Interact

Automated analysis of 1D and 2D NMR titration experiments - version 1.1

http://github.com/MetaSys-LISBP/Interact

User Manual

Version 1.1

Authors: Pierre Millard (pierre.millard@insa-toulouse.fr)
Guy Lippens (glippens@insa-toulouse.fr)

Group MetaSys, LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

http://metasys.insa-toulouse.fr

Table of contents

I. Introduction

1. Software description

Interact is a scientific software designed for the analysis of 1D and 2D NMR titration experiments. Interact i) performs peak picking and spectra annotation, ii) process each signal of interest in each experiment to extract different parameters (chemical shifts, linewidths, intensity, angle), and iii) integrates the results of the different experiments to infer thermodynamic, kinetic or structural information on the system under study.

2. Licensing

The original version of Interact was developed in the MetaSys team in the LISBP, Toulouse, France.

II. Installation

The software was developed on Windows and can be used on Windows, Linux or MacOS platforms. To use Interact, you’ll need some dependencies listed below.

1. Dependencies

To use Interact you must have TopSpin (3.0 or higher) installed.

Interact also requires Python (2.7+, 3.0 or higher) and modules:

- numpy
- matplotlib
- lmfit

If you are not used to install system wide environments like Python, ask some help from your local computer service. We don’t provide support for installation.

2. Installation

Unpack the content of `Interact_X.Y.zip` (where `X.Y` is the version number) somewhere on your disk, and copy the file `interact.py` in the TopSpin Python programs directory (by default: `<TopSpin installation directory>/exp/stan/nmr/py/user`).

3. Test of installation

To check that Python and the required modules are correctly installed, run the following command in TopSpin:

```
interact --test
```

A message will indicate if the test is successfully passed or not. If an error is returned, check the path of the system's Python interpreter (see Section II.2.1) and verify that the required modules are correctly installed.

III. Methods

1. Signal analysis

Number of dimension(s) of NMR experiments is detected automatically. For each signal of interest, the following parameters are extracted in each spectrum:

- Chemical shift in each dimension
- Intensity
- Full width at half maximum (FWHM) in each dimension
- Rotation angle (only for 2D spectra when the Gaussian model with rotation is used, see below)

Chemical shifts and intensity are extracted (using the TopSpin peak picking routines), and the spectra are annotated. Only the peak with highest intensity is picked and annotated in the defined window.
For 1D spectra, peak resolution is estimated using TopSpin routines. For 2D spectra, peak resolution is estimated in each dimension by fitting to a 2D model. The following models are available:

- 2D Gaussian model with rotation (default):

$$A_{\omega F2,\omega F1} = I \cdot e^{-\left(a\cdot(\omega F2-\omega F2_0)^2-2\cdot b\cdot(\omega F2-\omega F2_0)\cdot(\omega F1-\omega F1_0)+c\cdot(\omega F1-\omega F1_0)^2\right)}$$

with $a = \frac{\cos(\varphi)^2}{2\cdot\sigma_{F2}{}^2} + \frac{\sin(\varphi)^2}{2\cdot\sigma_{F1}{}^2}$; $b = -\frac{\sin(2\cdot\varphi)}{4\cdot\sigma_{F2}{}^2} + \frac{\sin(2\cdot\varphi)}{4\cdot\sigma_{F1}{}^2}$; and $c = \frac{\sin(\varphi)^2}{2\cdot\sigma_{F2}{}^2} + \frac{\cos(\varphi)^2}{2\cdot\sigma_{F1}{}^2}$

where $A_{\omega F2,\omega F1}$ is the signal amplitude at the chemical shift ($\omega F2$, $\omega F1$) for a peak centered at ($\omega F2_0$, $\omega F1_0$) with intensity I and full width at half maximum of σ_{F2} and σ_{F1} in the corresponding dimension, and φ is the clockwise rotation angle.

- 2D Lorentzian model:

$$A_{\omega F2,\omega F1} = \frac{2\cdot I \Big/ \left(\pi\cdot\sqrt{\sigma_{F2}{}^2+\sigma_{F1}{}^2}\right)}{\left(\frac{\omega F2-\omega F2_0}{\sigma_{F2}/2}\right)^2 + \left(\frac{\omega F1-\omega F1_0}{\sigma_{F1}/2}\right)^2 + 1}$$

- 2D Gaussian model without rotation:

$$A_{\omega F2,\omega F1} = \frac{I}{2\cdot\pi\cdot\sigma_{F2}\cdot\sigma_{F1}} \cdot e^{-\left(\frac{(\omega F2-\omega F2_0)^2}{2\cdot\sigma_{F2}{}^2}+\frac{(\omega F1-\omega F1_0)^2}{2\cdot\sigma_{F1}{}^2}\right)}$$

Plots are generated for visual inspection of fitted and experimental 2D spectra (see Section IV.3).

2. Data integration

For each signal of interest, the parameters extracted from the different experiments are integrated to facilitate data analysis and interpretations.

The **dissociation constant *Kd*** is estimated for both 1D and 2D experiments by fitting, using a two-state interaction model described by the following function:

$$E_L = E_{\max} \cdot \frac{P + L + Kd - \sqrt{(P + L + Kd)^2 - 4 \cdot P \cdot L}}{2 \cdot P}$$

where E_L is the euclidean distance of the chemical shift caused by the ligand at concentration L, E_{max} is the maximal euclidean distance (*i.e.* observed at ligand saturation), P is the protein concentration, and Kd is the dissociation constant. Euclidean distance E_L is calculated as follow:

$$E = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (c_i \cdot d\omega Fi_L)^2}$$

where N is the number of dimension, $d\omega Fi_L$ is the change of chemical shift (in ppm) in dimension Fi induced by the ligand at concentration L, and c_i is a weighting factor equal to γ_X/γ_H (where γ_H is the proton's gyromagnetic ratio and γ_X is the gyromagnetic ratio of the nuclei observed in the corresponding dimension).

For 2D spectra, **the slope and angle of ligand-induced changes of chemical shifts** are estimated by linear regression:

$$d\omega F1 = a \cdot d\omega F2 + b$$

and

$$\theta = \tan^{-1}\left(a \cdot \frac{c_{F2}}{c_{F1}}\right)$$

where a is the slope, b is the intercept, θ is the angle, $d\omega F1$ ($d\omega F2$) is the change of chemical shift in dimension F1 (F2), and c_{F1} (c_{F2}) is a weighting factor equal to γ_X/γ_H (where γ_H is the proton gyromagnetic ratio and γ_X is the gyromagnetic ratio of the nuclei observed in the corresponding dimension).

Standard errors on the parameters are calculated from the estimated covariance matrix, and plots are generated for visual inspection of fitted and experimental data (see Section IV.3).

IV. User manual

1. Input data

Interact requires as input a set of 1D or 2D spectra acquired on samples prepared with a unique protein concentration and different concentrations of ligand.

Spectra must be pre-processed (i.e. Fourier-transformed, phased, baseline-corrected and aligned) in TopSpin before running Interact.

The different experiments must be located in the same folder. *Expnos* must be numbered as ***XYYY*** (where *X* may be any number and *YYY* denotes the ligand concentration in the corresponding sample). Pre-processed 1D or 2D spectra must be in the first *procno* of each *expno*.

2. Interact usage

2.1. Commands and options

Interact can be run using the following TopSpin command:

```
interact <options>
```

where `<options>` is a list of options separated by a white character. Each option starts with a double dash (`--`). The available options are:

--test return a message indicating if Python and the required modules are found

--nopp skip peak picking & spectra annotation; only integrate previously processed data

--noint skip data integration; only perform peak picking & spectra annotation

--fwhm estimate full width at half maximum (skipped when `--nopp` is provided)

--upd update existing result files when other signals are (re)processed (otherwise result files are rewritten silently and previous results are lost)

--opt display a window to modify the following parameters:

- path of the Python interpreter installed on the system, depending on your operating system and how it has been set up:
  · empty – default value – if there is no need to type the 'python' command in front of a Python script name to run it (this works on Windows platforms)
  · just `python` if Python is in the path (this works on Unix platforms)
  · the complete path of the Python interpreter installed on the system, e.g. `"C:/path/python/python.exe"`, including quotation marks if the path contains spaces (this works both on Unix or Windows platforms)
- model used to fit 2D spectra (available models are `gaussian2Drot` – default value –, `gaussian2D`, or `lorentzian2D`; see Section III)
- initial value of FWHM for fitting 2D spectra, which depends on the spectrometer frequency and on the experiment to process (we recommend to keep it small, 0.02 by default)
- list of nuclei-specific coefficients (i.e. weighting factors c_X equal to γ_X/γ_H, where γ_H is the proton gyromagnetic ratio and γ_X is the gyromagnetic ratio of the nuclei observed in the corresponding dimension, as detailed in Section III), provided as a Python dictionary

"path/to/database/file" tabulated text file containing information on the signals to process in batch mode, which must comply with the following format:

# lines can be commented using the '#' symbol				
# signal name and peak picking window (upper (p) and lower (m) bounds in F1 and F2 dimensions)				
# Name	F1m	F1p	F2m	F2p
A_227	128.3	129.1	8.3	8.4
K_32	115.6	116.6	8.05	8.135

2.2. Automatic processing of a single signal

- Define the window that contains the signal of interest:
  - · open one of the spectra
  - · add the other spectra to the active window in layered (“multiple display”) mode
  - · adjust the window to see the peak of interest in all the spectra
  - · leave the “multiple display” mode
- Run Interact:
  - · run the command `interact <options>` (see Section II.2.1 for the complete list of options)
  - · enter the signal name (used for spectra annotation), the total number of experiments to process, the protein concentration (used for Kd estimation, or estimated by fitting if the value is 0), and confirm the information provided
- Processing results will be displayed and saved in the `res` subdirectory of the TopSpin experiments folder (see Section IV.3 for details on the output files)

2.3. Automatic processing of several signals

- Create a tabulated text file gathering information on all the signals to process (i.e. signal names and window boundaries in F1 – and F2 – dimensions, which should be defined as detailed in Section IV.2.2)
- Run Interact with the command `interact <options>`, where `<options>` must contains the database file gathering processing information (see Section II.2.1 for the complete list of options)
- Processing results will be displayed and saved in the `res` subdirectory of the TopSpin experiments folder (see Section IV.3 for details on the output files)

2.4. Additional situations

To perform only peak picking (without data integration), use the `--noint` argument.

To perform only data integration (without peak picking), use the `--nopp` argument. Data contained in the file `_pp.txt` of the `res` folder will be used as input data (see Section IV.3 for details). This file can therefore be edited manually to add/remove/modify data used for fitting. In this case, the ligand-induced changes of chemical shifts ($d\omega F1$ and $d\omega F2$) and euclidean distance (E_L) are automatically recalculated from the chemical shifts in F1 (and F2).

Note: when several rounds of processing are performed, or if some signals are reprocessed (e.g. with a different window), use the `--upd` argument to append the new results to the existing result files, otherwise these files will be rewritten silently and the previous results will be lost.

3. Output files

The folder `res` is created in the data subdirectory containing all the experiments, and results are stored in the following files (where *xxx* is the name of the signal and *yyy* denotes the *expno*):

_pp.txt summary of peak picking results:

PeakName	name of the signal
PeakID	ID of the peak in TopSpin
Expno	*expno* of the spectrum
Ligand conc.	concentration of the ligand
F1	F1 chemical shift (ppm)
F2	for 2D spectra only, F2 chemical shift (ppm)
dwF1	difference of chemical shifts in F1 compared to the reference spectrum, which is the first *expno* (ppm)
dwF2	for 2D spectra only, difference of chemical shifts in F2 compared to the reference spectrum, which is the first *expno* (ppm)
Euclidean dist.	euclidean distance (compared to the reference, which is the first *expno*)
Intensity	peak intensity
resF1, resF1_sd	FWHM in F1 dimension ± sd (ppm)
resF2, resF2_sd	for 2D spectra only, FWHM in F2 dimension ± sd (ppm)
phi, phi_sd	for 2D spectra only when the 2D Gaussian model is used for fitting, signal rotation angle ± sd

_fit.txt summary of data integration results:

KD, KD_sd	dissociation constant ± sd
P, P_sd	protein concentration ± sd
dmax, dmax_sd	maximal euclidean distance (i.e. at ligand saturation) ± sd
a, a_sd	for 2D spectra only, slope ± sd
b, b_sd	for 2D spectra only, ordinate value ± sd
theta, theta_sd	for 2D spectra only, angle ± sd

Detailed results on individual signals/experiments can be found in the following files:

/tmp/*xxx*_*yyy*_2Dfit.py for 2D spectra only, Python code generated by Interact to fit the spectra and estimate FWHM in each dimension

/tmp/*xxx*_*yyy*_2Dfit_res.txt for 2D spectra only, fitting results (estimated parameters and confidence intervals)

/tmp/*xxx*_*yyy*_2Dfit_res.pdf for 2D spectra only, plot of simulated vs experimental spectra

/tmp/*xxx*_fit.py	Python code generated by Interact to integrate the different experiments (calculation of Kd, slope, etc)
/tmp/*xxx*_fit_res.txt	data integration results (estimated parameters and confidence intervals)
/tmp/*xxx*_fit.pdf	plot of simulated vs experimental data for data integration

4. Error and warning messages

Error messages are explicit. After correcting the problem, rerun Interact.

V. License for Interact software

See the file license.txt in the Interact folder.