WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

May 31, 2020 – 10:25 pm BST

PDB ID	:	6H8C
Title	:	Structure of the human GABARAPL2 protein in complex with the UBA5 LIR motif
Authors	:	Huber, J.; Loehr, F.; Gruber, J.; Akutsu, M.; Guentert, P.; Doetsch, V.; Rogov, V.V.
Deposited on	:	2018-08-02

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
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:

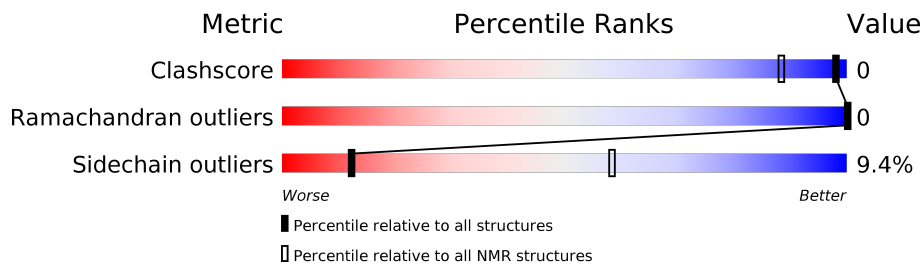|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	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 94%.

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		9.4%

Worse — *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	116	92%　　　　　　　　　　　　　　• 5%
2	B	19	47%　　　　　　　53%

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: *closest to the average*.

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:3-A:112, B:340-B:348 (119)	0.27	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, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20
2	3, 15

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2185 atoms, of which 1088 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Gamma-aminobutyric acid receptor-associated protein-like 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	116	1899	610	953	157	174	5	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P60520
A	2	GLY	-	expression tag	UNP P60520

- Molecule 2 is a protein called Ubiquitin-like modifier-activating enzyme 5.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	19	286	94	135	23	33	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	330	GLY	-	expression tag	UNP Q9GZZ9
B	331	ALA	-	expression tag	UNP Q9GZZ9
B	332	MET	-	expression tag	UNP Q9GZZ9

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: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

92%   •   5%

M1 G2   H9   R14   K97   N113 T114 F115 G116

- Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47%   53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339   E348
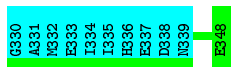
4.2 Scores per residue for each member of the ensemble

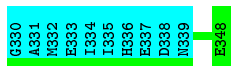Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2
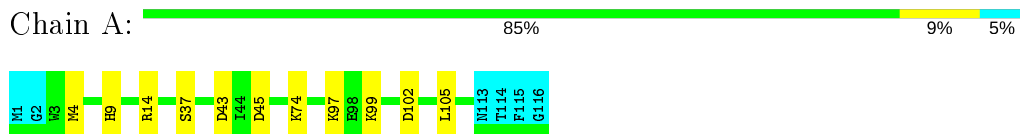
Chain A:

83%   12%   5%

M1 G2   D8 H9   H13 R14   R22   D27   K35   S39   I55   R65   E73   T90   K97 E98 K99   N113 T114 F115 G116

- Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47%   53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339   E348
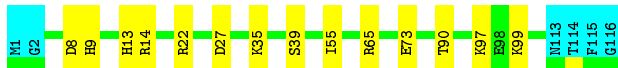
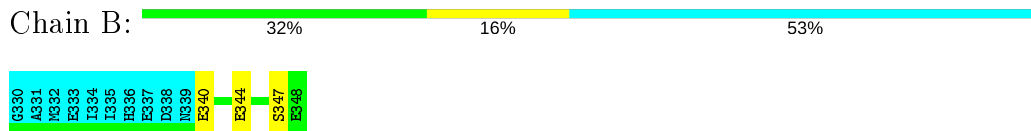#### 4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

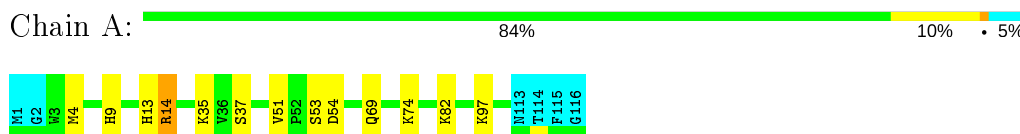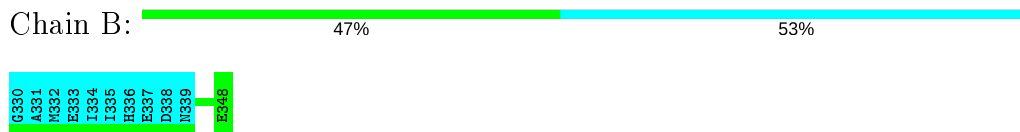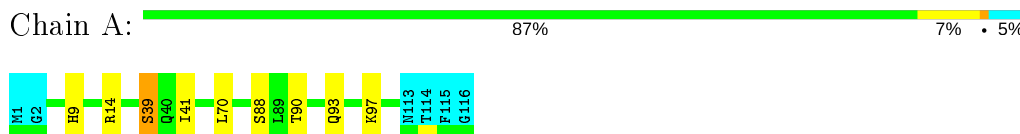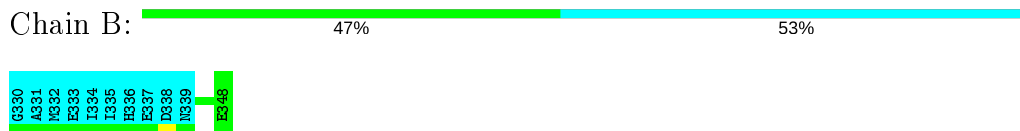85%   9%   5%

M1 G2 W3 M4 H9 R14 S37 D43 I44 D45 K74 K97 E98 K99 D102 L105 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

32%   16%   53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E340 E344 S347 E348

4.2.3 Score per residue for model 3

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

84%   10%   5%

M1 G2 W3 M4 H9 H13 R14 K35 V36 S37 V51 P52 S53 D54 Q69 K74 K82 K97 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47%   53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348

4.2.4 Score per residue for model 4

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

87%   7%   5%

M1 G2 H9 R14 S39 Q40 I41 L70 S88 I89 T90 Q93 K97 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47%   53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348

4.2.5 Score per residue for model 5

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2
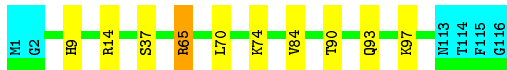

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

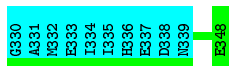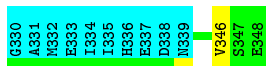#### 4.2.6 Score per residue for model 6

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

4.2.7 Score per residue for model 7

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

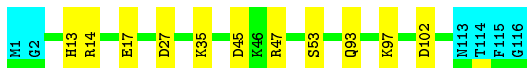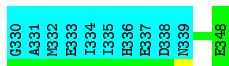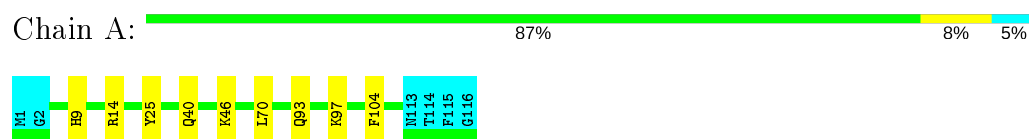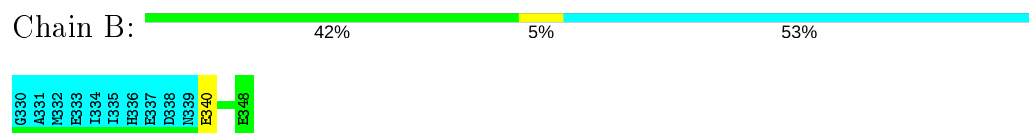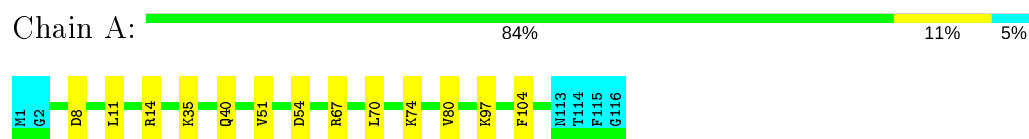

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5
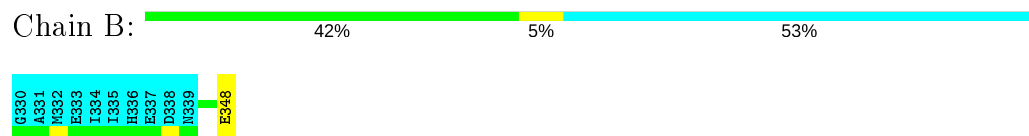
4.2.8 Score per residue for model 8

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 87% 8% 5%

M1 G2 H9 R14 Y25 Q40 K46 L70 Q93 K97 F104 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 42% 5% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E340 E348

4.2.9 Score per residue for model 9

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 84% 11% 5%

M1 G2 D8 L11 R14 K35 Q40 V51 D54 R67 L70 K74 V80 K97 F104 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 42% 5% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348

4.2.10 Score per residue for model 10

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 84% 11% 5%

M1 G2 D8 H9 R14 E17 D45 L70 K74 K82 T90 Q93 L94 Y95 E96 K97 D102 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 47% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348

4.2.11　Score per residue for model 11

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2
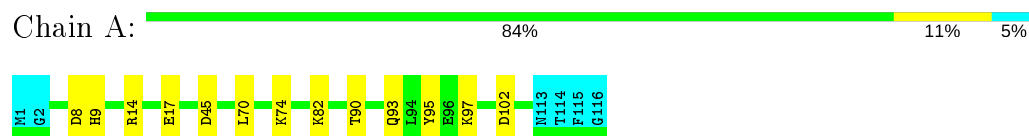
Chain A:

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

4.2.12　Score per residue for model 12

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5
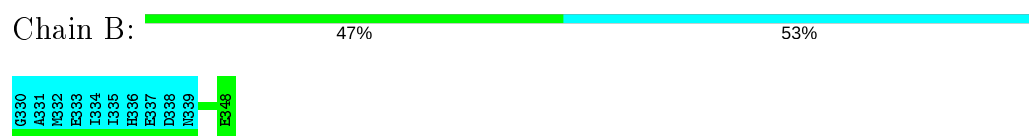
Chain B:

4.2.13　Score per residue for model 13

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2
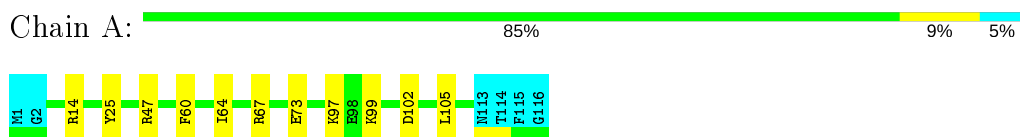
Chain A:

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5
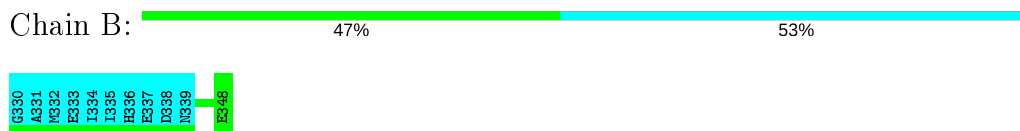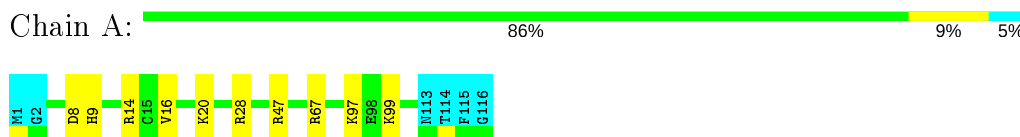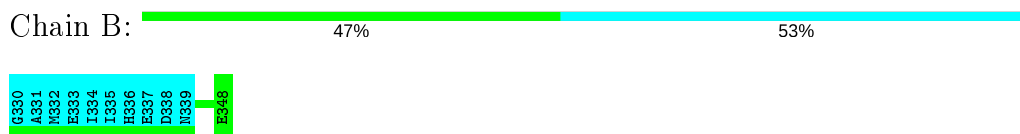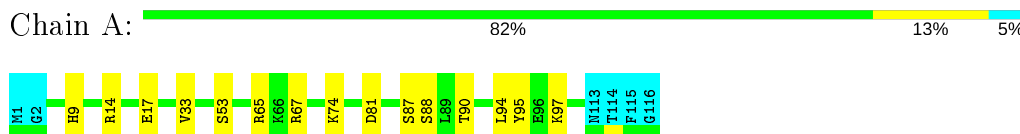
Chain B:

4.2.14 Score per residue for model 14

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

87% 8% 5%

M1 G2 W3 M4 L11 R14 E17 K20 K35 K48 K97 S110 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

42% 5% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E340 W341 E348
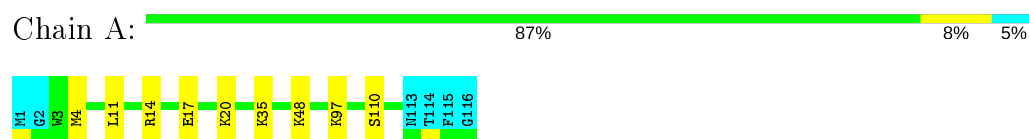
4.2.15 Score per residue for model 15

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

80% 15% 5%

M1 G2 H9 R22 D27 K35 V51 P52 S53 F60 I64 R65 V80 Q93 L94 Y95 E96 K97 E98 K99 D100 F104 Y109 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348
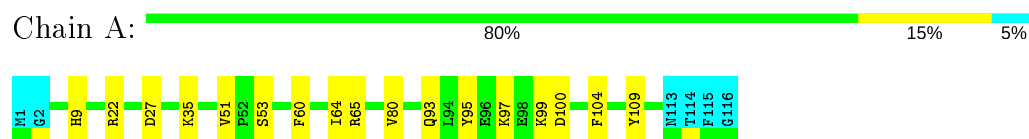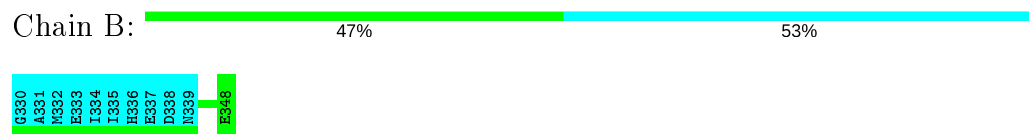
4.2.16 Score per residue for model 16

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

85% 9% 5%

M1 G2 H9 S10 L11 R14 K20 S37 V51 D54 D81 T90 K97 E98 K99 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

47% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348
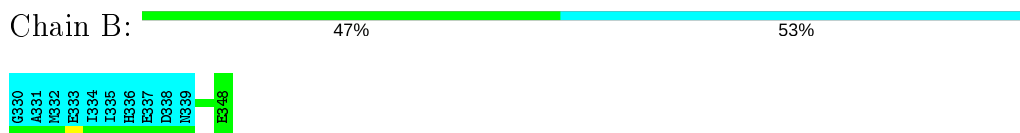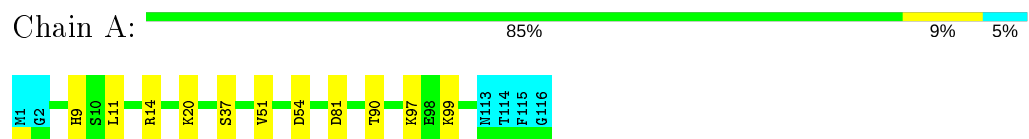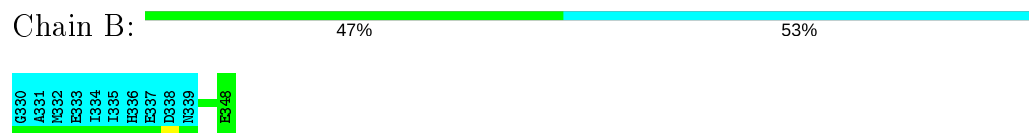
4.2.17 Score per residue for model 17

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 85% 9% 5%

M1 G2 H9 R14 Y25 V29 K35 K46 S53 S88 Q93 I94 Y95 E96 K97 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 47% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E348
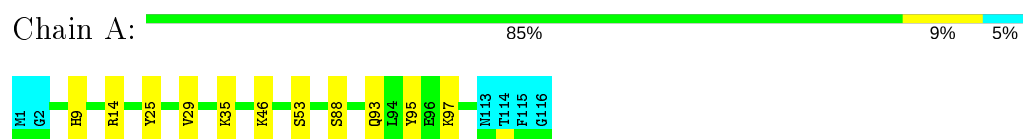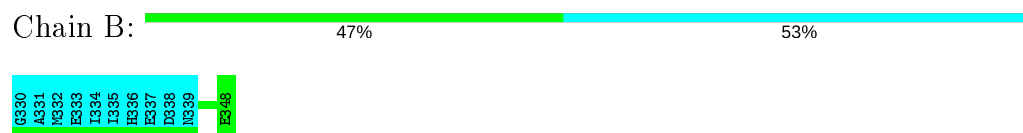
4.2.18 Score per residue for model 18

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 90% 5% 5%

M1 G2 H9 R14 K46 I55 I63 K97 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 42% 5% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 V346 S347 E348
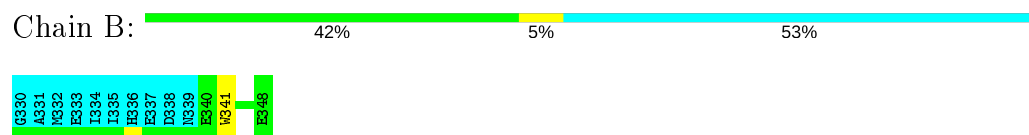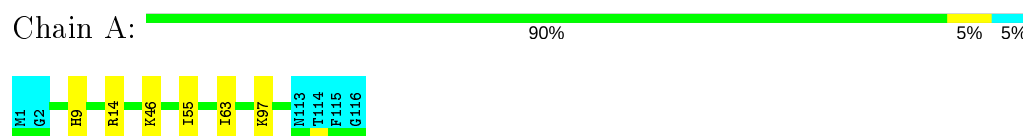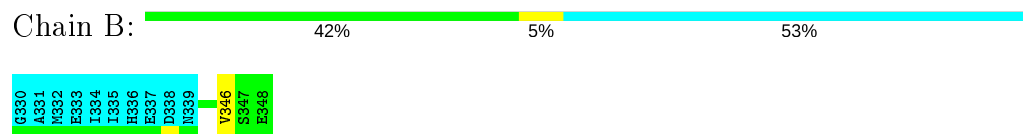
4.2.19 Score per residue for model 19

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A: 83% 12% 5%

M1 G2 D8 H9 R14 K35 V36 S37 G38 S39 K48 L70 P71 S72 E73 K74 T90 K97 D102 Y106 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B: 37% 11% 53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 E344 E348
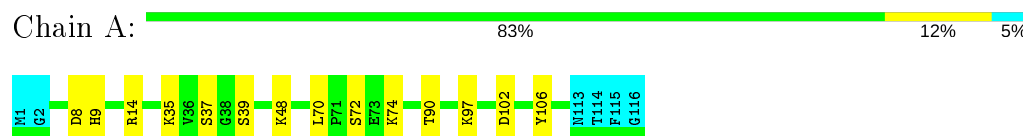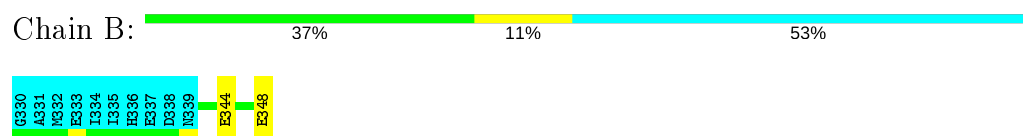
4.2.20　Score per residue for model 20

• Molecule 1: Gamma-aminobutyric acid receptor-associated protein-like 2

Chain A:

81%　　　　　　　　　　　　　　　　　　　14%　　　5%

M1 G2 K6 H9 R14 R22 R28 K35 K46 V51 I55 I63 I64 R65 L70 P71 S72 K82 K97 E98 K99 N113 T114 F115 G116

• Molecule 2: Ubiquitin-like modifier-activating enzyme 5

Chain B:

42%　　　　5%　　　　　53%

G330 A331 M332 E333 I334 I335 H336 E337 D338 N339 V346 S347 E348
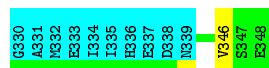
5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle 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 calculation	3.97
OPAL	refinement	

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	1765
Number of shifts mapped to atoms	1765
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	94%

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	904	916	916	1±1
2	B	74	67	67	0±0
All	All	19560	19660	19660	13

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:63:ILE:HG21	2:B:346:VAL:HG13	0.56	1.78	18	2
1:A:48:LYS:HE2	2:B:341:TRP:O	0.53	2.04	14	1
1:A:60:PHE:CZ	1:A:64:ILE:HD11	0.49	2.41	11	1
1:A:13:HIS:CD2	1:A:14:ARG:HH21	0.48	2.26	7	2
1:A:16:VAL:HG12	1:A:20:LYS:HE2	0.45	1.88	12	1
1:A:9:HIS:CD2	1:A:13:HIS:HD1	0.45	2.30	6	1
1:A:29:VAL:HG11	1:A:95:TYR:CG	0.43	2.48	17	1
1:A:63:ILE:HG21	2:B:346:VAL:CG1	0.43	2.44	6	1
1:A:39:SER:HB3	1:A:41:ILE:HG22	0.43	1.89	4	1
1:A:13:HIS:CD2	1:A:14:ARG:NH2	0.41	2.88	3	1
1:A:60:PHE:CE2	1:A:64:ILE:HD11	0.41	2.50	15	1

5.2 Torsion angles ⓘ

5.2.1 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	110/116 (95%)	106±1 (96±1%)	4±1 (4±1%)	0±0 (0±0%)	100	100
2	B	8/19 (42%)	7±1 (92±11%)	1±1 (8±11%)	0±0 (0±0%)	100	100
All	All	2360/2700 (87%)	2258 (96%)	102 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

5.2.2 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	101/105 (96%)	91±2 (90±2%)	10±2 (10±2%)	11	57
2	B	8/16 (50%)	8±1 (96±10%)	0±1 (4±10%)	32	81
All	All	2180/2420 (90%)	1975 (91%)	205 (9%)	12	58

All 53 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	97	LYS	20
1	A	14	ARG	15
1	A	9	HIS	14
1	A	35	LYS	9
1	A	90	THR	8
1	A	70	LEU	8
1	A	99	LYS	8
1	A	93	GLN	7
1	A	74	LYS	7
1	A	53	SER	6
1	A	8	ASP	6
1	A	51	VAL	5
1	A	37	SER	5
1	A	102	ASP	5
1	A	104	PHE	4
1	A	17	GLU	4
1	A	4	MET	4
1	A	46	LYS	4
1	A	39	SER	3
1	A	45	ASP	3
1	A	25	TYR	3
1	A	55	ILE	3
1	A	54	ASP	3
1	A	88	SER	3
1	A	27	ASP	3
1	A	11	LEU	3
1	A	82	LYS	3
1	A	40	GLN	2
1	A	81	ASP	2
1	A	73	GLU	2
1	A	105	LEU	2
2	B	340	GLU	2
1	A	20	LYS	2
1	A	67	ARG	2
1	A	100	ASP	2
1	A	80	VAL	2
1	A	65	ARG	2
1	A	72	SER	2
2	B	348	GLU	2
2	B	344	GLU	2
1	A	10	SER	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	28	ARG	1
1	A	110	SER	1
1	A	43	ASP	1
1	A	33	VAL	1
2	B	347	SER	1
1	A	47	ARG	1
1	A	48	LYS	1
1	A	84	VAL	1
1	A	87	SER	1
1	A	6	LYS	1
1	A	69	GLN	1
1	A	94	LEU	1

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 94% for the well-defined parts and 93% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *Star31_ corr.bmrb*

6.1.1 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	1765
Number of shifts mapped to atoms	1765
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

6.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	135	-0.30 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_\beta$	127	0.09 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	133	-0.12 ± 0.16	None needed (< 0.5 ppm)
^{15}N	126	-0.08 ± 0.40	None needed (< 0.5 ppm)

6.1.3 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 94%, i.e. 1440 atoms were assigned a chemical shift out of a possible 1540. 18 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	580/585 (99%)	231/233 (99%)	237/238 (100%)	112/114 (98%)
Sidechain	751/820 (92%)	469/482 (97%)	266/304 (88%)	16/34 (47%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	109/135 (81%)	67/71 (94%)	39/59 (66%)	3/5 (60%)
Overall	1440/1540 (94%)	767/786 (98%)	542/601 (90%)	131/153 (86%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 93%, i.e. 1595 atoms were assigned a chemical shift out of a possible 1714. 18 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	653/665 (98%)	259/265 (98%)	268/270 (99%)	126/130 (97%)
Sidechain	822/898 (92%)	513/527 (97%)	291/335 (87%)	18/36 (50%)
Aromatic	120/151 (79%)	74/80 (92%)	43/65 (66%)	3/6 (50%)
Overall	1595/1714 (93%)	846/872 (97%)	602/670 (90%)	147/172 (85%)

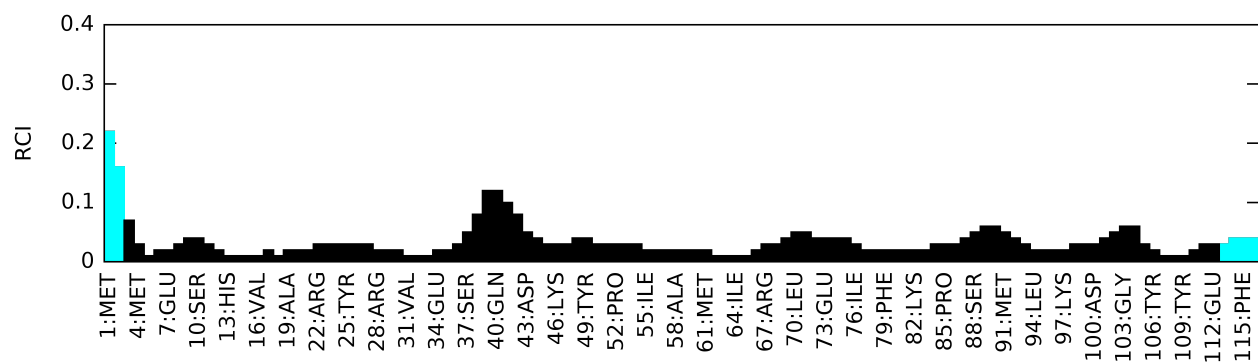### 6.1.4 Statistically unusual chemical shifts (i)

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	46	LYS	HD2	-0.68	2.76 − 0.46	-10.0
1	A	28	ARG	HH12	10.25	8.94 − 4.74	8.1
1	A	14	ARG	HH12	10.25	8.94 − 4.74	8.1
1	A	6	LYS	HE2	1.60	3.87 − 1.97	-6.9
1	A	47	ARG	HB2	0.06	3.15 − 0.45	-6.4
1	A	108	ALA	HA	6.57	6.46 − 2.06	5.2
1	A	106	TYR	HB2	1.02	4.76 − 1.06	-5.1

6.1.5 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: