wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 05:37 PM EDT

PDB ID	:	2LBU
BMRB ID	:	11064
Title	:	HADDOCK calculated model of Congo red bound to the HET-s amyloid
Authors	:	Schutz, A.K.; Soragni, A.; Hornemann, S.; Aguzzi, A.; Ernst, M.; Bockmann, A.; Meier, B.H.
Deposited on	:	2011-04-07

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

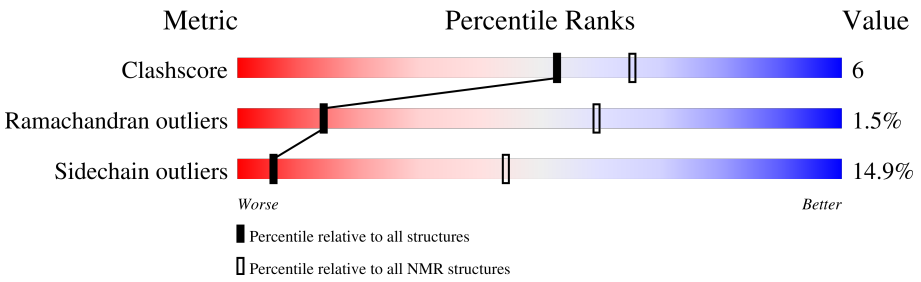
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLID-STATE NMR

The overall completeness of chemical shifts assignment is 7%.

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	79	80% 10% 10%
1	B	79	78% 11% 10%
1	C	79	68% 16% • • 10%
1	D	79	73% 15% • 10%
1	E	79	67% 18% 5% 10%

2 Ensemble composition and analysis (i)

This entry contains 4 models. Model 3 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:225-A:295, B:225-B:295, C:225-C:293, D:226-D:295, E:225-E:251, E:256-E:295 (348)	0.47	3

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. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3558 atoms, of which 747 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Small s protein.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	71	695	334	142	114	105	0
1	B	71	695	334	142	114	105	0
1	C	71	700	334	147	114	105	0
1	D	71	702	334	149	114	105	0
1	E	71	698	334	145	114	105	0

There are 35 discrepancies between the modelled and reference sequences:

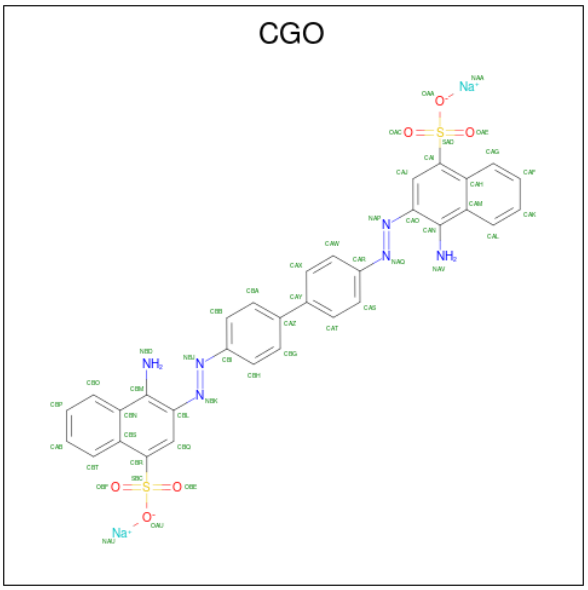
Chain	Residue	Modelled	Actual	Comment	Reference
A	217	MET	-	expression tag	UNP Q03689
A	290	HIS	-	expression tag	UNP Q03689
A	291	HIS	-	expression tag	UNP Q03689
A	292	HIS	-	expression tag	UNP Q03689
A	293	HIS	-	expression tag	UNP Q03689
A	294	HIS	-	expression tag	UNP Q03689
A	295	HIS	-	expression tag	UNP Q03689
B	217	MET	-	expression tag	UNP Q03689
B	290	HIS	-	expression tag	UNP Q03689
B	291	HIS	-	expression tag	UNP Q03689
B	292	HIS	-	expression tag	UNP Q03689
B	293	HIS	-	expression tag	UNP Q03689
B	294	HIS	-	expression tag	UNP Q03689
B	295	HIS	-	expression tag	UNP Q03689
C	217	MET	-	expression tag	UNP Q03689
C	290	HIS	-	expression tag	UNP Q03689
C	291	HIS	-	expression tag	UNP Q03689
C	292	HIS	-	expression tag	UNP Q03689
C	293	HIS	-	expression tag	UNP Q03689
C	294	HIS	-	expression tag	UNP Q03689
C	295	HIS	-	expression tag	UNP Q03689
D	217	MET	-	expression tag	UNP Q03689
D	290	HIS	-	expression tag	UNP Q03689

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	291	HIS	-	expression tag	UNP Q03689
D	292	HIS	-	expression tag	UNP Q03689
D	293	HIS	-	expression tag	UNP Q03689
D	294	HIS	-	expression tag	UNP Q03689
D	295	HIS	-	expression tag	UNP Q03689
E	217	MET	-	expression tag	UNP Q03689
E	290	HIS	-	expression tag	UNP Q03689
E	291	HIS	-	expression tag	UNP Q03689
E	292	HIS	-	expression tag	UNP Q03689
E	293	HIS	-	expression tag	UNP Q03689
E	294	HIS	-	expression tag	UNP Q03689
E	295	HIS	-	expression tag	UNP Q03689

- Molecule 2 is sodium 3,3'-(1E,1'E)-biphenyl-4,4'-diylbis(diazene-2,1-diyl)bis(4-aminonaphthalene-1-sulfonate) (three-letter code: CGO) (formula: $C_{32}H_{22}N_6Na_2O_6S_2$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	S
2	B	1	68	32	22	6	6	2

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Small s protein

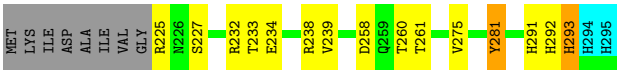
Chain A:

- Molecule 1: Small s protein

Chain B:

- Molecule 1: Small s protein

Chain C:



- Molecule 1: Small s protein

Chain D:

- Molecule 1: Small s protein

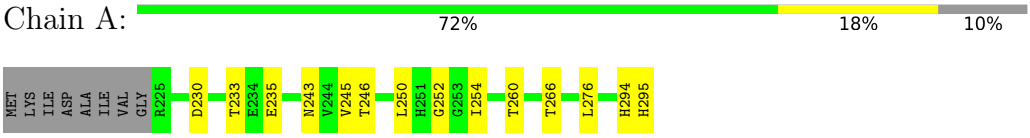
Chain E:

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

The representative model is number 3. Colouring as in section 4.1 above.

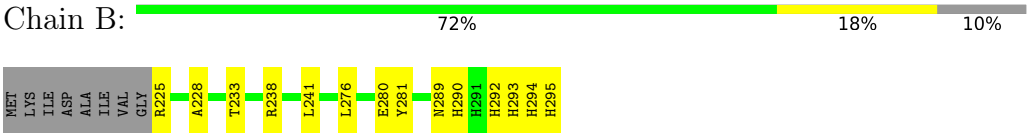
- Molecule 1: Small s protein

Chain A:



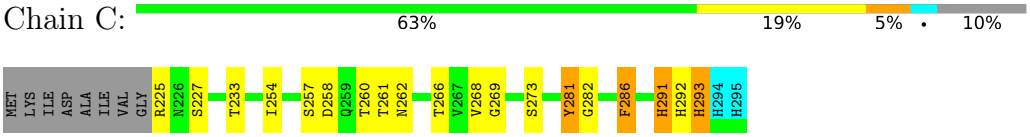
- Molecule 1: Small s protein

Chain B:



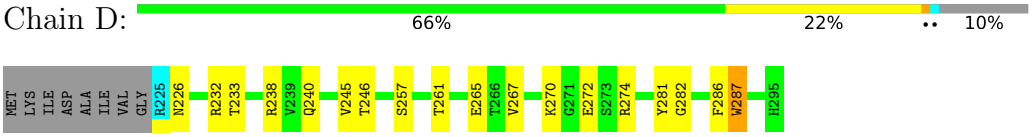
- Molecule 1: Small s protein

Chain C:



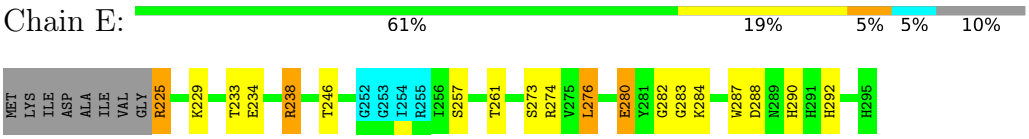
- Molecule 1: Small s protein

Chain D:



- Molecule 1: Small s protein

Chain E:



5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 4 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
HADDOCK	structure solution	
HADDOCK	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	326
Number of shifts mapped to atoms	310
Number of unparsed shifts	0
Number of shifts with mapping errors	16
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	7%

Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

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: CGO

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	553	142	528	2±1
1	B	553	142	528	6±3
1	C	532	141	514	8±2
1	D	542	143	516	10±2
1	E	526	136	498	9±2
2	B	46	22	22	6±2
All	All	11008	2904	10424	119

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 6.

5 of 87 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:1:CGO:HBT	2:B:1:CGO:OBE	0.87	1.69	4	2
1:E:280:GLU:HB2	1:E:288:ASP:HA	0.83	1.47	3	3
1:E:283:GLY:HA2	1:E:290:HIS:HB3	0.78	1.55	4	4
2:B:1:CGO:HBT	2:B:1:CGO:OAU	0.77	1.78	3	2
1:E:238:ARG:HD3	1:E:274:ARG:HG3	0.74	1.58	3	1

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	69/79 (87%)	54±4 (79±5%)	12±4 (18±6%)	2±1 (3±1%)	6 37
1	B	69/79 (87%)	56±2 (82±3%)	13±2 (18±3%)	0±0 (0±0%)	100 100
1	C	68/79 (86%)	57±0 (83±1%)	10±1 (15±1%)	1±1 (1±1%)	14 59
1	D	69/79 (87%)	55±1 (80±2%)	12±1 (18±1%)	2±1 (2±2%)	10 49
1	E	65/79 (82%)	52±1 (79±2%)	13±1 (20±2%)	0±0 (1±1%)	24 71
All	All	1360/1580 (86%)	1097 (81%)	242 (18%)	21 (2%)	14 59

5 of 12 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	254	ILE	4
1	C	293	HIS	3
1	A	250	LEU	2
1	A	287	TRP	2
1	D	257	SER	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	57/63 (90%)	50±1 (87±1%)	7±1 (13±1%)	8 49
1	B	57/63 (90%)	47±1 (82±2%)	10±1 (18±2%)	4 39
1	C	55/63 (87%)	44±2 (81±3%)	10±2 (19±3%)	4 36
1	D	56/63 (89%)	51±0 (91±1%)	5±0 (9±1%)	12 58
1	E	55/63 (87%)	46±1 (84±3%)	9±1 (16±3%)	5 42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1120/1260 (89%)	953 (85%)	167 (15%)	6 44

5 of 90 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	233	THR	4
1	A	276	LEU	4
1	B	238	ARG	4
1	B	276	LEU	4
1	C	233	THR	4

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 monosaccharides 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	CGO	B	1	-	48,51,53	1.68±0.01	10±0 (20±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	CGO	B	1	-	60,76,78	1.10±0.01	4±0 (6±0%)

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	CGO	B	1	-	-	0±0,22,26,32	0±0,6,6,6

5 of 10 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	B	1	CGO	CBL-NBK	4.25	1.33	1.42	4	4
2	B	1	CGO	CAO-NAP	4.24	1.33	1.42	3	4
2	B	1	CGO	CAN-CAM	3.73	1.39	1.44	2	4
2	B	1	CGO	CBM-CBN	3.69	1.39	1.44	1	4
2	B	1	CGO	CAR-NAQ	2.57	1.33	1.44	3	4

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	B	1	CGO	CAO-NAP-NAQ	3.43	122.48	114.56	3	4
2	B	1	CGO	CAG-CAH-CAI	2.90	119.80	123.60	4	4
2	B	1	CGO	CBT-CBS-CBR	2.76	119.98	123.60	2	4
2	B	1	CGO	CBL-NBK-NBJ	2.60	120.56	114.56	4	4

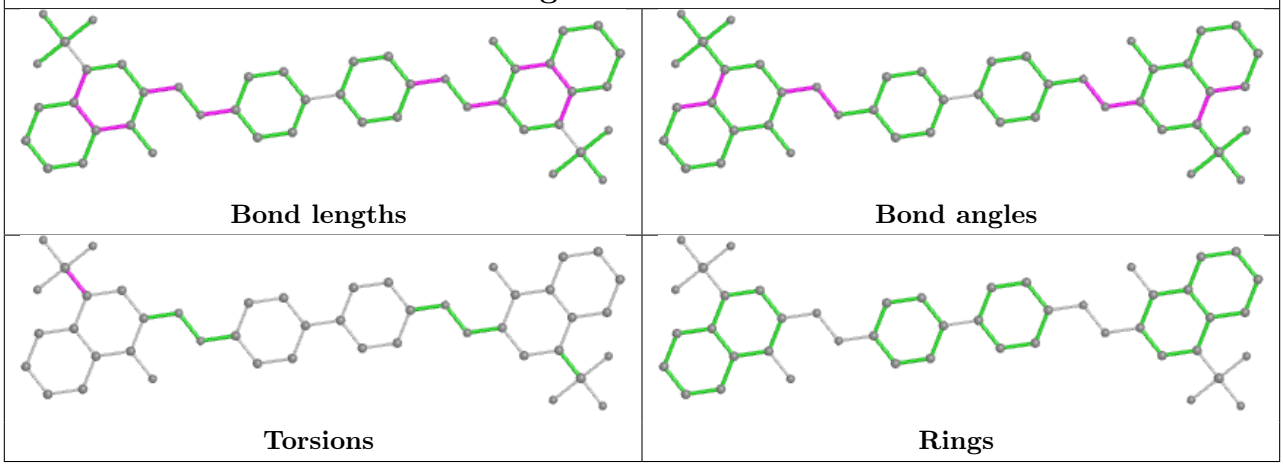
There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand CGO B 1



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 7% for the well-defined parts and 6% for the entire structure.

7.1 Chemical shift list 1

File name: working_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	326
Number of shifts mapped to atoms	310
Number of unparsed shifts	0
Number of shifts with mapping errors	16
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- No matching atom found in the structure. First 5 (of 16) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	222	ILE	C	175.1	0.2	1
1	A	222	ILE	CA	62.0	0.2	1
1	A	222	ILE	CB	38.3	0.2	1
1	A	222	ILE	CD1	13.0	0.2	1
1	A	222	ILE	CG1	27.5	0.2	1
1	A	222	ILE	CG2	18.7	0.2	1
1	A	222	ILE	N	128.9	0.2	9
1	A	223	VAL	C	175.95	0.2	9
1	A	223	VAL	CA	59.1	0.2	1
1	A	223	VAL	CB	32.4	0.2	1
1	A	223	VAL	CG1	22.2	0.2	9
1	A	223	VAL	CG2	19.7	0.2	9
1	A	223	VAL	N	125.4	0.2	1
1	A	224	GLY	C	178.5	0.2	9

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	224	GLY	CA	43.8	0.2	9
1	A	224	GLY	N	109.8	0.2	9

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	58	0.27 ± 0.30	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	48	-0.68 ± 0.20	Should be checked
$^{13}C'$	57	0.62 ± 0.18	Should be applied
^{15}N	57	-1.82 ± 0.44	Should be applied

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 7%, i.e. 305 atoms were assigned a chemical shift out of a possible 4637. 0 out of 50 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	163/1783 (9%)	0/739 (0%)	109/696 (16%)	54/348 (16%)
Sidechain	127/2455 (5%)	0/1580 (0%)	119/741 (16%)	8/134 (6%)
Aromatic	15/399 (4%)	0/207 (0%)	14/141 (10%)	1/51 (2%)
Overall	305/4637 (7%)	0/2526 (0%)	242/1578 (15%)	63/533 (12%)

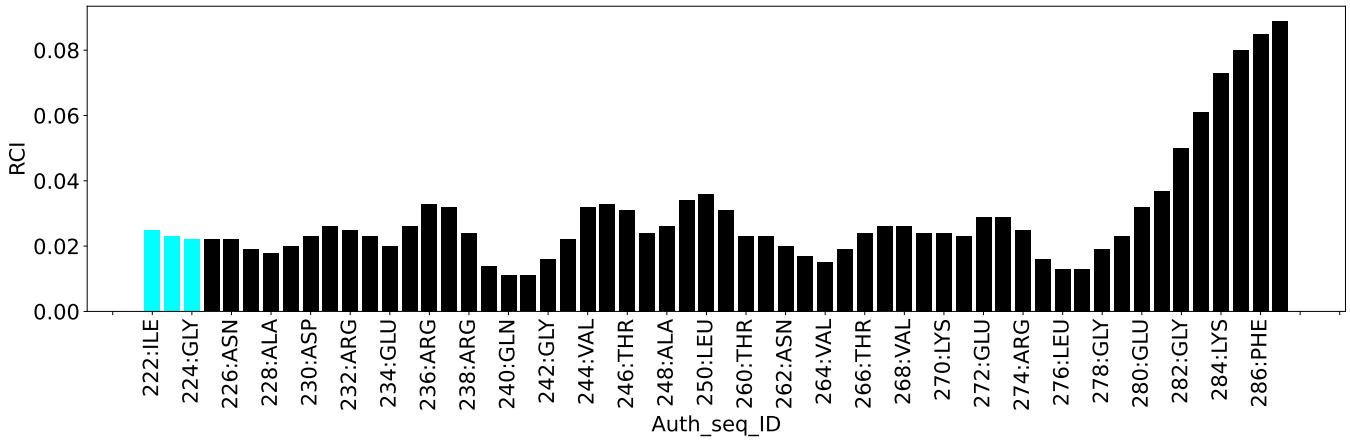
Note: This is a solid-state NMR structure, where hydrogen atoms are typically not assigned a chemical shift value, which may lead to lower completeness of assignment measure.

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: