wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 10:03 PM BST

PDB ID	:	2K2N
Title	:	Solution structure of a cyanobacterial phytochrome GAF domain in the red light-absorbing ground state
Authors	:	Cornilescu, C.C.; Cornilescu, G.; Ulijasz, A.T.; Vierstra, R.D.; Markley, J.L.
Deposited on	:	2008-04-04

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:

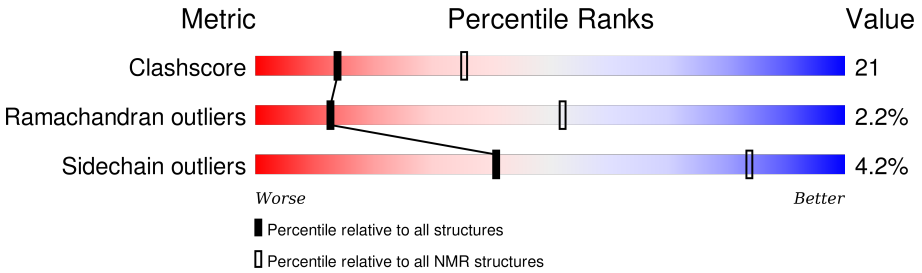
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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	172	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	CYC	203	20	-

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 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:31-A:113, A:134-A:202 (152)	0.35	19

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2776 atoms, of which 1390 are hydrogens and 0 are deuteriums.

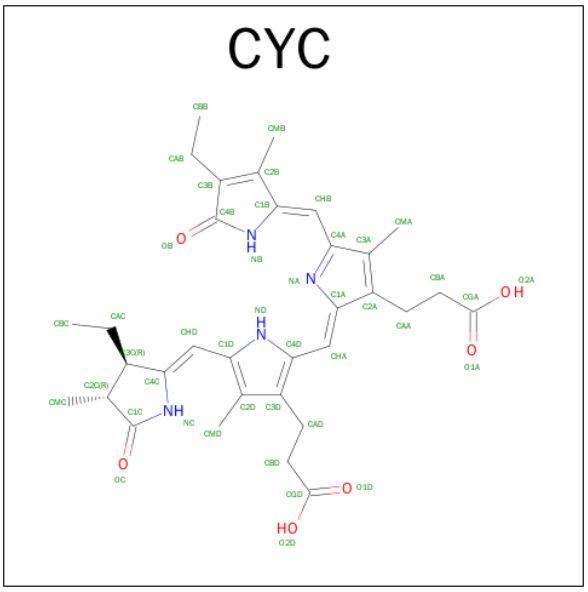
- Molecule 1 is a protein called Sensor protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	172	2695	850	1352	243	247	3	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	SER	-	EXPRESSION TAG	UNP Q2JIZ5
A	202	LEU	-	EXPRESSION TAG	UNP Q2JIZ5

- Molecule 2 is PHYCOCYANOBILIN (three-letter code: CYC) (formula: $C_{33}H_{40}N_4O_6$).



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
2	A	1	81	33	38	4	6

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: Sensor protein

Chain A: 63% 23% • 12%

L31 V39 R43 G47 T48 D49 R50 Y54 R55 F56 D57 V65 A66 E67 F82 I87 P88 A91 R92 F95 A98 Q99 V100 R101 D105 S112 I113 S114 Q115 P116 E117 S118 W119 G120 L121 S122 A123 R124 V125 P126 L127 G128 E129 P130 L131 Q132 R133 P134

V135 D136 P137 C138 H139 V140 L143 M146 M157 H158 H159 L162 W163 G164 L165 H169 H170 P173 Y176 L187 V191 L202

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

The representative model is number 19. Colouring as in section 4.1 above.

- Molecule 1: Sensor protein

Chain A: 62% 24% • 12%

L31 V39 R43 G47 R50 V51 K52 V53 Y54 D57 A66 E67 A68 R69 P75 S76 F82 P88 E89 E90 A91 F95 R96 L97 A98 Q99 V100 R101 S112 I113 S114 Q115 P116 E117 S118 W119 G120 L121 S122 A123 R124 V125 P126 L127 G128 E129 P130 L131 Q132

R133 C138 H139 V140 H141 Y142 L143 M146 S151 H158 H159 G164 L165 L166 V167 S168 H169 H170 Y176 S177 Q178 V183 V184 Q185 L186 L187 A188 V191 L200 S201 L202

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle molecular dynamics, internal variable dynamics.*

Of the 100 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
X-PLOR NIH	structure solution	
X-PLOR NIH	geometry optimization	
X-PLOR NIH	refinement	

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: CYC

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	1188	1199	1198	50±8
2	A	43	38	36	5±2
All	All	24620	24740	24680	1014

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 21.

5 of 450 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:91:ALA:HB2	2:A:203:CYC:HBD2	0.87	1.42	13	6
1:A:87:ILE:HD12	1:A:87:ILE:H	0.86	1.30	14	1
1:A:86:ASP:O	1:A:87:ILE:HG23	0.85	1.72	15	1
1:A:162:LEU:HD12	1:A:162:LEU:N	0.82	1.89	10	1
1:A:187:LEU:O	1:A:191:VAL:HG23	0.80	1.77	12	11

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	150/172 (87%)	138±3 (92±2%)	9±2 (6±2%)	3±1 (2±1%)	13 52
All	All	3000/3440 (87%)	2756 (92%)	179 (6%)	65 (2%)	13 52

5 of 14 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	100	VAL	20
1	A	99	GLN	14
1	A	88	PRO	11
1	A	70	GLY	3
1	A	173	PRO	3

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	129/146 (88%)	124±2 (96±1%)	5±2 (4±1%)	41 84
All	All	2580/2920 (88%)	2472 (96%)	108 (4%)	41 84

5 of 32 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	112	SER	17
1	A	136	ASP	7
1	A	57	ASP	7
1	A	54	TYR	7
1	A	170	HIS	6

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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	CYC	A	203	1	37,46,46	2.02±0.03	1±0 (2±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	CYC	A	203	1	44,67,67	2.53±0.05	3±0 (6±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	CYC	A	203	1	1±0,1,19,19	2±0 ,21,74,74	0±0,4,4,4

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	203	CYC	CHA-C1A	6.45	1.41	1.35	20	20

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
2	A	203	CYC	C4B-C3B-C2B	7.98	103.54	108.04	16	20
2	A	203	CYC	C1A-C2A-C3A	6.22	100.02	106.80	8	1
2	A	203	CYC	CAB-C3B-C4B	6.12	126.90	121.47	1	20
2	A	203	CYC	CAC-C3C-C4C	5.06	125.66	112.67	11	20

All unique chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	203	CYC	C2C	20

All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	203	CYC	C1B-CHB-C4A-NA	20
2	A	203	CYC	C1B-CHB-C4A-C3A	20
2	A	203	CYC	C4A-CHB-C1B-C2B	16
2	A	203	CYC	C4A-CHB-C1B-NB	11

There are no ring outliers.

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