1.93	8	1
1:A:341:LEU:CD1	1:A:343:TYR:CE1	0.41	3.03	1	1
1:A:337:ALA:HB1	1:A:348:LEU:CD1	0.41	2.46	6	1
1:A:341:LEU:HD11	1:A:343:TYR:CZ	0.40	2.51	1	1
1:A:319:GLU:CA	1:A:340:ILE:HD12	0.40	2.46	5	1
1:A:369:HIS:C	1:A:369:HIS:ND1	0.40	2.75	8	1

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	104/121 (86%)	88±1 (85±1%)	11±2 (10±2%)	5±1 (5±1%)	4	24
All	All	1040/1210 (86%)	879 (85%)	108 (10%)	53 (5%)	4	24

All 17 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	317	PRO	7
1	A	361	ASP	7
1	A	406	PRO	7
1	A	346	GLN	5
1	A	345	PRO	4
1	A	389	LYS	4
1	A	391	GLY	3
1	A	388	ALA	3
1	A	357	PHE	3
1	A	342	GLY	2
1	A	340	ILE	2
1	A	326	VAL	1
1	A	377	GLU	1
1	A	376	LYS	1
1	A	392	SER	1
1	A	352	SER	1
1	A	409	LYS	1

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	95/109 (87%)	78±3 (82±3%)	17±3 (18±3%)	4 38
All	All	950/1090 (87%)	779 (82%)	171 (18%)	4 38

All 48 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	343	TYR	10
1	A	368	LYS	10
1	A	373	LEU	10
1	A	389	LYS	9
1	A	324	PHE	7
1	A	402	SER	7
1	A	420	LEU	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	412	GLU	7
1	A	406	PRO	6
1	A	369	HIS	6
1	A	322	THR	5
1	A	341	LEU	5
1	A	357	PHE	5
1	A	354	TYR	5
1	A	345	PRO	4
1	A	397	LYS	4
1	A	401	PHE	4
1	A	362	HIS	4
1	A	384	TYR	4
1	A	387	ARG	4
1	A	400	TRP	4
1	A	405	ASN	3
1	A	318	THR	3
1	A	377	GLU	3
1	A	348	LEU	2
1	A	392	SER	2
1	A	329	LYS	2
1	A	364	ASN	2
1	A	317	PRO	2
1	A	413	TYR	2
1	A	331	VAL	2
1	A	349	LEU	2
1	A	336	ARG	2
1	A	380	LEU	2
1	A	418	ASN	2
1	A	394	VAL	1
1	A	419	THR	1
1	A	334	ASP	1
1	A	398	SER	1
1	A	416	SER	1
1	A	332	TYR	1
1	A	320	PHE	1
1	A	333	VAL	1
1	A	351	THR	1
1	A	396	LEU	1
1	A	409	LYS	1
1	A	335	GLN	1
1	A	411	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 monosaccharides 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

No chemical shift data were provided