Simplification traps

Heino Prinz

Received: 2 November 2011 / Accepted: 14 November 2011 / Published online: 7 December 2011
© Springer-Verlag 2011

Abstract When experiments are analyzed with simple functions, one gets simple results. A trap springs when experiments show deviations from the expected simplicity. When kinetic experiments do not follow exponential curves, they simply are not of the first or pseudofirst order. They can and have to be calculated on the base of plausible reaction schemes. When dose–response curves are analyzed with logistic functions (“4-parameter fit”) and give Hill coefficients different from one, this is an experimental result stating that more than one molecule is involved in eliciting the response. If one ignores that result, one usually finds forgiving referees, but one will lose real money when one tries to develop such an unspecific compound into a drug.

Keywords Numerical methods · Multiexponential fits · Dose–response curves · 4PL · Logistic function · Systematic deviations

Biochemical analysis requires the calculation of theoretical values. The current approach for enzyme kinetics, dose–response curves, binding kinetics or equilibrium binding studies is based on analytical solutions [3] with explicit functions like:

$$y = f(x). \quad (1)$$

x is the independent variable (for example, the time or the concentration of a compound) and the function $f(x)$ may be rather complex. Even though analytical solutions have been

developed over more than a century, they are limited to only few well-documented examples. Biochemical observations often reveal more complexity. Problems arise when the complexities are ignored or when simple analytical solutions are extended by unjustified means.

An entirely different approach to the calculation of theoretical values is possible with numerical methods, where any theoretical value y can be calculated from any reaction scheme. One simply translates a given scheme into a set of equations and lets the computer solve them.

$$x \rightarrow \text{computer with a set of instructions} \rightarrow y. \quad (2)$$

The procedure (Eq. (2)) also is a mathematical function [8] but not an explicit one. It is well-suited to a huge variety of biochemical reaction schemes because it does not impose limits. Unfortunately, numerical methods are taught in Mathematics departments but usually have no part of the curriculum in life sciences. An introductory textbook to numerical methods for life scientists has been published recently [7]. The main principles of numerical approaches are straight forward: one simply writes down plausible reaction schemes, translates them step by step into a set of ordinary or differential equations and let the computer solve them. This approach can be used for all reaction schemes independent of their complexity. Instructing the computer may seem to be an insurmountable obstacle but current high level computer languages are relatively easy to learn. Moreover, one only needs a limited set of instructions that are introduced step by step in the textbook [7]. GNU Octave, one of the high level languages covered there, is free, so that there are no additional costs involved. It is almost identical to MATLAB® and was used for the calculation of all figures shown below. Their source code can be found in the supplementary material.

This opinion paper gives no introduction to numerical solutions but wants to point out the problems arising with simplified analytical approaches. In many cases, biochemical

Electronic supplementary material The online version of this article (doi:10.1007/s12154-011-0069-3) contains supplementary material, which is available to authorized users.

H. Prinz (✉)
Max-Planck-Institut für molekulare Physiologie,
Otto-Hahn-Str. 11,
44227 Dortmund, Germany
e-mail: Heino.prinz@mpi-dortmund.mpg.de

routine analysis is done with software packages like GraFit® or Microrcal Origin® that provide numerous explicit functions. Some of these are inappropriate.

Multiexponential fits

Exponential curves are very common in nature and can be expected whenever the rate of a concentration decrease is proportional to the concentration itself. Radioactive decay and the discharge of a capacitor are just two typical examples from a large body of similar observations. In biochemistry, the dissociation of a ligand from a receptor is a first-order reaction that has an exponential curve as the analytical solution:

$$B = B_0 \cdot e^{-kt}. \quad (3)$$

B denotes the concentration of bound ligand, B_0 is the concentration of bound ligand at the beginning (time zero), t is the time (the independent variable), and k is the rate constant of the dissociation reaction. The simplicity of Eq. 3 is the reason why association kinetics usually is measured under condition of pseudofirst order, i.e., with one reaction partner (usually the ligand) in large excess, so that its free concentration does not significantly change during the reaction. All this is well understood. A simplification trap springs when kinetic experiments do not follow an exponential curve like Eq. 3 and when they are analyzed as a sum of exponential curves:

$$B = \sum B_0(i) \cdot e^{-k(i)t}. \quad (4)$$

i is an index, with $B_0(1)$ and $k(1)$, the amplitude and rate constant of the first element, $B_0(2)$ and $k(2)$ of the second element, and so forth. This approach is valid for small perturbations of equilibrium and had been used successfully for the calculation of T-jump and other relaxation experiments [10]. Binding kinetics generally cannot be regarded as small perturbations of an equilibrium, so that a sum of exponentials is not justified for the analysis. Another reason for trusting the sum of linear exponentials (Eq. (4)) may be a mix up with Fourier series, where indeed any periodic function can be written as a sum of exponentials. However, there, the exponents are imaginary, so that they are composed of sums of cosines and sinus, and only the exponential notation looks similar to Eq. 4:

