wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 10:47 pm BST

PDB ID	:	2KSC
Title	:	Solution structure of Synechococcus sp. PCC 7002 hemoglobin
Authors	:	Vuletich, D.A.; Falzone, C.J.; Lecomte, J.T.J.
Deposited on	:	2010-01-02

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:

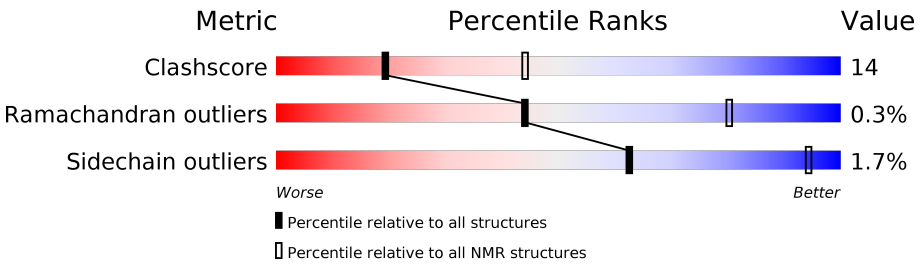
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
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 (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 88%.

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	123	57% 20% • 23%

2 Ensemble composition and analysis (i)

This entry contains 16 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:21-A:56, A:66-A:124 (95)	0.35	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 and 6 single-model clusters were found.

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

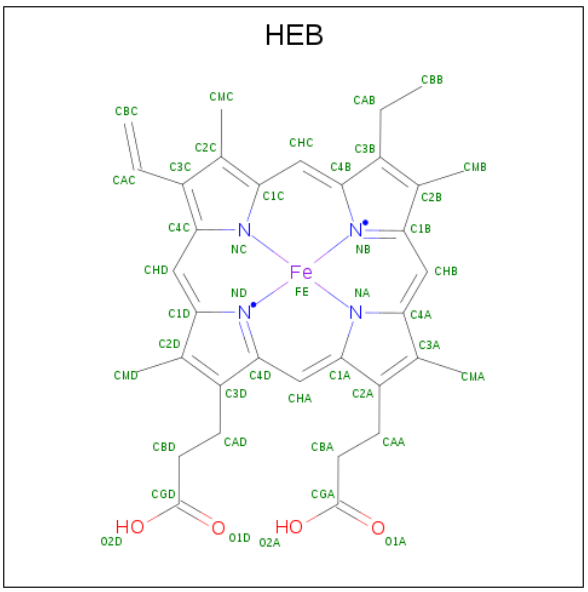
3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1990 atoms, of which 980 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cyanoglobin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	123	1916	605	949	172	187	3	0

- Molecule 2 is HEME B/C (three-letter code: HEB) (formula: $C_{34}H_{34}FeN_4O_4$).



Mol	Chain	Residues	Atoms					
			Total	C	Fe	H	N	O
2	A	1	74	34	1	31	4	4

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: Cyanoglobin

Chain A: 57% 20% • 23%

A2 S3 L4 Y5 E6 K7 L8 G9 G10 A11 A12 A13 V14 D15 L16 A17 V18 E19 K20 L26 R33 V36 D39 Q43 H46 Q47 F50 G56 G57 T58 D59 R60 F61 P62 G63 R64 S65 M66 R67 Q71 D72 L73 A77 T80 D81 V82 H83 Q102

I105 V109 G113 R118 N119 D120 V121 L122 N123 R124

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: Cyanoglobin

Chain A: 63% 15% 23%

A2 S3 L4 Y5 E6 K7 L8 G9 G10 A11 A12 A13 V14 D15 L16 A17 V18 E19 K20 F34 Q43 Q47 T52 Y53 A54 G57 T58 D59 R60 F61 P62 G63 R64 S65 M66 R67 Q71 A77 T80 D81 V82 H83 F84 S101 V109 G113 N123 R124

5 Refinement protocol and experimental data overview (i)

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

Of the 30 calculated structures, 16 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
X-PLOR NIH	structure solution	2.23
X-PLOR NIH	refinement	2.23

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1460
Number of shifts mapped to atoms	1429
Number of unparsed shifts	0
Number of shifts with mapping errors	31
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

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: HEB

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	765	745	745	21±5
2	A	43	31	31	3±1
All	All	12928	12416	12416	358

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 14.

5 of 117 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:LEU:HD13	1:A:27:ALA:N	0.94	1.77	16	1
1:A:26:LEU:O	1:A:26:LEU:HD22	0.82	1.73	16	1
1:A:73:LEU:HD11	2:A:125:HEB:HMD2	0.77	1.56	6	14
1:A:81:ASP:OD2	1:A:122:LEU:HD11	0.73	1.83	11	3
1:A:54:ALA:O	1:A:111:ILE:HG21	0.73	1.83	11	4

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	94/123 (76%)	92±2 (98±2%)	2±1 (2±1%)	0±1 (0±1%)	44 80
All	All	1504/1968 (76%)	1469 (98%)	31 (2%)	4 (0%)	44 80

All 4 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	123	ASN	1
1	A	113	GLY	1
1	A	82	VAL	1
1	A	101	SER	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	83/102 (81%)	82±1 (98±1%)	1±1 (2±1%)	62 94
All	All	1328/1632 (81%)	1306 (98%)	22 (2%)	62 94

5 of 7 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	47	GLN	11
1	A	26	LEU	4
1	A	42	LYS	2
1	A	102	GLN	2
1	A	96	GLN	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 carbohydrates in this entry.

LIGAND-GEOMETRY INFOmissingINFO

6.6 Other polymers (i)

There are no such molecules in this entry.

6.7 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 88% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

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	1460
Number of shifts mapped to atoms	1429
Number of unparsed shifts	0
Number of shifts with mapping errors	31
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	39

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Chain not found in structure. First 5 (of 31) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	HEB	HMB1	11.0	0.02	1
UNMAPPED	1	HEB	HBD1	0.46	0.02	2
UNMAPPED	1	HEB	HMD2	19.41	0.02	1
UNMAPPED	1	HEB	HMA1	13.27	0.02	1
UNMAPPED	1	HEB	HBB3	0.82	0.02	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	116	-0.99 ± 0.21	Should be applied
$^{13}C_{\beta}$	109	0.28 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	115	-0.70 ± 0.14	Should be applied

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	118	0.15 ± 0.27	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 88%, i.e. 1049 atoms were assigned a chemical shift out of a possible 1187. 14 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	475/475 (100%)	190/190 (100%)	190/190 (100%)	95/95 (100%)
Sidechain	494/614 (80%)	301/352 (86%)	175/229 (76%)	18/33 (55%)
Aromatic	80/98 (82%)	44/54 (81%)	34/40 (85%)	2/4 (50%)
Overall	1049/1187 (88%)	535/596 (90%)	399/459 (87%)	115/132 (87%)

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.

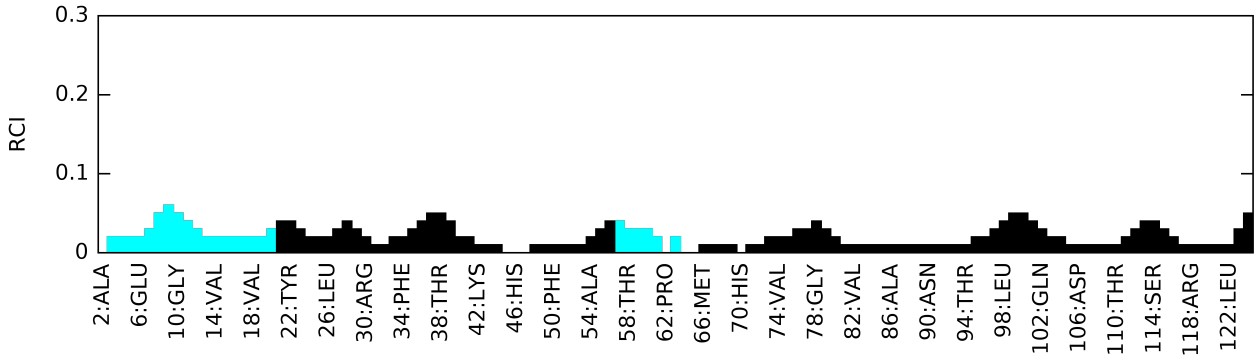
Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	46	HIS	HB2	11.84	4.91 – 1.31	24.2
1	A	70	HIS	HB2	9.92	4.91 – 1.31	18.9
1	A	70	HIS	HB3	9.45	5.00 – 1.10	16.4
1	A	46	HIS	HB3	9.20	5.00 – 1.10	15.8
1	A	121	VAL	HG12	4.15	2.13 – -0.47	12.8
1	A	121	VAL	HG13	4.15	2.13 – -0.47	12.8
1	A	121	VAL	HG11	4.15	2.13 – -0.47	12.8
1	A	46	HIS	CA	82.50	68.24 – 44.74	11.1
1	A	50	PHE	HD2	3.76	8.56 – 5.56	-11.0
1	A	50	PHE	HD1	3.76	8.56 – 5.56	-11.0
1	A	46	HIS	HA	9.41	6.81 – 2.41	10.9
1	A	121	VAL	HG22	3.78	2.20 – -0.60	10.6
1	A	121	VAL	HG21	3.78	2.20 – -0.60	10.6
1	A	121	VAL	HG23	3.78	2.20 – -0.60	10.6
1	A	69	ALA	HB3	-0.81	2.61 – 0.11	-8.7
1	A	69	ALA	HB2	-0.81	2.61 – 0.11	-8.7
1	A	69	ALA	HB1	-0.81	2.61 – 0.11	-8.7
1	A	50	PHE	HE2	4.54	8.69 – 5.49	-8.0
1	A	50	PHE	HE1	4.54	8.69 – 5.49	-8.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	87	ILE	HG21	-1.01	2.13 – -0.57	-6.6
1	A	87	ILE	HG22	-1.01	2.13 – -0.57	-6.6
1	A	87	ILE	HG23	-1.01	2.13 – -0.57	-6.6
1	A	87	ILE	HD11	-1.23	2.13 – -0.77	-6.6
1	A	87	ILE	HD12	-1.23	2.13 – -0.77	-6.6
1	A	87	ILE	HD13	-1.23	2.13 – -0.77	-6.6
1	A	70	HIS	CA	71.70	68.24 – 44.74	6.5
1	A	47	GLN	HA	6.96	6.41 – 2.11	6.3
1	A	66	MET	HB3	4.09	3.70 – 0.30	6.1
1	A	121	VAL	HA	7.49	7.09 – 1.29	5.7
1	A	73	LEU	HD13	-0.80	2.16 – -0.64	-5.6
1	A	73	LEU	HD12	-0.80	2.16 – -0.64	-5.6
1	A	73	LEU	HD11	-0.80	2.16 – -0.64	-5.6
1	A	87	ILE	HB	0.23	3.24 – 0.34	-5.4
1	A	33	ARG	HG3	0.02	3.00 – 0.10	-5.3
1	A	47	GLN	HB2	3.35	3.30 – 0.80	5.2
1	A	79	LEU	HD11	-0.69	2.16 – -0.64	-5.2
1	A	79	LEU	HD13	-0.69	2.16 – -0.64	-5.2
1	A	79	LEU	HD12	-0.69	2.16 – -0.64	-5.2
1	A	121	VAL	HB	3.61	3.59 – 0.39	5.1

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: