wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:16 AM BST

PDB ID	:	2ROZ
Title	:	Structure of the C-terminal PID Domain of Fe65L1 Complexed with the Cytoplasmic Tail of APP Reveals a Novel Peptide Binding Mode
Authors	:	Li, H.; Koshiba, S.; Tochio, N.; Watanabe, S.; Harada, T.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2008-04-25

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:

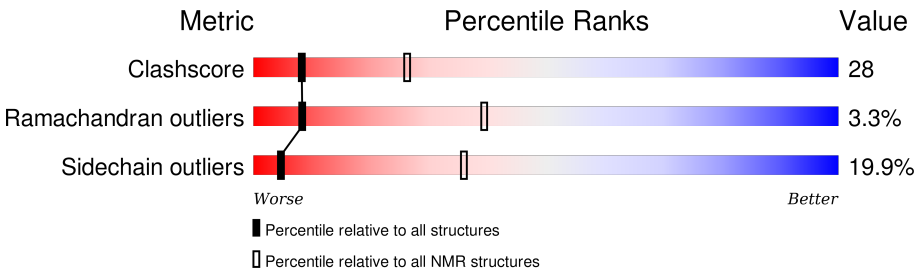
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal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

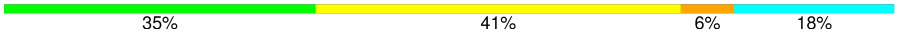
The overall completeness of chemical shifts assignment is 90%.

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	32	
2	B	136	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 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:11-A:20, B:15-B:102, B:108-B:130 (121)	0.19	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 4 clusters and 7 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2529 atoms, of which 1233 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called peptide from Amyloid beta A4 protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	32	512	166	244	46	54	2	0

- Molecule 2 is a protein called Amyloid beta A4 precursor protein-binding family B member 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	136	2017	638	989	174	206	10	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	GLY	-	EXPRESSION TAG	UNP Q9DBR4
B	2	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	3	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	4	GLY	-	EXPRESSION TAG	UNP Q9DBR4
B	5	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	6	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	7	GLY	-	EXPRESSION TAG	UNP Q9DBR4
B	131	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	132	GLY	-	EXPRESSION TAG	UNP Q9DBR4
B	133	PRO	-	EXPRESSION TAG	UNP Q9DBR4
B	134	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	135	SER	-	EXPRESSION TAG	UNP Q9DBR4
B	136	GLY	-	EXPRESSION TAG	UNP Q9DBR4

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

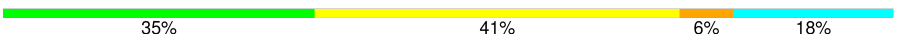
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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.

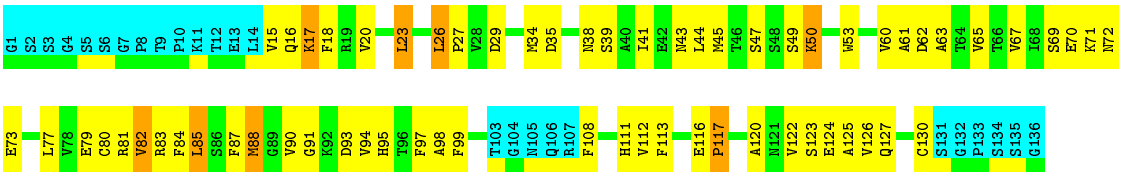
- Molecule 1: peptide from Amyloid beta A4 protein

Chain A:

- Molecule 2: Amyloid beta A4 precursor protein-binding family B member 2

Chain B:



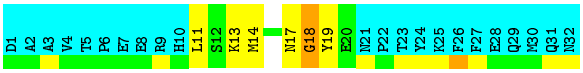


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

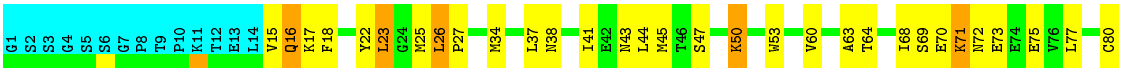
- Molecule 1: peptide from Amyloid beta A4 protein

Chain A:



- Molecule 2: Amyloid beta A4 precursor protein-binding family B member 2

Chain B:



R81
V82
R83
F84
L85
S86
F87
M88
G89
V90
D93
V94
H95
T96
F97
A98
F99
T103
G104
N105
Q106
R107
F108
E109
C110
H111
V112
F113
W114
C115
E116
P117
V122
A125
V126
C130
S131
G132
P133
S134
S135
G136

5 Refinement protocol and experimental data overview (i)

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

Of the 100 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	1.0.8
CYANA	refinement	1.0.8

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 10236
Number of chemical shift lists	1
Total number of shifts	1775
Number of shifts mapped to atoms	1775
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

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)

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	82	78	78	6±2
2	B	866	837	837	50±6
All	All	18960	18300	18300	1057

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 28.

5 of 259 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:63:ALA:HB1	2:B:82:VAL:HG21	1.03	1.27	13	4
1:A:11:LEU:HD11	2:B:120:ALA:HB2	0.93	1.38	14	16
2:B:122:VAL:O	2:B:126:VAL:HG23	0.87	1.69	2	20
2:B:63:ALA:HB1	2:B:82:VAL:HG11	0.86	1.46	6	15
2:B:26:LEU:HD13	2:B:27:PRO:HD2	0.80	1.52	14	20

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	10/32 (31%)	7±2 (66±15%)	3±1 (29±14%)	0±1 (4±6%)	6 30
2	B	111/136 (82%)	88±2 (79±2%)	20±2 (18±2%)	4±1 (3±1%)	8 40
All	All	2420/3360 (72%)	1888 (78%)	452 (19%)	80 (3%)	8 39

5 of 16 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	117	PRO	20
2	B	17	LYS	14
2	B	71	LYS	13
2	B	47	SER	6
1	A	18	GLY	5

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	9/29 (31%)	7±1 (74±11%)	2±1 (26±11%)	3 24
2	B	98/117 (84%)	79±3 (81±3%)	19±3 (19±3%)	5 37
All	All	2140/2920 (73%)	1714 (80%)	426 (20%)	5 36

5 of 52 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	88	MET	20
2	B	93	ASP	20
2	B	23	LEU	20
2	B	26	LEU	20
2	B	85	LEU	19

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)

There are no ligands in this entry.

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 90% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 10236

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	1775
Number of shifts mapped to atoms	1775
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	127	-0.11 ± 0.19	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	121	-0.02 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	120	0.22 ± 0.14	None needed (< 0.5 ppm)
^{15}N	120	0.36 ± 0.33	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 90%, i.e. 1304 atoms were assigned a chemical shift out of a possible 1452. 0 out of 23 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	563/597 (94%)	238/238 (100%)	218/242 (90%)	107/117 (91%)
Sidechain	638/738 (86%)	407/430 (95%)	219/278 (79%)	12/30 (40%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	103/117 (88%)	57/63 (90%)	44/50 (88%)	2/4 (50%)
Overall	1304/1452 (90%)	702/731 (96%)	481/570 (84%)	121/151 (80%)

7.1.4 Statistically unusual chemical shifts (i)

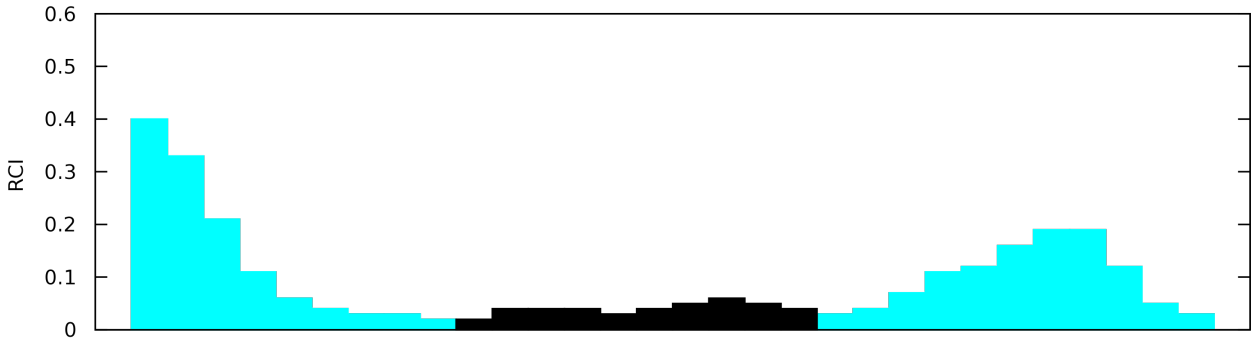
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
2	B	33	GLY	H	3.38	11.63 – 5.03	-7.5
2	B	23	LEU	HG	-0.88	3.16 – -0.14	-7.2
2	B	50	LYS	HA	1.31	6.46 – 2.06	-6.7
2	B	95	HIS	HB3	0.65	5.00 – 1.10	-6.2
2	B	86	SER	HB3	2.26	5.25 – 2.45	-5.7

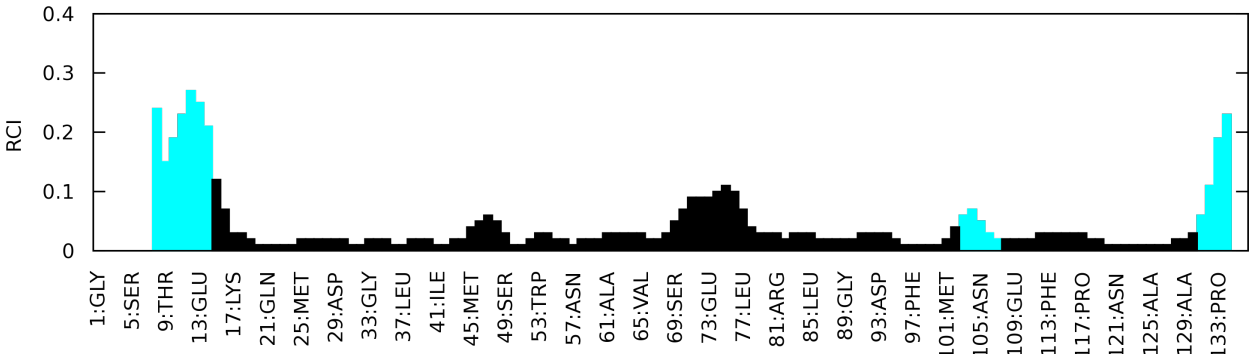
7.1.5 Random Coil Index (RCI) plots (i)

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:


RCI
0.4
0.3
0.2
0.1
0
1:GLY
5:SER
9:THR
13:GLU
17:LYS
21:GLN
25:MET
29:ASP
33:GLY
37:LEU
41:ILE
45:MET
49:SER
53:TRP
57:ASN
61:ALA
65:VAL
69:SER
73:GLU
77:LEU
81:ARG
85:LEU
89:GLY
93:ASP
97:PHE
101:MET
105:ASN
109:GLU
113:PHE
117:PRO
121:ASN
125:ALA
129:ALA
133:PRO