Effective experimental design: enzyme kinetics in the bioinformatics era

Emma F. Murphy, Steven G. Gilmour and M. James C. Crabbe

Acquiring details about the kinetic parameters of enzymes is crucial to both drug development and clinical diagnosis. The correct design of an experiment is crucial to collecting data suitable for analysis, modelling and deriving the correct information. As classical design methods are not targeted to the more complex kinetics now frequently studied, further work is required to estimate parameters of such models with low variance. This review examines the different options available to produce major gains in information, productivity and the accuracy of each experiment.

***Emma F. Murphy and M. James C. Crabbe**
Bioinformatics, Division of Cell and Molecular Biology
School of Animal and Microbial Sciences
The University of Reading
Whiteknights, Reading
Berkshire, UK RG6 6AJ
tel: +44 1189 317385
fax: +44 1189 310180
*e-mail: e.f.murphy@reading.ac.uk
Steven G. Gilmour
School of Mathematical Sciences
Queen Mary, University of London, Mile End Road
London, UK E1 4NS

▼ The availability of high-throughput technologies in the post-genomic era means that generating data for biological catalysis in an efficient and rational way is no longer a major problem. Instead, the focus has shifted onto another issue: integrating informatics and statistics to make use of the available information and answer important biological questions [1]. Obtaining details about the kinetic parameters of enzymes (novel or wild-type) is crucial to several research fields, including drug development, clinical diagnosis and biotechnology. For example, they could be essential in the prediction of the toxic and metabolic effects of drugs.

With this in mind, the fundamental starting-point must be obtaining correct experimental data. It will never be possible to make good data from a badly designed experiment; incorrect designs can lead to poor and/or insufficient measurements, which can, in turn, result in misleading parameter estimates. Here, we demonstrate how the integration of statistics and bioinformatics can produce major gains that are quantifiable in terms of information, productivity and accuracy of each experiment.

Enzyme kinetics

Parameter knowledge: an essential research and drug development tool

Enzymes have long been considered valuable natural products given their wide range of applications in medical, scientific and academic research fields. However, their use has significantly increased in recent years, being boosted in particular by the possibilities of genetic modification and protein engineering. Ultimately, this has led to the ability to create specialized enzymes in large quantities, with properties not possessed by natural substances.

Information about kinetics is necessary for the evaluation of any enzyme used in medical research, clinical diagnosis, pharmaceutical research or drug development, because it provides essential information about how an enzyme will behave or respond in given situations. For example, it is essential in areas of drug development for understanding pharmacokinetics, interactions within the formulation, and for stability testing. High-throughput screening is needed to establish the metabolic stability, kinetic parameters and prediction of metabolic clearance, and the information can be used to develop *in vitro* models that aid in the initial design. Such pharmacokinetic information can be scaled up to *in situ* models of the enzyme activity in model organs.

Some examples of situations where accurate knowledge and methods for determining kinetic parameters are crucial to drug development include:

- Metabolism of drugs by enzymes *in vivo* and the inhibition or induction of drug-metabolizing enzymes by drugs themselves or by altered enzymes in disease states. For example, disease conditions of the liver and the administration of drugs that are cytochrome P450 enzyme-inducers

1359-6446/02/$ – see front matter PII: S1359-6446(02)02384-X

or -inhibitors can influence the systemic clearance of the drug. Therefore, some knowledge of the influence on the kinetics of this enzyme system is very often essential in the early stages of the drug discovery process.

- Many drugs are themselves enzymes and it is important to establish their molecular interactions with other molecules, enzymes and drugs *in vivo*; for example, the use of thiopurine methyltransferases and HIV proteases.
- Common drug–drug interactions can be understood in terms of alterations in metabolism, which are associated primarily with changes in the activity of certain enzymes. Kinetic parameters that describe metabolism-based drug interactions can be used to predict the pharmacokinetic consequences of exposure to one or multiple drugs; for example, clinical consequences such as drug toxicity and ineffectiveness.
- Although certain enzymes might have unaltered levels in the blood of an individual with a particular disease, they might have an altered activity. Knowledge of K_m and V_{max} and methods for their accurate determination are therefore required for diagnosis.
- Kinetic models are needed before the mode-of-action of an inhibitor can be determined. Inhibitors are becoming increasingly important in research and so understanding how they work and how to target them is essential. Much work on the search for specific enzyme targets and inhibitors can be found in the literature [2,3].

As target discovery [identifying and validating suitable drug targets for therapeutic intervention (largely enzymes)] and lead discovery (uncovering the molecules that act on those targets) are now the main components of drug research, it is appropriate that some methodological and statistical work is being applied to the study of complex enzyme-kinetic models. Methods for optimizing enzymatic experimental methods are important for the decision-making process in the corporate sector and enable accurate comparison of parameters from one experiment to another.

Methods of analysis

Enzyme kinetic parameters are derived from rate equations based on mathematical treatment of data from enzyme-catalysed reactions. That is, once the kinetic characteristics of an enzyme are known, a model, with its associated equations, can be developed for the number of substrates binding at the active site. Kinetics are typically determined using a steady-state approximation and by following initial rates measured at different substrate concentrations. For most reseachers, the method introduced by Briggs and Haldane [4] remains the best method for the analysis of kinetic data [5]. However, it is also possible to study kinetics using single progress curves over time. Although this method is still used, it does have several problems; its success is highly dependent on the curve containing enough information to estimate the parameters with a low error, and the kinetics might be affected by substrate depletion, product accumulation, pH changes, and so on, over time.

In kinetic experiments, as a result of experimental error, the measured data will rarely solve the model equation exactly. The statistical approach to this problem is to include a random component in the model to account for all the unknown sources of variation (the error term).

The simplest enzyme–substrate complexes display Michaelian (rectangular hyperbolic) kinetics [6]; in this situation, the maximal velocity with respect to substrate concentration (V_{max}), and the Michaelis constant (K_m), are the parameters to be estimated. However, many native and engineered enzymes display non-Michaelian kinetics, having mechanisms which lead to higher-order rational polynomial functions.

Most enzyme kinetic data are analysed by graphical methods fitting a statistical model. Parameters of the Michaelis–Menten equation are often estimated by 'ordinary least squares' using the linearized form:

$$\frac{[S]}{v} = \frac{K_m}{V_{max}} + \frac{1}{V_{max}}[S] \qquad \text{[EQN 1]}$$

where [S] is substrate concentration and v is velocity.

This is also useful in identifying non-Michaelian behaviour, although there is some uncertainty among biochemists on its use for estimating the parameters of these models [6,7]. Several other more statistically sound procedures have been recommended and used. These include: non-linear least squares analyses [8]; the use of least absolute deviations (or L1-norm) regressions on the linearized form of the model [9,10]; Tukey's biweight regression [11]; and the TBS/PS (transform both sides/power of S) model [12]. Currie concluded that these are preferable to graphical methods [13] but none of these methods is readily extendable to higher-order rational polynomial models [14]. For more details on kinetic analysis there are many general reviews and computer programs [6,15–17].

Classical enzymology versus Bayesian experimental design

Experimental design is about how to obtain the optimum information required for the analysis, to find a model for the data and an error term to account for imperfections. In enzyme kinetics, there are many design issues that need to be considered for optimum efficiency, including the substrate range and individual concentrations used, and the number of replicates and data points that can provide the necessary information to fit and discriminate between models and obtain good parameter estimates. The importance of experimental design and its role in successful analysis is becoming increasingly recognized in pharmaceutical and commercial research [18–20].

Classical methods of experimental design for nonlinear models, obtained analytically or by simple computation from

the likelihood, produce highly variable results owing to their dependence on the initial parameter values chosen [21]. By contrast, a Bayesian design involves the use of pre-determined information in the form of distributions of parameter estimates, making it possible to combine information from different sources.

The conclusion that there should be some form of experimental design to improve the validity of the results is evident. For efficiency purposes, one wants to design to experiment, not experiment to design. In terms of which design is appropriate, there are both advantages and disadvantages of the classical and Bayesian approaches.

The advantages and disadvantages of classical enzymology

Classical design methods are frequently used, and computer programs aiding design are widely available and generally user-friendly. It is also often unnecessary to obtain a large amount of prior information before proceeding with experiments. If the methods of both design and subsequent analysis have already been established, this facilitates straightforward comparison of new results with earlier studies. For this reason, commercial research in particular will closely adhere to existing methodology. For example, pharmacokinetic profiles are measured in terms of plasma drug concentration versus time, and the ensuing analysis, by area-under-the-curve calculations, provides information on absorption, distribution, metabolism and elimination (ADME) parameters [22].

However, the classical approaches also have several drawbacks, which need to be considered. The design methodologies are based on the knowledge of experimental statistics and not on the science or the system itself. This can be of particular difficulty when more-complex kinetics are suspected as the classical systems were originally designed only for the simple first-order Michaelis–Menten kinetics. In addition, the designs are not iterative and thus provide poor efficiency and cost-effectiveness in industry. Both the design and data-fitting can also be largely dependent on the initial point parameters chosen. This can significantly influence both the data fitting and the results.

Advantages and disadvantages of Bayesian experimental design

Bayesian designs are based around prior distributions of parameter estimates and their variance, rather than on chosen single-point values. Importantly, they incorporate all previous scientific knowledge, which can be from a variety of sources, and not just knowledge of working designs and statistics. It is often necessary in any experimental design to obtain an order-of-magnitude of K_m; the Bayesian approach simply expands on this and uses existing information. The basis of Bayesian designs is that they are tailored to the type of kinetics – trends from simple kinetics are not just assumed for the study of more complex models. It is also an iterative approach, which increases both efficiency and cost-effectiveness.

A disadvantage of the Bayesian approach is that prior information about the parameters, kinetics and distribution of error might not always be easily available. In addition, the methodology is not well developed or established as-yet, and it's use currently requires much knowledge of the mathematics and statistics involved. There is, however, the potential to identify trends using Bayesian studies so that programs and databases can be written to counteract the need for so much knowledge and understanding. This issue is discussed later in this review.

Classical experimental design

There is currently limited evidence in the literature that considers the design of more-complex kinetic experiments. Authors either use established classical methods [23] or quote no defined method of choosing their design. Bayesian design receives little, if any, mention.

However, experimental design and the existence of prior knowledge are issues that have been raised in the past. For example, Cornish-Bowden [6] suggested that, in terms of substrate range, an optimum experimental design should be based around K_m when studying Michaelis–Menten kinetics. This idea is based on using an estimate of K_m; that is, prior knowledge.

Since this suggestion, there have been several attempts to look at the design of other simple Michaelis–Menten kinetic experiments. In particular, Duggleby *et al.* [24,25] have worked with time-course experiments (less accurate than steady-state courses over a range of substrate concentrations), and studied the optimum initial substrate concentrations to reduce to the smallest standard error for the Michaelis constant. Similar to Cornish-Bowden, Duggleby and colleagues again identified the importance of basing the design around K_m. However, this approach needs to be extended to more than simply considering the substrate range; for example, how many and which individual substrate concentrations should be used? These issues are crucial to the study of more-complex kinetics, and a good design is essential to saving on resources (such as time, money and materials) and in generating parameter estimates with low variance.

Bayesian experimental design

In effect, Bayesian data analysis is the method of making inferences from data using probability models for quantities we observe and about which we wish to learn. In a review of Bayesian experimental design, Chaloner and Verdinelli [26] concluded that it 'is an exciting and fast developing area of research' but 'it [is] regrettable…that so few case studies appear in the statistical literature'. They made the suggestion that Bayesian statistics could actually be applied to experiments, not simply used in their subsequent analysis, and identified a very important area for future research. Using a decision-theoretic

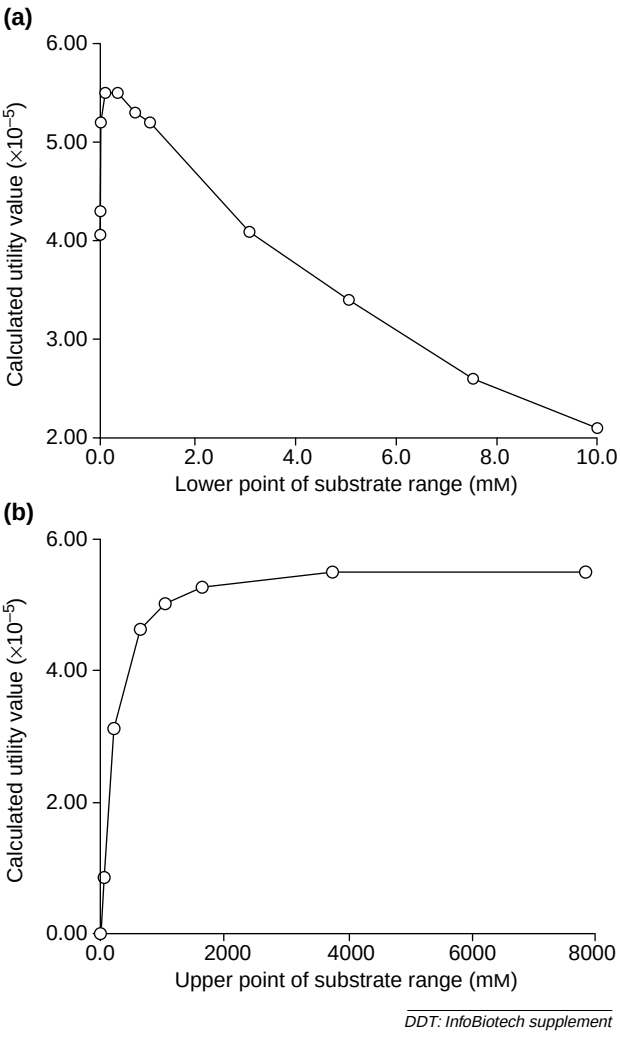

Figure 1. The effect of altering (a) the lower point and (b) the upper point of the substrate concentration range on the calculated utility function (example shown using data for a β-lactoglobulin dehydrogenase assay).

approach [27], a utility function can be specified, reflecting the purpose of the experiment (e.g. to reduce parameter variance), the design regarded as a decision problem, and then the design that maximizes the utility can be selected.

Little has been reported in the literature on the use of this proposal apart from one recent review [28], which used a Bayesian utility function to choose sample sizes to confirm the results from a clinical trial measuring protein expression levels in breast cancer subjects. However, as Clyde identified, this approach holds great potential and could mean dramatic improvements for experimental design in a very complex field.

An example of the effective use of Bayesian utility functions

We have taken several steady-state kinetic datasets and developed utility functions from the models used to fit the data [29–31]. This is based on Chaloner and Verdinelli's suggestion of using a Bayesian approach to design an experiment by specifying a utility function [26]. The design must be regarded as a decision problem and then the design maximizing the utility can be selected as the optimum. The prior knowledge is in the form of the parameter estimates, with known variance distribution from the initial experimental data. The prior distribution represents the uncertainty in the values, and this is what we aimed to minimize.

The utility functions were then written into a workbook with the aid of the Mathematica program [Version 4.1, Wolfram Research (http://www.wolfram.com)], as this procedure involves complex algebra. The prior distribution for the unknown parameters and the variance vector must first be written as a variance–covariance matrix. The utility function is then developed from expressions for the variance of the parameters in the kinetic model, incorporating the prior distribution. The variance expressions themselves are developed by differentiation of the model expressions with respect to each parameter. All of this involves differentiation and transposing, multiplying and inverting large matrices.

Once such a workbook has been constructed in Mathematica, the utility values for several designs can be systematically studied in terms of the substrate values measured. Little has been reported on the use of this valid proposal, yet our findings indicate dramatic ways to improve the study of enzymes. We have studied a variety of data to demonstrate the maximization in utility, minimization in variance, and subsequent improvements in experimental design that are possible. Interestingly, very similar trends are observed for each of several very different kinetic systems: the Michaelis–Menten equation (with trypsin) [29]; a three-parameter hyperbolic equation (with β-lactoglobulin dehydrogenase) [30]; and a more-complex four-parameter equation (with cytochrome P450 3A4) [31].

New experimental design rules for enzyme kinetics

From the study of Bayesian utility functions, and the evidence gathered, we now present an optimal and iterative Bayesian approach to designing an experiment:

- Estimate or obtain prior information about K_m (or the dissociation constant parameter) for the enzyme.
- Choose the substrate range: there are optimum lower and upper points above and below K_m. Figure 1 shows the trends observed for each dataset. For example, in all cases, the design requires the range to extend to 100 times the K_m prior value.
- Most importantly, the distribution of points in the range is crucial; it is not simply a matter of even or multiple increases. Indeed, at least 60% must be below the K_m (or plural if there is more than one dissociation constant) and 40% above. As shown in Table 1, this choice halves the variance found using a simple even-spread across the range.

- As might be expected, the more datapoints measured the better, but, more interestingly, how those points are made up is irrelevant. For example, there is no difference in using duplicates of six points, or twelve single points.

Crucial trends have been identified. Specifically, it is important to estimate K_m first. Although lengthy methods involving algebraic matrices have been used here to identify the optimum design, there is the possibility in the future to create a database of Bayesian information and algorithms for designing kinetic experiments according to prior information and aims.

Future impact of computational biology and experimental design on drug research

Although the example presented here is for enzyme kinetics, this concept of Bayesian design can, of course, be applied to other areas of study, such as receptor–ligand binding and immunoglobulin-binding kinetics. In effect, it could be very useful in any area where there is a requirement to determine catalytic parameters and/or constants.

The results show that the design of experiments in these types of study should be carefully considered and any prior information and optimal designs should be recorded. Together with the further study of some utility functions and some advances in computational biology and programming, it should be possible to pool information in the form of a database and eventually design programs to design experiments. Statistics and biology no longer have to remain separate entities.

These advances are crucial to drug design to study the effects of drugs in terms of their metabolic action, efficacy and toxic effects. Prior information is available in existing data and it is logical that it should be used to aid the design of the increasingly complex experiments required to advance drug design.

References

1 Sander, C. (2001) Challenges in 2001. *Bioinformatics* 17, 1–2

2 Chowdhury, S.F. *et al*. (1999) Design, synthesis and evaluation of inhibitors of trypanosoma leishmanial dihydrofolate reductase. *J. Med. Chem.* 42, 4300–4312

3 Owens, J. (2001) John Weinstein discusses information-intensive approaches to cancer drug discovery. *Drug Discov. Today* 6, 1145–1148

4 Briggs, G.E. and Haldane, J.B.S. (1925) A note on the kinetics of enzyme action. *Biochem. J.* 19, 338–339

5 Rudolph, F.B. and Fromm, H.J. (1996) Plotting methods for analysing enzyme raw data. In *Contemporary Enzyme Kinetics and Mechanism* (Purich, D.L., ed.), pp. 89–109, Academic Press

6 Cornish-Bowden, A. (1979) *Fundamentals of Enzyme Kinetics*, Butterworth and Co.

7 Bardsley, W.G. *et al*. (1995) Evaluation of model discrimination, parameter estimation and goodness-of-fit in non-linear regression problems by test statistics distributions. *Comp. Chem.* 19, 75–84

8 Wilkinson, G.N. (1961) Statistical estimations in enzyme kinetics. *Biochem. J.* 30, 324–332

9 Bliss, C.I. and James, A.T. (1966) Fitting the rectangular hyperbola. *Biometrics* 22, 573–602

10 Eisenthal, R. and Cornish-Bowden, A. (1974) A new procedure for estimating kinetic parameters. *Biochem. J.* 139, 715–720

11 Cornish-Bowden, A. and Endrenyi, L. (1981) Fitting of enzyme kinetic data without prior knowledge of weights. *Biochem. J.* 193, 1005–1008

12 Ruppert, D. *et al*. (1989) A transformation/weighting model for estimating Michaelis–Menten parameters. *Biometrics* 45, 637–656

13 Currie, D.J. (1982) Estimating Michaelis–Menten parameters: bias, variance and experimental design. *Biometrics* 38, 907–919

14 Crabbe, M.J.C. (1988) Simple graphical methods for use with complex ligand-binding and enzyme mechanisms. *FEBS Lett.* 235, 183–188

15 Bardsley, W.G. *et al*. (1996) Optimal design: a computer program to study the best possible spacing of design points for model discrimination. *Comp. Chem.* 20, 145–157

16 Crabbe, M.J.C. (1984) Microcomputers in the study of enzyme reactions. In *Microcomputers in Biology: A Practical Approach* (Ireland, C.R. and Long, S.P.), pp. 107–150, Oxford University Press

17 Wong, J.T. (1975) *Kinetics of Enzyme Mechanisms*, Academic Press

18 Ye, C. *et al*. (2000) Design of experiment and data analysis by JMP (SAS institute) analytical method validation. *J. Pharm. Biomed. Anal.* 23, 581–589

19 Lendrem, D. *et al*. (2001) Design of experiments in developmental chemistry: potential obstacles. *Org. Process Res. Dev.* 5, 324–327

20 Owen, M.R. *et al*. (2001) Efficiency by design: optimisation in process research. *Org. Process Res. Dev.* 5, 308–323

21 Heath, D. (1995) *An Introduction to Experimental Design and Statistics for Biology*, UCL Press

22 Gunaratna, C. (2001) Drug metabolism and pharmacokinetics in drug discovery: a primer for bioanalytical chemists, part II. *Current Separations* 19, 87–92

23 Martins, A.M. *et al*. (2001) *In situ* kinetic analysis of glyoxalase I and glyoxalase II in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 268, 3930–3936

24 Duggleby, R.G. and Clarke, R.B. (1991) Experimental designs for estimating the parameters of enzyme catalysed reactions. *Biochim. Biophys. Acta* 1080, 231–236

25 Gouder, C.T. *et al.* (1999) Parameter estimation using a direct solution of the integrated Michaelis–Menten equation. *Biochim. Biophys. Acta* 1429, 377–383

26 Chaloner, K. and Verdinelli, I. (1995) Bayesian experimental design: a review. *Stat. Sci.* 10, 273–304

27 Lindley, D.V. and Smith, A.F.M. (1972) Bayes estimates for the linear models. *J. Roy. Statist. Soc. Ser. B.* 34, 1–41

28 Clyde, M.A. (2001) Experimental design: a Bayesian perspective. In *International Encyclopedia of the Social and Behavioral Sciences* (vol. 8), pp. 5075–5081, Elsevier Science

29 Olsen, K. *et al*. (2000) Steady-state kinetics and thermodynamics of the hydrolysis of β-lactoglobulin by trypsin. *J. Agric. Food Chem.* 48, 3086–3089

30 Charlier, H.A. and Plapp, B.V. (2000) Kinetic cooperativity of human liver alcohol dehydrogenase γ_2. *J. Biol. Chem.* 275, 11569–11575

31 Shou, M. *et al*. (1999) Sigmoidal kinetic model for two co-operative substrate-binding sites in a cytochrome P450 3A4 active site: an example of the metabolism of diazepam and its derivatives. *Biochem. J.* 340, 845–853