

UCSF

UC San Francisco Previously Published Works

Title

Deconvolution Method for Specific and Nonspecific Binding of Ligand to Multiprotein Complex by Native Mass Spectrometry

Permalink

<https://escholarship.org/uc/item/2hm3k439>

Journal

Analytical Chemistry, 87(16)

ISSN

0003-2700

Authors

Guan, Shenheng
Trnka, Michael J
Bushnell, David A
[et al.](#)

Publication Date

2015-08-18

DOI

10.1021/acs.analchem.5b02258

Peer reviewed



Published in final edited form as:

Anal Chem. 2015 August 18; 87(16): 8541–8546. doi:10.1021/acs.analchem.5b02258.

Deconvolution method for specific and nonspecific binding of ligand to multi-protein complex by native mass spectrometry

Shenheng Guan^{a,b,*}, Michael J. Trnka^a, Alma L. Burlingame^a, David A. Bushnell^c, Philip J. J. Robinson^c, Roger D. Kornberg^c, Jason E. Gestwicki^{a,b}, and Stanley B. Prusiner^{b,d}

^aDepartment of Pharmaceutical Chemistry and Mass Spectrometry Facility, University of California, San Francisco, California 94158-2517, United States

^bInstitute for Neurodegenerative Diseases, University of California, San Francisco, California 94143-0518, United States

^cDepartment of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305

^dDepartment of Neurology, University of California, San Francisco, California 94143-0518, United States

Abstract

In native mass spectrometry, it has been difficult to discriminate between specific binding of a ligand to a multi-protein complex from the nonspecific interactions. Here, we present a deconvolution model that consists of two levels of data reduction. At the first level, the apparent association binding constants are extracted from the measured intensities of the target/ligand complexes by varying ligand concentration. At the second level, two functional forms representing the specific- and non-specific binding events are fit to the binding constants obtained from the first level of modeling. Using this approach, we found that an inverse power distribution described nonspecific binding of α -amanitin to yeast RNA polymerase II. Moreover, treating the concentration of the multi-protein complex as a fitting parameter reduced the impact of inaccuracies in this experimental measurement on the apparent association constants. This model provides an improved way of separating specific and non-specific binding to large, multi-protein complexes in native mass spectrometry.

Direct observations of ligand binding to protein complexes by native mass spectrometry offer a way to precisely measure association and/or dissociation constants. Since the discovery of native mass spectrometry¹, many protein/ligand complexes have been observed and characterized². In these experiments, measurements are made on pure protein/ligand complexes in an aqueous buffer containing volatile salt, such as ammonium acetate, through nanospray infusion. Even with significant purification efforts, unwanted adduct formation may still contribute to degradation of the observed spectra due to trace impurities, such as metal ions, other small molecules, and other proteins. In order to reduce the contribution of these unwanted adducts, collision activation in either electrospray source or collision cells is often applied. However, increased collisions may also cause dissociation of the bound

*Corresponding Author, sguan@cgl.ucsf.edu. .

ligand(s) and disruption of multi-protein complexes. Non-specific interactions are often difficult to deconvolve from these spectra. Because an increasing number of important drug targets are multi-protein complexes³, it is important to develop better ways of studying ligand binding to these systems in native mass spectrometry.

The goal of this study is to develop a data model to extract specific binding constants and discriminate them from nonspecific ones. Several approaches towards this goal have been proposed. For example, van der Rest and coworkers⁴ developed a statistical method to model specific binding as a binomial distribution, while modeling nonspecific binding as Poisson distribution (which was originally proposed by Wang, et al⁵). The modeling of nonspecific binding as a Poisson is statistically sound if the target is a large protein assembly and there are many nonspecific binding sites that are randomly chosen. However, treating specific binding with a statistical distribution may introduce a significant amount of error, especially in cases in which there is only one specific binding site. Shimon et al⁶ introduced a simpler model with a single nonspecific binding constant, and this method was used to study ATP binding to a multi-protein chaperone, GroEL⁷. However, because proteins, especially multi-protein complexes, are not perfectly spherical, binding of a ligand on the non-uniform surface would not be expected to be constant. In addition, if many ligands have already bound, additional nonspecific binding is likely to be affected (likely to be weaker). New methods are needed to overcome these issues.

In this study, we assume that the number of specific binding sites can be determined by how well the model fits to the experimental data. In addition, this value is often known from other studies, including those that employ isothermal calorimetry (ITC). The mass spectrometry algorithm is separated into two steps: first, the apparent association constants (containing contribution of both specific and nonspecific binding) are extracted together from the measured relative concentration (proportional to relative intensities) of target/ligand complex. Second, contributions from specific and nonspecific binding are separated by fitting them with different functional forms to the apparent association constants obtained from the first level of data extraction. To test this idea, we measured binding of alpha-amanitin to purified yeast RNA polymerase II. In addition, we treated the protein concentration as an unknown parameter and found that its value can be obtained by fitting the data model. This is an important observation because of the difficulties commonly encountered with accurately measuring the concentration of intact multi-protein complexes.

Experimental

Human FKBP12 was prepared as described⁸. Briefly, His-tagged human FKBP12 was expressed in *E. coli* and the protein purified using a nickel column. The protein concentration was measured using the BCA method, with bovine serum albumin as a standard. Rapamycin and α -amanitin were purchased (Sigma-Aldrich, St. Louis, MO). Yeast RNA polymerase II (Pol2) was prepared according to the previously described method⁹. Pol2 was buffer-exchanged to 100 mM ammonium acetate using a molecular weight cutoff centrifugation filter (YM-10 from Millipore, Billerica, MA). Samples for native mass spectrometry analysis were prepared by mixing pol II with α -amanitin to final concentrations of 1, 5, 10, 20, and 50 μ M. The samples were infused with glass static

infusion tips into an Exactive Plus EMR instrument (Thermo Scientific, Bremen, Germany). Spray voltage was typically 1500V; in-source and HCD collision energies were typically 35 and 150 respectively. Data were acquired with a mass resolving power of 8750. Other MS conditions were: micro scan count=10 and scans co-added=500-600. The raw data were exported and peak parameter extraction was performed using in-house developed charge state deconvolution script¹⁰.

Extraction of apparent association constants

The first level extraction model illustrated here is adapted from Zenobi and coworkers¹¹, with the following modifications: (a) only one ligand is considered and (b) arbitrary number of binding (N) is allowed, and (c) the total target concentration is treated as a fitting parameter.

Scheme 1 lists binding reactions leading to N -number of ligand (L) bound to a target (P). The apparent association constants ($Kap_j, j=1\dots,N$) are defined by

$$Kap_j = \frac{[PL_j]}{[PL_{j-1}][L]} \quad \text{Equation 1}$$

$j=1, 2, \dots, N$

Mass conservation for the target is defined by

$$[P]_T = [P] + \sum_{k=1}^N [PL_k] \quad \text{Equation 2}$$

and mass conservation for the ligand by

$$[L]_T = [L] + \sum_{k=1}^N k [PL_k] \quad \text{Equation 3}$$

$[P]_T$ and $[L]_T$ are the total concentrations of target and ligand, respectively. $[P]$ and $[L]$ are the free concentrations of target and ligand, respectively. $[PL_k], k=1, \dots, N$, are the concentrations of PL_k . Although experimentally $[P]_T$ can be a measured value, using BCA for example, it is treated in our model as a fitting parameter. This approach was taken because it is often difficult to accurately measure the concentration of proteins, especially multi-protein complexes. Therefore the first-level data model is reduced to an optimization procedure for extraction of $N+1$ parameters ($Kap_j, j=1\dots,N$ and $[P]_T$), given a series of (titration) experiments measuring $[PL_j]$ (relative to $[P]_T$), against several total ligand concentrations ($[L]_T$). One may be able to obtain analytical expressions for solutions of Equations (1-3) by eliminating $N+2$ intermediate variables $[P]$, $[L]$, and $[PL_j], j=1, \dots, N$, manually or with aid of math software packages, such as Mathematica or Matlab. We solve the problem directly with Python code implementing functional minimization algorithms (available in Supplemental Material).

Deconvolution of nonspecific binding from specific binding

We model the nonspecific binding as an inverse power distribution described by the following equation

$$Kns_j = \frac{\beta}{j^\gamma}, \quad j=1, \dots, N \quad \text{Equation 4}$$

The nonspecific binding is modeled by the inverse power function parameters of β , the coefficient and γ the power of the function. The rationales for choosing the inverse power distribution to represent the non-constant nonspecific binding constant distribution are: (1) the inverse power distribution is supported on the arguments of nonzero and nonnegative integers and (2) only one parameter (γ) is needed to define the shape of the distribution. The nonspecific binding strength is monotonically decreasing with the increase of number of ligands and the decreasing rate is determined solely by the value of γ . Therefore, only two parameters (β and γ) are needed to specify the nonspecific (non-constant) binding distribution. β is the scaling factor for the inverse power distribution.

If the target has only one specific binding site, the apparent association constants (Kap_j , $j=1, \dots, N$) obtained by the first-level of data modeling can be expressed as

$$\begin{aligned} Kap_1 &= Ks + Kn_1 = Ks + \beta \\ Kap_j &= Kn_j = \frac{\beta}{j^\gamma}, \quad j=2, \dots, N \end{aligned} \quad \text{Equation 5}$$

where Ks is the specific binding association constant (one specific site) and the first apparent association constant is the combination of the specific binding association constant and the contribution of the nonspecific binding distribution of one ligand.

The inverse power distribution model for nonspecific binding can be reduced to the constant nonspecific binding model by simply setting the power value of γ to zero. The physical interpretation of a constant nonspecific binding constant is all nonspecific binding sites are the same. Equation 5 now becomes

$$\begin{aligned} Kap_1 &= Ks + \beta \\ Kap_j &= \beta, \quad j=2, \dots, N \end{aligned} \quad \text{Equation 6}$$

Python code for deconvolution of specific binding constant from the nonspecific one can also be found in Supplemental Material.

Results and Discussion

Extraction of binding constant and target concentration

The simplest method to measure binding constant(s) of a ligand bound to a target is to keep the target solution concentration constant while titrating the ligand. Often the concentration of the ligand can be determined with much greater accuracy than the protein. In our experiments, we started with a model system of FKBP12 binding to the immunosuppressant, rapamycin. The concentration of rapamycin was readily determined through accurate weighing and NMR to determine purity. The concentration of purified FKBP12 was determined by BCA assay and purity estimated from SDS-PAGE to be >90% (data not shown). To explore binding, FKBP12 and its rapamycin complex were measured under native MS conditions (Figure 1). The dominant peaks for charge state 7 have two major

components, corresponding to the FKBP12 sequence containing a His-tag with and without a nickel ion. Nonspecific binding was not observed in this case, which was expected because of the high selectivity of rapamycin¹². To calculate the apparent association constant of FKBP12 and rapamycin, the relative abundances for the free and bound target were obtained by summation of intensities of target (FKBP12) 7+ ions with and without a nickel ion. Then, the target titration curves were constructed after normalization of abundances by the total intensities (either bound or free forms, shown in Figure 2a). Fitting of the experimental data with a single binding constant model yielded the association constant (K_a) of $1.02 \times 10^8 \text{M}^{-1}$ (or $K_d = 9.8 \text{ nM}$) and the target concentration was estimated to be $1.61 \mu\text{M}$. The concentration of FKBP12 estimated from the BCA method was only $1 \mu\text{M}$, nearly 40% lower. Because of the selectivity of the well-known rapamycin/FKBP12 complex, we conclude that the native MS titration is a superior way of accurately measuring effective (*e.g.* folded, intact) proteins concentration. In addition, a titration curve (Figure 2B) of the free ligand (rapamycin) can also be constructed and the ligand titration curve can be used to determine the concentration of the target without direct measurement of target and target/ligand complex. Although this approach may not be necessary for a relatively simple system, such as FKBP12, it is often much more difficult to deconvolve the effective concentration of multi-protein complexes.

Extraction of apparent association constants from relative abundances of ligand/target complexes

As a more complex example, we adapted α -amanitin binding to RNA polymerase II (Pol2). Titrations of the ligand into Pol2, followed by native MS were conducted (Figure 3). To submit the data to the model, relative abundances of the complexes between the target (Pol2) and different numbers of (α -amanitin) ligands were extracted from the raw spectra by use of an in-house software tool called PeakSeeker (described elsewhere¹⁰). Briefly, PeakSeeker detects overlapping peaks by counting zero crossings in second derivatives and deconvolutes raw data into individual Gaussians. Those individual Gaussians are grouped (fitted) into second-level Gaussian-distributed charge state envelopes. PeakSeeker effectively converts the complex and overlapping raw data into deconvoluted (charge state aggregated) spectral peaklists. The relative abundances are implicitly averaged among the charge states.

In the Pol2 complex, two charge state envelopes were detected corresponding to the full 12-component assembly (theoretical MW of 514,154 Da) and the 10-component assembly (rpb4/7, theoretical MW of 469,682 Da). For simplicity, the 12mer and its ligand-bound complexes are described below, but similar findings were observed for the 10mer complex. The relative abundances were formatted as a two dimensional array (Figure 4A), with the first dimension describing the total ligand concentrations and the second dimension describing the number of bound ligands from 0 to a maximal number. The observed maximal number of ligands also determined how many apparent association constants would be calculated. For the target/ligand complexes that were not observed, their relative abundance values were set to zero. In order to increase the robustness of the algorithms, the set of equations in Equation 1 was converted into

$$[PL_{j-1}] [L] Kap_j = [PL_j] \quad \text{Equation 7}$$

$$j=1, 2, \dots, N$$

They are solved together with Equations 2 and 3 for $[L]$, $[P]$, and $[PL_j]$, $j=1,2,\dots,N$, to given $[P]_T$, $[L]_T$, and Kap_j . The advantage for solving the multiple equations simultaneously is that there is no need to derive tedious analytical expressions. Fitting Kap_j ($j=1,2,\dots,N$) against the relative intensities was done using a standard nonlinear function minimization (Simplex algorithm based) procedure. Due to the complex topology of nonlinear minimization error function surfaces, initial guesses must be carefully chosen to ensure the fit was not a local minimum.

We found that there was excellent agreement between the experimental relative abundances (Figure 4A) and the simulated relative abundances (Figure 4B) using the fitted $Kaps$. The fitted Pol2 12mer concentration was estimated to be 6.1 μM .

Deconvolution of specific and nonspecific binding

Unlike the simple case of FKBP12/rapamycin, up to eight α -amanitin molecules were found bound to Pol2. In order to obtain the specific binding constant(s) accurately, contributions from nonspecific binding must be separated. Accordingly, application of the second-level deconvolution inverse power model of Equation 5 for the eight apparent association constants yields a Ks of $9.6 \times 10^5\text{M}$ ($K_d = 1.06 \mu\text{M}$). The nonspecific binding was modeled by two parameters of the power inverse function (equation 5 or 6): $\beta = 4.2 \times 10^5\text{M}$ and $\gamma = 1.8$ or

$$Kn_j = \frac{4.2 \times 10^5}{j^{1.8}}, \quad j=1, \dots, N \quad \text{Equation 8}$$

in which Kn_j have a unit of M. As shown in Figure 5A, the binding model characterized by the three parameters (Ks , β , and γ) fits experimental association binding constants distributions well. On the other hand, the constant nonspecific binding model of Equation 6 does not fit the same data as well (Figure 5B). The fitted values are the nonspecific association constant, $\beta = 4.2 \times 10^4\text{M}$ and the specific association constant, $Ks = 1.34 \times 10^6\text{M}$.

Traditionally, a Hill plot can be constructed of a ligand/protein binding event and the slope of curve is suggestive of cooperative or anti-cooperative binding. However, as demonstrated by Dyachenko and coworkers (8), the Hill coefficient is no longer a constant against the ligand concentration under native mass spectrometry conditions. Several data models have been developed to separate specific and nonspecific binding to solve this problem. Daubenfeld and coworkers modeled the experimentally observed data as a sum of binomial distributed specific binding and Poisson distributed nonspecific binding⁴. Their model predicts that the nonspecific events contribute ~38% of total binding of ADP to creatine kinase (CK). In our pol2/ α -amanitin experiments, deconvoluting apparent association constants with the inverse power function suggested that nonspecific binding accounted for 30% of total binding of α -amanitin to Pol2. The constant nonspecific binding model is

equivalent to that of Shimon and coworkers⁶, which does not fit the apparent association constants as well. In this case, the nonspecific binding accounts for merely 3% of the first binding event.

In the work of Daubenfeld and coworkers⁷, nonspecific binding events were modeled from the relative intensities as a Poisson distribution. Instead, our data model operated on the apparent association constants, giving us flexibility to choose functional forms. Indeed, the inverse power function that we chose has desirable properties. For example, if the power parameter (γ) is zero, the function becomes a constant (a single nonspecific binding constant). Physically, this implies an infinite and uniform large surface for ligands to bind and the model is equivalent to the Langmuir adsorption isotherm. If the power parameter (γ) is negative, association binding constants grow with the number of ligand. The model is then describing a process of ligand polymerization in which ligand/ligand interaction initiated by binding to the target molecule grows stronger. Second, because the target surface available for nonspecific binding is likely to be heterogeneous, with the binding to the first ligand likely to be stronger than that of the second and later ligands. Microscopically, this is intuitive because the first binding event occupies one of the available positions. Of course, the rate of the decrease, encoded in the power parameter of the inverse power function, depends on the nature the target/ligand interaction.

To explore this concept further, we also examined other functional forms, such as the exponential function (result shown in Supplemental Material). The exponential function fits better than the constant model but poorer than that of inverse power. The exponential decay with the number of ligands is too fast and the association constants for large numbers of ligands are selectively discriminated against. Therefore, it seems more appropriate to use polynomial rate of decay.

Theoretically, we should be able to extract specific and nonspecific binding functions and the total target concentration from the relative concentrations of target/ligand complexes directly or in a single stage. In doing so, the ratio of the number of fitting parameters over the number of raw data points is indeed quite small (four over 45) in the case of the inverse power nonspecific binding function or (three over 45) in case of the constant nonspecific model. However, the dependence of the extraction parameters to the raw data (the relative concentrations) becomes more complicated and the fitting process can easily be trapped to a local minimum. Dividing the model into two stages also allows for the inspection of intermediate data of the association binding constants and provides visual aids for a better understanding of binding mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by

- A. (to S.B.P.) grants from the National Institutes of Health (AG021601, AG002132, AG010770) as well as by gifts from the Sherman Fairchild Foundation and G. Harold and Leila Y. Mathers Charitable Foundation.
- B. (to A. L. B.) the Bio-Organic Biomedical Mass Spectrometry Resource at UCSF supported by the Biomedical Technology Research Centers program of the NIH National Institute of General Medical Sciences, NIH NIGMS 8P41GM103481. The ThermoScientific Exactive Plus EMR instrument was supported by NIH NIGMS 8P41GM103481, NIH 1S10OD016229, and UCSF 2013 ETAC Shared Equipment Grant. Initial native MS experiments were also supported by ThermoScientific.
- C. (to R. D. K.) grants from the National Institutes of Health (GM49985, GM36659, and AI-21144)
- D. (to J. E. G.) grants from the National Institutes of Health (NS059690 and GM109896)

References

1. Katta V, Chait BT. Observation of the Hemoglobin Complex in Native Myoglobin by Electrospray-Ionization Mass Spectrometry. *J Am Chem Soc.* 1991; 113:8534–8535.
- 2(a). Loo JA. Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom Rev.* 1997; 16(1):1–23. [PubMed: 9414489] (b) Heck AJ, Van Den Heuvel RH. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev.* 2004; 23(5):368–89. [PubMed: 15264235] (c) Sharon M, Robinson CV. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem.* 2007; 76:167–93. [PubMed: 17328674]
- 3(a). Thompson AD, Dugan A, Gestwicki JE, Mapp AK. Fine-tuning multiprotein complexes using small molecules. *ACS Chem Biol.* 2012; 7(8):1311–20. [PubMed: 22725693] (b) Arkin MR, Tang Y, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol.* 2014; 21(9):1102–14. [PubMed: 25237857]
4. Daubenfeld T, Bouin AP, van der Rest G. A deconvolution method for the separation of specific versus nonspecific interactions in noncovalent protein-ligand complexes analyzed by ESI-FT-ICR mass spectrometry. *J Am Soc Mass Spectrom.* 2006; 17(9):1239–48. [PubMed: 16793278]
5. Wang W, Kitova EN, Klassen JS. Nonspecific protein-carbohydrate complexes produced by nanoelectrospray ionization. Factors influencing their formation and stability. *Anal Chem.* 2005; 77(10):3060–71. [PubMed: 15889894]
6. Shimon L, Sharon M, Horovitz A. A method for removing effects of nonspecific binding on the distribution of binding stoichiometries: application to mass spectroscopy data. *Biophys J.* 2010; 99(5):1645–9. [PubMed: 20816078]
7. Dyachenko A, Gruber R, Shimon L, Horovitz A, Sharon M. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc Natl Acad Sci U S A.* 2013; 110(18):7235–9. [PubMed: 23589876]
8. Stankunas K, Bayle JH, Havranek JJ, Wandless TJ, Baker D, Crabtree GR, Gestwicki JE. Rescue of degradation-prone mutants of the FK506-rapamycin binding (FRB) protein with chemical ligands. *Chembiochem.* 2007; 8(10):1162–9. [PubMed: 17525916]
9. Robinson PJ, Bushnell DA, Trnka MJ, Burlingame AL, Kornberg RD. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc Natl Acad Sci U S A.* 2012; 109(44):17931–5. [PubMed: 23071300]
10. Lu, J.; Michael, TT.; Burlingame, AL.; Guan, S. Improved Peak Detection and Deconvolution of Native Electrospray Mass Spectra of Large Protein Complexes. unpublished work
11. Wortmann A, Jecklin MC, Touboul D, Badertscher M, Zenobi R. Binding constant determination of high-affinity protein-ligand complexes by electrospray ionization mass spectrometry and ligand competition. *J Mass Spectrom.* 2008; 43(5):600–8. [PubMed: 18074334]
12. Clackson T, Yang W, Rozamus LW, Hatada M, Amara JF, Rollins CT, Stevenson LF, Magari SR, Wood SA, Courage NL, Lu X, Cerasoli F Jr, Gilman M, Holt DA. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc Natl Acad Sci U S A.* 1998; 95(18):10437–42. [PubMed: 9724721]

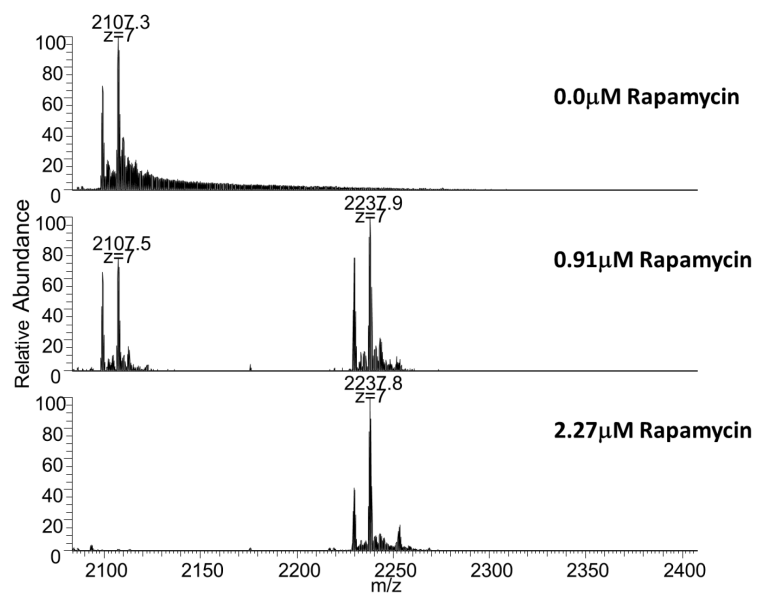


Figure 1. Native mass spectra of FKBP12 and Rapamycin. FKBP12 contains a his-tag and both ions with and without Ni atom were observed. Nonspecific binding products for the charge state of 7 were not observed.

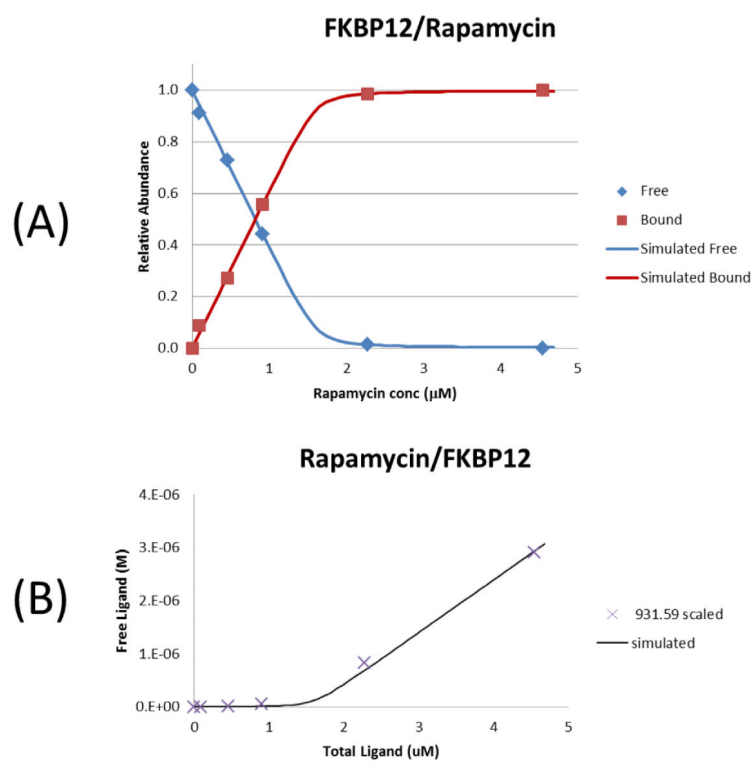


Figure 2. Titration curves of FKBP12 against Rapamycin. A. complex of FKBP12 and Rapamycin of charge state 7. Red curve: FKBP12/Rapamycin relative intensity; blue curve: free FKBP12 relative intensity. B. Rapamycin (at m/z 936.6)

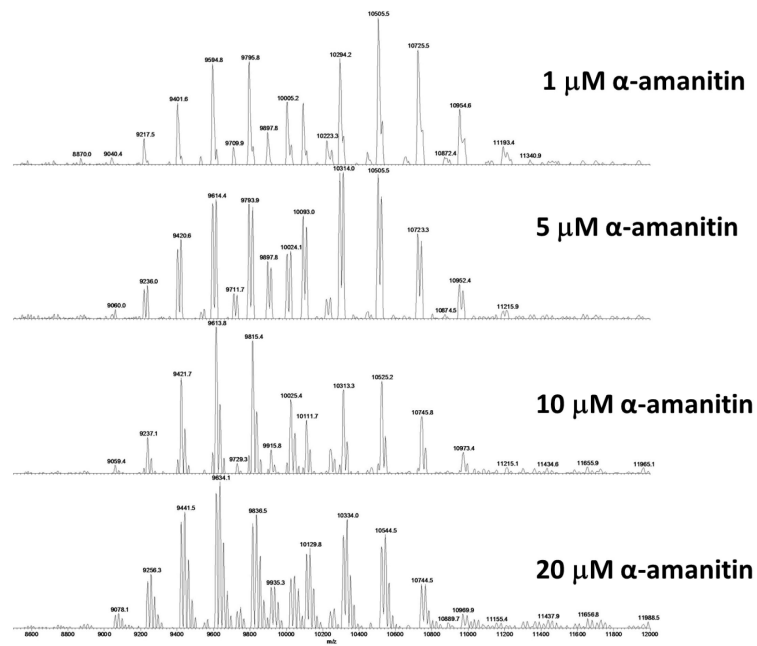


Figure 3.
Native MS spectra of Yeast RNA polymerase II (pol2) and α -amanitin.

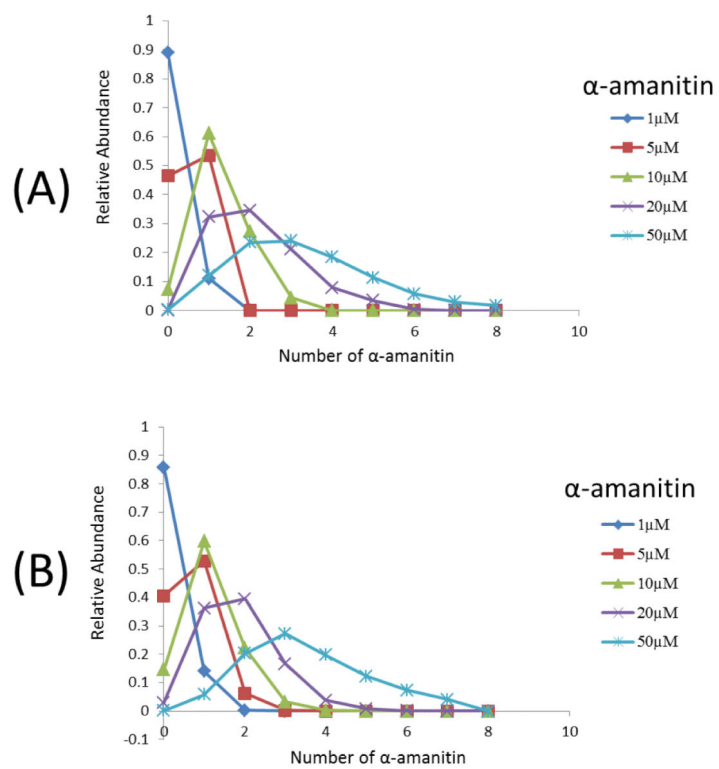


Figure 4. Titration curves of Yeast RNA polymerase II (pol2) and α -amanitin. A. Experimental. B. predicted titration curves using the fitted association binding constants.

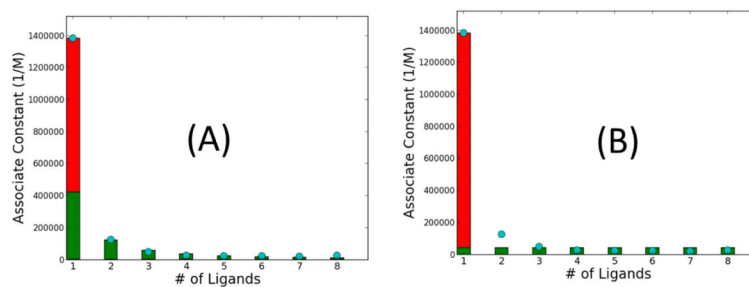
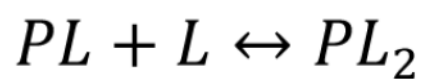
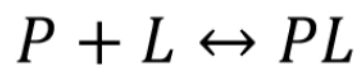
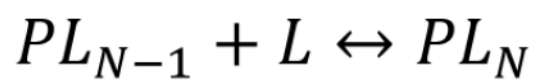


Figure 5. Deconvolution of specific binding constant from nonspecific constants. A. nonspecific binding is modeled by inverse power distribution (Equation 5). B. nonspecific binding is modeled by a constant (Equation 6). Blue circles: apparent association constants; Green bars: contributions of nonspecific binding; Red bar: specific binding component.



...



Scheme 1.