



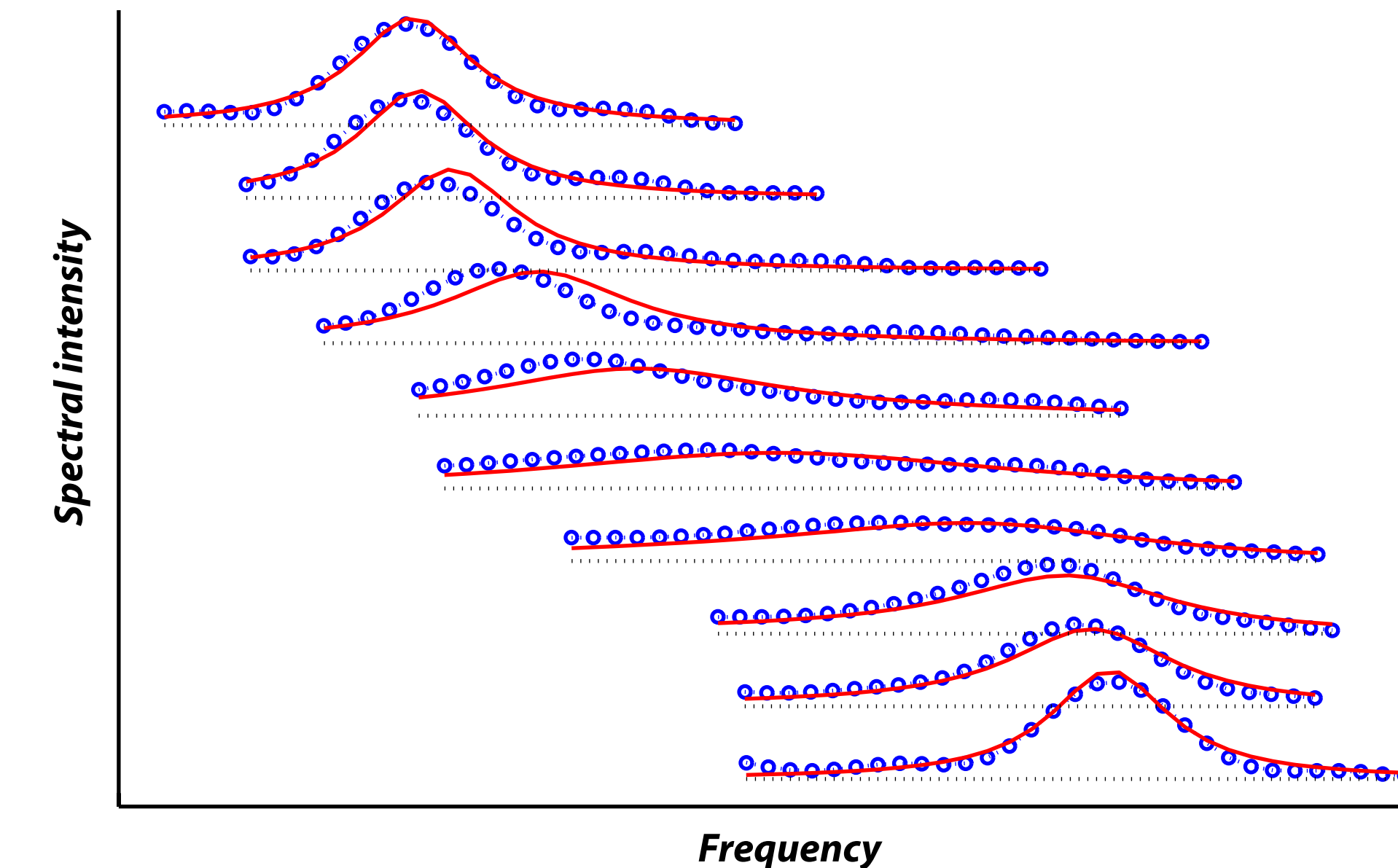
LineShapeKin

NMR Line Shape Analysis Software for Studies of Protein-Ligand Interaction Kinetics

<http://lineshapekin.net>

Evgenii L. Kovrigin

Department of Biochemistry, Medical College of Wisconsin
8701 Watertown Plank Rd., Milwaukee, WI 53226
email: ekovrig@mcw.edu



Interactions of proteins with their binding partners are characterized by the specific rates of association and dissociation (k_{on} and k_{off}). Binding kinetics is traditionally studied using rapid mixing techniques, which require highly specialized equipment not generally available. Nuclear Magnetic Resonance spectroscopy is uniquely suited for extraction of the kinetic information under equilibrium physiological conditions using standard NMR probes and hardware.

Recording a series of HSQC spectra while titrating protein with the ligand is routinely done to follow binding interactions with the site-specific resolution. Equilibrium binding constant, K_d , is readily computed from the chemical shift changes of affected nuclei. Analysis of NMR line shapes, in turn, produces the kinetic constants, k_{on} and k_{off} . However, despite frequent use of HSQC titrations in protein-ligand interaction studies, interaction kinetic constants are rarely computed. Combination of equilibrium and kinetic information extracted from HSQC spectra is essential for understanding of the mechanisms of protein-ligand interactions. A user-friendly software package **LineShapeKin** has been developed to assist analysis of the peak line shapes in HSQC spectra and extract constants of association/dissociation kinetics.

LineShapeKin software package integrates functionality of **SPARKY** NMR assignment program, **Python** programming language and **MATLAB**[®] technical computing environment. Major features of the **LineShapeKin** package include:

1. Sparky extension that allows interactive extraction of the line shape data from a series of 2D spectra.
2. Matlab notebooks that perform data fitting and graphic presentation of the results.

LineShapeKin allows fitting of the titration data on both per-residue basis (independent binding events) as well as globally (all resonances responding to the single binding event). Any subset of resonances may be fit as a group representing a binding sub-site. Akaike's Information Criterion is used to assess the likelihood of the alternative hypotheses.

LineShapeKin is freely available to academic users from <http://lineshapekin.net>

Goals

Make NMR line shape analysis easily applicable to routine HSQC titration data:

1. Utilize popular spectral analysis software to do extraction of line shapes from 2D HSQC spectra (Sparky and NMRView).
2. Utilize high-level programming environment (MATLAB) to keep the code easy to understand and modify by the end user.

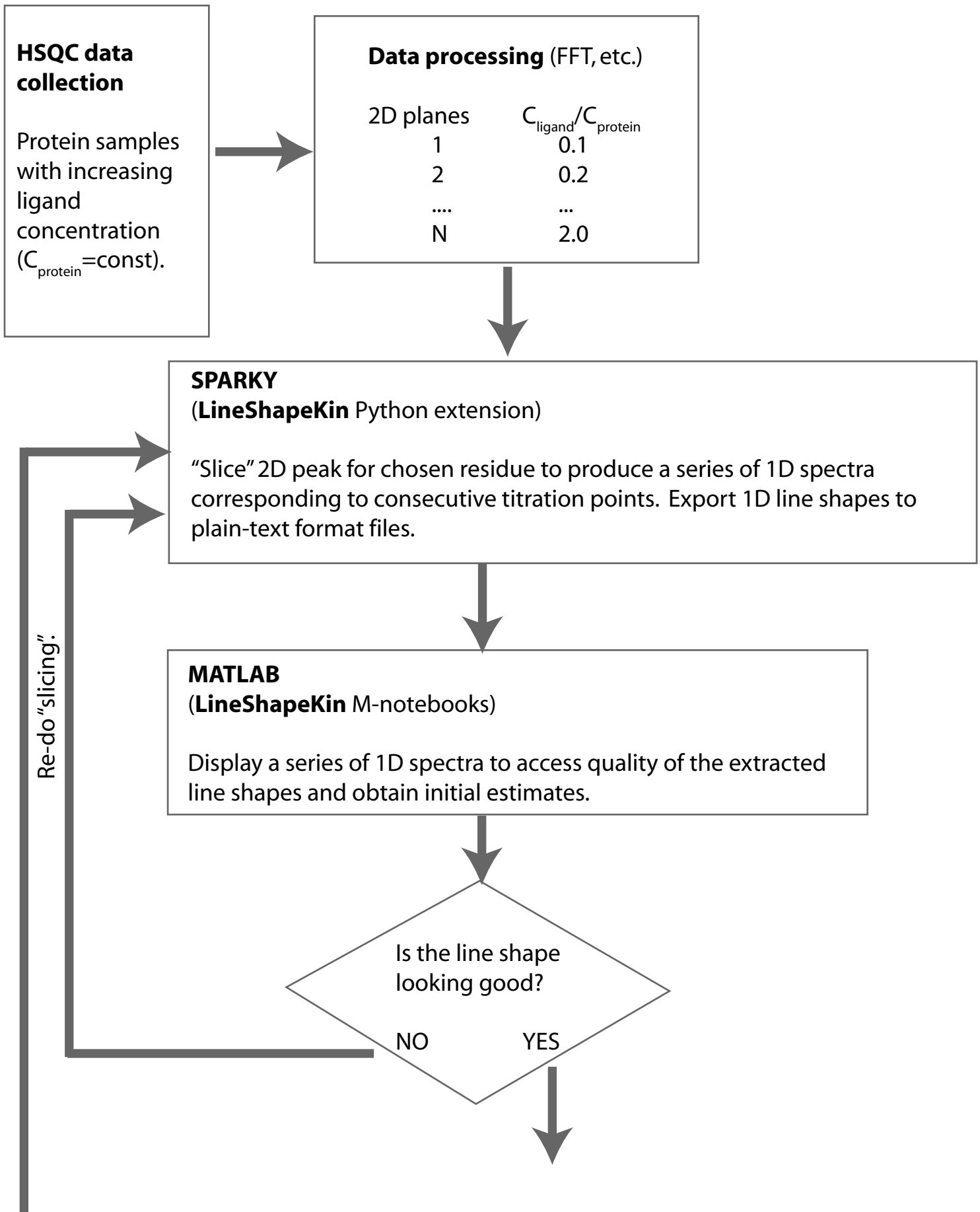
Current capabilities

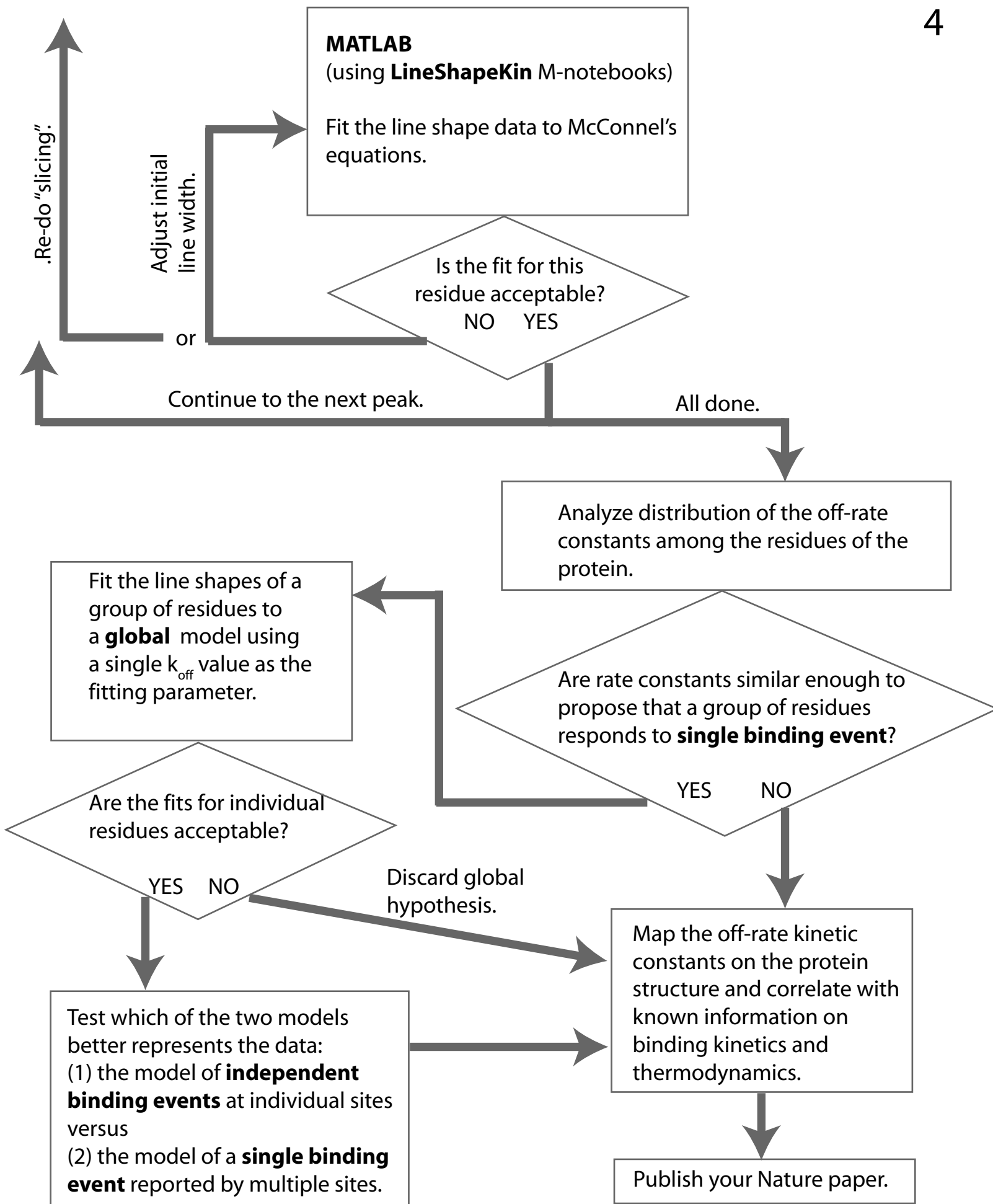
LineShapeKin performs:

1. Interactive extraction of peak slices from 2D HSQC spectra.
2. Fitting of a titration series to obtain site-specific off-rate constants, k_{off} and equilibrium constants, K_d .
3. Fitting of the end-state peak position and ligand concentration (optional).
4. Statistical testing of models with different variable parameters.
5. Global fitting of a group of residues affected by the same binding event.
6. Testing of a likelihood of the individual vs. global binding models using Akaike's Information Criterion.

Current site (computer) requirements

1. Linux/OS X/UNIX/Windows operating system.
2. Python, <http://www.python.org>
3. Sparky - NMR Assignment and Integration Software, <http://www.cgl.ucsf.edu/home/sparky/>
4. MATLAB, Mathworks, <http://www.mathworks.com/>





Considerations on the dissociation constant, K_d

5

LineShapeKin allows for fitting of K_d values. However, this is the least accurate parameter obtained from NMR titrations. If possible, one should determine K_d by an independent technique.

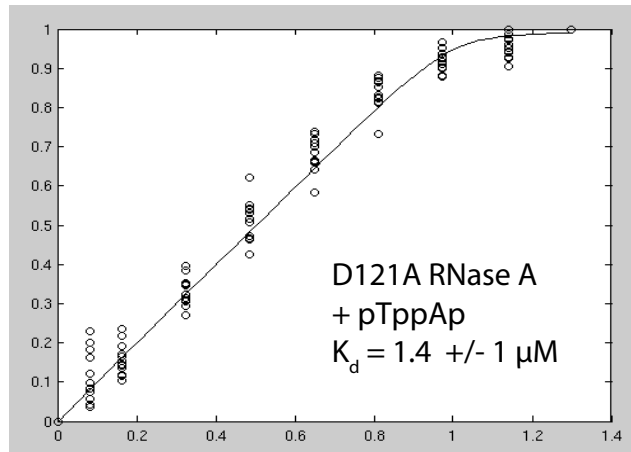
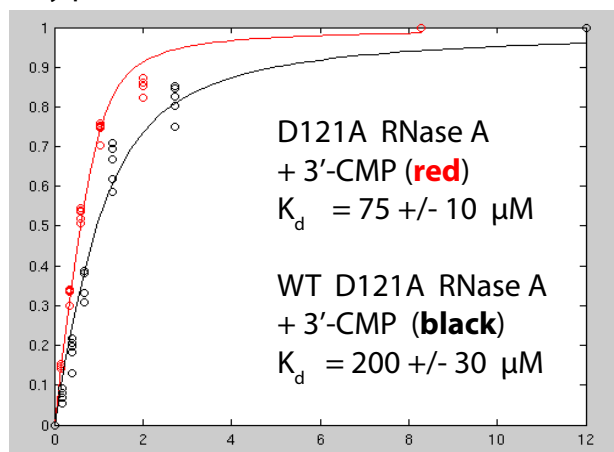
For a simple binding equilibrium $P + L \rightleftharpoons PL$, our ability to accurately measure dissociation constant depends upon both protein concentration in NMR tube and the value of dissociation constant. Conventional criterion, "c-value" is calculated as: $c = C_{\text{protein}} * n / K_d$, where n is reaction stoichiometry. Accurate determination of K_d is possible when $1 < c < 1000$.

NMR titration experiments usually employ high protein concentrations. Therefore optimality of titration conditions depend on K_d as (assume $n=1$):

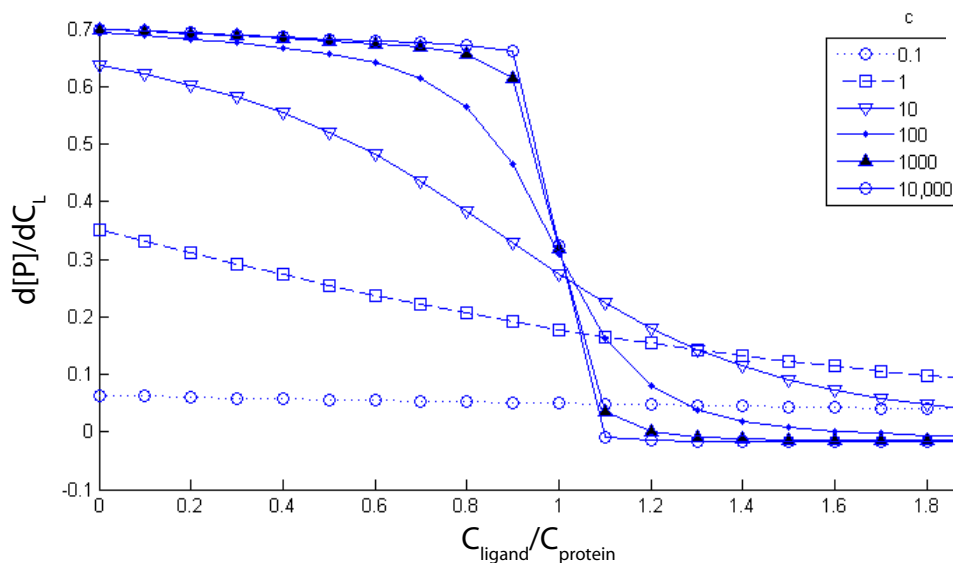
K_d	c (1mM protein sample)	c (0.1 mM)
10 mM	0.1	0.01
1 mM	1	0.1
100 μM	10	1
10 μM	100	10
1 μM	1000	100
100 nM	10,000	1000

- measurable K_d range shaded

Essentially, low c-values require very high $C_{\text{ligand}}/C_{\text{protein}}$ ratios to achieve 'complete' conversion to PL form. On the other end of scale, high c-values require very small additions of ligand near equivalence point and very precise chemical shift measurements.



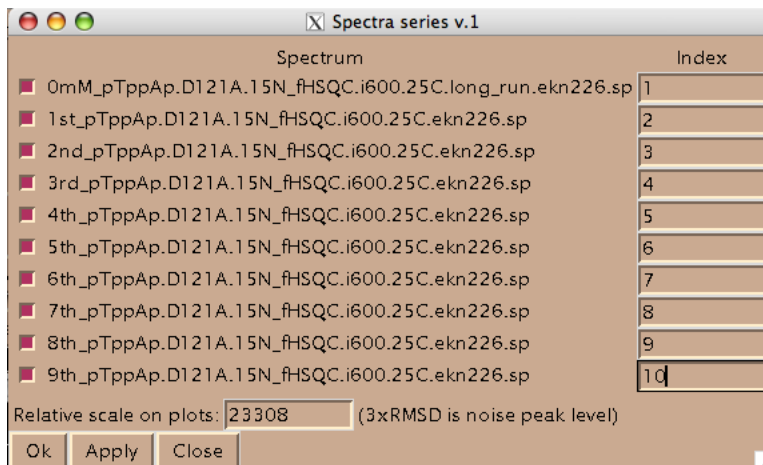
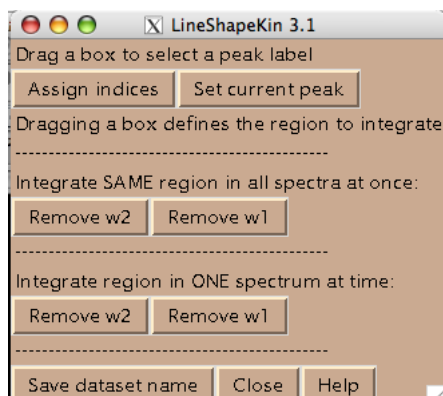
The simulated derivative of a titration curve for different c-values illustrates these important issues:



Extraction of the line shapes from spectral data

6

1. Process HSQC datasets with exponential window function (to preserve natural lorentzian lineshape, EM in nmrPipe). Zero-fill sufficiently to have ca 15-20 data points per peak envelope.
2. Invoke custom Sparky extension for extraction of 1D slices (ee) and assign consecutive indices to titration points:



3. Slices are defined by dragging a rectangle around the peak. Then one of the dimensions of the rectangle is integrated (summed) resulting in a 1D 'slice' of the peak. The program goes through all loaded spectra and performs the same integration.

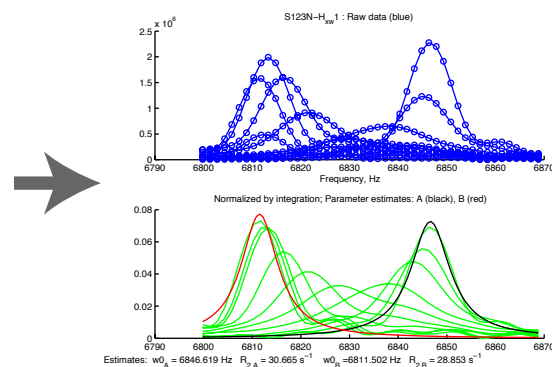
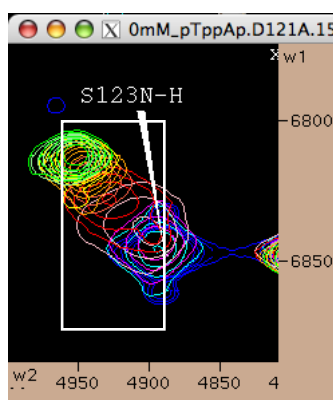
SPARKY



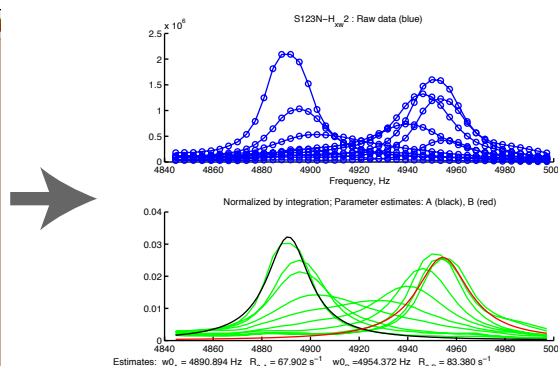
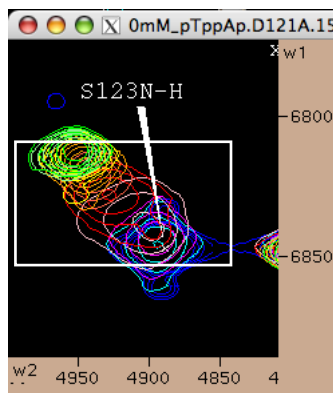
MATLAB

Integration of w_2 dimension

Top portion of the MATLAB figure corresponds to the raw spectral data (blue circles). The program normalizes peak areas to unity (lower portion, green lines) and estimates starting parameters. Black and red traces show fitting of the first and the last titration points with the estimates for A and B state peak positions and their line widths.

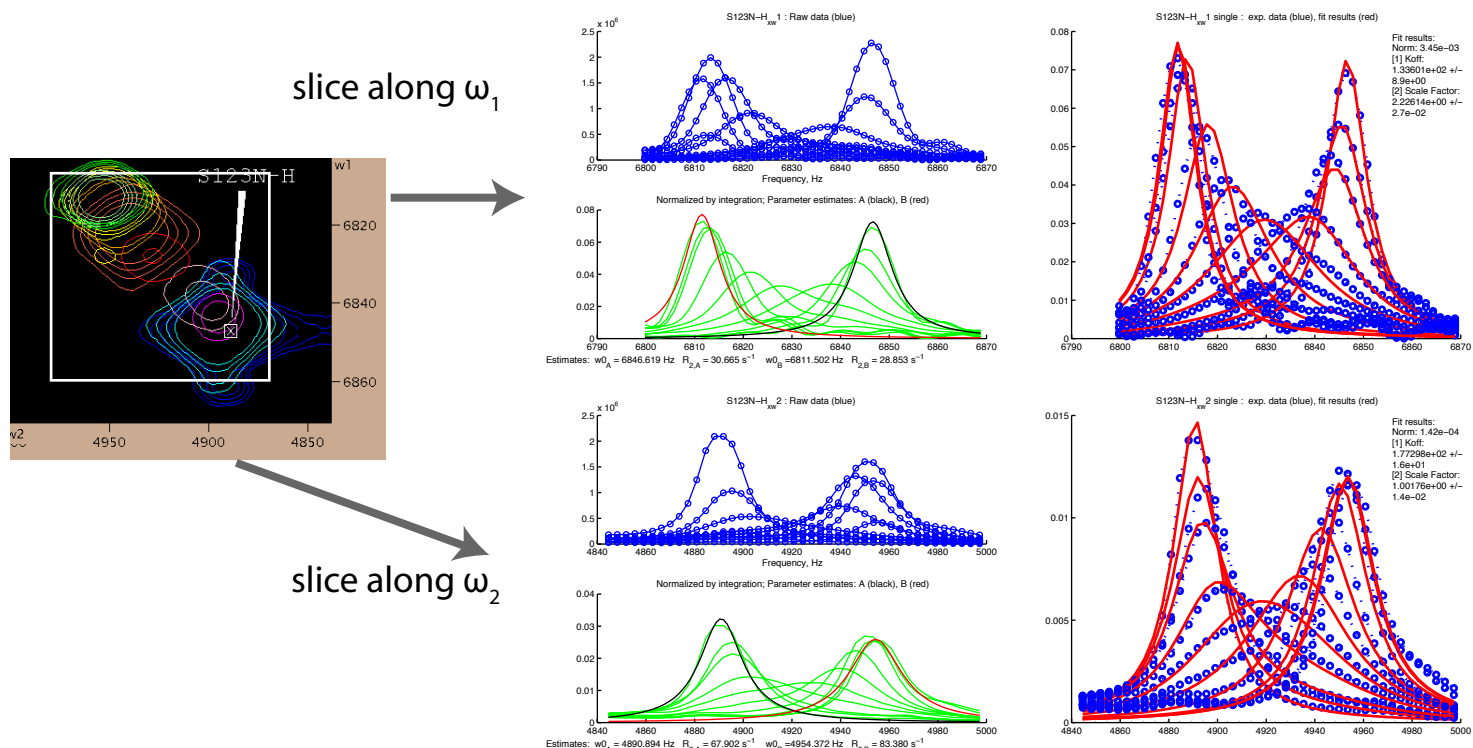


Integration of w_1 dimension

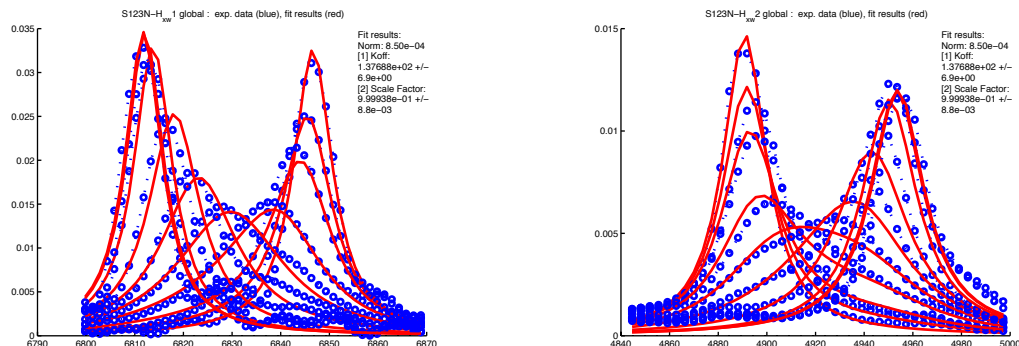


Line shapes extracted via Sparky LineShapeKin extension are fit using **LineShapeKin** M-notebooks.

Step 1. Individual fitting of different dimensions of the same peak. Here we set $K_d = 1.4 \mu\text{M}$ and obtain: $S123(\omega_1) K_{\text{off}} = 134 \pm 10 \text{ s}^{-1}$, and $S123(\omega_2) K_{\text{off}} = 180 \pm 20 \text{ s}^{-1}$.

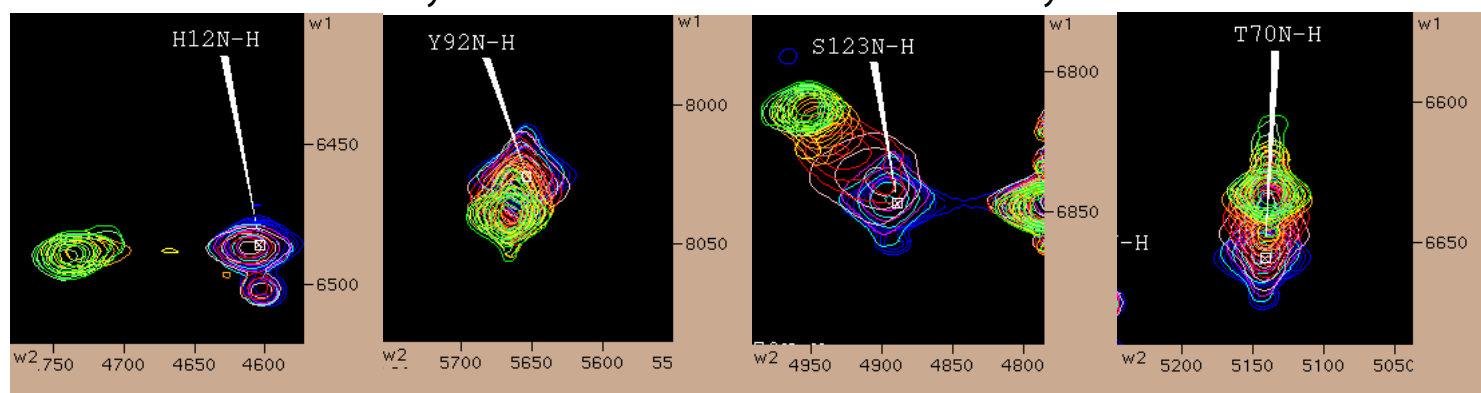


Step 2. It is likely that the amide nitrogen and proton of S123 experience the same chemical exchange process caused by ligand binding. We may fit them together and obtain the off-rate constant with better accuracy: $K_{\text{off}} = 138 \pm 7 \text{ s}^{-1}$



High resolution of HSQC spectra allows for calculation of off-rate constants for every individual site in a protein. Various molecular scenarios are possible ranging from many individual binding interactions to multiple nuclei reporting on the same binding events. **LineShapeKin** allows analysis of independent binding events as well as global binding processes.

Step 1. We select a number of residues, which we propose report on the same binding event. First we fit them individually to establish variation in the individually fit constants.



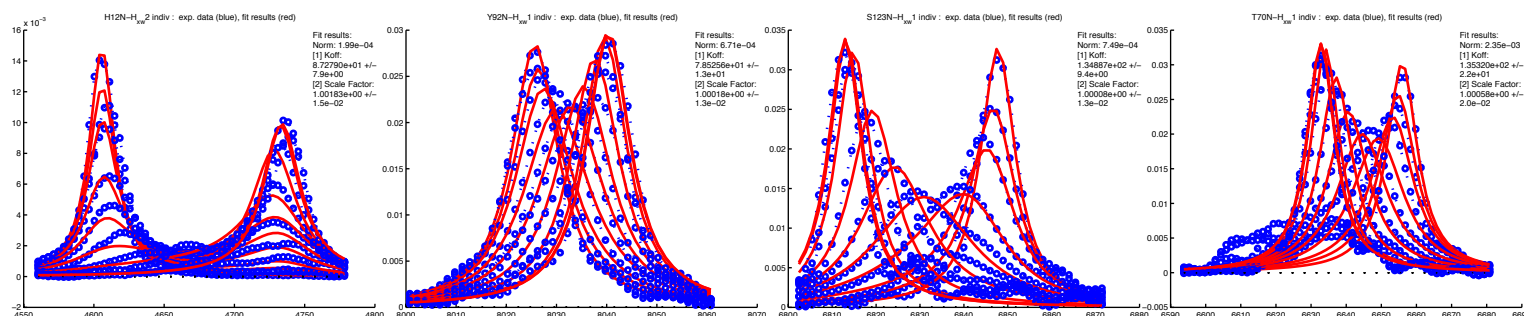
Individual fitting of these residues results in:

$$87 \pm 8 \text{ s}^{-1}$$

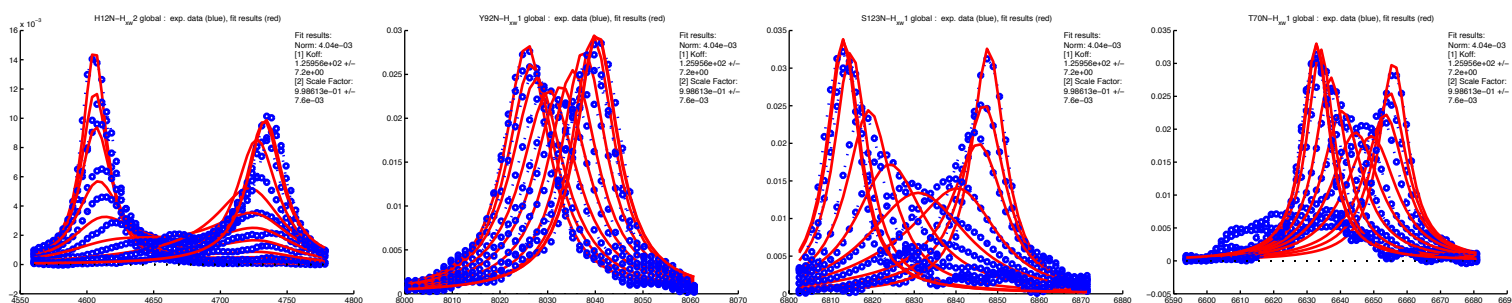
$$78 \pm 13 \text{ s}^{-1}$$

$$134 \pm 10 \text{ s}^{-1}$$

$$135 \pm 23 \text{ s}^{-1}$$



Step 2. Now we will treat them as a group that responds to the same binding event and fit them globally to a single K_{off} . Result is $127 \pm 8 \text{ s}^{-1}$:



Key issue: we need to judge whether the *global binding model* or a model involving *individual different binding interactions* is more correctly describing this data?

Hypothesis testing using Akaike Information Criterion 9

LineShapeKin uses Akaike's Information Criterion (AIC) to determine which model is *more likely* to be correct and *how much more likely*. Unlike the F test, Akaike's method *does not require models be 'nested'*, they may be totally *unrelated*. Similarly to Fisher hypothesis testing, here we assess whether using a model with a larger number of fitting parameters brings us *more than proportional* reduction in deviation of fitted curve from experimental values.

Akaike's Information Criterion is calculated as:

$$AIC = N \ln \frac{SS}{N} + 2K \quad (1)$$

where N is a number of data points, SS is sum of the squares of the residuals from fitting and K is a number of fitting parameters plus one. N must be significantly greater than K.

With smaller data sizes, when N is not much greater than K (but, at least, 2 greater), a second-order (corrected) Akaike's information criterion is more accurate:

$$AIC_c = AIC + \frac{2K(K+1)}{N-K-1} \quad (2)$$

AICSc-s themselves are not informative, it is the absolute difference between the two AICc that defines how much more likely one model is correct relatively to another. We use evidence ratio calculated as

$$\text{Evidence ratio} = \frac{\text{probability that model 1 is correct}}{\text{probability that model 2 is correct}} = \frac{1}{e^{-0.5(\Delta AIC_c)}}, \quad (3)$$

where $\Delta AIC_c = AIC_2 - AIC_1$

Applying AICc to the previously shown data yields:

1. Individual fitting of the proton and nitrogen dimensions of S123 versus simultaneous fitting to a single off-rate constant:

1st Model (global fitting), AICc=-194
2nd Model (individual fitting), AICc=-187
Evidence ratio = 26

Hypothesis that proton and nitrogen of S123 report on the same binding event is 26x more likely correct than an alternative one that states they report on separate exchange processes.

2. Four residues in RNase responding to a binding of the specific inhibitor pTppAp:

1st Model (global fitting), AICc=-361
2nd Model (individual fitting), AICc=-344
Evidence ratio = 4100

In accord with near-perfect fitting of data with the global model, these residues report on the same binding process (global hypothesis is 4100x more likely to be correct).

Summary

1. **LineShapeKin** is an easy-to-use NMR line shape analysis tool for studies of binding equilibria in proteins. The custom extension for SPARKY spectral analysis software allows interactive extraction of line shapes from HSQC titration series. MATLAB-based notebooks perform fitting, display and statistical analysis of the data on reaction kinetics. The package is fully modifiable and easily extendible to the more complicated binding models.

2. **LineShapeKin** is interfaced with the popular spectral analysis software. Current version of the package utilizes SPARKY for spectral data handling and display. Future versions of **LineShapeKin** will interface to NMRView with the goal to make kinetic information via NMR line shape analysis easily accessible for biomolecular NMR community.

3. Spectral overlap remains a significant hurdle in NMR data analysis. When analyzing titration series this problem is made worse by simultaneous shifting of a number of neighboring peaks. **LineShapeKin** allows to define individual extraction regions for HSQC spectra corresponding to different titration points thus increasing number of resonances amenable to line shape analysis.

4. One of the most important issues in science is availability of free academic software. Currently, **LineShapeKin** relies on efficient MATLAB computing environment to develop a full featured code while keeping its scientific content not obscured by details of programming implementation. However, using MATLAB to run **LineShapeKin** requires licensing from MathWorks, Inc. It is possible to develop a compiled version of **LineShapeKin** to relieve end user from necessity of obtaining and licensing MATLAB.

Acknowledgements

Evgenii Kovrigin acknowledges Dr. Roger Cole, Dr. James Kempf and Dr. Patrick Loria for contributions to the code at an early stage of the package development, Dr. Linda Nicholson and Soumya De for testing the package and helpful feedback.

Brief theoretical background on the NMR line shape analysis

Consider the elementary reaction:

$$P + L \xrightleftharpoons[k_{off}]{k_{on}} PL$$

The equation of motion for the transverse magnetization in the rotating frame will be:

$$\frac{dM^+(t)}{dt} = (i\Omega - R + K)M^+(t),$$

where M is transverse magnetization vector for the protein spins in free and bound forms, Ω is a diagonal matrix of their Larmor frequencies in the rotating frame, R - diagonal matrix of their transverse relaxation rates, and K is the rate matrix for the reaction.

The rate matrix K is

$$K = \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix},$$

where $k_2 = k_{off}$, while k_1 is a pseudo-first order rate constant incorporating equilibrium concentration of a ligand:

$$k_1 = k_{on}[L]$$

Intensity of the frequency domain spectrum is proportional to a total magnetization and may be obtained for the steady-state reaction conditions from:

$$I(\nu) \propto \text{Re} \left[[1, 1] (i(\Omega - \nu) - R + K)^{-1} \begin{bmatrix} p_P \\ p_{PL} \end{bmatrix} \right],$$

where p_P and p_{PL} are equilibrium mole fractions of free and complexed protein. One can readily compute their values for every ligand concentration from knowledge of the equilibrium binding constant.

For more details, refer to [1-3].

1. Binsch, G., *Top. Stereochem.*, 1968. **3**: p. 97-191.
2. McConnell, H., *J. Chem. Phys.*, 1958. **28**: p. 430-431.
3. Cavanaugh J, F.W., Palmer III AG, Skelton NJ, *Protein NMR Spectroscopy: Principles and Practice*. 2006: Academic Press. 587.