WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

Feb 13, 2017 – 01:34 am GMT

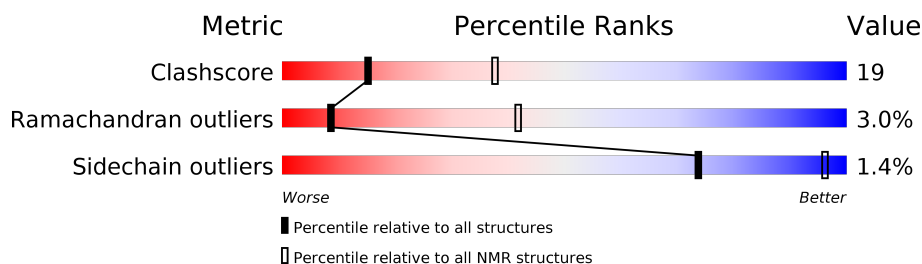|		
PDB ID	:	2MWG
Title	:	Full-Length Solution Structure Of YtvA, a LOV-Photoreceptor Protein and Regulator of Bacterial Stress Response
Authors	:	Jurk, M.; Bardiaux, B.; Schmieder, P.
Deposited on	:	2014-11-07

---

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 is 22%.

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	Percentile Ranks	Value
Clashscore		19
Ramachandran outliers		3.0%
Sidechain outliers		1.4%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	261	68% 28% • •
1	B	261	68% 26% • •

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2   Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 1 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:1-A:261, B:1-B:261 (522)	0.15	1

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, 4, 5, 6, 7, 9
2	8, 10
3	2, 3

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 8372 atoms, of which 4218 are hydrogens and 0 are deuteriums.
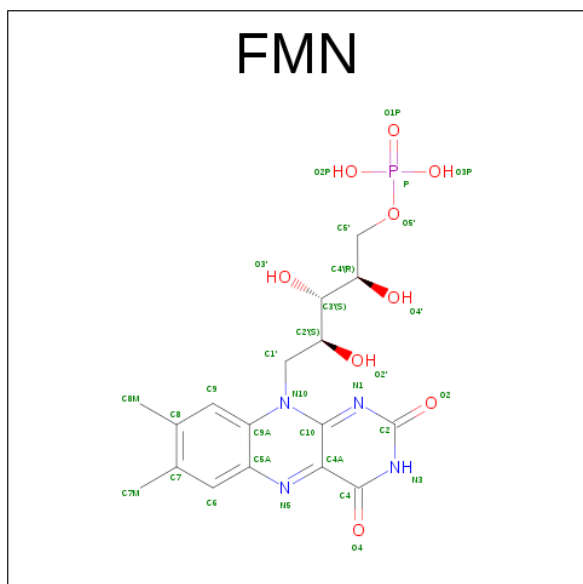
- Molecule 1 is a protein called Blue-light photoreceptor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	261	4136	1302	2090	333	403	8	0
			Total	C	H	N	O	S	
1	B	261	4136	1302	2090	333	403	8	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP O34627
B	1	GLY	-	EXPRESSION TAG	UNP O34627

- Molecule 2 is FLAVIN MONONUCLEOTIDE (three-letter code: FMN) (formula: $C_{17}H_{21}N_4O_9P$).
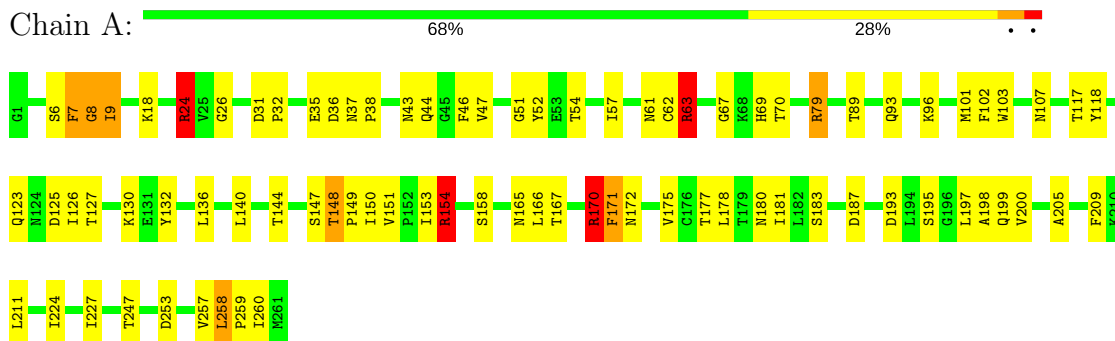

