Canadian Journal of Chemistry

Research Article

A simulation-guided "swapping" protocol for NMR titrations to study protein–protein interactions

Nicole Dcosta[a], **Megan Black**[a], **and Rui Huang**[a,b]

[a]Department of Chemistry, University of Guelph, Guelph, ON N1G 2W1, Canada; [b]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON N1G 2W1, Canada

Corresponding author: **Rui Huang** (email: rhuang08@uoguelph.ca)

Abstract

Solution nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for characterizing protein–protein interactions. NMR-monitored titration proves to be effective in determining dissociation constants, particularly for K_d values in the micromolar to millimolar range. In conventional NMR titrations, planning for optimal titration conditions requires a prior estimate of K_d. In addition, a highly concentrated ligand stock solution is often required, posing challenges when the ligand exhibits limited solubility and stability at elevated concentrations. To overcome these constraints, we propose a simulation-guided "swapping" protocol for NMR titrations. Guided by simulations of the binding curves, two samples, one with zero and the other with maximum ligand concentration, but both containing identical protein concentrations, initiate the titration. Using a "swapping" strategy, intermediate ligand concentrations in between those of the two initial samples are generated without the need of a concentrated ligand stock, while maintaining constant protein concentration. More importantly, this protocol facilitates estimation of K_d by early titration points and allows on-the-fly optimization of the titration points. The proposed approach enhances the efficiency of NMR titrations and provides a straightforward means to optimize the experimental conditions for the titrations.

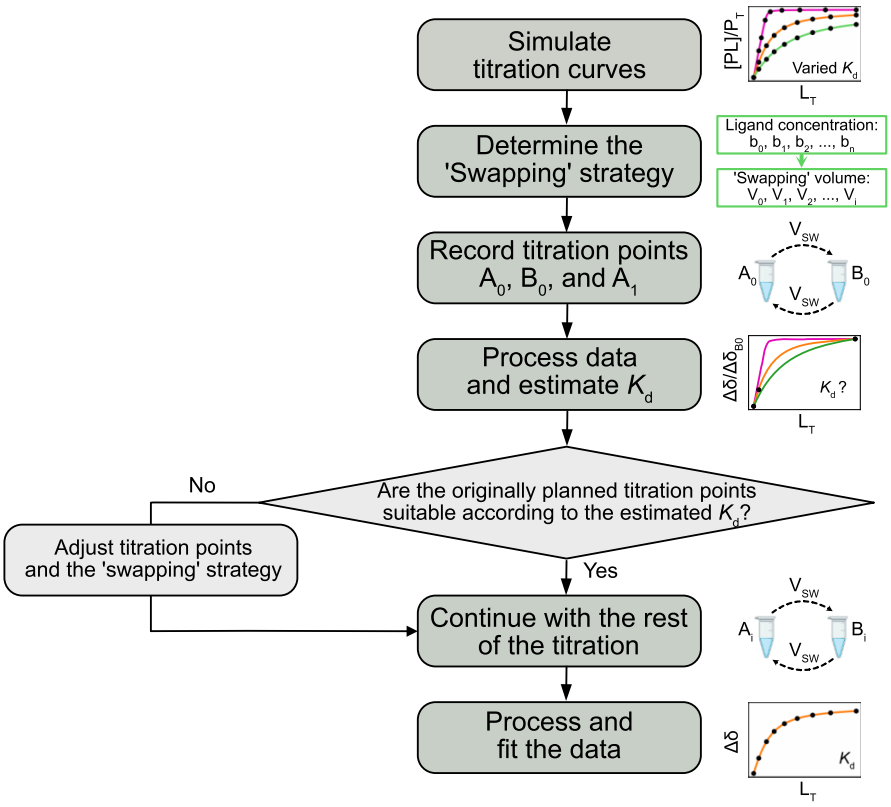
Key words: NMR titration, protein–protein interaction, NMR, simulation, p47, ubiquitin

1. Introduction

Protein–protein interactions play a fundamental role in diverse biological processes, ranging from enzymatic catalysis and signal transduction to gene regulation and immune responses.[1,2] Nuclear magnetic resonance (NMR) spectroscopy is a unique and powerful technique that provides atomic-level insights into the structure and dynamics of protein–protein complexes.[3–8] For complexes with moderate-to-low affinities (K_d in the μM to mM range), NMR is particularly instrumental in mapping binding interfaces and determining dissociation constants (K_d).[5,9–12] One commonly employed NMR technique for studying protein–protein interactions is an NMR-monitored titration. This method involves monitoring changes of the NMR signals, such as chemical shift, intensity, or linewidth, as the concentration of a binding partner (ligand) increases. Site-specific changes of the NMR signals upon ligand binding are used to elucidate the binding interface. In addition, NMR observables exhibit a direct correlation with the concentrations of free and bound states of the protein, which facilitates extraction of the dissociation constant through nonlinear least squares fitting of the binding curves.

Conventional NMR titrations between two proteins are carried out by sequential addition of an unlabeled protein titrant (ligand) into an isotopically labeled protein titrate while recording NMR signals at each addition. Therefore, a highly concentrated ligand stock (usually 5–20 times that of the final ligand concentration[12]) is needed to ensure that (i) the final ligand concentration is sufficiently high to nearly saturate all of the binding sites on the protein so that an accurate K_d can be extracted,[11,13] and (ii) dilution of the labeled protein is minimized to prevent a significant reduction in signal intensity throughout the progression of the titration, which could lead to difficulties in measuring the endpoints. However, in practice, the concentration of the ligand stock could be constrained by the solubility and stability of the ligand at elevated concentrations, especially when dealing with protein ligands and instances of relatively weak binding. Furthermore, to strategically plan for the optimal ligand concentrations in the titration, prior knowledge of an estimated K_d is often needed. Consequently, a "pre-titration" with fewer data points has been suggested to obtain a K_d estimate, followed by optimization of the titration conditions.[13,14] However, this process is time-consuming and cost ineffective. In a previous study, an alternative titration method involving co-varying ligand and protein concentrations during the titration was proposed and demonstrated higher precision.[15] By either serial diluting a highly concentrated complex sample or simultaneously decreasing the protein concentration and increasing the ligand concentration dur-

Fig. 1. Schematics of the workflow for a simulation-guided NMR titration using a "swapping" strategy.



ing the titration, sufficient saturation levels were achieved without a concentrated ligand stock.[15] In this approach, the titration data were presented in a three-dimensional space with protein and ligand concentrations as independent variables, while K_d and chemical shift perturbation (CSP) between the free and bound state ($\Delta\delta_{max}$) were fitted to search for the surface in the 3D space where the titration points resided.[15]

Here, we propose a simulation-guided "swapping" method for NMR titrations to study protein–protein interactions. This method overcomes the limitations of the conventional protocol while still facilitating straightforward data fitting and representation (Fig. 1). The protocol starts with the simulation of the potential binding curves, which aids in the initial planning of the titration points. Subsequently, two initial titration samples are prepared with identical concentrations of the labeled protein; one contains no ligand, while the other contains the maximum concentration of the ligand. Using a "swapping" strategy, titration points with ligand concentrations in between the two initial samples are generated, eliminating the need for a highly concentrated ligand stock and ensuring constant protein concentration throughout the titration. More importantly, this approach allows for a quick estimation of K_d using early titration points, enabling the optimization of the distribution of the remaining titration points based on the estimated K_d in real-time. To enhance accessibility, we introduce a Jupyter Notebook in Google Colaboratory, providing a user-friendly interface for simulating titration curves, designing the "swapping" strategy, and estimating K_d from early titration points. Finally, we present an NMR titration example between [U-^{15}N] labeled p47$_{1–94}$ (residues 1–94) (10.6 kDa) and unlabeled human ubiquitin (8.9 kDa) to demonstrate the efficacy of this new protocol.

