wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:30 PM BST

PDB ID	:	1EKY
Title	:	MODEL STRUCTURE FROM NON-NOE BASED NMR STRUCTURE CALCULATION
Authors	:	Hus, J.C.; Marion, D.; Blackledge, M.
Deposited on	:	2000-03-10

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:

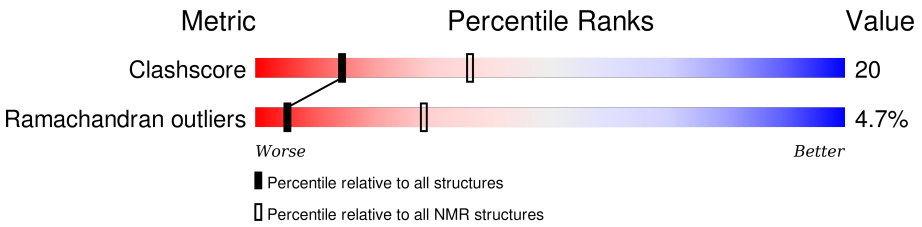
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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

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	129	58% 13% 29%

2 Ensemble composition and analysis (i)

This entry contains 32 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 15 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:4-A:48, A:77-A:123 (92)	0.57	20

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 4 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 16, 17, 18, 19, 20, 21, 23, 25, 26, 27, 29, 30, 32
2	8, 24
3	15, 31
4	12, 28
Single-model clusters	22

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 900 atoms, of which 268 are hydrogens and 0 are deuteriums.

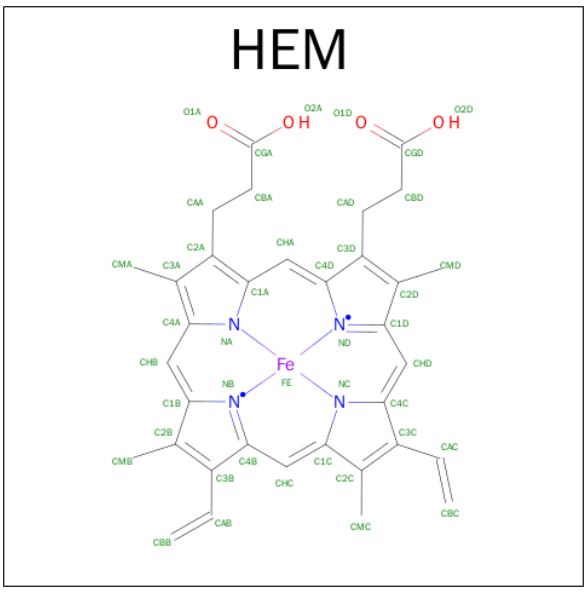
- Molecule 1 is a protein called CYTOCHROME C'.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	129	899	373	268	129	129	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	GLY	LYS	VARIANT	UNP P00147
A	30	ALA	SER	VARIANT	UNP P00147
A	38	VAL	ALA	VARIANT	UNP P00147
A	42	LYS	ALA	VARIANT	UNP P00147
A	76	ALA	THR	VARIANT	UNP P00147
A	79	ASP	ALA	VARIANT	UNP P00147
A	89	HIS	ASN	VARIANT	UNP P00147
A	90	GLU	ASP	VARIANT	UNP P00147
A	93	GLY	ALA	VARIANT	UNP P00147
A	94	ALA	GLU	VARIANT	UNP P00147
A	104	GLY	ALA	VARIANT	UNP P00147
A	105	ALA	THR	VARIANT	UNP P00147

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



Mol	Chain	Residues	Atoms	
			Total	Fe
2	A	1	1	1

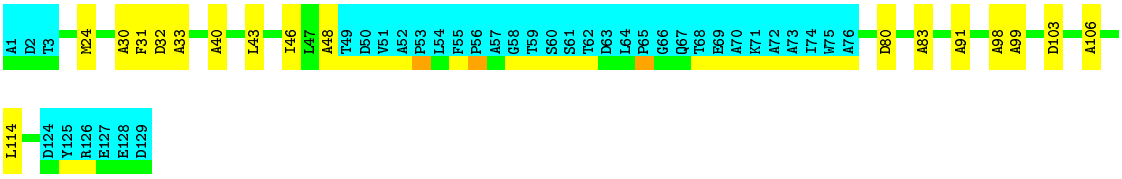
4 Residue-property plots (i)

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: CYTOCHROME C'

Chain A: 58% 13% 29%

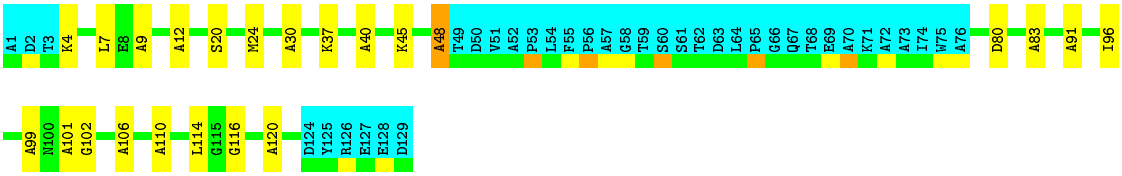


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 20. Colouring as in section 4.1 above.

- Molecule 1: CYTOCHROME C'

Chain A: 53% 17% • 29%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *in-house modification of discover code.*

Of the 300 calculated structures, 32 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
DISCOVER	structure solution	modified version of discover
DISCOVER	refinement	modified version of discover

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: HEM

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	0.31±0.01	0±0/448 (0.0±0.0%)	0.52±0.03	0±0/620 (0.0±0.0%)
All	All	0.31	0/14336 (0.0%)	0.52	1/19840 (0.0%)

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	47	LEU	N-CA-C	-6.43	93.64	111.00	8	1

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	448	196	268	14±4
All	All	14368	6272	8576	457

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 20.

5 of 120 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:44:GLU:O	1:A:48:ALA:HB2	1.23	1.33	19	2
1:A:24:MET:O	1:A:28:ALA:CB	1.14	1.95	27	10
1:A:20:SER:O	1:A:24:MET:CB	1.03	2.07	29	10
1:A:36:ALA:HB1	1:A:99:ALA:HB1	1.00	1.26	6	1
1:A:106:ALA:O	1:A:110:ALA:CB	0.96	2.14	7	9

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	92/129 (71%)	77±3 (83±3%)	11±3 (12±3%)	4±2 (5±2%)	5 28
All	All	2944/4128 (71%)	2455 (83%)	352 (12%)	137 (5%)	5 28

5 of 13 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	30	ALA	26
1	A	33	ALA	21
1	A	32	ASP	15
1	A	103	ASP	13
1	A	31	PHE	13

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	0	-	-	-
All	All	0	-	-	-

There are no protein residues with a non-rotameric sidechain to report.

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)

Of 1 ligands modelled in this entry, 1 is modelled with single atom - leaving 0 for Mogul analysis.

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