Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 08:38 PM BST

PDB ID	:	2I8F
Title	:	Solution Conformation of the H47A Mutant of Pseudomonas stutzeri ZoBell Ferrocytochrome c-551
Authors	:	Liang, Q.; Timkovich, R.
Deposited on	:	2006-09-01

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
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:

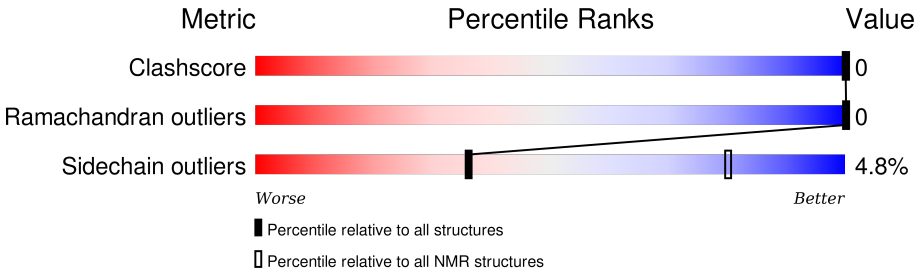
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 49%.

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	82	96% .

2 Ensemble composition and analysis

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1275 atoms, of which 637 are hydrogens and 0 are deuteriums.

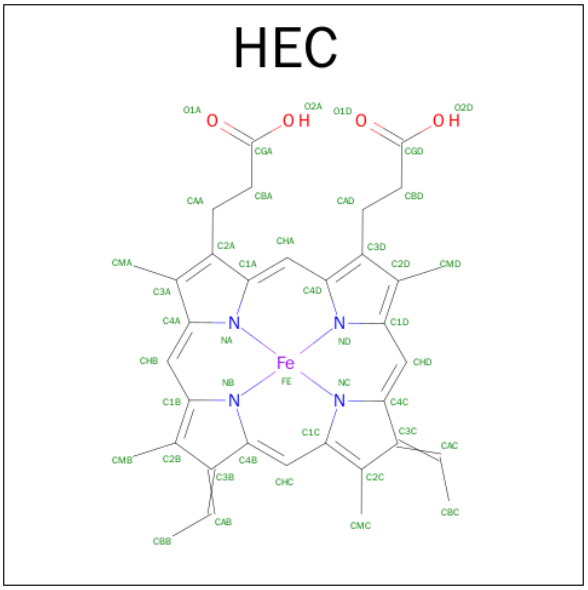
- Molecule 1 is a protein called Cytochrome c-551.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	82	1200	377	605	99	115	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	47	ALA	HIS	ENGINEERED	UNP P00101

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



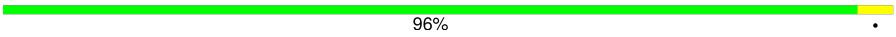
Mol	Chain	Residues	Atoms					
			Total	C	Fe	H	N	O
2	A	1	75	34	1	32	4	4

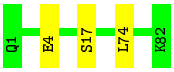
4 Residue-property plots (i)

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: Cytochrome c-551

Chain A:





5 Refinement protocol and experimental data overview

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

Of the ? calculated structures, 1 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
ARIA	refinement	2.0 alpha
CNS	refinement	1.1

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)	BMRB entry 7296
Number of chemical shift lists	1
Total number of shifts	596
Number of shifts mapped to atoms	564
Number of unparsed shifts	0
Number of shifts with mapping errors	32
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	49%

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: HEC

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

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
All	All	638	637	635	-

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 -.

There are no clashes.

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	80/82 (98%)	71 (89%)	9 (11%)	0 (0%)	100	100
All	All	80/82 (98%)	71 (89%)	9 (11%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	62/62 (100%)	59 (95%)	3 (5%)	36 80
All	All	62/62 (100%)	59 (95%)	3 (5%)	36 80

All 3 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	74	LEU
1	A	17	SER
1	A	4	GLU

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

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	HEC	A	83	1	24,50,50	1.21	0 (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	HEC	A	83	1	19,82,82	2.65	2 (10%)

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	HEC	A	83	1	-	0,6,54,54	0,0,8,8

There are no bond-length outliers.

All angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	83	HEC	CBB-CAB-C3B	8.01	109.83	127.34
2	A	83	HEC	CBC-CAC-C3C	7.30	111.39	127.34

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.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 49% for the well-defined parts and 49% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 7296

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	596
Number of shifts mapped to atoms	564
Number of unparsed shifts	0
Number of shifts with mapping errors	32
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	17

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

- Chain not found in structure. All 32 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	HEM	HMA2	3.424	0.03	1
UNMAPPED	1	HEM	HBD2	3.419	0.03	2
UNMAPPED	1	HEM	HBB1	1.908	0.03	1
UNMAPPED	1	HEM	HMA1	3.424	0.03	1
UNMAPPED	1	HEM	HMC1	3.709	0.03	1
UNMAPPED	1	HEM	HMB2	3.77	0.03	1
UNMAPPED	1	HEM	HBC2	2.446	0.03	1
UNMAPPED	1	HEM	HBC3	2.446	0.03	1
UNMAPPED	1	HEM	HBD1	2.637	0.03	2
UNMAPPED	1	HEM	HHC1	9.776	0.03	1
UNMAPPED	1	HEM	HMC2	3.709	0.03	1
UNMAPPED	1	HEM	HMB3	3.77	0.03	1
UNMAPPED	1	HEM	HAA1	3.927	0.03	2
UNMAPPED	1	HEM	HBB3	1.908	0.03	1

Continued on next page...

Continued from previous page...

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	HEM	HAB1	5.992	0.03	1
UNMAPPED	1	HEM	HMC3	3.709	0.03	1
UNMAPPED	1	HEM	HMD1	3.308	0.03	1
UNMAPPED	1	HEM	HMD2	3.308	0.03	1
UNMAPPED	1	HEM	HBA1	2.637	0.03	2
UNMAPPED	1	HEM	HBB2	1.908	0.03	1
UNMAPPED	1	HEM	HBC1	2.446	0.03	1
UNMAPPED	1	HEM	HAD1	4.242	0.03	2
UNMAPPED	1	HEM	HHA1	9.277	0.03	1
UNMAPPED	1	HEM	HMB1	3.77	0.03	1
UNMAPPED	1	HEM	HHD1	9.311	0.03	1
UNMAPPED	1	HEM	HBA2	3.231	0.03	2
UNMAPPED	1	HEM	HHB1	9.178	0.03	1
UNMAPPED	1	HEM	HMA3	3.424	0.03	1
UNMAPPED	1	HEM	HAC1	6.21	0.03	1
UNMAPPED	1	HEM	HAA2	4.436	0.03	2
UNMAPPED	1	HEM	HAD2	4.597	0.03	2
UNMAPPED	1	HEM	HMD3	3.308	0.03	1

7.1.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

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 49%, i.e. 452 atoms were assigned a chemical shift out of a possible 921. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	156/396 (39%)	156/157 (99%)	0/164 (0%)	0/75 (0%)
Sidechain	276/485 (57%)	276/284 (97%)	0/188 (0%)	0/13 (0%)
Aromatic	20/40 (50%)	20/21 (95%)	0/16 (0%)	0/3 (0%)
Overall	452/921 (49%)	452/462 (98%)	0/368 (0%)	0/91 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 49%, i.e. 452 atoms were assigned a chemical shift out of a possible 921. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	156/396 (39%)	156/157 (99%)	0/164 (0%)	0/75 (0%)
Sidechain	276/485 (57%)	276/284 (97%)	0/188 (0%)	0/13 (0%)
Aromatic	20/40 (50%)	20/21 (95%)	0/16 (0%)	0/3 (0%)
Overall	452/921 (49%)	452/462 (98%)	0/368 (0%)	0/91 (0%)

7.1.4 Statistically unusual chemical shifts (i)

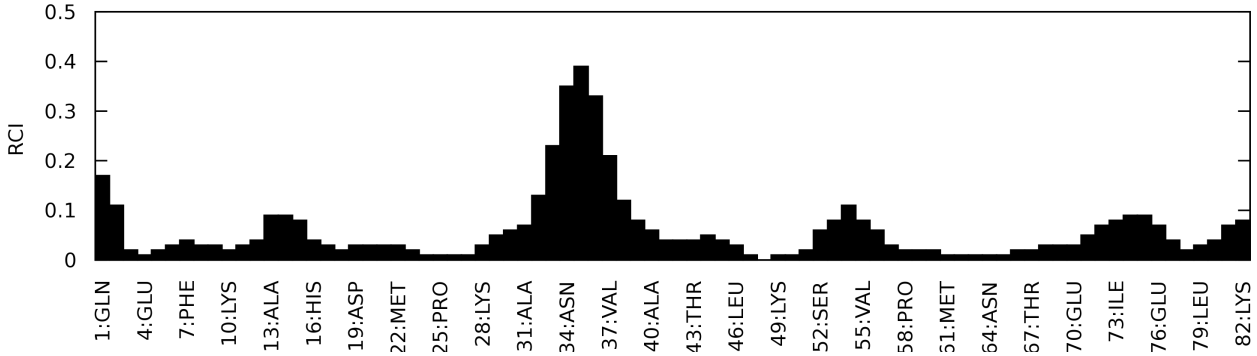
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	61	MET	HG2	-3.67	4.23 – 0.63	-16.9
1	A	16	HIS	HD2	0.64	9.28 – 4.78	-14.2
1	A	16	HIS	HE1	0.79	10.53 – 5.43	-14.1
1	A	61	MET	HB2	-2.76	3.73 – 0.33	-14.1
1	A	61	MET	HE1	-2.91	4.28 – -0.52	-10.0
1	A	61	MET	HE2	-2.91	4.28 – -0.52	-10.0
1	A	61	MET	HE3	-2.91	4.28 – -0.52	-10.0
1	A	24	GLY	HA3	0.19	5.80 – 2.00	-9.8
1	A	64	ASN	HD21	3.19	9.74 – 4.94	-8.6
1	A	61	MET	HB3	-0.90	3.70 – 0.30	-8.5
1	A	61	MET	HG3	-0.54	4.30 – 0.50	-7.7
1	A	25	PRO	HG2	-0.07	3.48 – 0.38	-6.4
1	A	27	LEU	HD12	-0.82	2.16 – -0.64	-5.6
1	A	27	LEU	HD13	-0.82	2.16 – -0.64	-5.6
1	A	27	LEU	HD11	-0.82	2.16 – -0.64	-5.6
1	A	25	PRO	HG3	0.11	3.56 – 0.26	-5.4
1	A	16	HIS	HB2	1.18	4.91 – 1.31	-5.4

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:


RCI
0.5
0.4
0.3
0.2
0.1
0
1:GLN
4:GLU
7:PHE
10:LYS
13:ALA
16:HIS
19:ASP
22:MET
25:PRO
28:LYS
31:ALA
34:ASN
37:VAL
40:ALA
43:THR
46:LEU
49:LYS
52:SER
55:VAL
58:PRO
61:MET
64:ASN
67:THR
70:GLU
73:ILE
76:GLU
79:LEU
82:LYS