wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 02:24 AM BST

PDB ID	:	2M7V
Title	:	Green Light-Absorbing State of TePixJ, an Active Cyanobacteriochrome Domain
Authors	:	Cornilescu, G.; Cornilescu, C.C.; Burgie, S.E.; Walker, J.M.; Markley, J.L.; Ulijasz, A.T.; Vierstra, R.D.
Deposited on	:	2013-05-01

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:

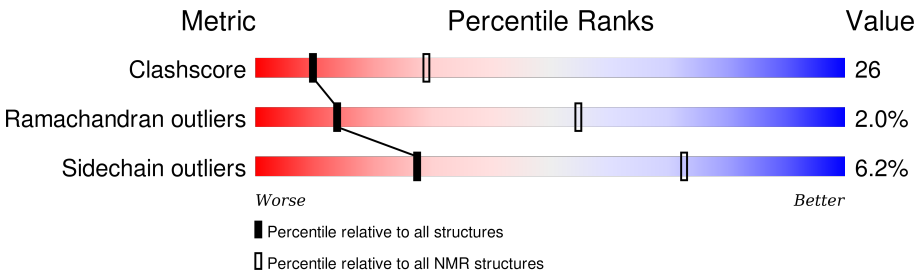
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 is 71%.

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	165	52% 27% 5% 16%

2 Ensemble composition and analysis (i)

This entry contains 10 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 2 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:439-A:492, A:503-A:541, A:546-A:590 (138)	0.24	2

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 3 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2712 atoms, of which 1349 are hydrogens and 0 are deuteriums.

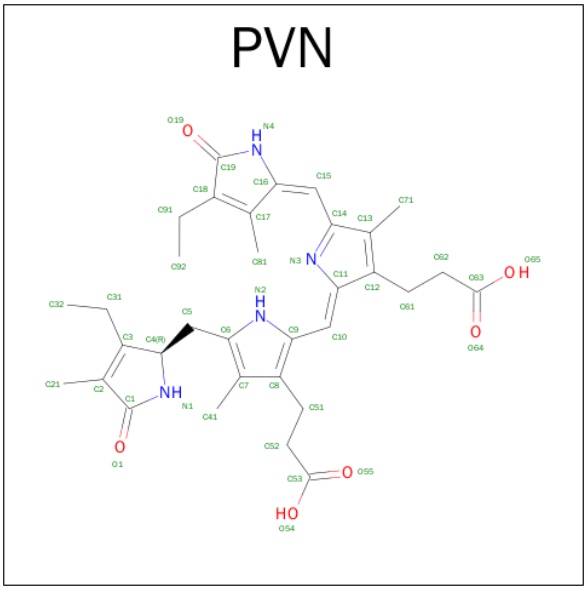
- Molecule 1 is a protein called Methyl-accepting chemotaxis protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	165	2632	836	1312	231	248	5	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	429	MET	-	EXPRESSION TAG	UNP Q8DLC7
A	555	ALA	CYS	ENGINEERED MUTATION	UNP Q8DLC7
A	592	SER	-	EXPRESSION TAG	UNP Q8DLC7
A	593	LEU	-	EXPRESSION TAG	UNP Q8DLC7

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



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
2	A	1	80	33	37	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: Methyl-accepting chemotaxis protein

Chain A: 52% 27% 5% 16%

M429 A430 A431 V432 Q433 L434 S435 E436 L437 R438 T446 L447 V448 A449 L455 A456 C457 D458 Y463 A464 F465 D466 D467 N468 T472 E476 S477 V478 A485 Q488 V489 I490 E491 D492 P493 C494 F495 R496 E497 H498 W499 V500 E501 A502 Y503 R504 Q505 G506 R507 A510 D513

I514 L519 Q522 H523 L524 L527 R528 P529 L530 K531 V532 R533 A534 N535 M540 V541 I542 D543 D544 Q545 L549 L550 I551 A552 H553 Q554 P558 R559 Q560 W561 Q562 E563 I564 F569 A573 L580 H584 Q588 A591 S592 L593

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

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: Methyl-accepting chemotaxis protein

Chain A: 52% 27% • 16%

M429 A430 A431 V432 Q433 L434 S435 E436 L437 R438 D439 R440 T446 L447 V448 A449 L455 A456 C457 D458 R459 Y463 A464 F465 D466 D467 N468 T472 V478 P483 D492 P493 C494 F495 R496 E497 H498 W499 V500 E501 A502 Y503 R504 Q505 A510 D513 I514 F515 K516 C522

H523 L524 L527 R528 P529 L530 K531 V532 R533 A534 N535 M540 V541 I542 D543 D544 Q545 L549 L550 I551 A552 H553 Q554 E557 P558 R559 Q560 W561 Q562 F569 A573 L580 E581 R582 L583 H584 F585 L586 E587 Q588 A591 S592 L593

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, torsion angle dynamics.*

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum.*

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	refinement	

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)	2m7v_cs.str
Number of chemical shift lists	1
Total number of shifts	1546
Number of shifts mapped to atoms	1546
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	71%

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: PVN

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	1103	1102	1099	58±7
2	A	43	37	37	6±2
All	All	11460	11390	11360	586

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 26.

5 of 241 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:463:TYR:CE1	1:A:551:ILE:HD11	0.84	2.08	4	5
1:A:510:ALA:HB1	1:A:561:TRP:CD1	0.75	2.16	4	10
1:A:535:ASN:HD21	2:A:600:PVN:H713	0.73	1.44	3	5
1:A:490:ILE:HD13	1:A:490:ILE:N	0.72	2.00	4	1
1:A:528:ARG:N	1:A:529:PRO:CD	0.71	2.52	4	10

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	138/165 (84%)	123±1 (89±1%)	12±2 (9±1%)	3±1 (2±1%)	14 55
All	All	1380/1650 (84%)	1234 (89%)	118 (9%)	28 (2%)	14 55

5 of 7 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	523	HIS	10
1	A	519	LEU	6
1	A	466	ASP	4
1	A	490	ILE	4
1	A	469	TYR	2

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	118/141 (84%)	111±1 (94±1%)	7±1 (6±1%)	27 73
All	All	1180/1410 (84%)	1107 (94%)	73 (6%)	27 73

5 of 16 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	588	GLN	10
1	A	562	GLN	9
1	A	524	LEU	7
1	A	522	CYS	7
1	A	561	TRP	7

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	PVN	A	600	1	37,46,46	2.65±0.01	2±0 (5±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	PVN	A	600	1	41,67,67	3.44±0.01	5±0 (11±1%)

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	PVN	A	600	1	-	0±0,20,74,74	0±0,4,4,4

All 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	A	600	PVN	C10-C11	10.52	1.44	1.35	10	10
2	A	600	PVN	C3-C2	5.40	1.40	1.34	6	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	600	PVN	C91-C18-C19	11.23	131.43	121.47	2	10
2	A	600	PVN	C19-C18-C17	9.63	102.61	108.04	3	10
2	A	600	PVN	C5-C6-C7	6.52	122.18	130.95	8	10
2	A	600	PVN	C15-C16-N4	5.86	138.75	126.15	1	10
2	A	600	PVN	C15-C16-C17	5.03	117.14	126.92	10	7

There are no chirality outliers.

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	600	PVN	C14-C15-C16-C17	3
2	A	600	PVN	C14-C15-C16-N4	3

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)

The completeness of assignment taking into account all chemical shift lists is 71% for the well-defined parts and 65% for the entire structure.

7.1 Chemical shift list 1

File name: 2m7v_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

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	1546
Number of shifts mapped to atoms	1546
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	10

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	138	-0.30 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	134	0.06 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	128	-0.02 ± 0.09	None needed (< 0.5 ppm)
^{15}N	128	0.20 ± 0.33	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 71%, i.e. 1239 atoms were assigned a chemical shift out of a possible 1738. 17 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	592/682 (87%)	237/272 (87%)	240/276 (87%)	115/134 (86%)
Sidechain	585/930 (63%)	345/538 (64%)	224/345 (65%)	16/47 (34%)

Continued on next page...

Continued from previous page...

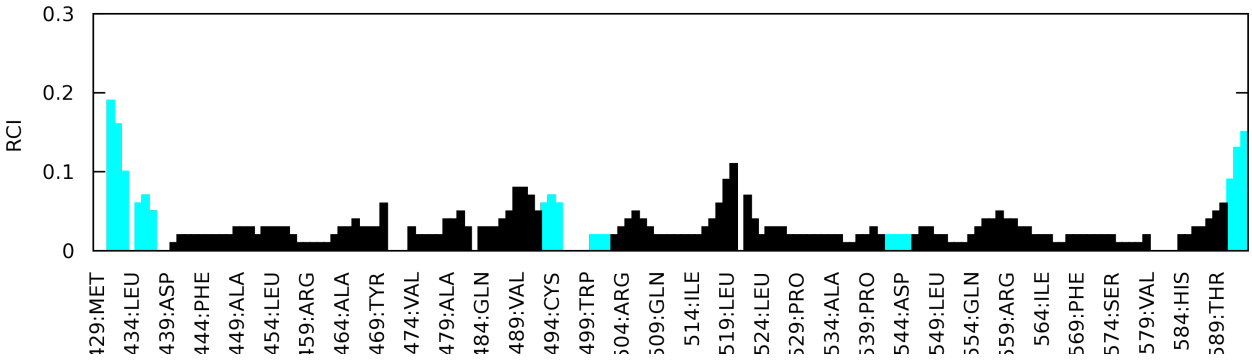
	Total	^{1}H	^{13}C	^{15}N
Aromatic	62/126 (49%)	35/66 (53%)	26/52 (50%)	1/8 (12%)
Overall	1239/1738 (71%)	617/876 (70%)	490/673 (73%)	132/189 (70%)

7.1.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	450	LYS	CD	41.90	34.86 – 23.06	11.0
1	A	536	LEU	HB2	-1.02	3.32 – -0.08	-7.8
1	A	559	ARG	HG2	-0.17	2.92 – 0.22	-6.4
1	A	555	ALA	HB3	-0.10	2.61 – 0.11	-5.9
1	A	555	ALA	HB2	-0.10	2.61 – 0.11	-5.9
1	A	555	ALA	HB1	-0.10	2.61 – 0.11	-5.9
1	A	450	LYS	HD3	2.94	2.75 – 0.45	5.8
1	A	555	ALA	H	11.55	11.19 – 5.19	5.6
1	A	450	LYS	HD2	2.86	2.76 – 0.46	5.4
1	A	455	LEU	HG	-0.15	3.16 – -0.14	-5.0

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: