wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 11:13 pm BST

PDB ID	:	2LNH
Title	:	Enterohaemorrhagic E. coli (EHEC) exploits a tryptophan switch to hijack host F-actin assembly
Authors	:	Aitio, O.; Hellman, M.; Skehan, B.; Kesti, T.; Leong, J.M.; Saksela, K.; Permi, P.
Deposited on	:	2011-12-28

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 (i)) were used in the production of this report:

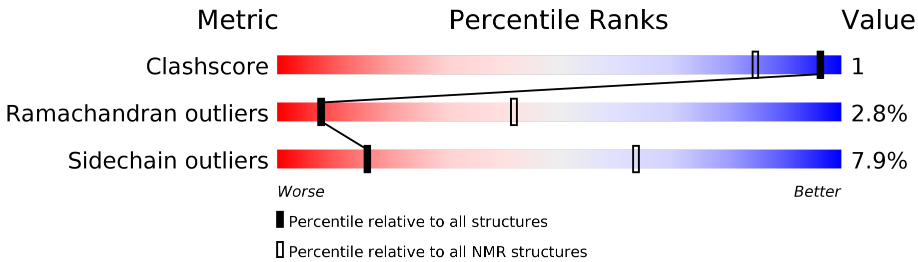
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 84%.

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	65	68% 12% 20%
2	B	67	69% 6% 25%
3	C	48	63% 6% 29% •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:12-A:63, C:133-C:153 (73)	0.12	4
2	B:73-B:80, B:86-B:127, C:159-C:170 (62)	0.10	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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
2	8, 20

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2782 atoms, of which 1371 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Neural Wiskott-Aldrich syndrome protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	65	993	320	483	86	102	2	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP O00401

- Molecule 2 is a protein called Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	67	1065	344	532	89	99	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	66	GLY	-	EXPRESSION TAG	UNP Q9UHR4
B	67	SER	-	EXPRESSION TAG	UNP Q9UHR4
B	68	HIS	-	EXPRESSION TAG	UNP Q9UHR4

- Molecule 3 is a protein called Secreted effector protein EspF(U).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	C	47	724	230	356	71	65	2	0

There is a discrepancy between the modelled and reference sequences:

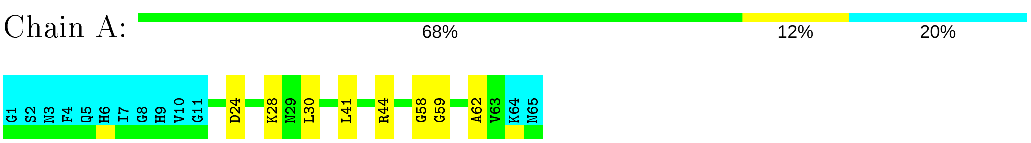
Chain	Residue	Modelled	Actual	Comment	Reference
C	132	GLY	-	EXPRESSION TAG	UNP P0DJ89

4 Residue-property plots (i)

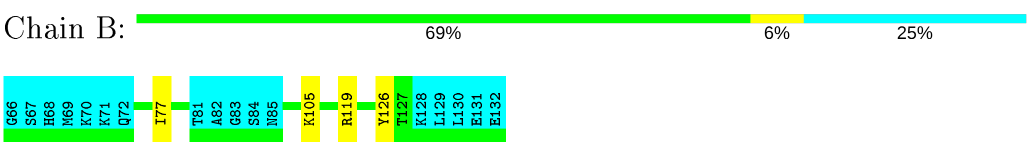
4.1 Average score per residue in the NMR ensemble

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: Neural Wiskott-Aldrich syndrome protein



- Molecule 2: Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1

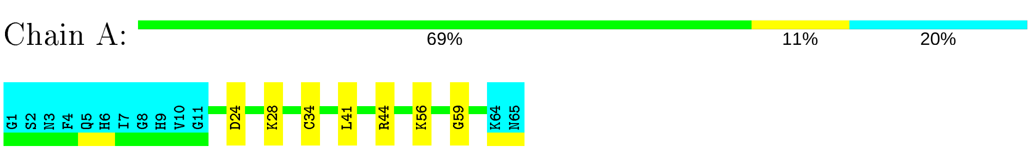


- Molecule 3: Secreted effector protein EspF(U)

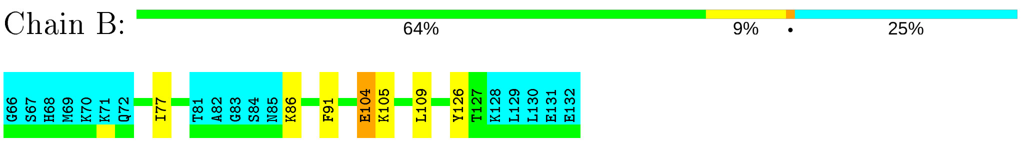
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Neural Wiskott-Aldrich syndrome protein



- Molecule 2: Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1



- Molecule 3: Secreted effector protein EspF(U)

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
AMBER	refinement	8.0
CYANA	refinement	2.1

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1921
Number of shifts mapped to atoms	1921
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	412	393	393	2±1
2	B	405	401	401	0±0
3	C	255	254	253	1±1
All	All	21440	20960	20940	35

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 1.

5 of 8 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:LEU:HD11	3:C:140:LEU:HD22	0.57	1.77	12	6
1:A:46:THR:HG23	3:C:149:ILE:HG23	0.50	1.84	8	5
1:A:58:GLY:O	1:A:62:ALA:HB3	0.49	2.07	3	12
1:A:30:LEU:HD21	3:C:140:LEU:HD22	0.48	1.84	10	1
1:A:31:PHE:CD2	1:A:41:LEU:HD21	0.44	2.47	9	5

5.2 Torsion angles (i)

5.2.1 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	52/65 (80%)	47±1 (90±1%)	4±1 (7±2%)	1±0 (2±1%)	9	48
2	B	50/67 (75%)	49±1 (97±2%)	1±1 (3±2%)	0±0 (0±0%)	100	100
3	C	32/48 (67%)	29±1 (91±3%)	0±1 (1±2%)	3±1 (8±2%)	2	14
All	All	2680/3600 (74%)	2494 (93%)	112 (4%)	74 (3%)	8	42

5 of 6 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	59	GLY	20
3	C	162	ALA	20
3	C	161	PRO	20
3	C	150	GLN	5
3	C	159	ILE	5

5.2.2 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	46/56 (82%)	41±1 (89±3%)	5±1 (11±3%)	10	55
2	B	44/58 (76%)	41±1 (93±2%)	3±1 (7±2%)	18	66
3	C	27/40 (68%)	26±1 (95±3%)	1±1 (5±3%)	29	78
All	All	2340/3080 (76%)	2154 (92%)	186 (8%)	16	63

5 of 29 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)
2	B	105	LYS	20
1	A	28	LYS	20
2	B	77	ILE	20
1	A	24	ASP	19
1	A	44	ARG	17

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 79% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	1921
Number of shifts mapped to atoms	1921
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	16

6.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	171	-1.02 ± 0.14	Should be applied
$^{13}C_\beta$	160	-0.58 ± 0.10	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	148	0.49 ± 0.31	None needed (< 0.5 ppm)

6.1.3 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 84%, i.e. 1394 atoms were assigned a chemical shift out of a possible 1666. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	507/649 (78%)	254/257 (99%)	134/270 (50%)	119/122 (98%)
Sidechain	750/863 (87%)	467/509 (92%)	278/323 (86%)	5/31 (16%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	137/154 (89%)	71/82 (87%)	62/64 (97%)	4/8 (50%)
Overall	1394/1666 (84%)	792/848 (93%)	474/657 (72%)	128/161 (80%)

6.1.4 Statistically unusual chemical shifts ⓘ

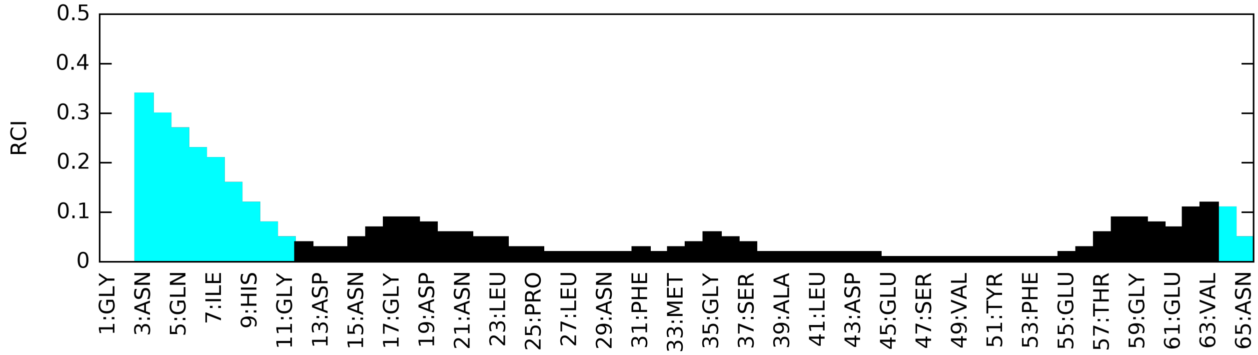
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
3	C	163	PRO	HG3	-1.15	3.56 – 0.26	-9.3
3	C	163	PRO	HB2	-0.62	3.82 – 0.32	-7.7
3	C	163	PRO	HB3	-0.71	3.81 – 0.21	-7.6
3	C	163	PRO	HD3	0.94	5.52 – 1.72	-7.1
2	B	110	TYR	HB3	0.20	4.75 – 0.95	-7.0
3	C	162	ALA	HB1	-0.22	2.61 – 0.11	-6.3
3	C	162	ALA	HB3	-0.22	2.61 – 0.11	-6.3
3	C	162	ALA	HB2	-0.22	2.61 – 0.11	-6.3
3	C	162	ALA	HA	1.64	6.46 – 2.06	-6.0
3	C	167	ALA	HB2	-0.03	2.61 – 0.11	-5.5
3	C	167	ALA	HB3	-0.03	2.61 – 0.11	-5.5
3	C	167	ALA	HB1	-0.03	2.61 – 0.11	-5.5
2	B	76	THR	HG23	-0.13	2.29 – -0.01	-5.5
2	B	76	THR	HG22	-0.13	2.29 – -0.01	-5.5
2	B	76	THR	HG21	-0.13	2.29 – -0.01	-5.5
3	C	163	PRO	HG2	0.28	3.48 – 0.38	-5.3

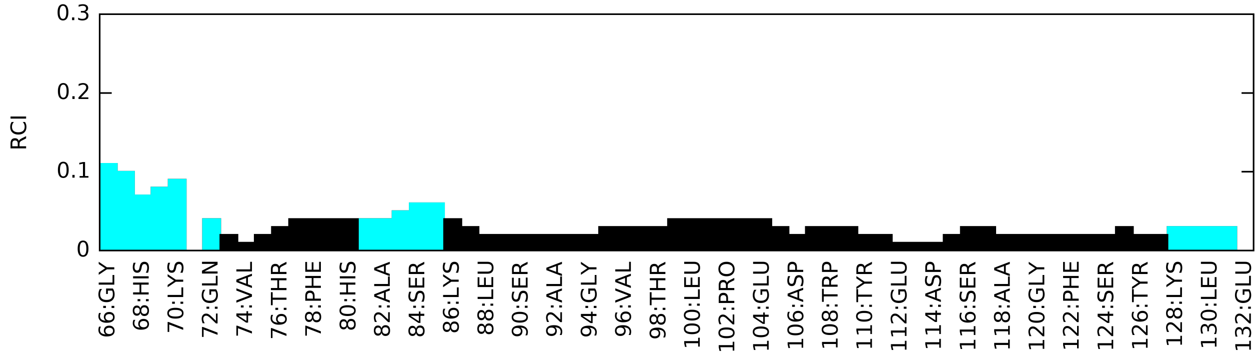
6.1.5 Random Coil Index (RCI) plots ⓘ

The images below report *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:



Random coil index (RCI) for chain B:



Random coil index (RCI) for chain C: