Novel Experimental Design for Steady-state Processes: A Systematic Bayesian Approach for Enzymes, Drug Transport, Receptor Binding, Continuous Culture and Cell Transport Kinetics

M. James C. Crabbe[1,*], Emma F. Murphy[1] and Steven G. Gilmour[2]

[1]*School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AJ, UK,* [2]*School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London, E1 4NS, UK*

Abstract: We demonstrate that a Bayesian approach (the use of prior knowledge) to the design of steady-state experiments can produce major gains quantifiable in terms of information, productivity and accuracy of each experiment.

Developing the use of Bayesian utility functions, we have used a systematic method to identify the optimum experimental designs for a number of kinetic model data sets. This has enabled the identification of trends between kinetic model types, sets of design rules and the key conclusion that such designs should be based on some prior knowledge of the kinetic model.

We suggest an optimal and iterative method for selecting features of the design such as the substrate range, number of measurements and choice of intermediate points. The final design collects data suitable for accurate modelling and analysis and minimises the error in the parameters estimated. It is equally applicable to enzymes, drug transport, receptor binding, microbial culture and cell transport kinetics.

Key Words: Kinetics, experimental design, bayesian design, kinetic parameters, prior knowledge, parameter variance, drug discovery.

INTRODUCTION

In the evaluation of any enzyme to be used in research, obtaining information about the kinetic parameters is crucial. There are several research fields and applications where they are widely used, including: medical and pharmaceutical research; clinical diagnosis; and drug development. Enzyme kinetics, drug transport kinetics, cell transport kinetics, microbial culture kinetics and receptor binding kinetics are examples of key kinetic systems about which information is fundamental to making advancements in drug discovery.

The fundamental starting point in collecting kinetic information for these purposes must be to obtain the correct experimental data. Design of the experiment is necessary because it will never be possible to make good data from a badly designed experiment. The incorrect design can lead to poor and/or insufficient measurements, which can, in turn result in misleading parameter estimates. Through careful design, major gains can be made that are quantifiable in terms of information, productivity and accuracy of each experiment [1].

Kinetic parameters are derived from rate equations based on mathematical treatment of data from enzyme-catalysed reactions. When 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. By measuring the initial rates at different substrate concentrations, kinetics are typically determined using a steady-state approximation. This method introduced by Briggs and Haldane [2] remains the best method, for most researchers, to analyse kinetic data [3].

Michaelian (rectangular hyperbolic) kinetics [4] are characteristic of the simplest complexes, for example enzyme-substrate complexes. However, many native and novel processes display more complex kinetics, having mechanisms which lead to higher order rational polynomial functions. The analysis of the kinetic data is usually performed by fitting statistical models using graphical methods and the parameters are often estimated by the ordinary least squares method. Experimental error always affects the model fitting and, for this reason, random components are included in the models to account for all the unknown sources of variation (namely, the error term). For more details on kinetic analysis, there are many general reviews and computer programs [4,5,6,7].

Experimental design is about how to obtain the optimum information required for the analysis of the kinetics and to find a model for the data. The features of the design to be considered include the substrate range; what individual substrate concentrations should be measured; the number of data points required; and whether replicates of measurements are useful. The design must ensure that the data provides the necessary information to fit and discriminate between models and obtain good parameter estimates. The importance of experimental design and its role in successful

*Address correspondence to this author at the School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AJ, UK; Tel: +44(01189)317385; Fax: +44(01189)310810; E-mail: m.j.c.crabbe@reading.ac.uk

1573-4064/05 $50.00+.00

analysis is becoming increasingly recognised in pharmaceutical and commercial research [8,9,10].

Classical methods of experimental design [11] are based on knowledge of the experimental statistics and not on the science of the system itself. This can present a problem when more complex kinetics are suspected, as such systems were originally designed only for the simple first-order kinetics. The methodologies are obtained analytically or by simple computation from the likelihood, produce highly variable results owing to their dependence on the initial parameter values chosen.

Bayesian designs are based around prior distributions of the model parameter estimates and their variance, rather than on chosen single-point values. This makes it possible to combine information from different sources and means each design can be tailored to the type of kinetics in question, i.e. trends for simple kinetics are not just assumed for the study of more complex models. The iterative approach of a Bayesian method also increases the efficiency and cost-effectiveness of the experiment. Bayesian methodology is not well developed or established as yet and its use currently involves much knowledge of the mathematics and statistics involved. However, there is 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.

Experimental design and the existence of prior knowledge are issues that have been previously raised in the literature. For example, Cornish-Bowden [4] suggested that, in terms of substrate range, an optimum experimental design should be based around K_M when studying Michaelis Menten enzyme kinetics. In response to this suggestion there have been several attempts to look at the design of Michaelis Menten experiments. In particular, Duggleby and colleagues [12,13] identified the importance of basing the choice of the initial substrate concentration in time-course experiments on the K_M. This design choice minimised the standard error for the Michaelis constant.

Other issues would be 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. There is currently limited evidence in the literature that the design of more complex experiments has been considered. Authors either use established classical methods [14] or quote no defined method of choosing their design. Bayesian design receives little mention.

Bayesian data analysis can be defined as a method of making inferences from data using probability models for quantities we observe and about which we wish to learn. It was Chaloner and Verdinelli [15] that, in a review of experimental design, suggested Bayesian statistics could be applied to experiments and not simply used in their subsequent analysis. Their idea was to use a decision-theoretic approach to specify a utility function reflecting the purpose of the experiment. The best design is that optimising the utility value. Only one example of the use of this proposal can be found in a review [16] of work using a Bayesian utility function to choose sample sizes to confirm the results from a clinical trial measuring protein expression levels in breast cancer subjects.

More recently we developed a set of rules and procedures for Bayesian experimental design for enzyme kinetics, starting with a very rough prior K_m estimate [17]. It was shown that a number of features of the design, including the substrate range and individual concentrations measured greatly affect the success and error in subsequent analysis of the data. A systematic series of Bayesian studies enabled us to identify trends and sets of design rules for both simple and complex kinetics. We review this here, and show that these rules are applicable to a wide variety of biological steady-state processes [18] important in drug discovery.

MATERIALS AND METHODS

Bayesian Utility Functions

The design is regarded as a decision problem and then the design maximising 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 of the parameters and this is what we aimed to minimise.

Utility functions were 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 distributions for the unknown parameters and the variance vector were first written as a variance-covariance matrix. The utility function was then developed from expressions for the variance of the parameters in the kinetic model, incorporating the prior distribution [17}.

The general form of a utility function is:

$$U = 1/[c_1\{V(P_1)\} + c_2\{V(P_2)\}+.....+c_p\{V(P_p)\}] , \qquad (1)$$

where U is the utility value for a model equation with p parameters; $\{V(P_i)\}$is the variance expression derived from the variance-covariance matrix for the model parameter P_i; and, c_i is the weight for each parameter expression (where i = 1,2,…,p), chosen subjectively to reflect the relative importance of estimating that parameter.

An example of the derivation of the utility function for the Michaelis-Menten model equation is outlined below. The simplest example is shown, as the functions are very long and complex for four parameter models. Taking the Michaelis-Menten model:

$$v = \frac{V_{max}[S]}{K_M + [S]} + \varepsilon \; ; \quad \varepsilon \sim N)0, \sigma^2) \qquad (2)$$

where V_{max} and K_M are the unknown parameters; [S] is the substrate concentration and v is the velocity. If the model equation is termed function f, then differentiation of the function with respect to each parameter gives:

$$\frac{df}{dV_{max}} = \frac{[S]}{K_M + [S]} \text{ and } \frac{df}{dK_M} = \frac{-V_{max}[S]}{(K_M + [S])^2}$$

By transposing and multiplying together the variance matrix of the variance expressions for the two parameters and then taking the inverse, the variance-covariance matrix is obtained. Using the expressions taken from this for the variance of each parameter, the utility function is written:

$$U = 1 / (C_1.\{V(V_{max})\} + C_2.\{V(K_M)\}), \quad (3)$$

where:

$$\{V(V_{max})\} = ((\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^4) / ((\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^2)*(\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^4)-(\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^3)^2)); \quad (4)$$

and:

$$\{V(K_M)\} = ((\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^2) / (V_{max}{}^2*((\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^2)*(\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^4)-(\Sigma^n_{j=1}([S_j]^2)/(K_M + [S_j])^3)^2))) \quad (5)$$

C_1 is the weighting constant in front of the expression for the variance of V_{max} and C_2 is that for K_M. C_1 and C_2 are calculated as described from the relative magnitudes of the prior values of the parameters. For example if $K_M = 0.53$mM and $V_{max} = 0.0318$mM/min then:

$C_1 = (0.53/0.0318)^2 = 280$, and C_2 is $(0.53/0.53)^2 = 1$.

It follows that the V_{max} should have the largest weighting in the function as a small error in its estimation would have more of an effect on its value, as it is considerably smaller.

Once the workbook had been constructed in Mathematica with the correctly weighted utility function for the study, the utility value for each experimental design could be calculated. This was achieved by inserting the chosen set of individual substrate concentrations measured in the theoretical experiment and the prior parameter values. Systematic calculations for multiple designs enabled observations to be made about the changes in and optimisation of the utility value. The aim was to identify the key features of the optimum design, including the substrate range and the distribution and choice of the points measured within that range. The actual Utility values calculated for each enzyme system are characteristic of that particular one, that is the values cannot be compared between different examples. The utility values facilitate the identification of the optimum design for that enzyme and kinetic model. The trends and conclusions about the design are the comparable features between examples.

Each utility study was performed to demonstrate the maximisation in utility, minimisation in variance and optimisation of design possible. This is a lengthy method used to search for the experimental design which minimises the variance of each model parameter estimated and it would not be used routinely.

Computational Simulation of Kinetic Data

In order to confirm and test the results of the work with the Utility functions and convert them into terms more specific to general understanding, kinetic data for each model was computationally simulated and fitted for each designed substrate value set. Simulation was achieved using a program written into a Mathematica workbook (Version 4.1, Wolfram Research) [17]. The program was then used to generate a list of random error values from a normal distribution. Adding one error term to each expected velocity provided a set of simulated data to later plot and fit to the kinetic model.

For example, the following is the theory behind the simulation of data for a Michaelis Menten enzyme:

If the number of substrate values being used is termed i and i = 1,….,n; then the model equation can be written:

$$v_i = \frac{V_{max}S_i}{S_i + K_M} + \varepsilon_i \quad (6)$$

The set of substrate values (S_1,….,S_n) in the design and the prior values of V_{max} and K_M are input into the program. The set of expected velocity values, with no error terms added, is calculated.

The error term ε has a normal distribution:

$\varepsilon_i \sim N(0,1)$ with a mean of zero and standard deviation of 1.

A list of n number error values (_1,……..,_n) is then generated from a normal distribution. Each error value must be multiplied by the standard deviation known for the prior V_{max} to correct from the standard deviation of 1. These error values are added to the equation to give a simulated set of data v_1,……,v_n. The end result is a set of v and s values to fit and analyse as if real and estimate V_{max} and K_M.

Modelling of Simulated Data Using Simfit Package

All simulated kinetic data was fitted using a Windows package called Simfit (version 5.40, release 4.005), which is available at http://www.biomed.man.ac.uk/simfit. This is a comprehensive curve-fitting program which can fit a number of functions. There is a large library of compiled models but also a facility to supply the users own mathematical models [17].

RESULTS

Bayesian Studies of the Design of Experiments for an Enzyme with Typical Michaelis Menten Kinetics

For the purpose of this study we used kinetic data for Glyoxalase I (EC 3.1.2.6) [14]. The Michaelis Menten kinetics of this enzyme have been widely characterised and

reported in literature [19]. In this case, therefore, we have accurate prior values for K_M (0.53mM) and Vmax (0.0318 mM/min).

The substrate range used markedly affected the Utility value. Systematic studies showed that there are definite optimum lower and upper range points below and above the K_M which maximise the Utility [1]. For other Michaelis Menten kinetic experiments the trend is the same: the range must extend from half the K_M concentration to one hundred times the K_M concentration. The key point at this stage is that knowledge of K_M facilitates the design of the optimum experiment.

After fitting the data, the standard errors in the K_M and Vmax obtained were used to compare each experimental design. Errors were minimised when 60% of the substrate concentrations measured were below/on the K_M concentration. The error markedly increased if the point distribution was lowered to less than 40% below K_M: the error was then greater than 13% for K_M (over five times higher than the optimum design) [17]. This suggests the importance of enough points below the K_M for its effective estimation. The error was acceptable in the range of 50 to 70% data point distribution, but there was a clear optimum.

Selecting the Total Number of Data Points and Replicates and Positions of Intermediate Substrate Concentrations in the Range to be Measured

We produced simulations using five substrate concentrations measured in triplicate, to provide five mean values for fitting the models. The decision to use this system of points came about from an extensive study of the effect of altering the number of points and replicates. Figure **1** shows the effect of increasing the number of individual data points measured on the error in the fitted values of K_M and Vmax. The error is observed to decrease with increasing data points to fit but reaches a minimum plateau with five points. Increasing the number of individual points further than five will yield no further improvement in reducing the error of the parameter estimates. However, it is possible to reduce the error further by measuring these five points in triplicate. Figure **2** shows that the percentage error in the K_M fitted value can be further reduced from 11.4% to 2.4% by measuring in triplicate and then fitting the mean values. The error is reduced with the use of duplicates but reaches a minimum plateau with the use of triplicates and increasing the number of repetitions any further does not improve the experiment. Therefore, only five points, each measured in triplicate are needed to fit a Michaelis Menten model accurately to a curve and as long as they are well chosen then increasing the number of points will simply be of no benefit and a waste of resources.

So far, the optimum substrate range, the distribution of points across the range and the number of points to measure have been determined in this study. The first measurement is made at the lower point of the range (half the K_M value) and the last at the upper point (one hundred times the K_M). As the optimum design is with 60% of the data points below/on the K_M concentration, this in turn means that there must be three measurements below/on the K_M and two in the upper part of the range. The question posed is: does it matter what substrate concentration the other measurements are made at?

The middle (third) point should be ideally made at the K_M concentration. The error in both fitted parameters is minimised at this concentration and increases either side of it. The choice of the second point (obviously in between the lower range point and the K_M concentration) also affects the outcome of the experiment. The error is minimised when the measurement is made at the concentration of 0.3mM. which is one quarter the distance towards the K_M . Again the errors increase markedly either side of this optimum point choice. The fourth measurement can be made up to 20% below the upper value [17].

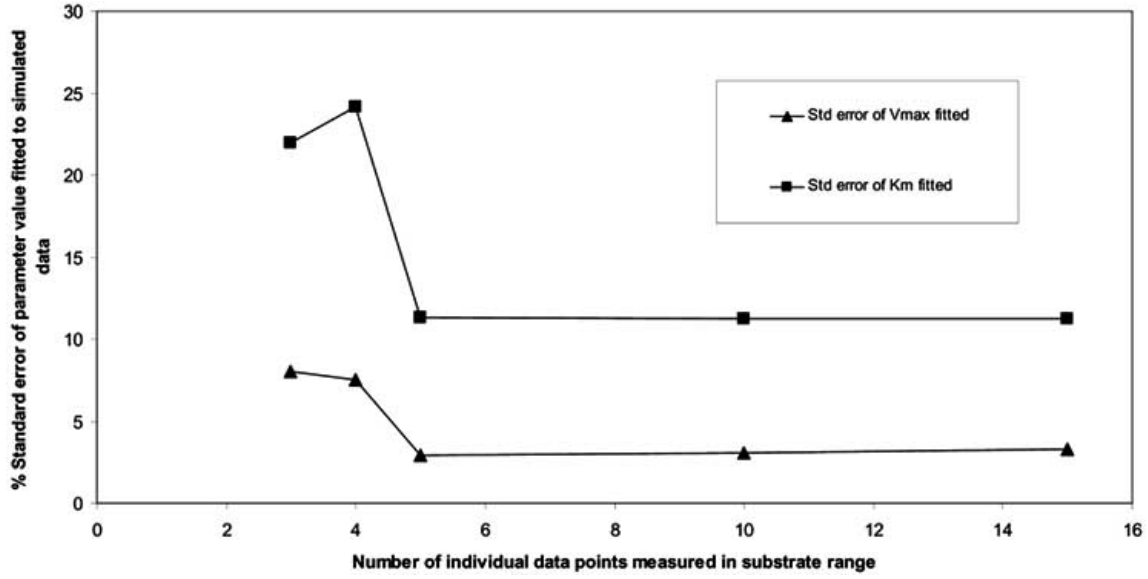

Fig. (1). Glyoxalase I Michaelis Menten kinetics: The effect of increasing the number of individual substrate concentrations measured on the percentage standard error of the fitted values of K_M (■) and V_{max} (▲). Modified from [17].

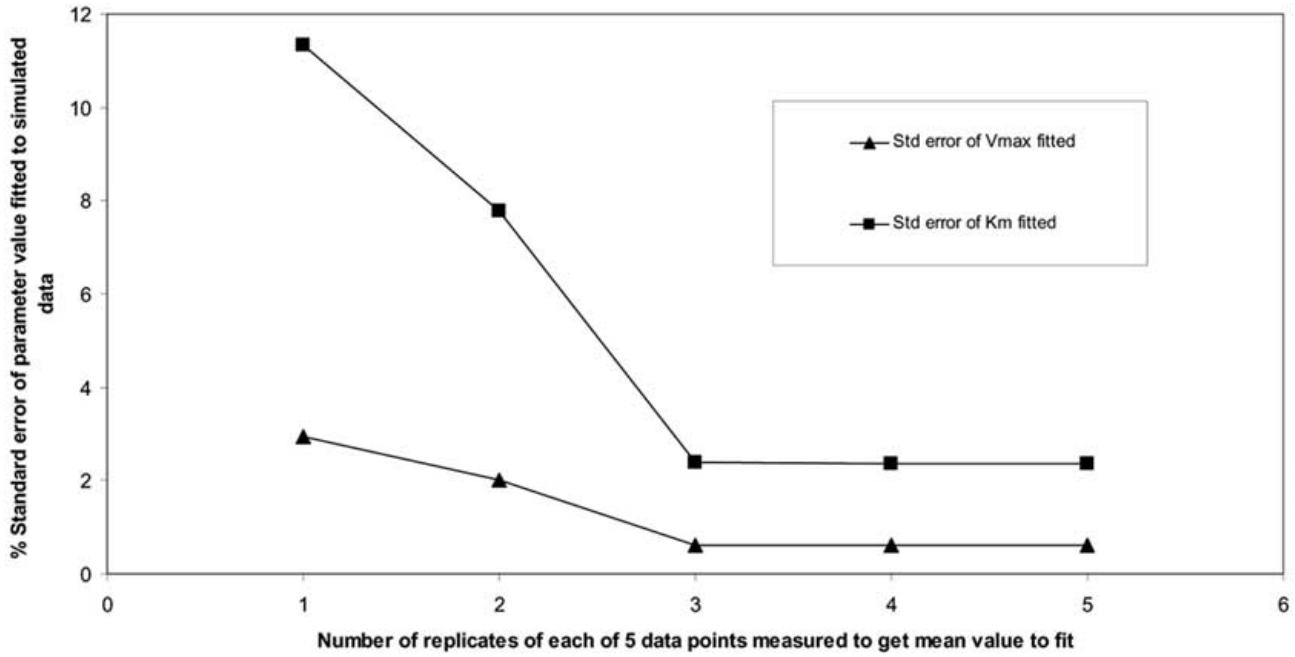

Fig. (2). Glyoxalase I Michaelis Menten kinetics: The effect of measuring replicates of each substrate concentration (then using a mean value for fitting) on the percentage standard error of the fitted values of K_M (■) and V_{max} (▲).Modified from [17].

Overall it seems that the choice of each of the intermediate substrate concentrations to measure is as important as the choice of range and percentage distribution across that range. There is an optimum for each of these data points.

The Prior Information Needed to Obtain an Efficient Design and the Concluded Bayesian Design Rules

In many cases the prior K_M estimate is not likely to be accurate or indeed it may be close to a guess if very little information or knowledge is readily available. In order to make practical use of these rules we must, therefore, pose the question as to whether this single experiment of measurements made in triplicate at five substrate concentrations will be successful whatever the prior estimate of K_M? Obviously one step is likely to be insufficient as we have previously shown that each stage in the design is based on the K_M value and there is a clear optimum for each choice so any significant changes to those will in turn decrease the accuracy in the estimated parameters.

Table **1** shows what happens when the experiment is designed on the basis of different K_M prior estimates (in varying degrees of poorness). We designed each experiment based on the discussed rules and using each K_M prior estimate. We then simulated data for the designed set of substrate concentrations, with the knowledge that the actual K_M is 0.53mM. Simulation was performed for both five concentrations measured in triplicate and five single concentrations. The fitting of this data gave us the new estimates of K_M and their errors. The results show that

Table 1. Michaelis Menten Kinetics: The Effect of Designing an Experiment Using the Discussed Optimum Bayesian Rules but Basing the Design on Varying Degrees of Poor Prior Estimates of K_m (Actual K_m Value is 0.53mM). Modified from [17].

Prior K_m estimate (mM) used to design experiment	Results of fitting data simulated for the optimum Bayesian design for 5 substrate concentrations		Results of fitting data simulated for the optimum Bayesian design for 5 substrate concentrations in triplicate (15 measurements)	
	Fitted K_m value	% standard error of K_m	Fitted K_m value	% standard error of K_m
0.2	0.46	22.2	0.47	12.8
0.4	0.53	23.7	0.56	4.1
0.53 (actual)	0.51	11.4	0.53	2.4
3	0.42	39.6	0.45	21.2
10	0.86	42.5	0.76	32.0

although the fitted values of K_M are not exact and the errors are high if the prior is more than 20% out, they are much more refined than the initial guess/estimate. For example, when the K_M is estimated initially to be 10mM (20 times greater than the actual), and the experiment designed with respect to that value, then the new estimate is 0.76mM which is only 1.5 times out. The error does, however, remain very high. This suggests that the Bayesian design is good enough to refine even a very poor estimate, i.e. the rules still work.

Measuring five single points is enough to refine the estimate. The use of triplicate measurements as predicted is useful in reducing the error but does not actually affect the parameter estimate. Therefore, we propose that a five-point experiment designed on the basis of the very rough prior is enough to gain a better estimate of K_M. Then a repeat five-point experiment is necessary to further refine or check the new estimate of K_M. Once these two short initial experiments have been carried out, the new value of K_M can be taken and the experiment redesigned and the five points measured in triplicate to obtain final estimates and their errors. The final step is to present the parameter estimates with the minimum error. Table **2** shows the design rules for simple, two-parameter models.

Bayesian Studies of the Design of Experiments for Enzymes with Complex, Multiple Parameter Equation Kinetics

Using the same approach as described in the previous section, we conducted a detailed Bayesian investigation to identify the key points in designing these complex experiments [17]. One important conclusion is that our experimental design can allow differentiation between the fittings of different four parameter complex models. Table **3** shows the design rules for complex four-parameter models.

Cell Transport Kinetics: Using a Bayesian Experimental Design to Improve the Accuracy and Efficiency of Model Parameter Estimation

Kinetic data for the transport of the radiolabelled amino acid 3-[^{123}I]iodo-L-α-methyl tyrosine ([^{123}I]IMT) in human GOS3 glioma cells [19] was used. The use of [^{123}I]IMT is a promising tool for the diagnosis and monitoring of brain tumours (both in the non-invasive grading of gliomas and the delineation of tumour extent) using single-photon emission tomography (SPECT). Critically, it has been established that there is a specific transport of [^{123}I]IMT across the blood-brain barrier and it is metabolically stable and not incorporated into proteins. However, in order to optimise its medical use, more work is needed to obtain precise kinetic details of its uptake in human glioma cells. Such details aid in the identification of the cell systems that mediate the transport and any differences between cell lines and types. This knowledge is essential to aid further *in vitro* studies and help identify suitable model cells.

The cell transport kinetics of [^{123}I]IMT are consistent with Michaelis-Menten type kinetics:

$$v = \frac{V_{max}[S]}{K_m + [S]} \quad (7)$$

where v is the transport velocity of [^{123}I]IMT; [S] is the concentration of [^{123}I]IMT; V_{max} is the maximum transport velocity; and K_m is the apparent Michaelis Menten constant. The latter two parameters are the unknown kinetic parameters to be estimated. The uptake of [^{123}I]IMT (in terms of its initial transport rate) was measured over a range of [^{123}I]IMT concentrations from 2.5 to 50_M. Eight experimental points were performed in triplicate and each

Table 2. Design Rules for Michaelis-Menten Type (2 Parameter) Kinetics. Modified from [18].

Experiment design decision	Bayesian rule
1) Start with a prior estimate of K_m.	Use prior information for an estimate or a rough guess.
2) Choose the substrate range for measurements.	The range should extend from half the K_m concentration up to 100 times it.
3) Choose the total number of data points to measure.	Measure at five different substrate concentrations in triplicate (the 1st concentration is the lower point of the range and the 5th is the upper point).
4) Choose the middle (3rd) concentration to measure at.	This measurement should be made at the K_m concentration.
5) Choose the 2nd concentration to measure at.	This concentration should be the value which is _ the distance towards the K_m concentration from the first point (the lower range point).
6) Choose the 4th concentration to measure at.	This measurement should be made at the concentration which is 20% below the upper point.
7) Perform the experiment to obtain data. Fit the two-parameter model to the kinetic data (plotting the mean of each of the data points measured in triplicate) to obtain parameter estimates and errors.	N.B. Take an iterative approach if the prior estimate of K_m that the design is based on is known to be poor. That is, carry out the experiment at five concentrations only (i.e. no triplicates) and fit to obtain a better estimate of K_m. Redesign based on new K_m estimate and perform a second five-point experiment. Redesign again and for the third experiment perform the experiment for the five points in triplicate to obtain final low variance parameter estimates.

Table 3. Design Rules for Complex Kinetics (Four Parameters). Modified from [18].

Experiment design decision	Bayesian rule
1) Start with a prior estimate of the ratio of the two K_m values (if more than one).	Use prior information for an estimate or a rough guess.
2) Choose the substrate range for measurements.	The range should extend from 10 times below the K_m ratio concentration up to 100 times above it.
3) Choose the total number of data points to measure.	Measure at twenty-five different substrate concentrations (no replicates) (the 1st concentration is the lower point of the range and the 25th is the upper point).
4) Choose the middle (13th) concentration to measure at.	This measurement should be made at the K_m ratio concentration.
5) Choose the 2nd concentration to measure at.	This concentration should be the value which is _ the distance towards the K_m ratio concentration from the first point (the lower range point).
6) Choose the 12th concentration (point below the K_m ratio) to measure at.	This measurement should be made at the concentration which is 10% below the K_m ratio value.
7) Choose the 14th concentration (point above the K_m concentration)	This measurement should be made at the concentration which is five times the K_m ratio concentration.
8) Choose the 24th concentration (point below the upper range point).	This measurement should be made at the concentration which is 20% below the upper range point.
9) Choose the intermediate points – points 3 to 11 and 15 to 23.	These should be as evenly spaced as possible between the pre-decided points.
10) Perform the experiment to obtain data. Fit the four-parameter model to the kinetic data to obtain parameter estimates and errors.	N.B. Take an iterative approach if the prior estimate of K_m that the design is based on is known, or thought, to be poor. That is, carry out the experiment at ten concentrations only (still selecting the individual points according to their position within the ten points) and fit to obtain a better estimate of K_m. Redesign based on the new K_m estimate and perform a second ten-point experiment. Redesign again and for the third experiment perform the experiment for the twenty-five points to obtain final low variance parameter estimates.

experiment was carried out at least twice with similar results. After fitting the data a K_m of 20.1 ± 1.5_M and a maximum transport velocity (V_{max}) of 34.8 ± 1.9 nmol/mg protein/ 10 min were calculated.

It is possible to improve this experimental design using an iterative and systematic Bayesian approach. That is, following our rules for designing the optimum experiment for a system with two-parameter equation kinetics, it is possible to make marked gains in terms of the efficiency and the accuracy of the parameters estimated.

With a prior estimate of K_m of 20.1μM, the concentration range specified by the rules is from 10 to 2000μM with measurements taken at five different concentrations in triplicate. Our Bayesian design offered a marked improvement [18], and the use of this design will result in more effective parameter estimation. When designing by the Bayesian rules the standard errors in the parameter estimates are considerably less than if obtained via a 'classical' route.

Bayesian Experimental Design is a Generic Approach: Examples of its Applicability to the Study of Continuous Culture, Drug Transport and Receptor Binding Kinetics

In addition to the study of enzyme and cell transport kinetics, the specified Bayesian rules and method of design are applicable to the study of a number of other important kinetic systems. Continuous culture, drug transport and receptor binding kinetics are three further examples.

The analysis of equilibrium data for receptor-ligand binding helps to identify the mechanism and specificity of the interaction of a receptor with a ligand. Such knowledge is also essential in the search for inhibitors for medical or research applications. In our final example of a kinetic system, we used an experiment looking at the binding of the ligand 10,11-[^{3}H]dihydroxy-N-n-propylnorapomorphine ([^{3}H]NPA) to the D2-dopamine receptor[¶]. The kinetics were consistent with second-order Michaelis-Menten type kinetics. This means they were described by a four-parameter equation. The total binding was measured for a range of [^{3}H]NPA concentrations. The original design used in this study was to measure binding at twelve [^{3}H]NPA concentrations in triplicate over a concentration range of 0.006 to 15.6 pM. The Bayesian design, suggested by the rules for a four-parameter equation and based on the prior K_{d1}/K_{d2} ratio of 0.06 (the ratio of the two equilibrium constants), was twenty five concentrations measured once over the range of 0.0058 to 5.770 pM. The

[¶] data provided by p.g.strange@reading.ac.uk

individual concentrations measured within that range were chosen according to the rules. The utility calculated for the Bayesian design is nearly two times higher. The percentage standard errors for the parameters estimated by fitting the data obtained by simulating the Bayesian design are also considerably lower than for the original. Again a higher degree of accuracy was obtained using the Bayesian method of design.

DISCUSSION

In the first instance, we used an example of a Michaelis Menten enzyme to demonstrate the findings of this Bayesian study. These results were both easy to understand and served as an initial base about which to begin observing trends and methods in optimising the designs. We then showed that these general trends are carried through to help design experiments for enzymes with more complex kinetics and equation parameters. Finally, we applied our new experimental design processes to drug transport, receptor binding, microbial culture and cell transport kinetics. The overall aim of this work was to obtain a set of rules for designing kinetic experiments with some very rough prior knowledge. In applying the lengthy statistical work here, the rules generated enabled the collection of data suitable for accurate analysis and modelling without the problem of insufficient and/or poor data sets.

The importance of basing a design for enzyme on the K_M became apparent when looking at the other choices to be made about the data points used within the chosen substrate range. Firstly, a trend emerged relating to the percentage of total data points measured that are below and above the K_M. The Utility value rose to a peak at an optimum choice of 60% points below (or on) the K_M, and falls off either side of this but more markedly when reduced to less than 60% [17]. That is the optimum data point distribution is when sixty percent of the data points fall below or on the K_M substrate concentration value but above half the K_M(the starting point of the range). This was demonstrated for a fifteen data point design of five points measured in triplicate.

In all these areas, whether the kinetics are simple or complex, Bayesian design rules (Tables **2** and **3**) produced significantly more cost-effective parameters with lower error values, resulting from fewer experiments [18].

We have shown that our Bayesian rules for the systematic design of kinetic experiments can be successfully applied to a number of different systems. The rapid evaluation of kinetics proves critical in many processes, and the effective utilisation of this method increases the efficiency of these experiments. The optimum data is collected following the rules and the iterative approach provides a more cost-effective route to accurate parameter estimates, as fewer data points are needed.

This method is timely to aid the design of the increasingly complex kinetic experiments in rapidly advancing research fields and the rules logically use existing data and information. The concept is easily transferred to any area of study requiring the analysis of steady-state kinetics.

The use of utility functions and the generation and fitting of simulated data is too lengthy to realistically use as a direct design method. Therefore, our final method of design needs only a very rough prior estimate of the K_M (or the ratio of K_M values) and knowledge of the number of parameters in the kinetic model equation that the data is to be fitted to. Based on this prior information and the rules, the following features of the experimental design can be chosen: the substrate range; the total number of substrate concentrations to take measurements at; the distribution of data points across the range; and, the choice of individual important points in that range. In addition, as the number of measurements required is based on the number of model parameters to be fitted rather than the actual equation, it is possible to design and also differentiate through fitting between kinetic models with the same number of parameters.

It is both clear that prior information is available in existing data and logical that it should be used to aid the design of the increasingly complex experiments required to advance drug design, clinical diagnosis and biotechnology applications. The use of prior knowledge and a Bayesian approach in this study has uncovered methods of design and showed that a careful approach can greatly improve the accuracy of the experiment. Together with further study it should be possible to pool information in the form of a database and eventually design programs to design experiments. In this way statistical and biological information can be pooled to provide a user-friendly approach to experimental design.

ACKNOWLEDGEMENTS

We would like to thank the BBSRC for funding of this project, and Professor P. Strange and Dr. D. Roberts, University of Reading, for providing receptor-ligand equilibrium data.

REFERENCES

[1] Murphy E.F.; Gilmour S.G.; Crabbe M.J.C. *Drug Discovery Today*, **2002**, *7*, (Suppl.), S187-S191.
[2] Briggs G.E.; Haldane J.B.S. *Biochem. J*., **1925**, *19*, 338-339.
[3] Rudolph F.B.; Fromm H.J. In: *Contemporary enzyme kinetics and mechanism*, Purich D.L. Ed.; Academic Press, San Diego, USA, **1996**, pp.89-109.
[4] Cornish-Bowden A. *Fundamentals of Enzyme Kinetics*. Butterworth and Co. Ltd., London, **1979**.
[5] Bardsley W.G.; Wood, R.M.W.; Melikhova, E.M. *Computers and Chemistry*, **1996**, *20*, 145-157.
[6] Crabbe M.J.C. In: *Microcomputers in Biology: A practical approach*; C.R.Ireland and S.P.Long, Eds.; Oxford University Press, **1984**, pp.107-150.
[7] Wong J.T. *Kinetics of enzyme mechanisms*. Academic Press, London, **1975**.
[8] Ye C.; Liu J.; Ren F.; Okafo N. *J. Pharm. Biomed. Anal*., **2000**, *23*, 581-589.
[9] Lendrem D.; Owen M.; Godbert S. *Organic Process Research and Development*, **2001**, *5*, 324-327.
[10] Owen M.R.; Luscombe C.; Lai-Wah L.; Godbert S.; Crookes D.L.; Embiata-Smith D. *Organic Process Research and Development*., **2001**, *5*, 308-323.
[11] Heath D. An introduction to experimental design and statistics for biology. UCL Press, London, **1995**.
[12] Duggleby R.G.; Clarke R.B. *Biochimica et Biophysica Acta*., **1991**, *1080*, 231-236.

[13] Gouder C.T.; Sonnad J.R.; Duggleby R.G. *Biochimica et Biophysica Acta*., **1999**, *1429*, 377-383.
[14] Martins A.M.; Mendes P.; Codeiro C.; Freire A.P. *Eur.J.Biochem.*, **2001**, *268*, 3930-3936.
[15] Chaloner K.; Verdinelli I. *Stat.Sci.* **1995**, *10*, 273-304.
[16] Clyde, M.A. *Int. Encyc. Social and Behavioural Sciences*, **2001**, 8, 5075-5081.
[17] Murphy, E.F.; Gilmour, S.G; Crabbe, M.J.C. *J. Biochem. Biophys. Methods* **2003**, *55*, 155-178.
[18] Murphy, E.F.; Gilmour, S.G; Crabbe, M.J.C. *FEBS Lett.* **2004**, *556*, 193-198.
[19] Riemann, B.; Kopka, K.; Stogbauer, F.; Halfter, H.; Ketteler, S.; Vu Phan, T.Q.; Franzius, C.; Weckesser, M.; Bernd Ringelstein, E.; Schober, O. *Nuclear Medicine and Biology*, **2001**, *28*, 293-297.

Received: 26 May, 2004 Accepted: 27 August, 2004