

MAGIC



Methyl Assignment

by Graphing Inference Construct

Please cite:

Monneau, Y.R., Rossi, P., Bhaumik, A., Huang, C., Jiang, Y., Saleh, T., Xie, T., Xing, Q. and Kalodimos, C.G., Automatic methyl assignment in large proteins by the MAGIC algorithm, *J. Biomol. NMR* (2017)

Download program:

<https://github.com/NMRsoftware/MAGIC>

DISCLAIMER: Magic-v1.0 is distributed free of charge, for entertainment purposes only and AS-IS with no warranty whatsoever, expressed or implied. Assignments are NOT certified for accuracy. Neither the authors nor their institutions will be held responsible for use or misuse of the program. We consider validation of the results using independent experimental data mandatory.

Installation and Basic Operation

Handbook:

Requirements:



Needs python2.7 - Anaconda5.0.0

distribution (anaconda2):

<https://www.anaconda.com/download/>

In addition, please obtain and install the most recent version of NMRAM-Sparky for use during the preparation and validation steps:

<http://www.nmrfam.wisc.edu/nmrfam-sparky-distribution.htm>

List of mandatory files:

2D peak list

3D peak list

PDB file

Fasta file

Magic_v1.0.py

generate.py

start.txt

The proper format for each file is described in the detailed procedure below.

Protocol:

1. Put all mandatory files within the working directory and launch the terminal
2. Generate the 2D peak list with methyl type and geminal information using the *generate.py* script (>*python generate.py -h* for further information)
3. Once the new 2D peak list is created, check and edit it as your convenience

4. Edit the *start.txt* file in order to specify the files pathway, labeling scheme, etc.
5. Start the calculation by typing:

```
>python Magic_v1.0.py start.txt
```

STEP 1: file preparations

MAGIC needs the files to be in proper format. Both 2D and 3D peak lists are in sparky format **without any header**. The PDB file does not need to have protons for assignment calculation but it is convenient to prepare the PDB with proton initially for later use with the validation scripts within NMRFAM Sparky. The 3D peak list needs peak height. The primary sequence of the actual construct has to be in the one letter code.

The 2D peak list:

?-?	26.611	1.203	The first column is for peak label.
?-?	26.326	0.846	The second column is the ¹³ C
?-?	26.274	0.689	frequencies. The third column is the
?-?	26.077	0.765	¹ H frequencies.
?-?	25.799	0.697	
?-?	25.264	0.813	
?-?	25.031	0.920	
?-?	24.707	0.760	
?-?	24.750	0.458	
?-?	24.696	0.899	
?-?	24.527	1.369	
?-?	24.474	0.597	
?-?	24.202	0.966	
?-?	24.157	0.767	

=>Needs to be converted with the generate.py script as shown below.

The 3D CCH-NOESY peak list:

The first column is for peak label. The second one is for NOE ¹³C frequencies. The third one is for edited ¹³C frequencies. The fourth one is for ¹H

?-?-?	12.643	22.716	0.841	99709132800
?-?-?	12.645	22.133	0.905	56579497984
?-?-?	12.649	14.573	1.908	34486972416
?-?-?	12.664	14.940	0.421	23421962240
?-?-?	12.666	26.010	1.140	22552233984
?-?-?	12.986	25.045	0.736	80192913408
?-?-?	12.987	21.084	0.755	61704085504
?-?-?	12.990	14.322	0.807	112064831488
?-?-?	13.002	26.253	0.789	75430666240
?-?-?	13.116	16.456	1.886	13009507328
?-?-?	13.148	25.158	0.617	45412483072
?-?-?	13.171	21.453	0.779	158998593536
?-?-?	13.180	13.747	0.631	130473730048
?-?-?	13.187	16.257	1.578	162703540224
?-?-?	13.228	18.813	0.831	29239457792
?-?-?	13.353	18.773	0.836	28295053312
?-?-?	13.381	17.551	2.120	188872458240
?-?-?	13.391	17.299	0.574	42611273728

frequencies. The fifth one is for peak heights. Peaks are picked with 'kr' within Sparky and then manually cleaned up as done for any structure determination run with automated NOESY assignment.

NOTE: All diagonal peaks have to be purged prior to running.

The PDB file:

The PDB file has to be in the format delivered by the PDB website. The coordinate only information needs to be present (i.e. no headers, HETATM, CONNECT, etc). For structure that does not contain proton, it is recommended to add them using MolProbity webserver if you plan to use 3D NOESY involving amides for validation purposes (<http://molprobity.biochem.duke.edu/>). See below for how to use the validation script pdb2noe.py within NMRFAM Sparky.

The construct amino acid sequence file:

This file is important because it enables MAGIC to take in account methyls that are not included within the PDB file. This file has to be cleaned up for anything but the sequence of one-letter-code amino acids.

The parameter file (start.txt):

The parameter file contains all the information MAGIC needs to conduct an assignment run. **Line 3:** the 2D peak list file name; **line 5:** the 3D peak list file name; **line 7:** the PDB file name; **line 9:** the methyl list generated by generate.py script; **line 11:** the labeling scheme (*=>do not use semicolon at the end of this line*); **line 14:** the numerical code for the type of VL labeling; **line 16:** the ppm score

```
1 ##### Input #####
2 ##2d reference spectrum:
3 new_Chmc_IMLsVs_noOL.list
4 ##3d NOESY:
5 FlhAC_cch_noesy_IMLsVs_437p.list
6 ##pdb:
7 kakah4.pdb
8 ##SEQ:
9 seq.auto
10 ##labeling:
11 I,CD1;L,CD2;V,CG2;M
12 ##### parameters #####
13 ##stereospecific=2;double methyl=1;mono-methyl=0:
14 2
15 ##ppm tolerance (13C,13C,1H):
16 0.1 0.1 0.01
17 ##score threshold factor:
18 1
19 ##Distance threshold:
20 7 10
21 ##Score tolerance for one-by-one step (off or on):
22 off
23 ##Area of expected conformational changes:
24
25 ###END#####
```

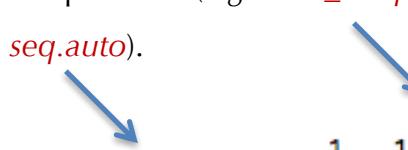
threshold factor, with 1 the standard value, higher and lower values than 1 makes the exclusion of assignments more or less stringent; **line 20**: the lower and higher distance thresholds to define methyl network; **line 22**: the polishing step of MAGIC procedure only keep assignments with the highest score if set to 'off', otherwise, the same score tolerance function apply to that step if set to 'on'; **line 24**: *UNDER CONSTRUCTION*

STEP 2: Generate 2D peak list

To generate the 2D peak list with all mandatory format and information, *i.e.* methyl type and peak labeling, and optionally the geminal pairing, the script *generate.py* has to be used as following:

```
> python generate.py [2D peak list file name]
[construct sequence file name, fasta format] [labeling, e.g. AILMTV]
[starting sequence number, e.g. 48]
[rename peaks: y or n]
[optional: [CCH short mixing time peak list]]
e.g. > python generate.py hmqc.list seq.fasta AILMV 48 y
```

The script outputs two files that are used by MAGIC, the formatted 2D peak list (*e.g.* *new_hmqc.list*) and the construct methyl list (*e.g.* *seq.auto*).



L48	1	19.311	1.394	A
V52	2	24.873	0.815	L ;3
A53	3	23.265	0.793	ALV ;2
L61	4	21.161	0.928	VA ;5
A64	5	20.318	0.928	VA ;4
A65	6	19.438	1.366	A
L73	7	18.832	1.358	A
L74	8	18.943	1.353	A
A75	9	24.997	0.892	L ;10
L84	10	23.600	0.862	LAV ;9;12
V86	11	24.866	0.863	L ;12
A87	12	23.354	0.817	LAV ;11;10
L88				
V92				

Within the 2D peak list, the three first columns are obviously for peak labels, ^{13}C and ^1H frequencies. The fourth one is for the definition of methyl type (one letter code) and the fifth one, the geminal pairing. Note that before each new putative geminal pair it is mandatory to put a semicolon.

=> *Needs careful user review! The information in the newly created 2D peak list is treated as 100% correct by MAGIC. Particularly critical checking has to be done on methyl type definition.*

Correctness of bootstrapped assignments (if any) need to be absolute.

STEP 3:

Edit the starting file and RUN:

The parameter file *start.txt* has to be properly filled to avoid any trouble (file pathways, labeling and setting parameters – see p. 4, *The parameter file*).

To start the calculation, type the following command:

> *python Magic_v1.0.py start.txt*

The Output files:

After the run starts, MAGIC generates one time-stamped folder, inside are three folders:

1. The input folder containing all input files, for your record
2. The output folder, continuously edited, containing the current assigned 2D and 3D peak lists along with the mapping file to visualize into PDB structure the quality factors.
3. The run folder, which contains all intermediate files used by the software

=> *The contents of the generated ‘run’ directory can be quite large and can be removed at the end of the calculation.*

Description of output files:

Inside the output folder, the 2D and 3D peak lists are updated all along the run. The anatomy of both peak lists is detailed below:

The assigned 2D peak list:

```
A119CB-HB  21.845  0.843  8.78  1.0  {'A119CB': 224.991}
A121CB-HB  24.513  1.066  14.88  1.0  {'A121CB': 224.991}
A155CB-HB  18.3    1.492  4.44   1.0  {'A155CB': 224.991, 'A156CB': 212.723}
A156CB-HB  18.402  1.468  8.67   1.0  {'A155CB': 212.723, 'A156CB': 224.991}
```

1st column: best assignment (highest score)

2nd and 3rd column: ¹³C and ¹H chemical shifts

4th column: the sum of all peak-peak connection confident scores for each peak

5th column: the NOE assignment completeness of the strip related to each peak

6th column: the list of all possible assignment with their scores

The assigned 2D peak list:

```
L184CD*-L232CD*-HD* 24.426  22.482  0.135  0.5  d=4.1  0
L184CD*-L232CD*-HD* 27.995  22.482  0.135  5.0  d=4.1  1
L228CD*-L232CD*-HD* 26.068  22.482  0.135  5.0  d=5.8  1
L228CD*-L232CD*-HD* 20.883  22.482  0.135  3.8  d=5.8  1
L232CD*-L232CD*-HD* 25.385  22.482  0.135  2.0  d=2.5  1
```

1st column: best assignment (highest score)

2nd, 3rd and 4th column: NOE ¹³C, ¹³C and ¹H chemical shifts

5th column: the peak-peak connection confident score

6th column: the methyl-methyl (carbon-carbon) distance from PDB file

7th column: a numerical code to display whether this connection was used for peak clustering or not

Validation:

After the run is complete the user can test the result for accuracy from within NMRFAM-Sparky:

1. Paste the assigned 2D and 3D peak list back into the spectra using the command 'rp'.
2. Assignments can be validated against a series of 3D NOESY such as HmCmHm, H_NCmHm, N-CmHm and CmNH_N type amide 3D NOESY data using the pdb2noe.py (type 'SN' in Sparky to launch the graphic interface). For details see the manuscript.



ENJOY!