2. Theory and methods

2.1. A simple 1:1 binding model and chemical exchange

For a simple 1:1 binding model, described by P + L ⇌ PL, the fractional population of the bound protein, p_b, is given by the following equation:

$$(1) \qquad p_b = \frac{[PL]}{P_T} = \frac{P_T + L_T + K_d - \sqrt{(P_T + L_T + K_d)^2 - 4P_T L_T}}{2P_T}$$

where [PL] is the concentration of the complex, P_T and L_T are the total concentrations of the protein (e.g., [U-^{15}N] labeled titrate) and the ligand (e.g., unlabeled titrant), respectively, and K_d is the dissociation constant.

In an NMR titration, assuming the concentration of the protein remains constant while the concentration of the ligand increases, the apparent change of the NMR spectra during the titration depends on the chemical shift difference between free and bound states of the protein, $\Delta\omega$ (in radian per second), in comparison to the exchange rate, k_{ex}, given by

$$(2) \qquad k_{ex} = k_{on}\,[L] + k_{off}$$

where [L] is the concentration of free ligand, and k_{on} and k_{off} are the rate constants of the forward and reverse reactions, respectively. In the case of fast exchange ($k_{ex} \gg \Delta\omega$), a single resonance will be observed throughout the titration with progressive chemical shift changes. The observed chemical shift of a resonance, δ_{obs}, is the population-weighted average between the free and bound state, given by

$$\delta_{obs} = \delta_f(1 - p_b) + \delta_b p_b \quad (3)$$

where δ_f and δ_b are the chemical shifts of a resonance in the free and ligand-bound states, respectively. From eq. 3, one can easily deduce that the CSP, $\Delta\delta$, relative to the free state, is proportional to p_b, as shown by

$$\Delta\delta = \delta_{obs} - \delta_f = (\delta_b - \delta_f)\, p_b = \Delta\delta_{max} p_b \quad (4)$$

in which $\Delta\delta_{max}$ is the CSP between the bound and free state. Therefore, based on eqs. 1 and 4, CSP ($\Delta\delta$) is a function of the protein and ligand concentrations (P_T and L_T), the dissociation constant (K_d), and the maximum CSP ($\Delta\delta_{max}$). By fitting $\Delta\delta$ with the known P_T and L_T values, one can extract K_d and $\Delta\delta_{max}$ as fitted parameters. In the slow exchange regime ($k_{ex} \ll \Delta\omega$), no CSPs will be observed during the titration. Instead, two separate resonances corresponding to the free and bound states will be present simultaneously with their intensities proportional to the concentrations of the free and bound protein, respectively, as described by the following equations:

$$I_b = I_{0,b} p_b \quad (5)$$

$$I_f = I_{0,f}(1 - p_b) \quad (6)$$

where I_f and I_b are the intensities of the free and bound peaks for a given resonance, respectively, $I_{0,f}$ is the intensity of the free peak in the free state, $I_{0,b}$ is the intensity of the bound peak in the completely bound state, and p_b is the fractional bound population of the protein. Thus, combining eqs. 1 and 5 (and/or 6), one can extract the K_d and $I_{0,b}$ (and/or $I_{0,f}$) by fitting the intensities of the bound (and/or free) peaks as a function of the protein and ligand concentrations in the titration. In the intermediate exchange regime ($k_{ex} \sim \Delta\omega$), linewidth increase is usually observed during the titration. In some cases, the line broadening can become so severe that the intensity of the resonance decreases below the detection level. At the same time, CSP, if observable, does not follow a simple population-weighted average.[11] In this study, we will focus only on the fast and slow exchange regimes where K_d can be readily derived from fitting chemical shift or resonance intensity in a 1:1 binding model.

2.2. Protein sample preparation

The expression plasmid of $p47_{1–94}$ was generated from the full-length *Mus musculus* p47 in a pET29 vector as previously described.[16] The expression gene of human ubiquitin was encoded in a pET29 vector, which contains an N-terminal His_6-tag and a tobacco etch virus cleavage site between the His_6-tag and the protein. Both proteins were expressed in Escherichia coli BL21(DE3) cells (New England Biolabs) and purified following a protocol that was previously described in detail.[16] For [U-^{15}N] labeled $p47_{1–94}$, cells were grown in M9 minimal media with $^{15}NH_4Cl$ as the sole nitrogen source, as previously described.[16] Purified proteins were exchanged into a buffer containing 25 mM Bis-Tris (pH 6.6), 25 mM NaCl, 1 mM EDTA, and 10% D_2O using an Amicon centrifugal concentrator (Millipore) with a molecular weight cutoff of 3 kDa before NMR measurement. The purities of both of the proteins were assessed to be >95% based on analyses of the SDS-PAGE results, and the molecular weight of the samples were confirmed using mass spectrometry (Fig. S1). The concentrations of both of the proteins were determined by UV absorbance at 280 nm with theoretical molar extinction coefficients predicted from the primary sequences of the proteins using Expasy ProtParam tool (https://web.expasy.org/protparam/).

2.3. Simulation of the binding curves

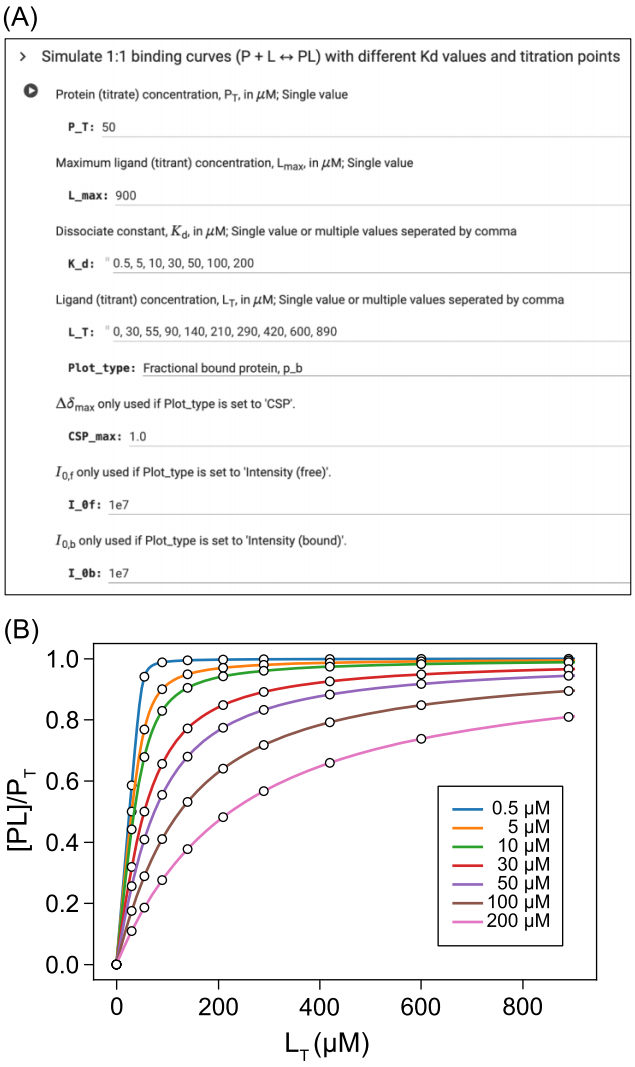
Simulation of the titration curves was performed using an in-house Python script and made available on Google Colaboratory (link). The Colab interface takes inputs of the total protein concentration (P_T), the maximum ligand concentration (L_{max}), and an array of dissociation constants (K_d) to simulate the fractional bound population of the protein (p_b) as a function of the total ligand concentration (L_T) (Fig. 2A). A list of ligand concentrations (L_T) are also inputted to simulate the scattered titration points (Fig. 2A). The binding curves can also be plotted with CSP ($\Delta\delta$) and the intensities of the bound or free peaks (I_b or I_f) on the *y*-axis, in cases of fast and slow exchange regimes, respectively. This is achieved by adjusting the "Plot_type" in the Colab interface to "CSP", "Intensity (free)" or "Intensity (bound") and inputting the values of $\Delta\delta_{max}$, $I_{0,f}$, or $I_{0,b}$, respectively (Fig. 2A).

To simulate the binding curve for the titration of [U-^{15}N] labeled $p47_{1–94}$ with ubiquitin, we input P_T of 50 μM, L_{max} of 900 μM, and a list of K_d values including 0.5, 5, 10, 30, 50, 100, and 200 μM (Fig. 2A). Figure 2B shows the simulated binding curves in solid lines. Data points with a list of ligand concentrations (L_T)—0, 30, 55, 90, 140, 210, 290, 420, 600, and 890 μM, were simulated as well (Figs. 2A and 2B, white circles) and planned as experimental titration points. The choices of the total protein and ligand concentrations are discussed in the "Results and discussion" section.

2.4. Design of the "swapping" strategy

Figure 3A shows the schematics of the "swap" titration. The titration starts with the preparation of two samples of identical volumes (e.g., 500 μL), A_0 and B_0. Sample A_0 contains [U-^{15}N] labeled protein in its free form ($c_0 = 0$ μM), while sample B_0 contains the same concentration of the labeled protein and the maximum concentration of the ligand (c_n). These two samples are subjected to NMR data acquisition and are then removed from the NMR tubes for the "swapping" process. A certain volume, $V_{SW,1}$, of the A_0 and B_0 samples is exchanged to generate the next pair of samples, A_1 and B_1. One or both of the next two samples are subsequently subjected to NMR data collection. This process is repeated until all the ligand concentrations are generated, and the corresponding NMR

Fig. 2. Simulation of the titration curves. (A) Screenshot of the Colab user interface (version 1) for generating the simulated binding curves. (B) Simulated binding curves using the Python script shared on Colab. The protein concentration (P_T) is set to 50 μM and the ligand concentration (L_T) ranges from 0 to 890 μM. The dissociation constant (K_d) ranges from 0.5 to 200 μM. The ligand concentrations for the planned titration points are set at 0, 30, 55, 90, 140, 210, 290, 420, 600, and 890 μM (white circles).



spectra are collected. The concentration of the ligand after each "swapping" is determined by the following equations:

$$c_i = c_{i-1}\frac{V_T - V_{SW,i}}{V_T} + c_{n-i+1}\frac{V_{SW,i}}{V_T} \quad (7)$$

$$c_{n-i} = c_{i-1}\frac{V_{SW,i}}{V_T} + c_{n-i+1}\frac{V_T - V_{SW,i}}{V_T} \quad (8)$$

in which c_{i-1}, c_i, c_{n-i+1}, and c_{n-i} represent the ligand concentrations in samples A_{i-1}, A_i, B_{i-1}, and B_i, respectively, $V_{sw,i}$ denotes the "swap" volume used to generate samples A_i and B_i, and V_T is the total sample volume (Fig. 3A). The ligand concentration progressively increases from A_0 to A_i and decreases from B_0 to B_i. With a pre-determined list of ligand concentrations, for each step in the titration, the "swap" volume ($V_{sw,i}$) is calculated using an in-house Python script shared on Colab (link). The Colab interface takes the list of the planned ligand concentrations (L_T) and the total sample volume (V_T) as inputs (Fig. 3B, top), and outputs a list of "swapping" volumes ($V_{sw,i}$) and the ligand concentrations in samples A_i and B_i after each "swap" (Fig. 3B, dashed-lined box).

To carry out the "swapping" procedure, two 5 mm NMR tubes, two 1.5 mL microcentrifuge tubes, two 12 inch glass pipettes, each labeled "A" or "B", will be used for the series of A_i and B_i samples, respectively, throughout the titration (Fig. S2). The initial samples A_0 and B_0 are prepared in 1.5 mL microcentrifuge tubes (labeled "A" and "B") and transferred to the respective 5 mm NMR tubes using 12 inch glass pipettes for NMR data acquisition. Afterwards, samples are transferred back to the respective microcentrifuge tubes using the same glass pipettes. "Swapping" is carried out in the microcentrifuge tubes ("A" and "B") with the aid of a third "middle" tube: a certain volume of sample ($V_{sw,i}$) is transferred from "A" to the "middle" tube, followed by transferring the same volume of sample from "B" to "A" and then from the "middle" tube to "B", which generates the next pair of samples. The samples are then gently mixed by tipping the microcentrifuge tubes and spun down at 18510 × *g* for 1 min before transferring them back to the same NMR tubes using the same glass pipettes for the next data acquisition. Note that to minimize the volume loss during transfer, and due to evaporation, the glass pipettes are placed in a vertical direction when not used with their tips pointing down (Fig. S2).

In the titration of [U-^{15}N] labeled $p47_{1-94}$ with ubiquitin, we designed a set of ligand concentrations—0, 30, 55, 90, 140, 290, 420, 600, and 890 μM, with a total sample volume of 500 μL (Fig. 3B). Using the Python script, we determined the "swap" volumes to generate samples with the intended ligand concentrations (Fig. 3B, dashed-lined box). It is important to note that more samples are often generated from "swapping" than are needed due to the asymmetric distribution of ligand concentrations throughout the titration. Consequently, the script also outputs the list of samples intended for NMR data acquisition (Fig. 3B, bottom).

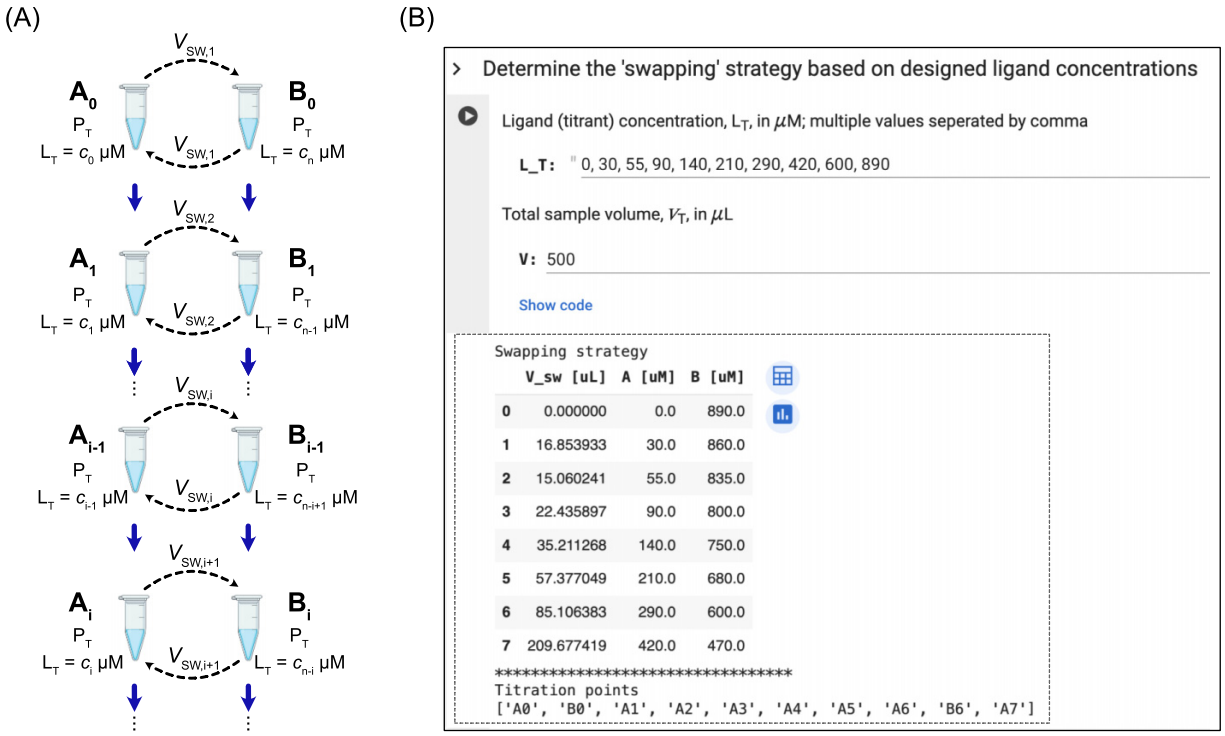
2.5. NMR experiments

All NMR experiments were performed on a 600 MHz Bruker Avance III NMR spectrometer equipped with a 5 mm TCI cryoprobe. All the ^{15}N-^{1}H HSQC spectra with sensitivity enhancement[17] were acquired at 25 °C with an interscan delay of 1.5 s, acquisition times of 39 and 64 ms, and 70 and 640 complex points, in the t_1 and t_2 dimensions, respectively. Carriers were positioned in the center of the water resonance (~4.673 ppm) for ^{1}H and at 119 ppm for ^{15}N. All NMR spectra were processed using NMRPipe[18] and visualized using NMRFAM-SPARKY.[19] A combined CSP, $\Delta\delta$, is calculated from ^{15}N-^{1}H HSQC spectra according to

$$\Delta\delta = \sqrt{\Delta\delta_H{}^2 + (0.14 \times \Delta\delta_N)^2} \quad (9)$$

in which $\Delta\delta_H$ and $\Delta\delta_N$ are chemical shift differences (ppm) in the ^{1}H and ^{15}N dimensions, respectively.[20]

Fig. 3. "Swap" titration. (A) Schematics of the "swap" titration. Two samples, A_0 and B_0, are initially generated such that both samples contain the same protein concentration (P_T) while sample A_0 contains no ligand ($c_0 = 0$ μM) and sample B_0 contains the maximum concentration of ligand (c_n). After NMR data acquisition, a total volume of $V_{sw,1}$ is exchanged between the two samples to generate the next set of samples, A_1 and B_1, and so on. The ligand concentration increases from A_0 to A_i and decreases from B_0 to B_i, while the protein concentration remains constant. The "swapping" volumes ($V_{sw,1}, V_{sw,2}$, …, $V_{sw,\,i}$) are calculated based on the list of ligand concentrations designed for the titration using the Python script shared on Colab. (B) Screenshot of the Colab user interface (version 1) to generate the "swapping" strategy. The input includes a list of ligand concentrations (L_T) and the total sample volume (V). The output includes a table that displays the "swapping" volumes (V_sw) and the ligand concentrations for the *A*- and *B*-series of samples after each "swap." A list of "Titration points" that indicates the samples with the intended ligand concentrations for NMR data acquisition is also given.



2.6. Estimation of K_d with early titration points

Utilizing the "swapping" strategy, two NMR samples were initially recorded, corresponding to titration points A_0 and B_0. Sample A_0 contained 50 μM [U-^{15}N] labeled p47$_{1–94}$, while sample B_0 consisted of 50 μM [U-^{15}N] labeled p47$_{1–94}$ and 890 μM unlabeled wild-type ubiquitin. After NMR data acquisition, 16.9 μL of sample were "swapped" between the samples A_0 and B_0, generating samples A_1 and B_1 with ligand concentrations of 30 and 860 μM, respectively. Sample A_1 subsequently underwent NMR data acquisition, and the NMR spectra for the three titration points – A_0, B_0, and A_1 – were processed and visualized (Fig. 4A).

In the fast exchange regime, the ratio between the CSPs of any titration point (A_i) and B_0 is given by

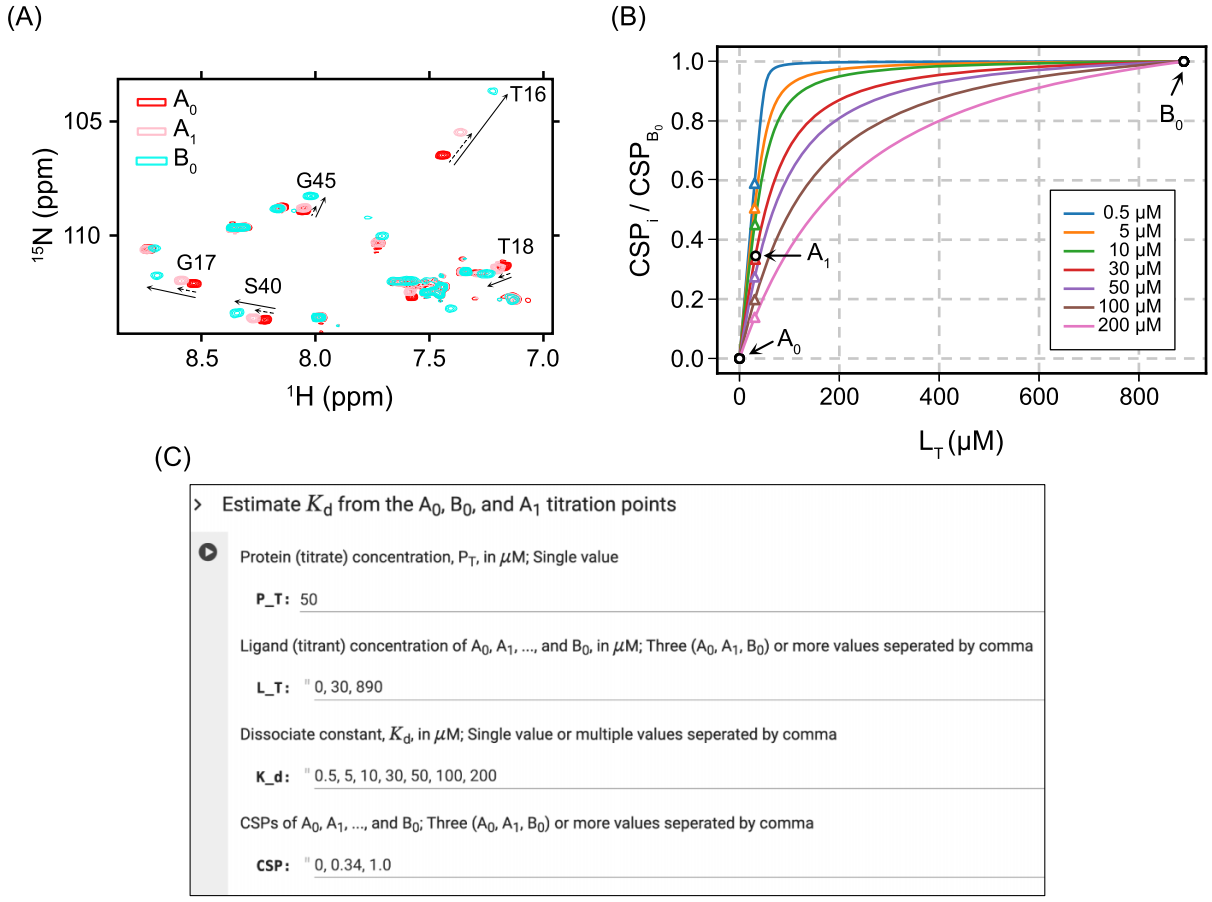
$$\frac{\Delta\delta_i}{\Delta\delta_{B_0}} = \frac{p_{b,i}}{p_{b,B_0}} = \frac{P_T + L_i + K_d - \sqrt{(P_T + L_i + K_d)^2 - 4P_T L_i}}{P_T + L_{B_0} + K_d - \sqrt{(P_T + L_{B_0} + K_d)^2 - 4P_T L_{B_0}}} \quad (10)$$

Given the concentration of the protein (P_T) and the maximum concentration of the ligand (L_{B_0}), the CSP ratio, $\frac{\Delta\delta_i}{\Delta\delta_{B_0}}$, can be simulated across the entire range of the ligand concentration (L_i) with various dissociation constants (K_d). Using an in-house Python available on Colab (link), we simulated the CSP ratio with P_T set at 50 μM, L_{B_0} at 890 μM, L_i ranging from 0 to 890 μM, and a list of K_d values (0.5, 5, 10, 30, 50, 100, and 200 μM) (Fig. 4B, solid lines). The user interface on Colab is shown in Fig. 4C. CSP ratios between A_1 and B_0 were calculated for a group of resonances (Fig. 4A), averaged, and then plotted alongside the simulations (Fig. 4B, white sphere). Colored triangles mark anticipated CSP ratios at the same ligand concentration with different K_d values for comparison, enabling the estimation of K_d. As shown in Fig. 4B, K_d was estimated to be ~30 μM based on the A_0, B_0, and A_1 points. The distribution of titration points was then re-evaluated on the simulated binding curve with a K_d of 30 μM (Fig. 2B, red line). The originally planned titration points were deemed suitable to represent the binding curve, and the titration proceeded with the "swapping" strategy outlined in Fig. 3B.

2.7. Mapping of the binding interface

CSPs between the first and last titration point (*i.e.*, A_0 and B_0) were used to identify the binding interface on p47$_{1–94}$ upon interacting with ubiquitin. In addition, the complex structure between p47 and ubiquitin was predicted by Al-

Fig. 4. Estimation of K_d from the first three titration points. (A) Selected region of ^{15}N-^{1}H HSQC spectra recorded on titration points A_0 (red), A_1 (pink) and B_0 (cyan), that contain 50 μM [U-^{15}N] labeled p47$_{1–94}$ at each point and 0, 30, and 890 μM unlabeled ubiquitin, respectively. The solid- and dashed-lined arrows indicate chemical shift perturbations (CSPs) for B_0 and A_1 (CSP_{B_0} and CSP_{A_1}), respectively, which are used for K_d estimation. (B) Simulation of CSP ratios between any titration point (CSP_i) and the B_0 point (CSP_{B_0}) as a function of the ligand concentration, given a K_d value of 0.5, 5, 10, 30, 50, 100, and 200 μM. Anticipated positions of the titration points A_0, A_1, and B_0 for each K_d value are marked by coloured triangles while the experimental data are presented in white circles. The position of the A_1 data point in comparison to the theoretical values (triangles) allows for the estimation of K_d. (C) Screenshot of the Colab user interface (version 1) to simulate the CSP ratios with different K_d values and to position the experimental data points on the graph.



phaFold Multimer,[21,22] based on the primary sequences of ubiquitin and residues 1–94 or 1–46 of p47 (Fig. S3; Fig. 5C). The interacting interface on p47 from CSP mapping was compared with that on the predicted complex structure. The structures were visualized and presented using UCSF ChimeraX.[23]

2.8. Fitting K_d using a 1:1 binding model

K_d was extracted by fitting each individual titration curve using a nonlinear least squares fit to the following equation:

$$\Delta\delta = \Delta\delta_{max} p_b = \Delta\delta_{max} \frac{P_T + L_T + K_d - \sqrt{(P_T + L_T + K_d)^2 - 4P_T L_T}}{2P_T} \quad (11)$$

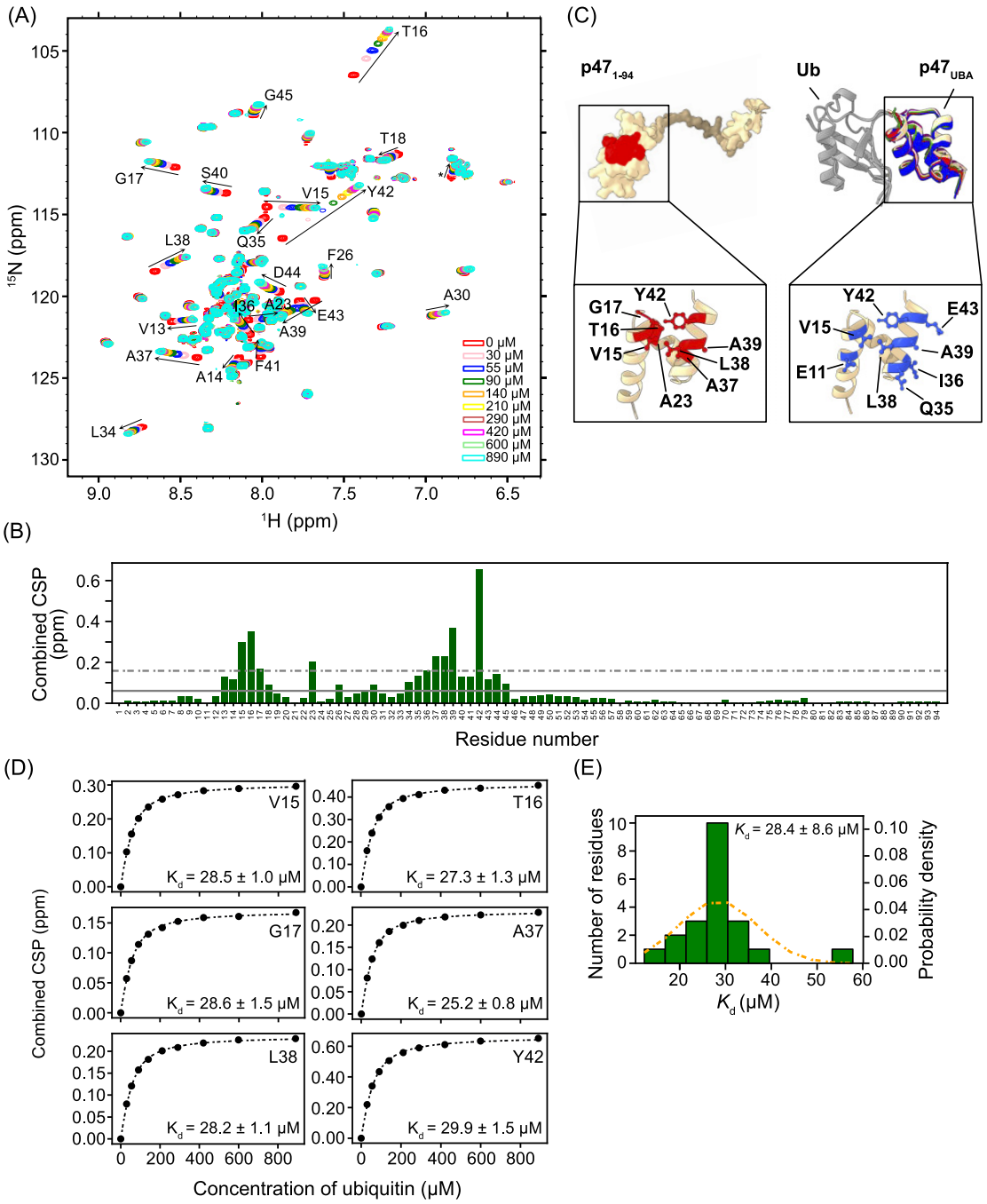
The fitting was performed using an in-house Python script (supplemental material). The uncertainty of the individual K_d value was obtained from the covariance matrix. For residues that exhibit combined CSPs higher than the average, the distribution of their K_d values was then fitted to a normal distribution to extract the average K_d and the standard deviation.

3. Results and discussion

3.1. Characterization of the interaction between p47$_{1–94}$ and ubiquitin using NMR titration

Previous studies have revealed that the N-terminal region of p47 contains a ubiquitin-associated (UBA) domain that is responsible for interacting and recruiting mono-ubiquitinated syntaxin 5 to the p47–p97 complex, ultimately mediating Golgi membrane fusion.[24–26] The UBA domain is a structurally–conserved ubiquitin-binding motif found in many proteins participating in diverse ubiquitin-dependent functions and pathways.[27,28] While the majority

Fig. 5. NMR titration between $p47_{1–94}$ and ubiquitin using the simulation-guided "swapping" protocol. (A) Superposition of ^{15}N-^{1}H HSQC spectra of 50 μM [U-^{15}N] labeled $p47_{1–94}$ in the presence of 0, 30, 55, 90, 140, 210, 290, 420, 600, and 890 μM unlabeled ubiquitin. Arrows indicate progressive chemical shift changes with increasing concentrations of ubiquitin. (B) Chemical shift perturbations (CSPs) of the residues in [U-^{15}N] labeled $p47_{1–94}$ between the first (A_0) and the last (B_0) titration point. The average of the combined CSP values and 1 S.D. above the average are indicated with solid and dashed lines, respectively. (C) Left: surface representation of $p47_{1–94}$, predicted by AlphaFold[47], with residues that show significant CSPs (1 S.D. above the average) coloured in red. The ubiquitin associated (UBA) domain (residue 1–46) is enlarged and shown in a cartoon presentation below, where residues with significant CSPs are shown as red sticks and annotated. Right: Predicted complex structure between $p47_{UBA}$ (residue 1–46) and ubiquitin by AlphaFold-Multimer[21,22]. The five best models are aligned by ubiquitin and shown in different colours for $p47_{UBA}$. Residues on the binding interface of $p47_{UBA}$ are shown as blue sticks and annotated in the enlarged cartoon representation below. (D) Titration curves of selected residues showing the combined CSP as a function of the ubiquitin concentration. Each curve was individually fitted to a simple 1:1 binding model to derive K_d. (E) Histogram of the K_d values from fitting individual titration curves of the residues with above-average CSPs (green bars). The histogram was then fitted to a normal distribution (orange dashed line) to obtain the average K_d (28.4 μM) and the standard deviation (8.6 μM).



of the UBA domains exhibit weak binding to monoubiquitin (K_d > 200 μM),[27,29–33] a few exceptions exist that demonstrate moderate affinities (K_d < 100 μM).[34–38] Notably, the UBA domain of p47 has been reported to preferentially bind mono-ubiquitinated substrates over poly-ubiquitin chains,[25] an uncommon characteristic compared to other UBA domains.[27,36,39,40] However, the exact binding affinity between the UBA domain of p47 and ubiquitin remains uncharacterized. In this study, we used an NMR titration to investigate the interaction between $p47_{UBA}$ and ubiquitin, aiming to quantitatively assess the binding affinity. This also serves as an example of the application using the simulation-guided "swapping" titration protocol.

A [U-^{15}N] labeled $p47_{1–94}$ construct, consisting of the UBA domain and a 48-amino acid-long linker, was subjected to a titration by unlabeled wild-type ubiquitin following the protocol described in the "Theory and methods" section. Figure 5A shows the ^{15}N-^{1}H HSQC spectra of $p47_{1–94}$ titrated by ubiquitin. Progressive chemical shift changes were observed for a set of resonances throughout the course of the titration, indicating a fast-exchange regime. CSPs between the free $p47_{1–94}$ (A_0 titration point) and $p47_{1–94}$ with the maximum concentration of ubiquitin (B_0 titration point) were plotted against the primary sequence of $p47_{1–94}$ (Fig. 5B). Residues exhibiting significant CSPs (1 standard deviation above the average) exclusively reside on the UBA domain. These residues collectively form a highly localized surface (Fig. 5C, left), suggesting a potential binding interface. The region showing significant CSPs in our study (Fig. 5B) — the C-terminus of α1 and most of α3 — aligns with a previous investigation between $p47_{UBA}$ and ubiquitin, although the relative magnitudes of the perturbations among different residues vary.[41] While a structural model of the complex between $p47_{UBA}$ and ubiquitin was previously proposed,[41] the lack of atomic coordinates for this model hinders detailed comparison between interfaces. Alternatively, we utilized AlphaFold Multimer to predict the complex structure between $p47_{UBA}$ and ubiquitin.[21,22] Interestingly, predictions for the complex involving $p47_{1–94}$ (including the linker) and ubiquitin failed to converge (Fig. S3). In contrast, using constructs of $p47_{1–46}$ (containing only the UBA domain) and ubiquitin resulted in five best structural models with an average RMSD of 1.3 Å (Fig. 5C, right). Since no significant CSPs were observed in the linker (Fig. 5B), we used the predicted structure between $p47_{1–46}$ and ubiquitin for comparison. As depicted in Fig. 5C, the interaction interface delineated by NMR CSPs is highly consistent with the binding interface inferred from the predicted complex structure.

For residues showing CSPs above the average (Fig. 5B, solid line), a 1:1 binding model (eq. 11) was employed to fit the individual titration profile and to extract the K_d value (Fig. 5D). Assuming a normal distribution of the K_d values, an average K_d of 28.4 ± 8.6 μM was obtained (Fig. 5E). This suggests that the UBA domain of p47 interacts with ubiquitin with moderate affinity, consistent with its physiological role in recruiting mono-ubiquitinated substrate for downstream processing.[24–26]

3.2. Simulation-guided "swap" titration

3.2.1. Choose the optimal concentrations of the protein and the ligand

In an NMR titration, the shape of the binding curve is dependent on both ligand and protein concentrations (Fig. S4). The protein concentration should be high enough so that NMR spectra can be obtained with sufficient signal-to-noise within reasonable experimental time. However, it should also be low enough to ensure close-to-saturation of the binding sites with the maximum ligand concentration so that the K_d value can be accurately extracted. A previous simulation study showed that, for 1:1 binding, the optimal protein concentration should be approximately 0.5 K_d,[12,13] and it is recommended not to exceed an order of magnitude higher than the K_d value.[11] Therefore, without prior knowledge of the K_d value, it is advisable to use a low protein concentration in case of moderate to tight binding. The lowest usable protein concentration is constrained by the signal-to-noise of the NMR spectra and depends on various factors such as the molecular weight of the protein and the complex, sensitivity of the spectrometer (e.g., whether it is equipped with a cryogenic probe), and the availability of acquisition time.

The maximum ligand concentration should be high enough to ensure that the binding sites are nearly saturated at the end of the titration; otherwise, errors in the fitted K_d values increase significantly.[13] In the absence of prior knowledge of an estimated K_d, it is recommended to use a high maximum ligand concentration, preferably ~1 mM or higher, to guarantee saturation levels. However, in practice this concentration may be limited by the solubility and stability of the ligand. It is worth noting that the "swapping" strategy eliminates the need of a ligand stock solution, as required in the conventional protocol for NMR titrations, and therefore allows for a higher maximum ligand concentration to be achieved, which is particularly advantageous for ligands with limited solubility.

3.2.2. Optimize the distribution of the titration points and the "swapping" strategy

Optimal distribution of the titration points along the binding curve is crucial for accurately determining the K_d value. A previous study has demonstrated that using 15–20 titration points, with the ligand concentration increasing by a constant factor rather than a constant increment, provides higher accuracy in the fitted K_d value.[13] Typically, the linear portion of the titration curve lacks substantial information about equilibrium states (e.g., Fig 2B, blue, in the range of L_T < 50 μM and L_T > 100 μM), while the curved section offers richer information for fitting. Therefore, optimal distribution of the titration points varies with the binding strength. For relatively tight binding, it is critical to include sufficient data points within the 3 equivalents of the protein concentration to recapitulate the features of the binding curve. Conversely, in the case of weak binding ($K_d \gg P_T$), titration points

can be more evenly spaced across the entire range of ligand concentrations. However, without prior knowledge of K_d, planning an appropriate distribution of titration points poses a challenge. To address this challenge, we introduced the "swapping" strategy. Titration point A_0, representing the free form of the protein, is recorded alongside titration point B_0 with the maximum ligand concentration, followed by titration point A_1 with the second lowest ligand concentration. Figure 4B demonstrates that the ratio of CSP at titration point A_1 to the last titration point (B_0) correlates with K_d. This correlation facilitates the estimation of K_d using the first three titration points and allows for the re-adjustment of titration point distribution according to the estimated K_d using simulation (Fig. 2).

Once the distribution of titration points in terms of ligand concentrations is established, the "swap" volumes for making the titration samples can be easily generated through the Python script or the Colab interface (Fig. 3B). "Swapping" is carried out following the protocol detailed in the "Theory and methods" section and Fig. S2. During the "swaps", ~2% volume reduction is estimated for the NMR samples per "swap" (Figs. S5A, S5B), which originates predominantly from residual sample in the glass transfer pipettes. However, the ligand concentrations are highly accurate with an average relative error of 2.6% throughout the titration (Figs. S5C, S5D). It is worth noting that to mitigate significant sample volume reduction which could affect NMR data acquisition (e.g., shimming), it is advised to limit the number of "swaps" to no more than 10 times. The use of Shigemi tubes with a plunger should be avoided to prevent sample volume loss, while in situations with limited sample volume, a Shigemi tube without the plunger can be utilized with a sample volume as low as 380–400 μL. To improve the efficiency of the titration, careful planning of the titration points is recommended to maximize the utilization of both A and B samples after each "swap" for NMR data acquisition whenever feasible.

3.2.3. Limitations of the strategy

The demonstrated simulation-guided "swapping" protocol is most useful in the fast exchange regime, where CSP is directly proportional to the fractional bound population of the protein. This approach can also be applied to the slow exchange limit, where the intensities of the free and/or bound peaks are analyzed to derive K_d according to eqs. 5 and 6. It is important to note that limitations may arise in the case of slow exchange due to limited NMR signal sensitivity to low concentrations of bound or free states at the beginning or end of the titration, respectively. Moreover, small K_d values, more common in the slow exchange regime, result in titration curves with less curvature and larger fitting uncertainties (Fig. 2A, blue). This method is unsuitable for intermediate exchange due to severe line broadening during titration progression. More importantly, CSP deviates from a simple weighted average as in fast exchange scenarios.[42,43] If severe linewidth increase is observed for early titration points (*i.e.*, A_1, A_2, etc.) in comparison to both the A_0 and B_0 points, it indicates intermediate exchange. In such cases, more intricate lineshape analysis,[44] or more advanced NMR experiments,[45,46] can be employed to extract the K_d value as well as kinetic parameters.

4. Conclusion

In conventional NMR titrations, selecting the optimal conditions for the titration, including protein and ligand concentrations and the distribution of titration points, typically relies on prior knowledge of an estimated K_d. It is also constrained by the necessity for a concentrated ligand stock to prevent significant dilution of the protein, or requires more complicated data fitting and representation methods if dilution is unavoidable. In this study, we introduce an innovative simulation-guided protocol integrated with a "swapping" strategy for NMR titrations. Our protocol allows for real-time K_d estimation using early titration points and enables dynamic adjustments to the distribution of titration points when necessary. Importantly, the "swapping" method eliminates the need for highly concentrated protein stock solutions as titrants, making it particularly suitable for proteins with limited solubility and stability at elevated concentrations. Meanwhile, this approach maintains a constant protein concentration throughout the titration, facilitating straightforward graphical representation and data fitting. To improve accessibility of this protocol, a Colab user interface was established for convenient access to the binding curve simulation, "swapping" strategy design, and K_d estimation from early titration points. Illustrated with an NMR titration between $p47_{1–94}$ and ubiquitin, our approach proves to be particularly effective for fast-exchange interactions on the NMR timescale.

Acknowledgement

The authors thank Felipe Perez from Signal 1 AI for assisting with the Colab script and the design of the new user interface (version 2).

Article information

History dates

Received: 8 February 2024
Accepted: 2 April 2024
Accepted manuscript online: 14 May 2024
Version of record online: 19 September 2024

Notes

This paper is part of a Special Issue entitled Celebrating 150 Years of Chemistry at the University of Guelph.

Data availability

All the scripts used in the paper are shared in the format of Jupyter notebook in the supplementary information.

Author information

Author ORCIDs

Nicole Dcosta https://orcid.org/0000-0002-4615-3091
Megan Black https://orcid.org/0000-0001-5006-5337
Rui Huang https://orcid.org/0000-0002-4064-6397

Author notes

Nicole Dcosta and Megan Black contributed equally to this work.

Author contributions

Conceptualization: RH
Data curation: ND, RH
Formal analysis: ND, MB, RH
Investigation: MB
Methodology: RH
Project administration: RH
Resources: RH
Software: RH
Supervision: RH
Visualization: ND, MB, RH
Writing – original draft: RH
Writing – review & editing: ND, MB

Competing interests

The authors declare no conflict of interest.

Funding information

This research was funded by the the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-05066 (to RH).

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/cjc-2024-0031.

References

(1) Nooren, I. M. A.; Thornton, J. M. *EMBO J.* **2003**, *22*(14), 3486. doi:10.1093/emboj/cdg359. PMID: 12853464.
(2) Burke, D. F.; Bryant, P.; Barrio-Hernandez, I.; Memon, D.; Pozzati, G.; Shenoy, A.; Zhu, W.; Dunham, A. S.; Albanese, P.; Keller, A.; Scheltema, R. A.; Bruce, J. E.; Leitner, A.; Kundrotas, P.; Beltrao, P.; Elofsson, A. *Nat. Struct. Mol. Biol.* **2023**, *30*(2), 216. doi:10.1038/s41594-022-00910-8. PMID: 36690744.
(3) Alderson, T. R.; Kay, L. E. *Cell*, **2021**, *184*(3), 577. doi:10.1016/j.cell.2020.12.034. PMID: 33545034.
(4) Rosenzweig, R.; Kay, L. E. *J. Am. Chem. Soc.* **2016**, *138*(5), 1466. doi:10.1021/jacs.5b11346. PMID: 26651836.
(5) Zuiderweg, E. R. P. *Biochemistry*, **2002**, *41*(1), 1. doi:10.1021/bi011870b. PMID: 11771996.
(6) Moraes, A. H.; Valente, A. P. *J. Magn. Reson. Open*, **2023**, *14–15*, 100093. doi:10.1016/j.jmro.2023.100093.
(7) Huang, C.; Kalodimos, C. G. *Annu. Rev. Biophys.* **2017**, *46*, 317. doi:10.1146/annurev-biophys-070816-033701. PMID: 28375736.
(8) Hu, Y.; Cheng, K.; He, L.; Zhang, X.; Jiang, B.; Jiang, L.; Li, C.; Wang, G.; Yang, Y.; Liu, M. *Anal. Chem.* **2021**, *93*(4), 1866. doi:10.1021/acs.analchem.0c03830. PMID: 33439619.
(9) Takeuchi, K.; Wagner, G. *Curr. Opin. Struct. Biol.* **2006**, *16*(1), 109. doi:10.1016/j.sbi.2006.01.006. PMID: 16427776.
(10) Vinogradova, O.; Qin, J. *Top. Curr. Chem.* **2012**, *326*, 35. doi:10.1007/128_2011_216. PMID: 21809187.
(11) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1. doi:10.1016/j.pnmrs.2013.02.001. PMID: 23962882.
(12) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*(4), 219. doi:10.1016/j.pnmrs.2007.04.001.
(13) Granot, J. *J. Magn. Reson.* **1983**, *55*(2), 216.
(14) Waudby, C. A.; Christodoulou, J.In *Intrinsically Disordered Proteins: Methods and Protocols*; B. B. Kragelund; K. Skriver, Eds.; Springer US: New York, NY, 2020; pp 477–504.
(15) Markin, C. J.; Spyracopoulos, L. *J. Biomol. NMR*, **2012**, *53*(2), 125. doi:10.1007/s10858-012-9630-9. PMID: 22534787.
(16) Kim, P.; Black, M.; Perez, F.; Huang, R. *Can. J. Chem.* **2024**, *102*(3), 122. doi:10.1139/cjc-2023-0160.
(17) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*(10), 10663. doi:10.1021/ja00052a088.
(18) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR*, **1995**, *6*(3), 277. doi:10.1007/BF00197809. PMID: 8520220.
(19) Lee, W.; Tonelli, M.; Markley, J. L. *Bioinformatics*, **2015**, *31*(8), 1325. doi:10.1093/bioinformatics/btu830. PMID: 25505092.
(20) Williamson, M. P.In *Modern Magnetic Resonance*; G. A. Webb, Ed.; Springer International Publishing: Cham, 2018; pp 995–1012.
(21) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. *Biorxiv*, **2022**, 2021. doi:10.04.463034.
(22) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nat. Methods*, **2022**, *19*(6), 679. doi:10.1038/s41592-022-01488-1. PMID: 35637307.
(23) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*(13), 1605. doi:10.1002/jcc.20084. PMID: 15264254.
(24) Rabouille, C.; Kondo, H.; Newman, R.; Hui, N.; Freemont, P.; Warren, G. *Cell*, **1998**, *92*(5), 603. doi:10.1016/S0092-8674(00)81128-9. PMID: 9506515.
(25) Meyer, H. H.; Wang, Y.; Warren, G. *EMBO J.* **2002**, *21*(21), 5645. doi:10.1093/emboj/cdf579. PMID: 12411482.
(26) Huang, S.; Wang, Y. *F1000Research*, **2017**, *6*, 2050. doi:10.12688/f1000research.11900.1. PMID: 29225785.
(27) Raasi, S.; Varadan, R.; Fushman, D.; Pickart, C. M. *Nat. Struct. Mol. Biol.* **2005**, *12*(8), 708. doi:10.1038/nsmb962. PMID: 16007098.
(28) Buchberger, A. *Trends Cell Biol.* **2002**, *12*(5), 216. doi:10.1016/S0962-8924(02)02269-9. PMID: 12062168.
(29) Mueller, T. D.; Kamionka, M.; Feigon, J. *J. Biol. Chem.* **2004**, *279*(12), 11926. doi:10.1074/jbc.M312865200. PMID: 14707125.
(30) Tse, M. K.; Hui, S. K.; Yang, Y.; Yin, S. T.; Hu, H. Y.; Zou, B.; Wong, B. C. Y.; Sze, K. H. *PLoS One*, **2011**, *6*(12), e28511. doi:10.1371/journal.pone.0028511. PMID: 22194841.
(31) Trempe, J. F.; Brown, N. R.; Lowe, E. D.; Gordon, C.; Campbell, I. D.; Noble, M. E. M.; Endicott, J. A. *EMBO J.* **2005**, *24*(18), 3178. doi:10.1038/sj.emboj.7600797. PMID: 16138082.
(32) Varadan, R.; Assfalg, M.; Raasi, S.; Pickart, C.; Fushman, D. *Mol. Cell*, **2005**, *18*(6), 687. doi:10.1016/j.molcel.2005.05.013. PMID: 15949443.
(33) Matta-Camacho, E.; Kozlov, G.; Trempe, J. F.; Gehring, K. *J. Mol. Biol.* **2009**, *386*(2), 569. doi:10.1016/j.jmb.2008.09.086. PMID: 18948116.
(34) Swanson, K. A.; Hicke, L.; Radhakrishnan, I. *J. Mol. Biol.* **2006**, *358*(3), 713. doi:10.1016/j.jmb.2006.02.059. PMID: 16563434.
(35) Zhou, Z.; Gao, H.; Zhou, C.; Chang, Y.; Hong, J.; Song, A.; Lin, D.; Hu, H. *Protein Sci.* **2008**, *17*(10), 1805. doi:10.1110/ps.036384.108. PMID: 18596201.
(36) Zhang, D.; Raasi, S.; Fushman, D. *J. Mol. Biol.* **2008**, *377*(1), 162. doi:10.1016/j.jmb.2007.12.029. PMID: 18241885.
(37) Ohno, A.; Jee, J.; Fujiwara, K.; Tenno, T.; Goda, N.; Tochio, H.; Kobayashi, H.; Hiroaki, H.; Shirakawa, M. *Structure*, **2005**, *13*(4), 521. doi:10.1016/j.str.2005.01.011. PMID: 15837191.
(38) Blankenship, J. W.; Varfolomeev, E.; Goncharov, T.; Fedorova, A. V.; Kirkpatrick, D. S.; Izrael-Tomasevic, A.; Phu, L.; Arnott, D.; Aghajan, M.; Zobel, K.; Bazan, J. F.; Fairbrother, W. J.; Deshayes, K.; Vucic, D. *Biochem. J.* **2009**, *417*(1), 149. doi:10.1042/BJ20081885. PMID: 18939944.

(39) Wilkinson, C. R. M.; Seeger, M.; Hartmann-Petersen, R.; Stone, M.; Wallace, M.; Semple, C.; Gordon, C. *Nat. Cell Biol.* **2001**, *3*(10), 939. doi:10.1038/ncb1001-939. PMID: 11584278.
(40) Bavies, G. C.; Ettenberg, S. A.; Coats, A. O.; Mussante, M.; Ravichandran, S.; Collins, J.; Nau, M. M.; Lipkowitz, S. *Oncogene*, **2004**, *23*(42), 7104. PMID: 15273720.
(41) Yuan, X.; Simpson, P.; McKeown, C.; Kondo, H.; Uchiyama, K.; Wallis, R.; Dreveny, I.; Keetch, C.; Zhang, X.; Robinson, C.; Freemont, P.; Matthews, S. *EMBO J.* **2004**, *23*(7), 1463. doi:10.1038/sj.emboj.7600152. PMID: 15029246.
(42) London, R. E. *J. Magn. Reson. Ser. A*, **1993**, *104*(2), 190. doi:10.1006/jmra.1993.1208.
(43) Sudmeier, J. L.; Evelhoch, J. L.; Jonsson, N. B.-H. *J. Magn. Reson.* **1980**, *40*(2), 377 .
(44) Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J. *Sci. Rep.* **2016**, *6*, 24826. doi:10.1038/srep24826. PMID: 27109776.
(45) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Biomol. NMR*, **2008**, *41*(3), 113. doi:10.1007/s10858-008-9251-5. PMID: 18574698.
(46) Kempf, J. G.; Loria, J. P. *Methods Mol. Biol.* **2004**, *278*, 185. PMID: 15317998.
(47) Tunyasuvunakool, K.; Adler, J.; Wu, Z.; Green, T.; Zielinski, M.; Žídek, A.; Bridgland, A.; Cowie, A.; Meyer, C.; Laydon, A.; Velankar, S.; Kleywegt, G. J.; Bateman, A.; Evans, R.; Pritzel, A.; Figurnov, M.; Ronneberger, O.; Bates, R.; Kohl, S. A. A.; Potapenko, A.; Ballard, A. J.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Clancy, E.; Reiman, D.; Petersen, S.; Senior, A. W.; Kavukcuoglu, K.; Birney, E.; Kohli, P.; Jumper, J.; Hassabis, D. *Nature*, **2021**, *596*(7873), 590. doi:10.1038/s41586-021-03828-1. PMID: 34293799.