SCIENTIFIC REPORTS
nature research

OPEN

NmrLineGuru: Standalone and User-Friendly GUIs for Fast 1D NMR Lineshape Simulation and Analysis of Multi-State Equilibrium Binding Models

Chao Feng[1], Evgenii L. Kovrigin[2] & Carol Beth Post[1,3*]

The ability of high-resolution NMR spectroscopy to readout the response of molecular interactions at multiple atomic sites presents a unique capability to define thermodynamic equilibrium constants and kinetic rate constants for complex, multiple-step biological interactions. Nonetheless, the extraction of the relevant equilibrium binding and rate constants requires the appropriate analysis of not only a readout that follows the equilibrium concentrations of typical binding titration curves, but also the lineshapes of NMR spectra. To best take advantage of NMR data for characterizing molecular interactions, we developed NmrLineGuru, a software tool with a user-friendly graphical user interface (GUI) to model two-state, three-state, and four-state binding processes. Application of NmrLineGuru is through stand-alone GUIs, with no dependency on other software and no scripted input. NMR spectra can be fitted or simulated starting with user-specified input parameters and a chosen kinetic model. The ability to both simulate and fit NMR spectra provides the user the opportunity to not only determine the binding parameters that best reproduce the measured NMR spectra for the selected kinetic model, but to also query the possibility that alternative models agree with the data. NmrLineGuru is shown to provide an accurate, quantitative analysis of complex molecular interactions.

NMR is a powerful technique to study molecular structure, dynamics and kinetics without the need for chemical modification. Due to its atom-level resolution, it is particularly useful to study molecular interactions involving multi-state equilibrium and can be used to simultaneously extract thermodynamics and kinetics parameters of each binding step[1–5]. In contrast, the readout of other label-free techniques, such as isothermal titration calorimetry, contains averaged information of multiple binding steps and/or provides estimates of only thermodynamics parameters[6,7].

To study binding interactions by NMR, a titration experiment is usually performed. At each titration point, an NMR-observable molecule (P) is mixed with its binding partner (L) at a certain ratio and 1D or 2D NMR spectra are recorded. The changes of chemical shifts, peak widths, and peak intensities observed during this process are then used to extract thermodynamics and kinetics parameters.

NMR titration data exhibit different behavior depending on the binding process and exchange regimes set by the relative values of the on/off kinetic rates and chemical shift frequencies of the various molecular species. For the fast-exchange regime, the observed chemical shifts are a weighted average of the two exchanging states and thus thermodynamic parameters can be extracted directly from the normalized chemical shifts as for any titration binding curve[1,8]. For the slow-exchange regime, the peak intensities of exchanging states are proportional to their population and thus can be used to extract thermodynamic parameters[1,6]. However, if the exchange falls in the intermediate regime or involves multiple binding steps or both thermodynamics and kinetics parameters are wanted, a lineshape analysis using the whole peak envelope is necessary.

[1]Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology, and Purdue Center for Cancer Research, Purdue University, West Lafayette, IN, 47907, USA. [2]Magnetic Resonance Research Center, Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN, 46556, USA. [3]Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907, USA. *email: cbp@purdue.edu

Name	Dependency	Extra requirements	Simulation functionality	Fitting functionality
TITAN[12]	NMRPipe MATLAB R2016b*	Spec. acqu. parameters 2D lineshape only	Scripts-based Up to 4 states	GUI-based Up to 4 states
IDAP[1,2]	MATLAB R2014a	1D lineshape only	Scripts-based Up to 6 states	Scripts-based Up to 3 states
LineShapeKin Simulation[1]	MATLAB R2011b	1D lineshape only	Scripts-based Up to 6 states	N/A
LineShapeKin[13]	MATLAB R2011b	1D lineshape only	N/A	Scripts-based Up to 2 states
NMRKIN[14]	NMRLab MATLAB?	1D lineshape only	Scripts-based Up to 3 states	Scripts-based Up to 3 states

Table 1. Summary of existing NMR lineshape software. The shown MATLAB version is known to work with the indicated software but may not be the only version. "?" indicates unknown version. "N/A": not available in the indicated software. *TITAN currently has standalone version for OS X and Linux, but "some features of TITAN are not available in the standalone application, and are currently only accessible from within MATLAB".

Although lineshape analysis can extract both thermodynamics and kinetics parameters for multi-state equilibrium binding models and all types of exchange regimes, the calculation steps are more complicated by involving both the equilibrium binding constants and kinetics rate constants. This complication limits the wide application of NMR lineshape analysis method.

Many published works involving NMR lineshape analysis utilized home-written scripts which cannot be easily adopted[2,3]. There are also several published software packages for NMR lineshape analysis which are summarized in Table 1. All the listed NMR lineshape analysis software packages require MATLAB installation, which is a commercial computing environment with expensive licensing fees. In addition, the MATLAB language syntax changes in different versions and older versions (required for older software) are harder to obtain and use as the compatible operating systems become obsolete. In addition, none of the listed software packages is fully based on a graphical user interface (GUI) for both lineshape simulation and fitting; running the script-based simulation or fitting requires a certain level of programming knowledge for MATLAB.

Based on our previous work in NMR lineshape analysis utilizing multi-state equilibrium models[3], we developed NmrLineGuru, a standalone and user-friendly NMR lineshape software containing six GUIs for simulating and fitting NMR lineshape with two-, three-, and four-state binding models. These GUIs have no dependency on other software; NmrLineGuru is developed in the MATLAB environment, but does not require separate installation of MATLAB. NmrLineGuru aims to be extremely user-friendly for non-experts and time-saving with the hope of promoting the application of NMR lineshape analysis in research.

Results and Discussion

Models and interface. NmrLineGuru aims to be extremely friendly for non-experienced users. It currently supports the most commonly used two-, three-, and four-state binding models. For each model, there are two single-window GUI applications, one for simulating and the other for fitting 1D NMR lineshape data. Figure 1 shows screenshots of the 2- and 3-state GUIs as examples (see Supporting Information Fig. S1 for the 4-state GUIs). The input fields on each GUI are grouped and arranged by logical order to facilitate use. Default parameters are filled in by example, and to enable one-click simulation or fitting.

Overview of the workflow. The general workflow of NmrLineGuru is shown in Fig. 2. Users are expected to provide only the minimal amount of information in the single-window GUI and the GUI will handle the subsequent steps automatically.

The simulation GUIs take the default or user-provided thermodynamic, kinetic, and resonance parameters to generate 1D NMR lineshape data for the user-specified concentrations of each species in a given titration series according to the selected model. The generated data can contain arbitrary noise and any number of points as requested by the user. The results will be automatically plotted, displayed, and saved in various formats (png, eps, fig, and txt) according to user preference.

The fitting GUIs read in the user-provided lineshape data to search for the best-match thermodynamic and kinetic parameters. The input can be data from the simulation GUIs, experimental data from 1D NMR spectra, or 1D slices from 2D HSQC experiments. The input format (detailed in online tutorials) is simple two-column text files that can be generated or converted from most NMR-spectrum software packages. For the users' convenience, a plugin for data export is provided and integrated in the NMRFAM[9] distribution of Sparky (Goddard TD & Kneller DG, University of California, San Francisco), the most popular NMR spectra visualization tool[9]. After reading the lineshape data and initial parameters, the fitting GUIs automatically normalize the input data, perform a Lorentzian fit to estimate resonance- parameters, iteratively search the best-fit dynamic and kinetic parameters within the user-specified range of values, perform Monte Carlo error estimation, and output the results as plots and tables. Automatic global fitting will be performed if multiple datasets are found in the input data directory.

Example 2-state simulations and fittings. The 2-state simulation GUI simulates the single-step binding process ($P + L \leftrightarrow PL$) with arbitrary parameters. Figure 3 shows example 2-state simulations using the prefilled default parameters (300 μM [P_{total}]; 0–900 μM [L_{total}]; 10 μM K_D; NMR frequencies $\omega_0 = 0$, 500 s^{-1} for P and PL, respectively; all line widths $d\omega = 100\ s^{-1}$) with different k_{off} values ranging from 0.01 to 100 × Δω (5–50000 s^{-1})

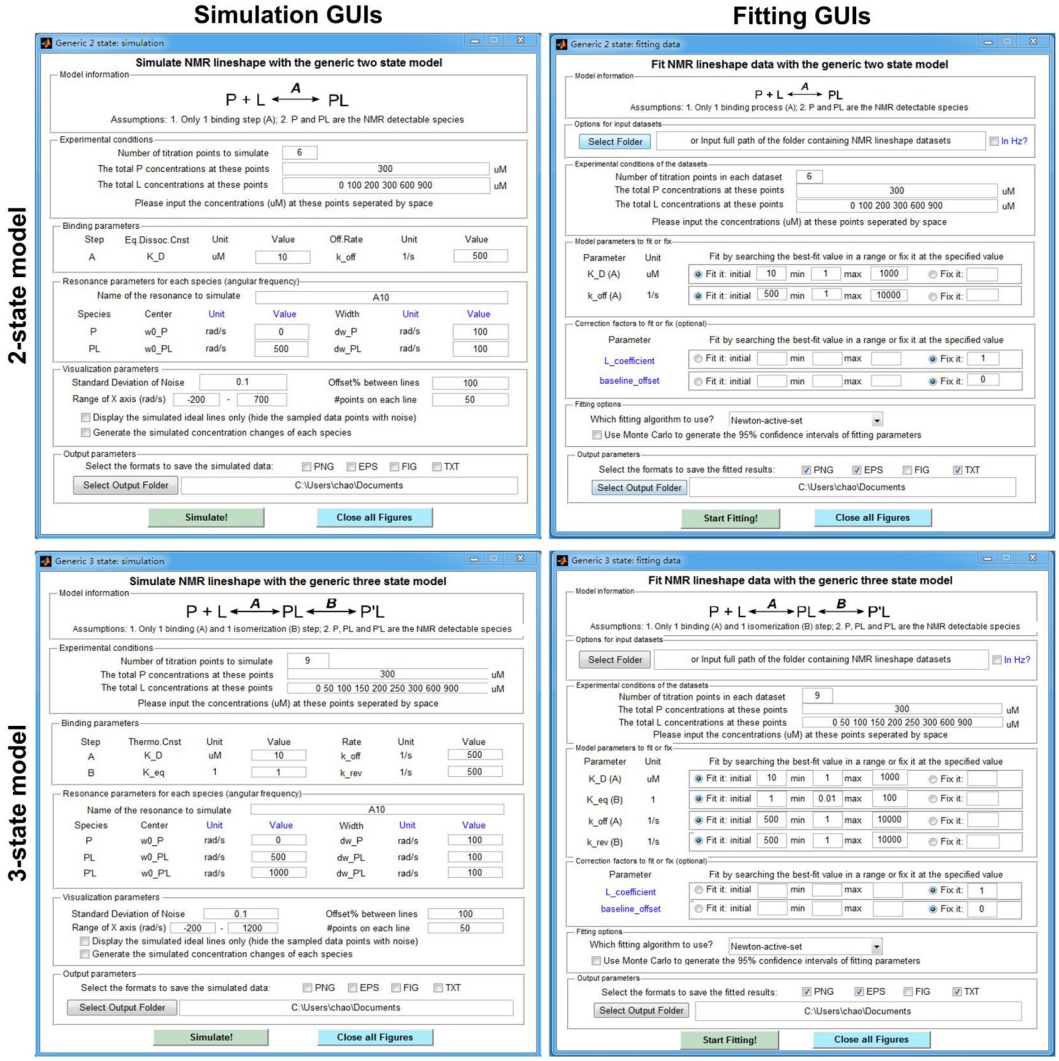

Figure 1. Example lineshape simulation and fitting GUIs. Shown are GUIs for the 2- and 3-state models. GUIs for the 4-state model are shown in Supporting Information Fig. S1. Each GUI is an independent single-window application. Example parameters are filled in and can be used for quick simulation or fitting. User-provided values enable customized simulations or fittings of NMR lineshape data.

to highlight the effects of exchange kinetics on NMR lineshape. For a 2-state binding system (free P and bound PL), the exchange rate is $k_{ex} = k_{on}[L] + k_{off}$, which is dominated by k_{off} at low ligand concentration. As expected, the generated lineshapes show canonical slow-exchange behavior (Fig. 3b) when $k_{off} \ll \Delta\omega$ and fast-exchange behavior (Fig. 3f) when $k_{off} \gg \Delta\omega$. For moderate k_{off} values ($0.1–10 \times \Delta\omega$), the resonances broaden to various extent and shows intermediate-slow or intermediate-fast behavior (Fig. 3c–e).

These simulated 2-state lineshape data (50 data points sampled on each line with random noise, and signal/noise ≈50; a typical signal/noise level of 50–100 was seen in our NMR titration studies, such as the 29 kD Syk tSH2 constructs[3,7]) are fitted with the 2-state fitting GUI (Fig. 3 and Table 2). In all cases, the GUI correctly determines the K_D value. k_{off} values are determined for the range of $0.1–10 \times \Delta\omega$, and cannot be determined for values $\geq 100 \times \Delta\omega$, which is above the fast-exchange limit (i.e. spectrum appearance becomes insensitive to the k_{off} values above the limit).

Example 3-state simulations and fittings. The 3-state simulation GUI can simulate any [L]-dependent binding followed by a [L]-independent step such as ligand isomerization or ligand-induced protein conformational change (Fig. 4a). Figure 4 displays example 3-state simulations (300 μM [P_{total}]; 0–900 μM [L_{total}]; 10 μM K_D; 1 K_{eq}; NMR frequencies $\omega_0 = 0$, 500, 1000 s^{-1} for P, PL and P′L, respectively; all line widths $d\omega = 100\,s^{-1}$) for fast or slow binding coupled with fast or slow isomerization (see Table 3 for k_{off} and k_{rev} values). For simplification, intermediate exchange is not illustrated for either step. For this 3-state binding system (free P, bound PL and isomerized P′L), the isomerization step lacks dependency on ligand concentration so species PL and P′L form at a constant ratio defined by the equilibrium constant $K_{eq} = [PL]/[P'L]$. Coupled with the [L]-dependent exchange between P and PL, this system has interesting behavior. When the isomerization step is slow exchange

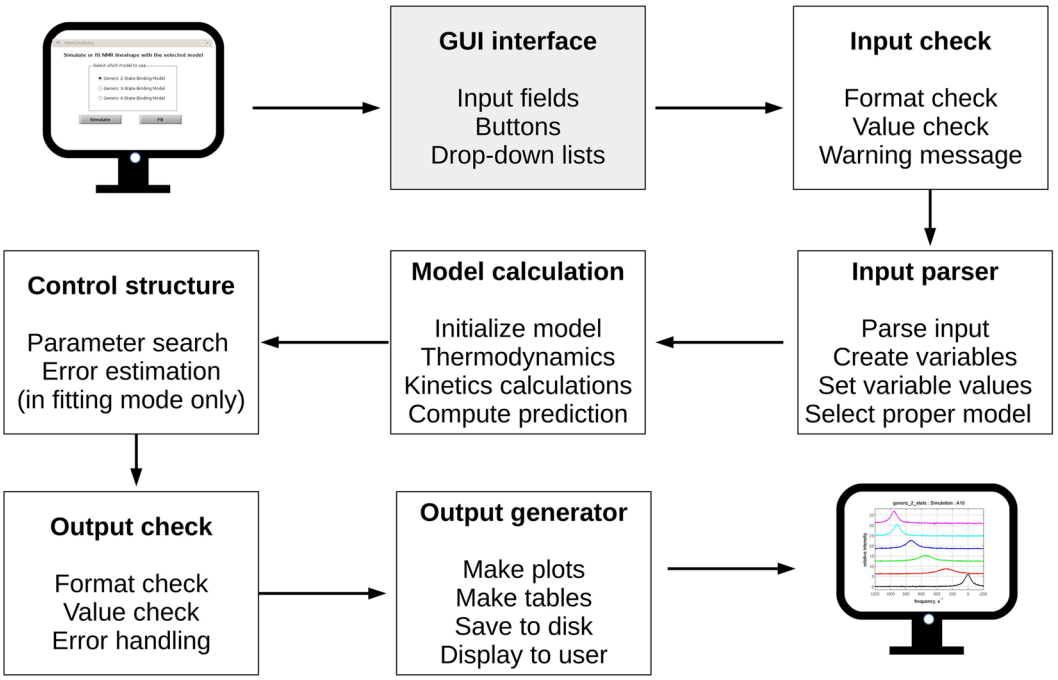

Figure 2. Workflow of NmrLineGuru. Users only need to interact with the graphical user interface (GUI) while all other steps are handled automatically by the software. The results are displayed as on-screen plots and written to the user-specified folder in chosen formats.

(Fig. 3c,e), PL and P′L resonances are separate peaks; the system can appear like 2-state exchange between P and PL or P′L depending on K_{eq}. When the isomerization step is fast exchange (Fig. 3d,f), PL and P′L resonances show up as one peak; the system looks like 2-state exchange between P and the PL/P′L-degenerate peak. The behavior therefore can be deceptive and careful inspection of the spectra is needed when investigating the detailed binding mechanism[1].

The combined fitting and simulation capability of NmrLineGuru enables facile consideration of different binding models. The simulated 3-state lineshape data (50 data points on each line with noise, signal/noise ≈50) were fitted with the 2- and 3-state fitting GUIs. The results are summarized in Table 3. As required, the 3-state model fits well with all the lineshape data (Supporting Information Fig. S2), and gives correct thermodynamic and kinetic parameters within 95% confidence intervals for most values. The difference in K_D estimated from resonance in Fig. 4c and the actual value is less than a factor of two, which is typically considered to be within experimental error in practice. Interestingly, the 2-state model agrees well with the lineshape data when the isomerization is in fast exchange (Fig. 4d,f and Supporting Information Fig. S3). When both steps are in slow exchange, and the resonances for P and PL are selected for the 2-state fitting, ignoring the P′L resonance, the lineshapes are well fitted (Fig. 4c and Supporting Information Fig. S3). But for the case where ligand binding is in fast exchange and isomerization is slow (Fig. 4e), fitting is poor even though the P′L resonance is ignored and deviations in fitting the NMR lineshapes are sufficient to suspect an incorrect model. Nevertheless, even for the cases where the lineshapes are well fitted, the K_D and k_{off} values are inaccurate, especially when both the binding and isomerization steps are in slow exchange (Table 3). In part, the apparent good fit of the lineshapes is due to the normalization step imposed in the fitting algorithm.

Example 4-state simulations and fittings. The 4-state simulation GUI can simulate any system with two independent or coupled binding sites and arbitrary binding affinities (Fig. 5a). Figure 5 shows example 4-state simulations for two independent binding sites with similar binding affinities (300 μM [P_{total}]; 0–900 μM [L_{total}]; 10 μM K_D for all binding steps; NMR frequencies $\omega_0 = 0$, 400, 600, 1000 s^{-1} for P, PL, LP and LPL, respectively; all line widths $d\omega = 100\ s^{-1}$), and disparate kinetic rate constants (see Table 4). Lineshape simulations are only shown for the limiting conditions of fast or slow exchange. When binding to both sites is in slow exchange, the NMR lineshape shows four resonances for the free, two singly ligated (LP and PL) and doubly ligated (LPL) forms of the protein (Fig. 5c), with the peak intensity for P gradually decreasing, that for LP and PL increasing and then decreasing, and that for LPL gradually increasing, which directly reflects the population change shown in Fig. 5b. When one of the two sites is in fast exchange, the NMR lineshape is dominated by the rapid kinetics for binding either free protein or protein ligated at the alternate site, and the spectrum looks like two fast-exchange binding processes in that the line frequency appears to follow the equilibrium concentration; however, the peak intensity for free P decreases and that for the ligated form increases (Fig. 5d,e). When both sites are in fast exchange, the NMR lineshape is collapsed into a single resonance at all titration points and looks like a 2-state binding system (Fig. 5f).

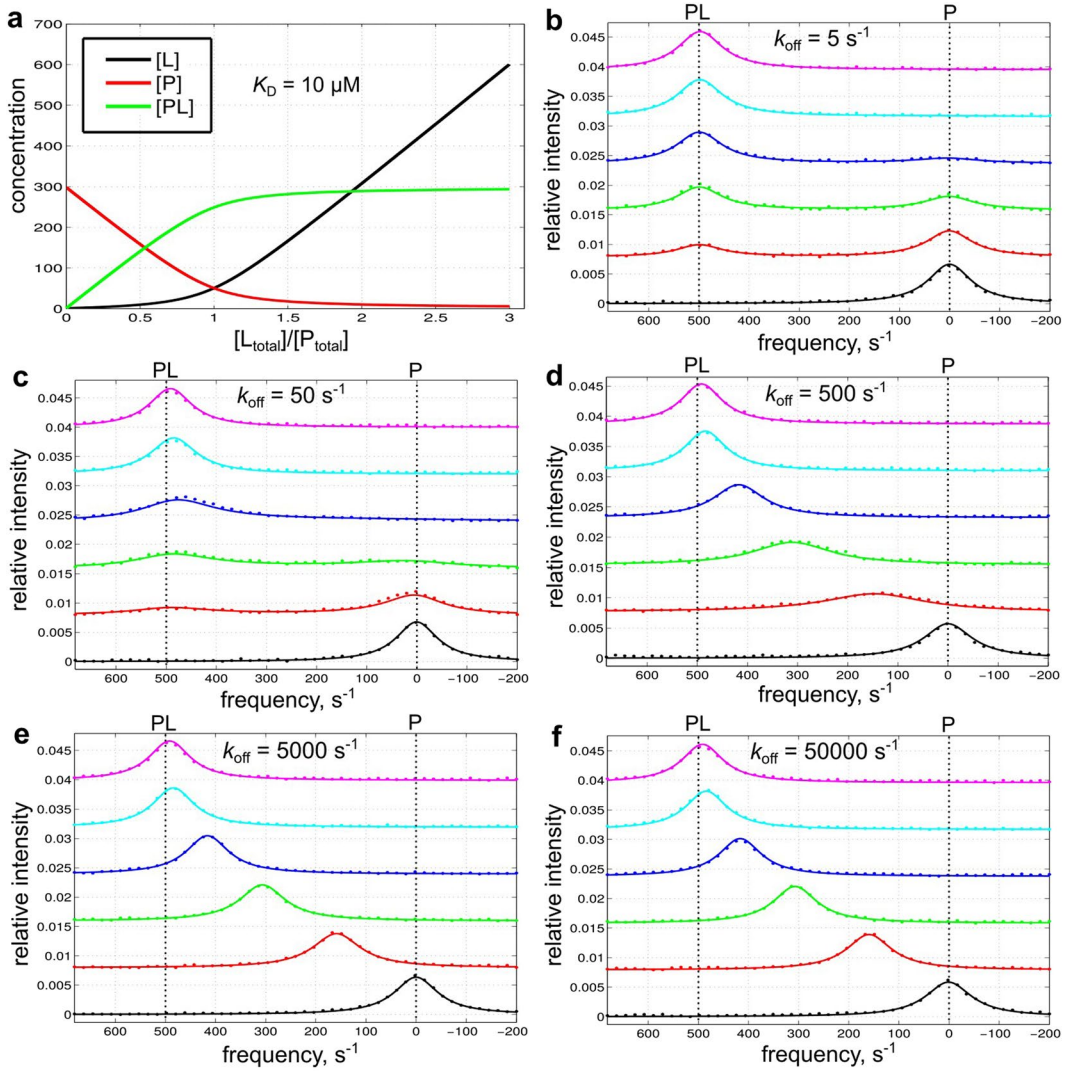

Figure 3. Example 2-state simulations. Simulations use a range of k_{off} values and the prefilled default values for other parameters: 300 μM [P_{total}]; 0, 100, 200, 300, 600, 900 μM [L_{total}]; 10 μM K_D; $\omega_0 = 0$, 500 s^{-1}; line width $d\omega = 100\ s^{-1}$. (**a**) Simulated concentration change of each species as a function of [L_{total}]/[P_{total}]. (**b**–**f**) Simulated lineshapes with different k_{off} values ranging from 0.01–100 × Δω (500 s^{-1}). The simulated data (50 data points on each line with noise, signal/noise ≈50) are shown as dots. These data are subsequently fitted with the 2-state fitting GUI and the fitted lineshapes are shown as smooth curves.

	True parameters		2-state fitting	
Figure	K_D (μM)	k_{off} (s^{-1})	K_D (μM)	k_{off} (s^{-1})
3b	10	5	11 ± 2	5.2 ± 0.7
3c	10	50	11 ± 1	45 ± 3
3d	10	500	10 ± 1	580 ± 40
3e	10	5000	10 ± 1	6300 ± 1500
3f	10	50000	10 ± 1	UB

Table 2. Summary of example 2-state fittings. The input data are from 2-state simulations shown in Fig. 3 as indicated, containing noise (signal/noise ≈50) and 50 data points on each line. The data are fitted with the 2-state fitting GUI. The parameter searching range is [1 nM, 1 mM] for K_D and [0.1 s^{-1}, 100000 s^{-1}] for k_{off}. The fitted parameter values are shown with 95% confidence intervals determined by Monte Carlo resampling. UB: not determined due to reaching the upper boundary.

These simulated 4-state lineshape data (50 data points on each line with noise, signal/noise ≈50) were fitted with the 2- or 4-state fitting GUI. The 2-state model cannot fit any of the lineshape data, not even those from Fig. 5f which look like 2-state exchange (Fig. 5e,f; Supporting Information Fig. S4). In contrast, the 4-state model

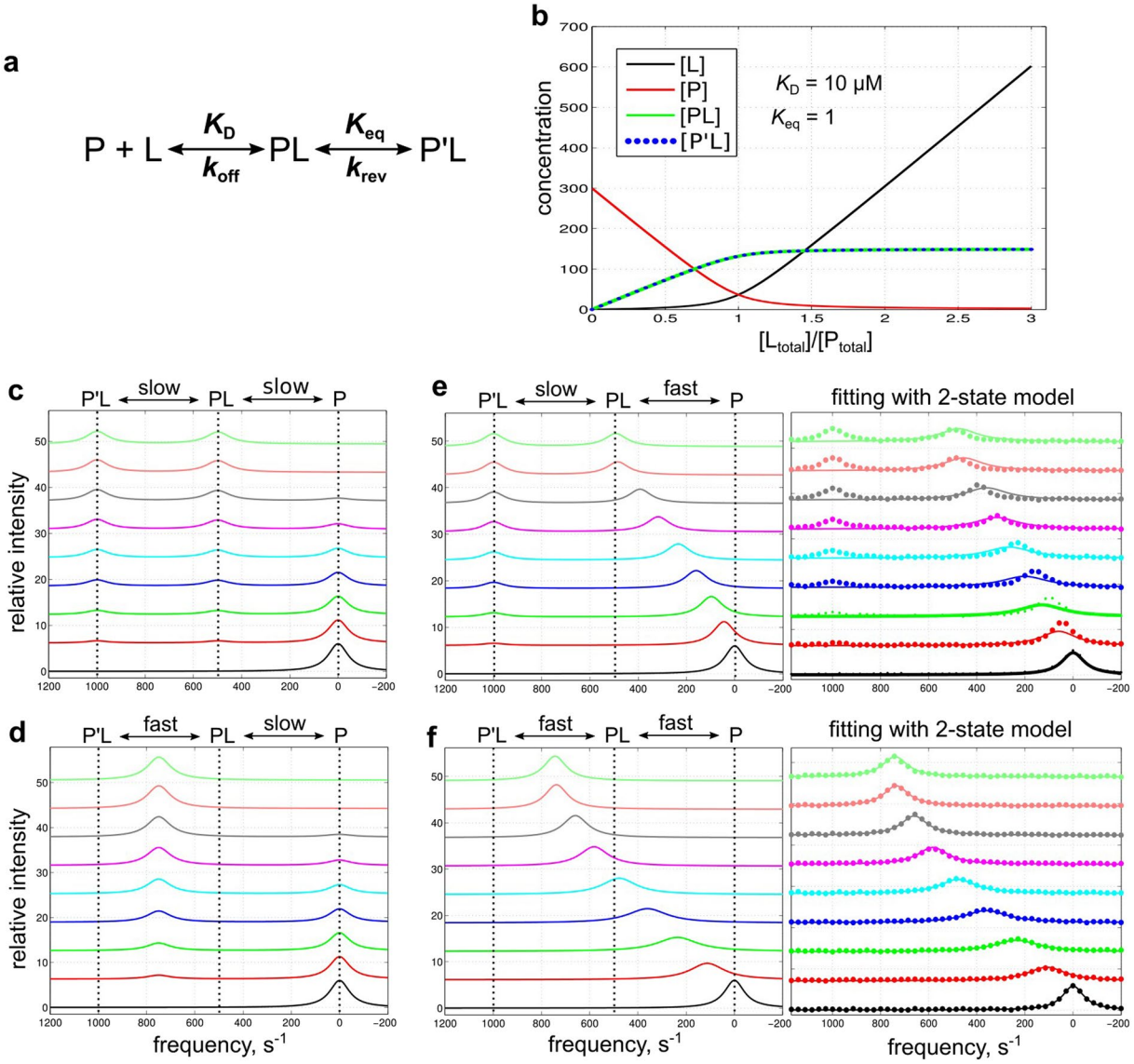

Figure 4. Example 3-state simulations. Simulations use a range of kinetic rates and the prefilled default values for other parameters: 300 μM [P_{total}]; 0, 50, 100, 150, 200, 250, 300, 600, 900 μM [L_{total}]; 10 μM K_D; 1 K_{eq}; $\omega_0 = 0$, 500, 1000 s^{-1}; line width $d\omega = 100\ s^{-1}$. (**a**) The 3-state binding model. (**b**) Simulated concentration change of each species as a function of [L_{total}]/[P_{total}]. (**c**–**f**) Simulated lineshape with different k_{off}/k_{rev} values (slow: 5 s^{-1}; fast: 5000 s^{-1}). Fifty data points on each simulated line are sampled with noisy (signal/noise ≈50) and subsequently fitted with the 3- or 2-state fitting GUI. For panels e and f, the sampled data points and fitted curves with the 2-state GUI are shown on the right as examples. See Supporting Information Figs S2 and S3 for all fittings.

	True parameters				3-state fitting				2-state fitting	
Figure	K_D (μM)	K_{eq}	k_{off} (s^{-1})	k_{rev} (s^{-1})	K_D (μM)	K_{eq}	k_{off} (s^{-1})	k_{rev} (s^{-1})	K_D (μM)	k_{off} (s^{-1})
4c	10	1	5	5	16 ± 3	0.98 ± 0.03	2.8 ± 1.5	5.5 ± 1.2	44 ± 4	2.0 ± 1.4
4d	10	1	5	5000	11 ± 2	1.0 ± 0.1	2.6 ± 1.5	3100 ± SV	4.5 ± 0.8	LB
4e	10	1	5000	5	9.7 ± 0.5	0.99 ± 0.02	4500 ± SV	9.3 ± 1.2	28 ± 1	750 ± 50
4 f	10	1	5000	5000	10 ± 1	0.99 ± 0.01	3300 ± SV	7100 ± SV	4.4 ± 0.2	1400 ± 100

Table 3. Summary of example 3-state fittings. The input data are from 3-state simulations shown in Fig. 4 as indicated, containing noise (signal/noise ≈50) and 50 data points on each line. The data are fitted with the 3- or 2-state fitting GUI. The parameter searching range is [1 μM, 1 mM] for K_D, [0.01, 100] for K_{eq}, and [1, 10000] for k_{off}/k_{rev}. The fitted parameter values are shown with 95% confidence intervals determined by Monte Carlo resampling. LB: not determined due to reaching lower boundary. SV: confidence interval not determined due to small variation of the fitted parameter from Monte Carlo runs.

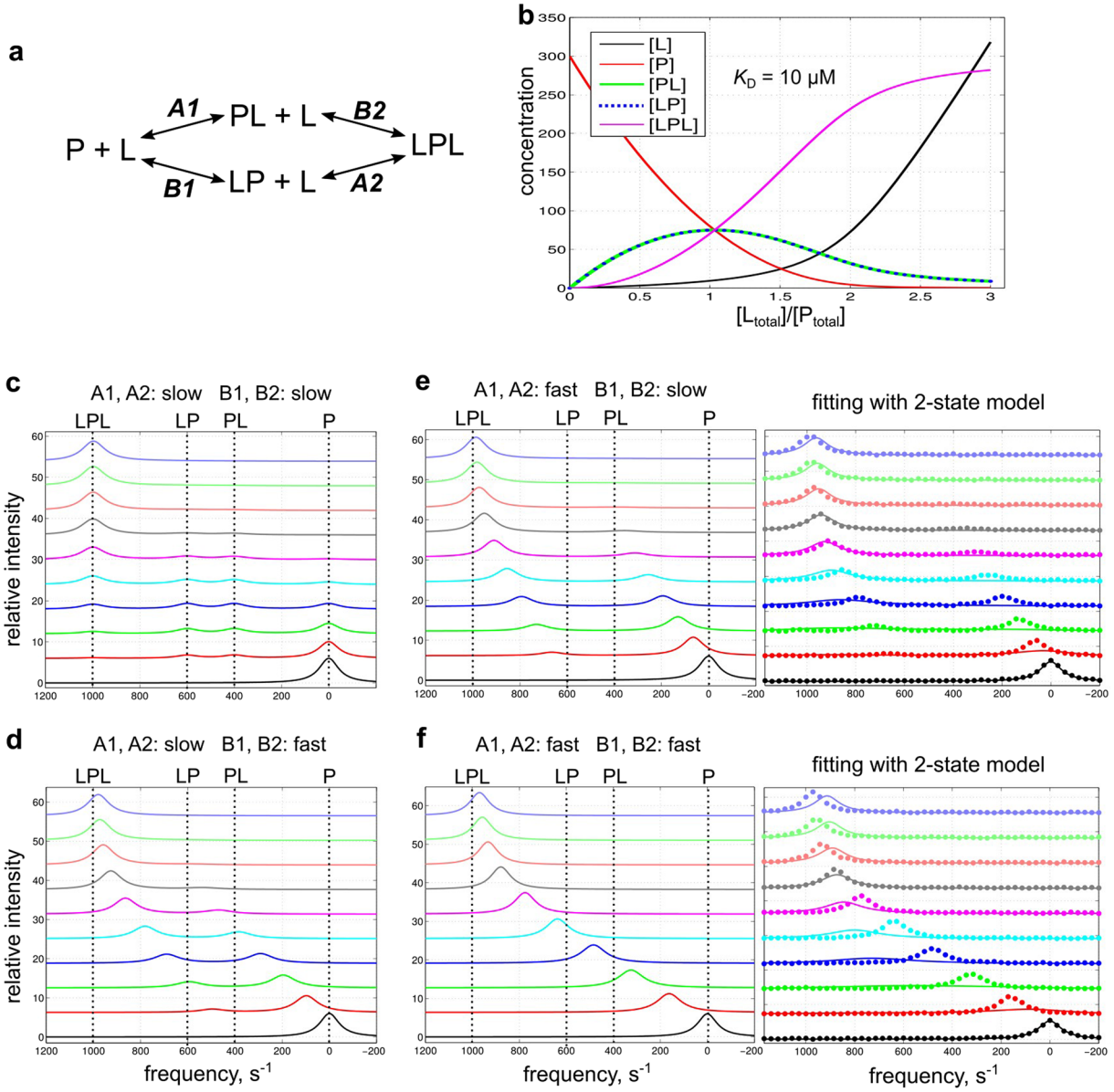

Figure 5. Example 4-state simulations. Simulations use a range of kinetic rates and the prefilled default values for other parameters: 300 μM [P_{total}]; 0, 100, 200, 300, 400, 500, 600, 700, 800, 900 μM [L_{total}]; K_D 10 μM for the four binding steps; $\omega_0 = 0$, 400, 600, 1000 s^{-1}; line width $d\omega = 100\ s^{-1}$. (**a**) The 4-state binding model. (**b**) Simulated concentration change of each species as a function of [L_{total}]/[P_{total}]. (**c**–**f**) Simulated lineshape with different k_{off} values (slow: 5 s^{-1}; fast: 5000 s^{-1}) for each binding step. Fifty data points on each simulated line are sampled with noise (signal/noise ≈50) and subsequently fitted with the 4- or 2-state fitting GUI. For panel (**e**,**f**), the sampled data points and fitted curves with the 2-state GUI are shown on the right as examples. See Supporting Information Figs S4 and S5 for all fittings.

fits well with all lineshape data (Supporting Information Fig. S5) and gives correct values for all K_D parameters and most of the k_{off} parameters (Table 4). There are two k_{off} parameters which cannot be determined from the 4-state fitting due to reaching the fast-exchange limit ("UB" in Table 4).

Experimental lineshape data for Syk tSH2 binding with N-IHP. The GUIs were used to examine data reported for the interaction between Syk tandem SH2 domain and tyrosyl phosphorylated peptides[3]. In previous work, the necessity and power of using NMR lineshape analysis to effectively extract dynamic and kinetic information was recognized[3]. NmrLineGuru is demonstrated here to analyze data on this interaction.

The Syk tandem SH2 domain fragment (tSH2) contains two SH2 domains and either domain can bind to a phosphotyrosine peptide, N-IHP [Ac-PD(pY)EPIRKG-NH_2], with a sequence derived from the CD3ε chain of the T cell receptor. tSH2 was titrated with N-IHP and the binding process was monitored by ^{15}N-^{1}H HSQC experiments (Fig. 6a,b). Although the lineshapes look like 2-state exchange, it is demonstrated that the 2-state model

Figure	True parameters			4-state fitting			2-state fitting	
	K_D (μM) A1 = A2 = B1 = B2	k_{off} (s^{-1}) A1 = A2	k_{off} (s^{-1}) B1 = B2	K_D (μM) A1 = A2 = B1 = B2	k_{off} (s^{-1}) A1 = A2	k_{off} (s^{-1}) B1 = B2	K_D (μM)	k_{off} (s^{-1})
5c	10	5	5	16 ± 2	5.0 ± 1.3	5.6 ± 1.4	70 ± 3	18 ± 1
5d	10	5	5000	10 ± 1	6.7 ± 0.7	UB	23 ± 1	300 ± 10
5e	10	5000	5	10 ± 1	2600 ± SV	6.7 ± 0.7	15 ± 1	200 ± 10
5f	10	5000	5000	10 ± 1	2077 ± SV	UB	39 ± 1	380 ± 20

Table 4. Summary of example 4-state fittings. The input data are from 4-state simulations shown in Fig. 5 as indicated, containing noise (signal/noise ≈50) and 50 data points on each line. The data are fitted with the 4- or 2-state fitting GUI. The parameter searching range is [1 μM, 1 mM] for all K_D values and [1 s^{-1}, 10000 s^{-1}] for all k_{off} values. The fitted parameter values are shown with 95% confidence intervals determined by Monte Carlo resampling. UB: not determined due to reaching upper boundary. SV: confidence interval not determined due to small variation of the fitted parameter from Monte Carlo runs.

cannot fit the data (Fig. 6c,d). The lineshapes of resonances from both domains (proton-dimension slices for G32, L37, H61, A74, F106, D175, G184, and G210; nitrogen-dimension slices for L52, Y73, L192, C205, and S244) are then globally fitted with the 4-state model assuming two independent binding sites and the 4-state model fits the data well (Fig. 6e,f). The fitted equilibrium dissociation constants are 730 ± 20 and 69 ± 4 μM and the fitted off rates are 1800 ± 300 and 380 ± 20 s^{-1} (mean ± SD of fitted values from two independent experiments), for the N- and C-terminal SH2 domains binding N-IHP, respectively.

Comparison of the fitted parameters by IDAP, TITAN, and NmrLineGuru. The experimental lineshape data[10] for E22A CsnN174 chitosanase binding with the chitosan hexamer substrate $(GlcN)_6$ were used to compare the fitting quality of NmrLineGuru to two other packages, IDAP[1,2] and TITAN[12]. Data from multiple peaks were globally fitted in all three cases. Example lineshapes and fittings by NmrLineGuru are shown in Fig. 7 and the fitted parameters are summarized in Table 5 (fitted values with IDAP and TITAN are from reference[10]).

The lineshape behavior shown in Fig. 7, especially that of W28, resembles the simulation in Fig. 4e whereby one peak shifts and decreases intensity, and another peak gradually appears. This behavior indicates that the underlying mechanism is fast ligand binding coupled with slow isomerization, consistent with the reported analysis[10]. Global fitting of the selected 1D spectral slices by NmrLineGuru gives similar thermodynamic and kinetic parameters to those from IDAP and TITAN; the differences shown in Table 5 are considered negligible. The off rate of the ligand binding step is too fast to be accurately determined, which for NmrLineGuru is reported as beyond fast-exchange limit.

Conclusion

We describe six standalone and user-friendly GUIs for both simulating and fitting 1D NMR lineshape data using the common 2-, 3-, or 4-state binding models. Accuracy in fitting the NMR data was demonstrated with simulated and experimental data, including global fitting of multiple peak lineshapes. These GUIs require the minimal amount of user input and handle most of the workflow in an automatic way. Aiming for non-experienced users, these GUIs can help to promote the wide use of the NMR lineshape analysis method, which is powerful and unique in studying dynamics and kinetics for multi-state binding systems.

Methods

Code development. The GUIs are developed in MATLAB R2014a (The MathWorks, Inc.) and compiled into standalone applications for both Windows and Linux. The low-level 1D NMR lineshape data I/O APIs (including the Sparky plugin) are from IDAP[1,2] (Integrative Data Analysis Platform, http://lineshapekin.net).

NMR lineshape for an exchange system. The NMR transverse magnetization for an equilibrium system of N spins in chemical exchange is generally described by the matrix form of the Bloch-McConnell equations[2,11]. The spectrum intensity at angular frequency ω is given by the sum of real components of the following complex vector **S**:

$$\mathbf{S} = \mathbf{\Omega}^{-1}\mathbf{P} \tag{1}$$

where **P** is a complex vector of length N and **Ω** is a $N \times N$ complex matrix. The elements of **P** are the relative spin populations, p_i, defined by the kinetic rate constants k_{ij}^*

$$\text{spin } i \underset{k_{ji}^*}{\overset{k_{ij}^*}{\leftrightarrows}} \text{spin } j \tag{2}$$

Ω is given by

$$\mathbf{\Omega} = \mathbf{M} - \mathbf{K} \tag{3}$$

M is a diagonal matrix with the following element M_{ii} for spin i

$$M_{ii} = R_2^i + i2\pi(\omega - \omega_0^i) \tag{4}$$

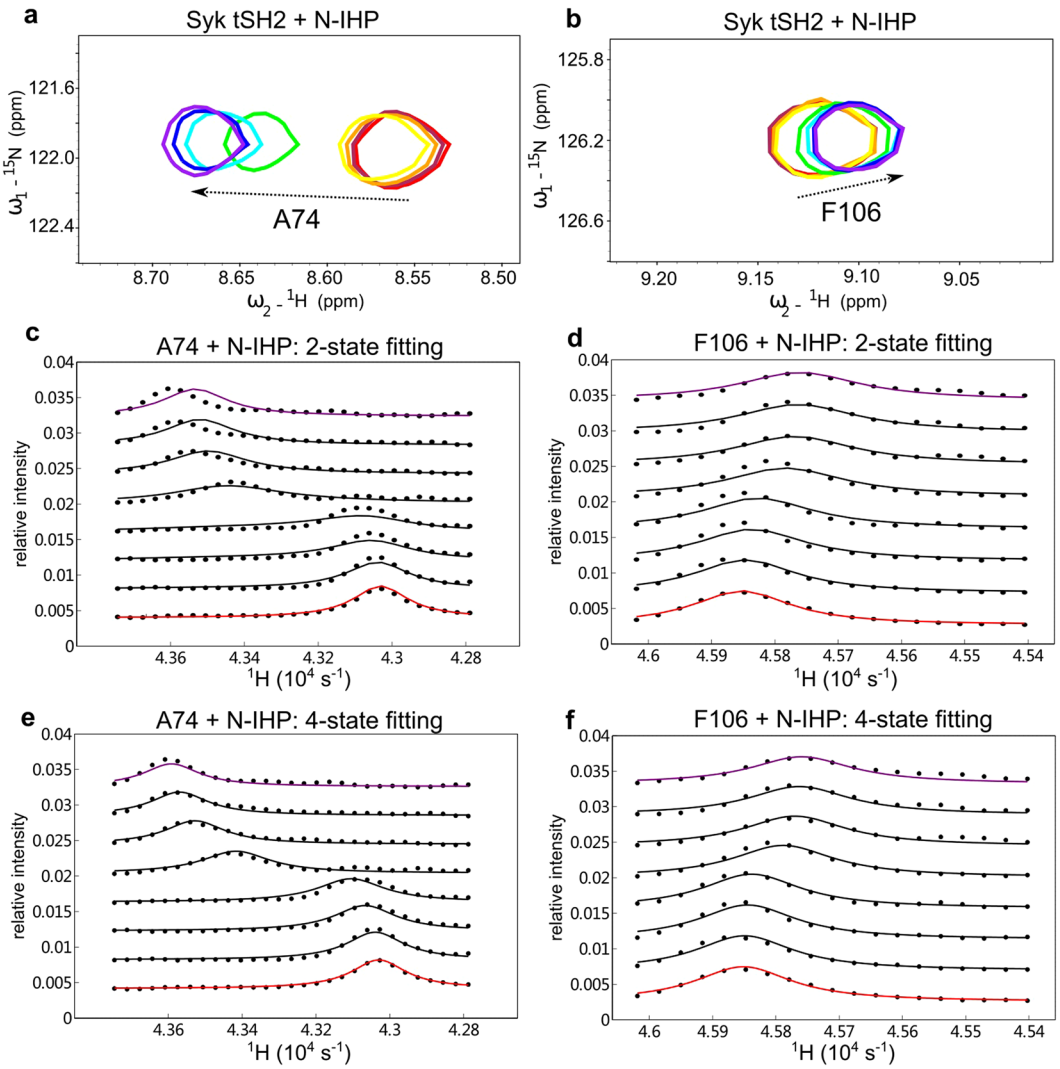

Figure 6. Example NMR ^{15}N-^{1}H HSQC chemical shift titration data and comparison of lineshape fitting with the 2-state and 4-state models. (**a**,**b**) Overlaid HSQC spectra for the unphosphorylated Syk tSH2 (~0.3 mM) titration with the peptide N-IHP [Ac-PD(pY)EPIRKG-NH_2]. Increasing [N-IHP]/[tSH2] ratios correspond to color changes: 0 (red), 0.2 (maroon), 0.6 (orange), 1.0 (yellow), 5.0 (green), 10.1 (cyan), 15.1 (blue), and 20.1 (purple). Zoomed regions with A74 and F106 are shown as examples. The arrows indicate the shift of peaks during the titration process. (**c**–**f**) Lineshape analysis for residues A74 and F106 in tSH2 plus N-IHP, with either the conventional 2-state model or the 4-state model. Circles: raw lineshape data. Lines: predicted lineshape with the indicated model (Red: the first titration point with no ligand. Purple: the last titration point with high concentration of ligand). The 4-state model fits well with the NMR data while the 2-state model does not.

where R_2^i and ω_0^i are the transverse relaxation rate and intrinsic resonance frequency of spin i. R_2^i and ω_0^i are provided by user input for simulations, or obtained by fitting the given resonance peaks to a Lorentzian function (assuming the peak centers at ω_0 with a line width $2R_2$), or searched as fitting parameters during fitting procedures. **K** is the $N \times N$ exchange matrix with the following elements from kinetic rate constants:

$$K_{ij} = k_{ji}^{*}(i \neq j) \tag{5}$$

$$K_{ii} = -\sum_{j \neq i}^{N} k_{ji}^{*} \tag{6}$$

Theory for 2-state binding and lineshape analysis. The 2-state binding model contains only one binding step between P and L to form PL, assuming P and PL are the NMR-detectable species. The system contains one equilibrium dissociation constant, K_D:

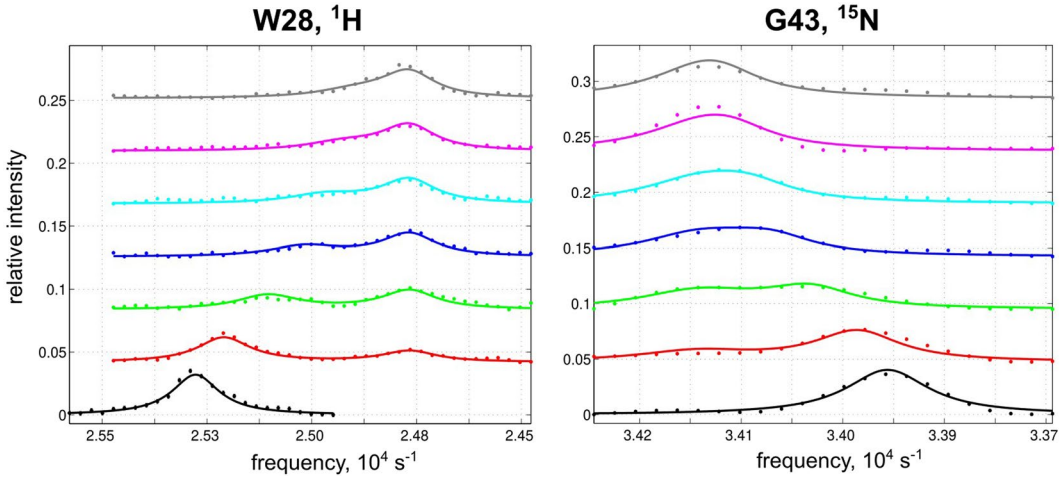

Figure 7. Example NMR ^{15}N-^{1}H HSQC 1D spectral slice data from E22A CsnN174 chitosanase (~0.08 mM) titration[10] with the chitosan hexamer substrate $(GlcN)_6$ and the lineshape fitting using NmrLineGuru. Increasing ligand/protein molar ratios correspond to color changes: 0 (black), 0.5 (red), 1 (green), 1.5 (blue), 2 (cyan), 3 (magenta), and 5 (gray). Lineshape analysis and global fitting were performed with the 3-state fitting GUI in NmrLineGuru for the following resonances: proton dimension for W28, G39, G153, and T157; nitrogen dimension for G43 and T157. Circles: raw lineshape data. Lines: predicted lineshape from global fitting with NmrLineGuru.

Software	Fitted parameters with the 3-state model			
	K_D (μM)	K_{eq}	k_{off} (s^{-1})	k_{rev} (s^{-1})
IDAP	42	0.26	30000	9
TITAN	36 ± 2	0.32	17000 ± 12000	3.6 ± 0.9
NmrLineGuru	29 ± 2	0.36 ± 0.01	UB	8.7 ± 0.9

Table 5. Comparison of the fitted parameters by IDAP, TITAN, and NmrLineGuru for the E22A CsnN174 chitosanase titration data[10]. The input data for IDAP and NmrLineGuru are the six 1D spectral slices described in Fig. 7 while TITAN uses nine amide NH 2D peaks[10] (residues N23, W28, A30, G39, G43, G47, G153, T157, and D232). The fitted values for IDAP and TITAN are from reference[10]. The fitted parameter values for NmrLineGuru are shown with 95% confidence intervals determined by Monte Carlo resampling. UB: not determined due to reaching the upper boundary (50000 s^{-1}).

$$K_D = \frac{[P][L]}{[PL]} \tag{7}$$

The total concentration of protein and ligand of each titration point is:

$$[P_{total}] = [P] + [PL] \tag{8}$$

$$[L_{total}] = [L] + [PL] \tag{9}$$

From Eqs (7–9), the relationship between [L], $[P_{total}]$, $[L_{total}]$, and K_D can be solved as:

$$[L]^2 + ([P_{total}] - [L_{total}] + K_D)[L] - K_D[L_{total}] = 0 \tag{10}$$

With given values of $[P_{total}]$, $[L_{total}]$, and varying values for K_D, the value of [L] is solved analytically. Once [L] is known, the concentrations of other species in the system are determined based on Eqs (7–9):

$$[P] = \frac{K_D[P_{total}]}{[L] + K_D} \tag{11}$$

$$[PL] = \frac{[L][P]}{K_D} \tag{12}$$

The NMR lineshape calculation requires an additional parameter to describe the 2-state system, the off rate k_{off}. The corresponding on rate is:

$$k_{on} = \frac{k_{off}}{K_D} \tag{13}$$

The predicted lineshape for a nucleus at a given titration point is described in Eqs (1–6). The following matrices are now

$$\mathbf{P} = \begin{bmatrix} [P] \\ [PL] \end{bmatrix} \tag{14}$$

$$\mathbf{K} = \begin{bmatrix} -k_{on}[L] & k_{off} \\ k_{on}[L] & -k_{off} \end{bmatrix} \tag{15}$$

Theory for the 3- state binding and lineshape. The 3-state binding model (Fig. 4a) contains an inter-molecular binding step between the free P and L to form PL, and an intra-molecular (concentration independent) isomerization step for PL to form P′L. These two steps are described by the equilibrium dissociation constant K_D and equilibrium isomerization constant K_{eq}:

$$K_D = \frac{[P][L]}{[PL]} \tag{16}$$

$$K_{eq} = \frac{[PL]}{[P'L]} \tag{17}$$

The total concentration of protein and ligand of each titration point is expressed as

$$[P_{total}] = [P] + [PL] + [P'L] \tag{18}$$

$$[L_{total}] = [L] + [PL] + [P'L] \tag{19}$$

From Eqs (16–19), the relationship between [L], $[P_{total}]$, $[L_{total}]$, K_D, and K_{eq} is

$$(K_{eq} + 1)[L]^2 + ((K_{eq} + 1)[P_{total}] - \left(K_{eq} + 1\right)[L_{total}] + K_D K_{eq})[L] - K_D K_{eq}[L_{total}] = 0 \tag{20}$$

With given values of $[P_{total}]$, $[L_{total}]$, K_D, K_{eq}, the value of [L] can be solved analytically. Once [L] is known, the concentrations of other species in the system could be determined as follows based on Eqs (16–19):

$$[P] = \frac{K_D\, K_{eq}\, [P_{total}\,]}{(K_{eq} + 1)[L] + K_D\, K_{eq}} \tag{21}$$

$$[PL] = \frac{[P][L]}{K_D} \tag{22}$$

$$[P'L] = \frac{[PL]}{K_{eq}} \tag{23}$$

The lineshape analysis requires two more parameters to describe the 3-state system: the off rate for the binding process (k_{off}) and the reverse rate for the isomerization process (k_{rev}). The corresponding on rate and forward rate are:

$$k_{on} = \frac{k_{off}}{K_D} \tag{24}$$

$$k_{fwd} = \frac{k_{rev}}{K_{eq}} \tag{25}$$

The predicted lineshape for a nucleus at a given titration point is described in Eqs (1–6). The following matrices are now

$$\mathbf{P} = \begin{bmatrix} [P] \\ [PL] \\ [P'L] \end{bmatrix} \tag{26}$$

$$\mathbf{K} = \begin{bmatrix} -K_{on}[L] & k_{off} & 0 \\ K_{on}[L] & -k_{off} - k_{fwd} & k_{rev} \\ 0 & k_{fwd} & -k_{rev} \end{bmatrix} \tag{27}$$

Theory for the 4- state binding and lineshape. The 4-state binding model (Fig. 5a) contains four binding steps described by four equilibrium dissociation constants:

$$K_D^{A1} = \frac{[P][L]}{[PL]} \tag{28}$$

$$K_D^{B1} = \frac{[P][L]}{[LP]} \tag{29}$$

$$K_D^{A2} = \frac{[LP][L]}{[LPL]} \tag{30}$$

$$K_D^{B2} = \frac{K_D^{B1} K_D^{A2}}{K_D^{A1}} \tag{31}$$

Note that only the first three equilibrium dissociation constants are independent parameters while the fourth becomes a function of the first three. The total concentration of protein and ligand of each titration point is expressed as

$$[P_{total}] = [P] + [PL] + [LP] + [LPL] \tag{32}$$

$$[L_{total}] = [L] + [PL] + [LP] + 2[LPL] \tag{33}$$

From Eqs (28–33), the relationship between [L], $[P_{total}]$, $[L_{total}]$, K_D^{A1}, K_D^{B1} and K_D^{A2} is

$$\begin{aligned} &K_D^{A1}[L]^3 + (2K_D^{A1}[P_{total}] - K_D^{A1}[L_{total}] + K_D^{A2}K_D^{B1} + K_D^{A1}K_D^{A2})[L]^2 \\ &+ ((K_D^{A2}K_D^{B1} + K_D^{A1}K_D^{A2})[P_{total}] \\ &-(K_D^{A2}K_D^{B1} + K_D^{A1}K_D^{A2})[L_{total}] + K_D^{A1}K_D^{A2}K_D^{B1})[L] - K_D^{A1}K_D^{A2}K_D^{B1}[L_{total}] = 0 \end{aligned} \tag{34}$$

With given values of $[P_{total}]$, $[L_{total}]$, K_D^{A1}, K_D^{B1} and K_D^{A2}, the value of [L] can be solved analytically with symbolic linear algebra software, such as Maxima (a Computer Algebra System, version 5.42.2, http://maxima.sourceforge.net). Once [L] is known, the concentrations of other species in the system are known from Eqs (28–33):

$$[P] = \frac{K_D^{A1} K_D^{A2} K_D^{B1}[P_{total}]}{K_D^{A1}[L]^2 + (K_D^{A2} K_D^{B1} + K_D^{A1} K_D^{A2})[L] + K_D^{A1} K_D^{A2} K_D^{B1}} \tag{35}$$

$$[PL] = \frac{[P][L]}{K_D^{A1}} \tag{36}$$

$$[LP] = \frac{[P][L]}{K_D^{B1}} \tag{37}$$

$$[LPL] = \frac{[P][L]^2}{K_D^{B1}K_D^{A2}} \tag{38}$$

The lineshape analysis requires four more parameters to describe the 4-state system: the off rates for the four binding process: $k_{off}^{A1}, k_{off}^{B1}, k_{off}^{A2}$, and k_{off}^{B2}. The corresponding on rate is

$$k_{on}^{i} = \frac{k_{off}^{i}}{K_D^{i}} \tag{39}$$

where *i* denotes any of the four binding steps. The predicted lineshape for a nucleus at a given titration point is described in Eqs (1–6). The following matrices are now

$$\mathbf{P} = \begin{bmatrix} [\mathrm{P}] \\ [\mathrm{PL}] \\ [\mathrm{LP}] \\ [\mathrm{LPL}] \end{bmatrix} \tag{40}$$

$$\mathbf{K} = \begin{bmatrix} -(k_{\mathrm{on}}^{\mathrm{A1}} + k_{\mathrm{on}}^{\mathrm{B1}})[\mathrm{L}] & k_{\mathrm{off}}^{\mathrm{A1}} & k_{\mathrm{off}}^{\mathrm{B1}} & 0 \\ k_{\mathrm{on}}^{\mathrm{A1}}[\mathrm{L}] & -k_{\mathrm{off}}^{A1} - k_{\mathrm{on}}^{\mathrm{B2}}[\mathrm{L}] & 0 & k_{\mathrm{off}}^{\mathrm{B2}} \\ k_{\mathrm{on}}^{B1}[\mathrm{L}] & 0 & -k_{\mathrm{off}}^{\mathrm{B1}} - k_{\mathrm{on}}^{\mathrm{A2}}[\mathrm{L}] & k_{\mathrm{off}}^{A2} \\ 0 & k_{\mathrm{on}}^{\mathrm{B2}}[\mathrm{L}] & k_{\mathrm{on}}^{\mathrm{A2}}[\mathrm{L}] & -k_{\mathrm{off}}^{B2} - k_{\mathrm{off}}^{\mathrm{A2}} \end{bmatrix} \tag{41}$$

Fitting-specific algorithms. The fitting GUIs search the best-fit values of unknown thermodynamic and kinetic parameters based on the input lineshape data. The unknown parameters (e.g. K_{D}, k_{off}) are assigned to the initial values from user selection and then iteratively refined by an optimization algorithm to match the experimental data, such that the following target function (sum of squared errors between the experimentally determined and the predicted lineshape data) is minimized:

$$\sum_{i,j,k} \left\{ \left[S_{i,j}^{pred}(\omega_k) a_i + b_i \right] - S_{i,j}^{expt}(\omega_k) \right\}^2 \tag{42}$$

where $S_{i,j}^{pred}(\omega_k)$ and $S_{i,j}^{expt}(\omega_k)$ are the predicted and experimental spectrum intensity of resonance i at titration point j at frequency ω_k, respectively; a_i and b_i are the intensity and baseline correction factors for resonance i to compensate the normalization errors, if any, introduced during the normalization process of the experimental lineshape data.

If selected by the user in the interface, fitting errors of the best-fit values are determined by Monte Carlo resampling, in which a random noise is generated based on the spectrum noise level input with the experimental data and added to each of the input lineshape data points. The resampling and fitting procedure is repeated a minimum of 50 times and until convergence to determine the 95% confidence intervals of the fitted parameters. To determine experimental errors, multiple experiments should be performed and processed independently. The variance from experimental errors is generally larger than that from fitting errors[3]. If multiple experimental datasets are available, results from their independent fitting should be used to determine uncertainty for the fitted parameters.

Data availability

The datasets generated during and/or analyzed during the current study are available in the online tutorials from NmrLineGuru GitHub wiki (https://github.com/stonefonly/NmrLineGuru/wiki) or from the corresponding author.

Code availability

The compiled GUIs and related documents are available from NmrLineGuru GitHub repo (https://github.com/stonefonly/NmrLineGuru). The software has been pre-installed in NMRbox (https://www.nmrbox.org). For academic use, this work should be cited. For commercial use or source code request, please contact us directly.

Received: 29 March 2019; Accepted: 14 August 2019;
Published online: 05 November 2019

References

1. Kovrigin, E. L. NMR line shapes and multi-state binding equilibria. *J. Biomol. NMR* **53**, 257–270 (2012).
2. Greenwood, A. I. *et al*. Complete determination of the Pin1 catalytic domain thermodynamic cycle by NMR lineshape analysis. *J. Biomol. NMR* **51**, 21–34 (2011).
3. Feng, C. & Post, C. B. Insights into the allosteric regulation of Syk association with receptor ITAM, a multi-state equilibrium. *Phys. Chem. Chem. Phys. PCCP* **18**, 5807–5818 (2016).
4. Furukawa, A., Konuma, T., Yanaka, S. & Sugase, K. Quantitative analysis of protein–ligand interactions by NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 47–57 (2016).
5. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Curr. Top. Med. Chem.* **3**, 39–53 (2003).
6. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137 (1989).
7. Feng, C., Roy, A. & Post, C. B. Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs: Entropic Allostery Dominates Syk Release. *Protein Sci.* **27**, 1780–1796 (2018).
8. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
9. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinforma. Oxf. Engl.* **31**, 1325–1327 (2015).
10. Shinya, S. *et al*. NMR line shape analysis of a multi-state ligand binding mechanism in chitosanase. *J. Biomol. NMR* **67**, 309–319 (2017).

11. Post, C. B. Characterization of enzyme-complex formation by analysis of nuclear magnetic resonance line shapes. *Methods Enzymol.* **240**, 438–446 (1994).
12. Waudby, C. A., Ramos, A., Cabrita, L. D. & Christodoulou, J. Two-Dimensional NMR Lineshape Analysis. *Sci. Rep*. **6** (2016).
13. Kovrigin, E. L. & Loria, J. P. Enzyme dynamics along the reaction coordinate: critical role of a conserved residue. *Biochemistry* **45**, 2636–2647 (2006).
14. Günther, U. L. & Schaffhausen, B. NMRKIN: simulating line shapes from two-dimensional spectra of proteins upon ligand binding. *J. Biomol. NMR* **22**, 201–209 (2002).

Acknowledgements
This work was supported by NIH R01GM039478.

Author contributions
C.F. developed the GUIs and fitting/simulation code, contributed to the research design and wrote the manuscript. E.K. developed 1D NMR lineshape data APIs (including the Sparky plugin) and contributed to the written manuscript. C.B.P. contributed to the research design and the written manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-52451-8.

Correspondence and requests for materials should be addressed to C.B.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.