$$B = \sum c_n \cdot e^{-i \cdot n \cdot t}. \quad (5)$$

For this Fourier series (Eq. (5)), i is not an index but an imaginary number. The index is n , which is used as the natural number in $i \cdot n \cdot t$. c_n are the Fourier coefficients that themselves are functions of the independent parameter t . In contrast to the Fourier series (Eq. (5)), the simple sum of linear exponential functions (Eq. (4)) is not a series of

orthogonal functions, so that the rate constants and amplitudes are correlated, which means that the selection of one number for one rate constant generally influences amplitudes and rate constants of the other elements of Eq. 4.

In order to illustrate the problems arising when one uses such multiexponential fits, a nonexponential curve was fitted to two linear exponentials. Figure 1 shows a hyperbolic binding curve (star) that was fitted to a sum of two exponential curves (solid line). The fit itself looks good, but of course such a fit is meaningless. There simply are no exponential elements in an equilibrium-binding curve.

The binding curve (star) was calculated with the program HypEx.m, provided in the supplementary material. Five percent random noise was considered. Most readers will agree that an exponential fit to equilibrium-binding curves is nonsense. However, why do journals accept multiexponential fits for complex kinetics? Why don't they insist on the calculation of differential equations? Some referees may not be aware that a numerical solution of differential equations can be performed by anybody at no additional costs. It just requires a (any) personal computer, public domain software and some instructions.

Additional binding sites and systematic deviations from a given theoretical fit

A more general problem lies in systematic aberrations from expected curve progressions. We all know that biochemical experiments are prone to experimental errors. Heterogeneous concentrations, protein activities, drug solubilities,

Fig. 1 Hyperbolic binding curve with $R_0=10 \mu\text{M}$, $KD=10 \mu\text{M}$ and a maximal free ligand concentration of $100 \mu\text{M}$. The data points were generated with 5% random noise. They were fitted with a sum of two exponential functions (red line)

temperatures and sometimes just the charges of an enzyme may lead to quite a variety of experimental curves. Therefore, it is a good lab practice to repeat the experiments several times and use error ranges for the final graphs. For linear relations, the goodness of fit can be assessed with the coefficient of determination R^2 , which basically is a normalized sum of squares.

$$R^2 = 1 - \Sigma(y_i - f_i)^2 / \Sigma(y_i - y_{ave})^2. \quad (6)$$

The numerical values of R^2 lie between 0 and 1. For example, when R^2 is 0.91, then 91% of the total variation in y can be explained by a linear relationship between the independent variables x and y . One generally would trust such data. One should be aware, however, that R^2 is a statistical parameter and that both systematic and statistical deviations will lead to a decrease in R^2 . Small systematic deviations may be better detected with the human eye, but one has to look out for them as the following example may illustrate:

Equilibrium binding of one ligand to one binding site or simple steady state enzyme kinetics follows a well-known hyperbolic function. When bound ligand or initial velocity of a simple enzymatic reaction is plotted versus the ligand or substrate concentration in a double reciprocal manner, a hyperbolic function would appear as a straight line. Figure 2 shows a double reciprocal plot for a ligand (star) binding to a protein to two sites with two markedly distinct affinities, one with an equilibrium dissociation constant of 20 and the other with 200 μM . This theoretical curve is indicated in green. The

“data” were derived from 10% random noise added to the calculated bound concentrations. A linear regression was performed to these “data”, and the result is indicated as the black line. The coefficient of determination R^2 (0.983) looks all right, but the systematic aberration is evident to the eye.

Figure 2 is just one theoretical example. For real experiments, such systematic deviations from simple binding curves are often observed, but usually are not followed up. We have made a point to follow this up [4] and found that indeed two distinct inhibitor-binding sites could be detected in the x-ray structures of a family 18 chitinase.

Dose–response curves

Most screening results, see for example, the NIH guidelines [2], are analyzed with the help of logistic functions. These mathematical functions had been developed for the description of (e.g., bacterial) population growth [9]. They are easy to calculate and produce symmetrical sigmoid curves when plotted in a half-logarithmic scale. The shape of these curves is similar to dose–response curves, whereby a given physiological response (membrane polarisation, heart rate, enzymatic activity or agonist binding, etc.) is plotted versus the logarithm of the applied concentration x . Only four parameters are required to determine the shape of such a logistic curve, namely Min for the baseline, Max for the amplitude, EC_{50} for the concentration of half-maximal response and n for the Hill coefficient.

$$\text{Response} = (\text{Max} - \text{Min}) / (1 + (EC_{50}/x)^n). \quad (7)$$

Fig. 2 An equilibrium binding curve was calculated (green line) for two independent sites with $R_0=1 \mu\text{M}$ and two different affinities ($KD_1=20$, $KD_2=200 \mu\text{M}$). The data were generated with 10% random noise for the bound concentrations and plotted in a double reciprocal manner (red stars). A linear regression (black line) was performed and the coefficient of determination R^2 was calculated as 0.9827

Fig. 3 A dose–response curve was calculated assuming cooperative binding to four sites (star). Analytical approximations are indicated for a logistic curve with $n=4$ (blue line) and a one-site binding curve (black line)

For these symmetric curves, EC_{50} is the value of half-maximal response. It corresponds to the inflection point. Logistic functions (Eq. (7)) correspond to hyperbolic binding curves if and only if $n=1$. For all other values of n , there is no comprehensible reason why a drug response curve should follow a logistic function. In the absence of computers, logistic functions were practical approximations for the description of cooperative effects under equilibrium conditions [1], but times have changed. Real dose–response curves usually are not symmetric [5], as shown in Fig. 3. For one-site binding or one-site competition, one would expect a hyperbolic binding or inhibition curve that is the same as a logistic curve with $n=1$ (black line in Fig. 3). Cooperative effects are encountered when there are multiple sites and the affinity of a subsequent ligand increases once a previous ligand has bound. It would also be encountered if the binding of a first ligand facilitates the binding of a second one [5]. Figure 3 shows such a cooperative curve (red stars). The curve is not symmetrical any more, and the maximal slope increases. For extreme cooperativity, one may envisage an “all or nothing” effect, so that either all sites are occupied or none. Only in this fictional case, one would encounter a logistic curve with a Hill coefficient as an integer value corresponding to the number of sites. Published Hill coefficients (e.g. <http://www.ncbi.nlm.nih.gov/sites/entrez>) typically give noninteger values that differ from one.

One explanation for large Hill coefficients is rather trivial [6]: most screening campaigns were aimed at enzyme inhibitors. Even if an inhibitor binds unspecifically, its binding will cause a slight conformational change of the target protein. This change will loosen the structure, so that a second unspecific inhibitor can bind. This again will loosen the structure and so forth, until the protein structure is sufficiently damaged and the enzyme activity decreases. This process may be called “multiple allosteric binding” [6], which is just the same as unspecific denaturation. How can an unspecifically denaturing agent be developed into a specific drug? The industry lost billions of dollars with that futile endeavor.

Conclusion

Today, the analysis of most biochemical experiments is performed with a few well-known analytical solutions,

usually for one-site binding. These simple solutions are inappropriate for complex reactions. For example, there is no justification for using sums of exponential functions when a complex kinetic reaction can be calculated from differential equations. There is no justification for using logistic functions for the analysis of dose–response curves. Whenever one observes systematic aberrations of experimental data from the calculated simple functions, one should follow this up. Otherwise, one may be caught in a trap: ignoring deviations in equilibrium binding, one may believe that a nonspecific ligand binds to only one site. Ignoring additional electron densities in the x-ray structure, one may identify only one site for this ligand. Ignoring the shape of dose–response curves, one may believe that a whole class of denaturing agents binds specifically to this site and overlook the controls. One may fool oneself and others as long as the papers get published, but one should be careful whenever drug development is concerned and real money is involved.

References

1. Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol (Lond)* 40:iv–vii
2. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP (2006) Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *PNAS* 103:11473–11478
3. Michaelis L (1912) *Einführung in die Mathematik für Biologen und Chemiker*. Springer, Berlin
4. Pantoom S, Vetter IR, Prinz H, Suginta W (2011) Potent family-18 chitinase inhibitors: x-ray structures, affinities and binding mechanisms. *J Biol Chem* 286:24312–24323
5. Prinz H, Schöniichen A (2008) Transient binding patches: a plausible concept for drug binding. *J Chem Biol* 1:95–104
6. Prinz H (2010) Hill coefficients, dose–response curves and allosteric mechanisms. *J Chem Biol* 3:37–44
7. Prinz H (2011) Numerical methods for the life scientist. Binding and enzyme kinetics calculated with GNU Octave and MATLAB. Springer, Heidelberg
8. Ruthing D (1984) Some definitions of the concept of function from Bernoulli, Joh. to Bourbaki, N. *Math Intell* 6(4):72–77
9. Verhulst P-F (1845) *Recherches mathématiques sur la loi d'accroissement de la population*. *Nouv. mém. de l'Académie Royale des Sci. et Belles-Lettres de Bruxelles* 18:1–41
10. Winkler-Oswatitsch R, Eigen M (1979) The art of titration. From the classical end-point procedures to the modern differential and dynamic analyses *Angew. Chem Int Ed* 18(1):20–49