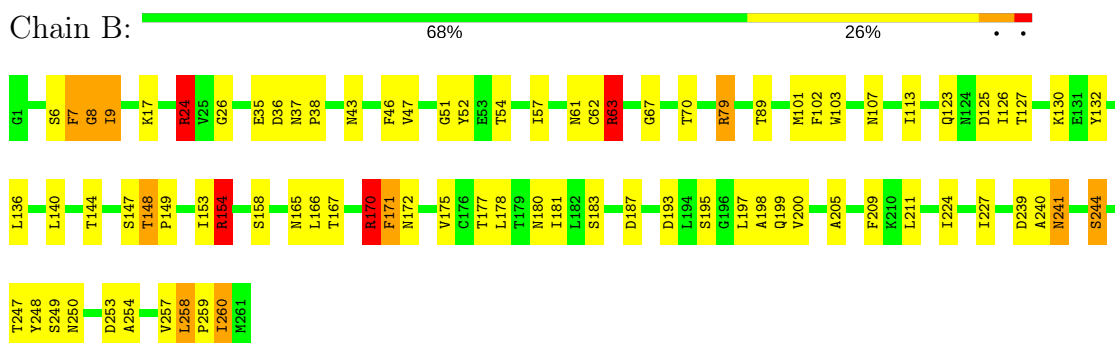| Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
2	A	1	50	17	19	4	9	1
			Total	C	H	N	O	P
2	B	1	50	17	19	4	9	1

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: Blue-light photoreceptor

- Molecule 1: Blue-light photoreceptor

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

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Blue-light photoreceptor

G1 A2 S6 F7 G8 I9 K18 R24 V25 G26 P32 E35 D36 N37 P38 N43 Q44 G45 F46 V47 G51 Y52 E53 T54 I57 N61 C62 R63 G67 K68 H69 T70 R79 T89 Q93 N94 Y95 M101 F102 W103 N107 I108 D109 T117

Y118 Q123 N124 D125 I126 T127 K130 E131 Y132 L136 L140 T144 S147 T148 P149 I150 V151 P152 I153 R154 S158 N165 L166 T167 R170 F171 N172 V175 L178 T179 N180 S183 D187 D193 L194 S195 G196 L197 A198 Q199 V200 A205 F209 K210

L211 I224 I227 D239 A240 T247 Y248 S249 N250 A254 V257 L258 P259 I260 M261

● Molecule 1: Blue-light photoreceptor

Chain B: 69% 26%

G1 S6 F7 G8 I9 K17 R24 V25 G26 E35 D36 N37 P38 N43 F46 V47 G51 Y52 E53 T54 I57 N61 C62 R63 G67 T70 R79 T89 M101 F102 W103 N107 I108 D109 I113 Q123 N124 D125 I126 T127 K130 E131

Y132 L136 L140 T144 S147 T148 P149 I153 R154 S158 N165 L166 T167 R170 F171 N172 V175 C176 T177 L178 T179 N180 I181 L182 S183 D187 D193 L194 S195 G196 L197 A198 Q199 V200 A205 F209 K210 L211 I224 I227 D239 A240 N241

S244 T247 Y248 S249 N250 A254 V257 L258 P259 I260 M261

5   Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *monte-carlo simulated annealing using torsion angle dynamics, 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
CS-Rosetta	structure solution	v3.2
X-PLOR NIH	structure solution	v2.30
X-PLOR NIH	refinement	v2.30
X-PLOR NIH	geometry optimization	v2.30

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2mwg_cs.str
Number of chemical shift lists	1
Total number of shifts	1610
Number of shifts mapped to atoms	1610
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	22%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: FMN

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	0.69±0.01	0±0/2077 (0.0±0.0%)	1.00±0.02	12±1/2815 (0.4±0.0%)
1	B	0.69±0.00	0±0/2077 (0.0±0.0%)	0.99±0.01	12±1/2815 (0.4±0.0%)
All	All	0.69	0/41540 (0.0%)	0.99	242/56300 (0.4%)

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	5.0±0.0
1	B	0.0±0.0	5.0±0.0
All	All	0	100

There are no bond-length outliers.

5 of 26 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	79	ARG	CD-NE-CZ	-11.86	107.00	123.60	1	10
1	B	79	ARG	CD-NE-CZ	-11.72	107.19	123.60	1	10
1	A	170	ARG	CD-NE-CZ	-11.40	107.64	123.60	1	10
1	B	170	ARG	CD-NE-CZ	-11.37	107.68	123.60	1	10
1	B	24	ARG	CD-NE-CZ	-11.24	107.86	123.60	1	10

There are no chirality outliers.

5 of 10 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	B	170	ARG	Sidechain	10
1	B	79	ARG	Sidechain	10
1	A	63	ARG	Sidechain	10
1	B	63	ARG	Sidechain	10
1	A	24	ARG	Sidechain	10

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	2046	2090	2087	79±5
1	B	2046	2090	2087	80±6
2	A	31	19	19	2±1
2	B	31	19	19	2±1
All	All	41540	42180	42120	1562

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 19.

5 of 268 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:B:197:LEU:CD2	1:B:199:GLN:O	1.11	1.99	7	10
1:A:197:LEU:CD2	1:A:199:GLN:O	1.07	2.03	7	10
1:B:197:LEU:HD22	1:B:199:GLN:O	0.97	1.55	5	10
1:A:165:ASN:O	1:A:170:ARG:NH1	0.94	2.01	6	10
1:B:103:TRP:CD1	1:B:130:LYS:NZ	0.93	2.36	9	10

6.3   Torsion angles ⓘ

6.3.1   Protein backbone ⓘ

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	259/261 (99%)	246±1 (95±0%)	6±1 (2±0%)	7±0 (3±0%)	10	45
1	B	259/261 (99%)	243±2 (94±1%)	8±2 (3±1%)	8±1 (3±0%)	8	40
All	All	5180/5220 (99%)	4892 (94%)	134 (3%)	154 (3%)	9	42

5 of 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	258	LEU	10
1	B	9	ILE	10
1	A	148	THR	10
1	B	148	THR	10
1	B	7	PHE	10

6.3.2 Protein sidechains ⓘ

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	234/234 (100%)	231±1 (99±0%)	3±1 (1±0%)	74	96
1	B	234/234 (100%)	231±1 (99±0%)	3±1 (1±0%)	74	96
All	All	4680/4680 (100%)	4615 (99%)	65 (1%)	74	96

5 of 10 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	171	PHE	10
1	A	54	THR	10
1	B	171	PHE	10
1	B	54	THR	10
1	B	154	ARG	7

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 ⓘ

2 ligands are 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	FMN	A	900	-	30,33,33	2.08±0.00	3±0 (10±0%)
2	FMN	B	900	-	30,33,33	2.08±0.00	3±0 (10±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	FMN	A	900	-	37,50,50	1.41±0.01	1±0 (2±0%)
2	FMN	B	900	-	37,50,50	1.41±0.01	1±0 (2±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	FMN	A	900	-	-	0±0,16,18,18	0±0,3,3,3
2	FMN	B	900	-	-	0±0,16,18,18	0±0,3,3,3

5 of 6 unique bond 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(Å)	Ideal(Å)	Models Worst	Total
2	B	900	FMN	C8M-C8	5.92	1.39	1.51	3	10
2	A	900	FMN	C7M-C7	5.89	1.39	1.51	7	10
2	A	900	FMN	C8M-C8	5.88	1.39	1.51	4	10
2	B	900	FMN	C7M-C7	5.86	1.39	1.51	4	10
2	B	900	FMN	C10-N1	5.20	1.40	1.33	1	10

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	900	FMN	C2-N1-C10	5.27	119.97	114.90	9	10
2	B	900	FMN	C2-N1-C10	5.26	119.97	114.90	7	10

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 22% for the well-defined parts and 22% for the entire structure.

7.1 Chemical shift list 1

File name: 2mwg_cs.str

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping ⓘ

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1610
Number of shifts mapped to atoms	1610
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	255	0.22 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_\beta$	239	0.89 ± 0.08	Should be applied
$^{13}C'$	255	-0.17 ± 0.15	None needed (< 0.5 ppm)
^{15}N	242	-0.45 ± 0.26	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 22%, i.e. 1417 atoms were assigned a chemical shift out of a possible 6388. 0 out of 94 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	994/2566 (39%)	242/1022 (24%)	510/1044 (49%)	242/500 (48%)
Sidechain	421/3476 (12%)	91/2012 (5%)	330/1334 (25%)	0/130 (0%)

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	2/346 (1%)	1/182 (1%)	0/150 (0%)	1/14 (7%)
Overall	1417/6388 (22%)	334/3216 (10%)	840/2528 (33%)	243/644 (38%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: