

Review

Experimental designs for the Adair model



J. López-Fidalgo*, M.M. Rodríguez-Hernández

Department of Mathematics, University of Castilla-La Mancha, Institute of Mathematics Applied to Science and Engineering, Avda. Camilo José Cela 3, 13071 Ciudad Real, Spain

ARTICLE INFO

Article history:

Received 31 March 2014

Received in revised form 12 July 2014

Accepted 18 July 2014

Available online 30 July 2014

Keywords:

Adair equation

Free ligand

Macro-molecule reactions

Optimal experimental design

Saturation rate

ABSTRACT

The Adair equation is used to model biological macro-molecule reactions. This equation relates the saturation rate to the free ligand concentration. But, the latter is not a variable completely under the control of the experimenter. The ligand is a random variable depending on an initial ligand added by the experimenter, which can be designed, but the dependence of the saturation rate on the initial ligand has not been considered in the literature. In this paper a transformed model based on the Adair model of first order (monomer) is derived in order to obtain a proper model that depends on the initial ligand. This model will allow proper fitting and optimal designs using the initial ligand. It will be called the transformed Adair model (TAM). Optimal designs as well as seven-point quasi-optimal designs forced to follow a harmonic, geometric or uniform progression, are computed. The parameters are estimated and compared for simulated data from these designs. A sensitivity analysis against the choice of nominal values of the parameters is also performed for the TAM. The analytic version of the transformed model is only possible for the first class model. But the good efficiencies of the optimal designs obtained directly from the monomer model for fitting the TAM justify doing something similar for the second order model (dimer). Designs were computed numerically in this case.

© 2014 Elsevier B.V. All rights reserved.

Contents

1. Introduction	133
2. Deriving the Adair model	134
2.1. Optimal design background	134
2.2. Outline of the paper	135
3. Optimal design for the monomer model	135
3.1. Optimal design for a transformed monomer model	136
3.2. Suboptimal sequence designs	137
3.3. Simulations to compare estimates and designs for both models	137
3.3.1. Example	137
3.4. Robustness with respect to the choice of the nominal value of the parameter	138
4. Optimal designs for the dimer model	138
5. Discussion	140
Conflict of interest	140
Acknowledgments	140
References	141

1. Introduction

The Adair model is used in chemical reactions once the balance between a biological macro-molecule and a ligand is reached. A ligand is usually a smaller molecule and sometimes another macro-molecule.

These types of reactions are usually reversible. Adair [1] proposed a sequential equation to model how the oxygen molecules are bound to subunits of hemoglobin. The hemoglobin is a quaternary structure protein, whose main function is the transport of oxygen. The sigmoidal shape of the curve describing the union of the hemoglobin to the oxygen means that it has a relatively low affinity to capture the first molecule. The increasing slope means that the binding of the first molecule of oxygen facilitates the binding of another oxygen molecule to the second subunit and so on. This is known as binding cooperation. The

* Corresponding author. Tel.: +34 926 295212; fax: +34 926 295361.

E-mail addresses: jesus.lopezfidalgo@uclm.es (J. López-Fidalgo), MMercedes.Rodriguez@uclm.es (M.M. Rodríguez-Hernández).

hemoglobin binds a fourth oxygen molecule with an affinity 100 times greater than the first one. Obtaining the balance constants is a fundamental issue (see e.g. [12]). The known antigen binding, such as viral antigens, bacterial antigens, and drugs, is of common use in biomedical practice through immunoassays. Tetin and Hazlett [15] considered the interaction antibody–ligand and the determination of the balance binding constant with different models, including the Adair model. The protein–ligand models are remarkably analogous to metal–ligand complexation models that are dependent on the initial ligand concentration [12]. However, the positive cooperativity behavior is unique to these biological systems.

In this paper, optimal experimental designs are computed in order to better fit the Adair model. The theory of optimal experimental designs has been developed considerably, especially after finding the celebrated equivalence theorem [9]. This theorem gives a powerful tool to check whether a particular design is D-optimal (equivalently G-optimal). Whittle [16] generalized this to a more general class of optimality criteria. The General Equivalence Theorem (GET) provides also a tool to construct optimal designs using different algorithms [5,17] for the popular Wynn–Fedorov algorithm. A suitable choice of the experimental conditions can improve inferences on the model. The Adair equation properly describes a particular microbiological phenomenon, in which a random error with zero mean needs to be added to condense the uncertainty. The variance of an observation (or the error) may be constant or dependent on the experimental conditions through a mathematical model, e.g. depending on the mean model.

2. Deriving the Adair model

Let Y be the saturation rate of a macro-molecule. This random variable is decomposed into a form depending on the concentration of the free ligand, $[L]$, plus a random error.

A typical experiment to describe the chemical process consists of introducing a solution of concentration of a macro-molecule or a protein, $[P_0]$, into a semi-impermeable dialysis bag. Then, the bag is put in a recipient with initial ligand concentration, $[L_0]$. Fig. 1 (left) shows the initial situation for the monomer case with just one binding site.

Then, the ligand molecules start penetrating the bag but the protein molecules cannot get out. The equilibrium is reached once there is equal free ligand concentration in and out of the bag, say $[L]$. Fig. 1 (right) displays this situation for the monomer. In order to measure this concentration the bag is withdrawn and the free ligand within the container is measured. If the protein has n binding sites and the ligand can bind the protein through 0, 1, 2, ..., n sites, the concentration of protein bound to the ligand will be denoted by $[M_n^i]$, with $i = 1, 2, \dots, n$ the number of sites of the protein bound to the ligand. Thus, in the interior of the bag there is the free ligand, $[L]$, the bound ligand (to the protein or macro-molecule), $[M_n^i]$, $i = 1, 2, \dots, n$, and the unbound protein, $[M_n^0]$.

Let r be the observed average number of busy sites, that is, the number of moles of the bound ligand per mol of protein, $0 < r < n$. The

saturation rate is defined as $Y = r/n \in [0, 1]$. The mean value of the busy sites is

$$r = \frac{[\text{Bound ligand}]}{[\text{Total of macro-molecule}]} = \frac{[M_n^1] + 2[M_n^2] + \dots + n[M_n^n]}{[M_n^0] + [M_n^1] + [M_n^2] + \dots + [M_n^n]} = \frac{K_1[L] + 2K_1K_2[L]^2 + \dots + nK_1\dots K_n[L]^n}{1 + K_1[L] + K_1K_2[L]^2 + \dots + K_1\dots K_n[L]^n},$$

where the asymptotic value of r is the average number of bound sites in each macro-molecule divided by the number of macro-molecule binding sites, n . Thus, the model can be formulated as

$$Y = \frac{K_1[L] + 2K_1K_2[L]^2 + \dots + nK_1\dots K_n[L]^n}{n(1 + K_1[L] + K_1K_2[L]^2 + \dots + K_1\dots K_n[L]^n)} + \varepsilon. \quad (1)$$

The parameters may be estimated using Least Squares Estimates (LSE), Maximum Likelihood Estimates (MLE) or by doing some transformation to linearize the model for the monomer. For more details see [6]. Tetin and Hazlett [15] stressed that either these linear transformations or the deletion of experimental data introduce uncontrolled bias in the estimation process. In order to use these linear equations they suggested using weighting optimization. Throughout this paper the MLE will be used assuming normality and so they are LSE. No linearization by transforming the equation will be considered in this paper.

Monod's equation is a classic microbiological model describing the kinetics of batch microbial growth. Another typical model used in this field to describe the kinetics of fast equilibrium of enzymatic systems and the analysis of data from drugs, neurotransmitters and assays with hormonal receptors is the popular Michaelis–Menten model [11]. The Monod equation has the same form as the Michaelis–Menten equation, but the Monod equation is empirical while the latter is based on theoretical considerations. López-Fidalgo et al. [10] computed optimal designs for population growth models.

The aim of this paper is to compute and compare suitable experimental designs for the Adair model for the saturation ratio with macro-molecules with one and two binding sites (monomer and dimer). An optimal design will provide both, good estimates and savings in the use of experimental resources [5]. The theory of optimal experimental design provides tools for computing good estimates from different points of view. One of the criticisms made to this theory is that optimal designs frequently claim for extreme points not allowing the detection of some features of the data as is the case of possible curvature. This is the reason why in this paper a greater number of different points in the design are forced through regular sequences as classes of harmonic, geometric or uniform progressions.

2.1. Optimal design background

For nominal values of the parameters locally optimal designs will be obtained [3] maximizing the determinant of the Fisher Information Matrix (FIM). For the monomer the information matrix is a scalar and an analytic expression is available for the one-point optimal design. For a dimer two parameters are in the model and therefore two or three points will be needed. For this case the equivalence theorem will be used to compute optimal designs.

An (exact) experimental design is a collection of points $[L]_1, \dots, [L]_N$ from a design space χ . Some of them may be repeated, which means several experiments are realized for the same experimental condition. The total number of observations is N and this number is limited by cost, time or feasibility restrictions. A good experimental design may reduce this number, and so as cost and time, preserving a good estimation of the model. Taking into account that possible replicates of some points are possible a probability measure may be defined identifying the

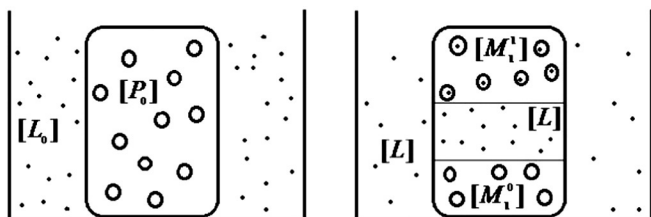


Fig. 1. The initial phase of the experiment (left) and the experiment at equilibrium (right) for the monomer.

probability of one point as the proportion of times it appears in the sample, say p_i , $i = 1, \dots, k$, for k different points in the sample. Using this idea Kiefer [8] introduced the concept of approximate design, ξ , as any probability measure defined on a design space \mathcal{X} , where the design points may be chosen. From a practical point of view, just designs with a finite support are feasible. Caratheodory's theorem states that for any FIM from an exponential family distribution there is always an associated design with no more than $m(m + 1)/2 + 1$ points in its support, where m is the number of parameters to be estimated in the model. Thus, there is always an optimal design with this characteristic. Then a realizable design has to be "approximated" as $N_i \approx N \times p_i$ replicates for the i -th experiment. Obviously, the greater is the sample size, N , the better are the estimates and inferences. Imhof et al. [7] showed for a particular optimality criterion and polynomial regression that exact, instead of approximate, designs need to be used for small sample sizes.

A general approximate design may be denoted as

$$\xi = \left\{ \begin{matrix} [L]_1 & \dots & [L]_k \\ p_1 & \dots & p_k \end{matrix} \right\},$$

where $\xi([L]_i) = p_i$ is the mass probability function defining the design measure.

The support of a design ξ is defined in the usual way

$$S_\xi = \{[L]_i \in \mathcal{X} : \xi([L]_i) > 0\},$$

and the information matrix associated to an approximate design ξ is

$$M(\xi; \theta) = \sum_{[L]_i \in \mathcal{X}} f([L]_i; \theta) f^T([L]_i; \theta) \xi([L]_i),$$

where

$$f([L]; \theta) = \frac{\partial \eta([L]; \theta)}{\partial \theta}$$

for a model $Y = \eta([L]; \theta) + \varepsilon$, where ε has zero mean and constant variance.

This is the FIM assuming normality of the observations,

$$M(\xi; \theta) = -E_\xi E_y \left[\frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \log l(y; x, \theta) \right] = E_\xi E_y \left[\frac{\partial \log l}{\partial \alpha_i} \frac{\partial \log l}{\partial \alpha_j} \right],$$

where l is the likelihood function and E_y , and E_ξ stands for the expectation with respect to the probability distribution of y and the probability measure defined by the design ξ . For non-linear models, this matrix depends on the parameters.

D-optimality is one of the most popular optimality criteria. A D-optimal design maximizes the determinant of the Information Matrix, $\max_\xi \det[M(\xi; \theta_0)]$. The statistical meaning is that it minimizes the volume of the confidence ellipsoid of the parameters of the model. The equivalence theorem establishes that this is also equivalent to minimizing the largest variance of the predictions on the design space if approximate designs are considered. In order to check whether a particular design, ξ^* , is D-optimal, the following inequality has to be verified,

$$f^T([L]; \theta) M^{-1}(\xi^*; \theta) f([L]; \theta) \leq m. \tag{2}$$

Moreover, the equality is reached at the design points.

The D-efficiency of a particular design ξ is defined as

$$\text{Eff}_D(\xi) = \left(\frac{\det M(\xi; \theta)}{\det M(\xi^*; \theta)} \right)^{1/m}.$$

This efficiency is always between 0 and 1 and it has an obvious mathematical meaning, the nearer to 1 the better is the design. But it

has also a statistical meaning, e.g. 60% efficiency means a save of 40% of the experiments if the optimal design is used.

For more details about optimal experimental design theory see e.g. [5,2].

2.2. Outline of the paper

In Section 3 optimal designs are computed for the Adair model (3). This would need to control the free ligand concentration, $[L]$. But this is a random variable, that is, for the same initial ligand concentration, $[L_0]$, different values of $[L]$ will be obtained if the experiment is replicated. The initial ligand, $[L_0]$, is what can be designed. An analytic relationship between $[L_0]$ and $[L]$ can be obtained for any Adair model. Using this relationship, the original model relating $[L]$ to the response can be transformed to a model relating $[L_0]$ to the response. This is the so called modified Adair model, which is more appropriate both for estimating the parameters and for designing the experiment (8). Unfortunately, only in the most simple model with one parameter a simple expression can be obtained. Then, optimal designs are computed for $[L_0]$ for that model. This will be made in Section 3.1. These designs have only one point in their support, which is not convenient from a practical point of view. Then, in Section 3.2 sub-optimal uniform, harmonic and geometric designs will be computed in order to provide designs with more than one point. These designs have more than one point and still have good efficiencies. Simulations will be performed in Section 3.3 in a way close to reality with all the designs obtained previously. Then empirical properties of the MLE for the different designs are checked. In Section 3.4 relative efficiencies for the different designs will be computed in order to study the robustness against the choice of the parameter of the transformed Adair model. The results obtained for the monomer are a reference for dealing with the Adair model with two or more parameters, where the analytic expression of a modified model is not available. In Section 4 numerical computations are made for obtaining optimal designs for the dimer model. Finally, in Section 5 a discussion is written about the results and possible future research.

3. Optimal design for the monomer model

In this section, ligand binding to a single specific site ($n = 1$) in the macro-molecule is considered. Fig. 1 shows this particular situation. The general model for any value of n was derived in Section 2. Therefore, the Adair equation for a single binding site, $n = 1$, is obtained from Eq. (1),

$$Y = \eta([L]; K) + \varepsilon = \frac{K[L]}{1 + K[L]} + \varepsilon, \tag{3}$$

where $\theta = K$ is the only parameter to be estimated and ε is a Gaussian random error with an expectation value of 0 and a constant variance. It is a non-linear model on the parameter K and the explanatory variable, $[L]$, is not completely under the control of the experimenter. Thus, another error should be added to this measurement. This fact is usually neglected when fitting the model. In this paper, the impact of doing so is considered, both for estimating the parameter and for designing the experiment. Obviously, the larger is the error, variance of $[L]$, the larger is the impact of neglecting it when fitting the model. This error can be adjusted using historical data or a mathematical deduction. If the interest is in prediction, the error in $[L]$ leads to bias in \hat{K} , but such \hat{K} is still effective together with the error present in $[L]$ to predict future responses y . As mentioned above, in this model estimating the balance constants is a fundamental issue. Therefore, considering $[L]$ as a random variable may matter here.

Actually, just the initial ligand introduced into the recipient, $[L_0]$, at the beginning of the experiment can be controlled. Then $[L]$ is only measured after the experiment is realized and may be a different quantity in two experiments with the same initial $[L_0]$. This means that the typical estimation of the parameter made through Least Square Estimates

(LSE) may not be adequate here. In the present paper this fact will be checked and optimal designs will be computed for the traditional estimation made with this model and for a proper estimation taking into account $[L]$ as a random variable. In both cases the FIM will depend on the parameter K . In particular, for the monomer case the model can be transformed analytically as a function of $[L_0]$. This will not be the case for $n > 1$.

In order to find the FIM for the Adair model the following derivative will be needed,

$$f([L]; K) = \frac{\partial \eta([L]; K)}{\partial K} = \frac{[L]}{(1 + K[L])^2}.$$

Since there is only one parameter, the information matrix, which is scalar in this case, is the square $f^2([L]; K)$. The optimal design to estimate the parameter will be a single point

$$[L]^* = \arg \max_{[L]} f([L]; K) = \arg \max_{[L]} \frac{[L]}{(1 + K[L])^2} = \frac{1}{K}.$$

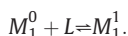
Thus, the best estimate of K is reached by using a unique value of the free ligand at $[L]^*$. There are two practical problems with this design. On the one hand more points are convenient to check the chance of using a more complex model. On the other hand, the experimenter cannot get this exact value of $[L]^*$, he or she can only fix a value of the initial ligand concentration $[L_0]^*$ to obtain approximately the value $[L]^*$. In any case, this design may be used as a reference to measure how good a particular design is. With the classical notation in this area,

$$\xi_{[L]}^* = \left\{ \frac{1}{K} \right\}.$$

Notice that $\eta(1/K; K) = \frac{1}{2}$, meaning that at the optimal point $1/K$ half of the maximum saturation rate is reached.

3.1. Optimal design for a transformed monomer model

There exists a relationship between $[L]$ and $[L_0]$, that allows us to transform the optimal design $[L]^*$ to a design $[L_0]^*$, which can be controlled experimentally. In this simple case a reversible reaction takes place,



When the equilibrium is reached, part of the initial ligand is bound to some parts of the macro-molecule, $[M_1^1]$. The rest of the concentration of the macro-molecule is not bound, $[M_1^0]$, and $[L]$ is then the concentration of the free ligand in the bag. Thus, the, so called, association constant is

$$K = \frac{[M_1^1]}{[M_1^0][L]}. \quad (4)$$

Moreover, the initial protein concentration, $[P_0]$, will be the sum of the free protein concentration, $[M_1^0]$, plus the bound protein concentration, $[M_1^1]$,

$$[P_0] = [M_1^0] + [M_1^1]. \quad (5)$$

An initial ligand concentration, $[L_0]$, is introduced in the recipient at the beginning of the experiment. After the chemical equilibrium is reached then (Fig. 1)

$$[L_0] = 2[L] + [M_1^1]. \quad (6)$$

Specifically, using Eqs. (4), (5) and (6), then

$$[L_0] = 2[L] + \frac{K[L][P_0]}{1 + K[L]}. \quad (7)$$

Plugging $[L]^* = 1/K$ into Eq. (7), the transformed optimal concentration of the initial ligand is $[L_0]^* = \frac{2}{K} + \frac{[P_0]}{2}$.

Thus, the quasi-optimal transformed design for the Adair model is

$$\xi_{[L_0]^*}^* = \left\{ \frac{2}{K} + \frac{[P_0]}{2} \right\},$$

which can be used in practice. Then the experiment is replicated at this particular value N times. Due to the uncertainty in relation $[L]$ to $[L_0]$, N different free ligand concentrations will be obtained. The advantage of doing so is that there is now a space for checking a different model with more parameters.

Using Eq. (7), the Adair equation can be transformed into a new equation that depends on the initial ligand instead of the free ligand, which cannot be controlled exactly by the experimenter. Then an optimal design will be obtained directly for the initial concentration of the ligand. In particular, two solutions are obtained from Eq. (7). One is always negative and the positive one is

$$[L] = \frac{1}{2} \left(\frac{-1}{K} + \frac{[L_0]}{2} - \frac{[P_0]}{2} + \sqrt{\frac{8K[P_0] + (2 + K[L_0] - K[P_0])^2}{2K}} \right).$$

Plugging it into model (3),

$$Y = \mu([L_0]; K) + \varepsilon \quad (8)$$

$$= \frac{2 + K([L_0] + [P_0]) - \sqrt{8K[P_0] + (2 + K[L_0] - K[P_0])^2}}{2K[P_0]} + \varepsilon. \quad (9)$$

It is again a non-linear model on the parameter $\theta = K$ to be estimated, now in function of a controllable variable, $[L_0]$.

Let

$$f_{mod}([L_0]; K) = \frac{\partial \mu([L_0]; K)}{\partial K} = \frac{2 + K[P_0] + KL_0 - \sqrt{8K[P_0] + (2 + K[L_0] - K[P_0])^2}}{K^2[P_0] \sqrt{8K[P_0] + (2 + K[L_0] - K[P_0])^2}}.$$

The optimal ligand initial concentration is then

$$[L_0]^{*mod} = \arg \max_{[L_0]} \frac{\partial f_{mod}([L_0]; K)}{\partial [L_0]} = \frac{2 + K[P_0]}{K} = \frac{2}{K} + [P_0].$$

With the classical notation,

$$\xi_{[L_0]^*}^{*mod} = \left\{ \frac{2 + K[P_0]}{K} \right\}.$$

Notice that

$$\mu\left(\frac{2 + K[P_0]}{K}; K\right) = \frac{2 + KP_0 - \sqrt{4 + 2KP_0}}{KP_0}$$

and $[L_0]^{*mod} > [L_0]^*$. Since the Adair model is an increasing function then

$$\mu\left(\frac{2 + K[P_0]}{K}; K\right) > \eta\left(\frac{1}{K}; K\right) = \mu\left(\frac{4 + K[P_0]}{2K}; K\right) = \frac{1}{2}.$$

3.2. Suboptimal sequence designs

In the transformed monomer model the optimal design is concentrated at a single point. This does not allow one to fit a different model with more parameters. Additionally, the experimenter wants to obtain points to assess the adequacy of the chosen model. In this section, some suboptimal designs with more than one point are computed. For that, the points of the design have to be forced to follow some particular law, e.g. to be in a uniform, geometric or harmonic progression (see e.g. [11,13,14]). The uniform law is likely the most common in practice [4]. Most of the results in this section are numerical.

The designs obtained are compared between them and with the universal optimal design. Thus, the researcher has different options to choose a proper design. In this section, only the transformed Adair model is used, since this is the correct option. Color blue seven-point designs will be considered in our example. This choice is in the median of the usual number of points in this type of experiments according to the literature reviewed by the authors, e.g. [6].

First, a harmonic series of points is considered in this way

$$[L_0]_1 = \frac{a}{7K}, [L_0]_2 = \frac{a}{6K}, [L_0]_3 = \frac{a}{5K}, [L_0]_4 = \frac{a}{4K}, \\ [L_0]_5 = \frac{a}{3K}, [L_0]_6 = \frac{a}{2K}, [L_0]_7 = \frac{a}{K},$$

where a has to be optimized. Then, a geometric series is considered fixing the last point $[L_0]_7$ as that obtained for the harmonic series. If this point is not fixed the seven points will collapse in one. Thus,

$$[L_0]_1 = \frac{a}{K} - r^6, [L_0]_2 = \frac{a}{K} - r^5, [L_0]_3 = \frac{a}{K} - r^4, [L_0]_4 = \frac{a}{K} - r^3, \\ [L_0]_5 = \frac{a}{K} - r^2, [L_0]_6 = \frac{a}{K} - r, [L_0]_7 = \frac{a}{K},$$

where the value of a is the optimal value for the harmonic series and r has to be optimized here.

If a uniform design is aimed, the seven points will be

$$[L_0]_1 = \frac{a}{7K}, [L_0]_2 = \frac{2a}{7K}, [L_0]_3 = \frac{3a}{7K}, [L_0]_4 = \frac{4a}{7K}, \\ [L_0]_5 = \frac{5a}{7K}, [L_0]_6 = \frac{6a}{7K}, [L_0]_7 = \frac{a}{K},$$

where here just the value of a has to be optimized.

For the transformed Adair model (8) the FIM at a particular point is

$$M_{mod}([L_0]; K) = f_{mod}([L_0]; K)^2 \\ = \left(\frac{2 + K[P_0] + K[L_0] - \sqrt{8K[P_0] + (2 + K[L_0] - K[P_0])^2}}{K^2[P_0] \sqrt{8K[P_0] + (2 + K[L_0] - K[P_0])^2}} \right)^2,$$

and for the seven points,

$$M_{mod}(\xi_7; K) = \sum_{i=1}^7 \frac{1}{7} M_{mod}([L_0]_i; K).$$

This matrix, scalar in our case, has to be maximized in order to obtain suboptimal designs forced to follow some regular series of points. Computations will be made later for a particular example.

3.3. Simulations to compare estimates and designs for both models

A simulation is performed in order to compare estimates and designs for the two different models considered above. Assume that experiments are going to be realized with initial ligand concentrations $[L_0]_i$, $i = 1, \dots, k$ for a fixed concentration of the protein, $[P_0]$. Ten

replicates will be made for each initial ligand concentration, that is 10 k experiments will be performed in total (70 if $k = 7$). Once the equilibrium is reached, the value of the saturated proteins' rate, $Y_{i1} \dots Y_{i,10}$, $i = 1, \dots, k$ will be observed.

For the initial ligand quantity, $[L_0]_i$, its corresponding saturation value for each replicate is generated adding a Gaussian random error with mean zero, and constant and known variance,

$$Y_{ij} = \mu([L_0]_i; K) + \varepsilon_i^Y, \quad \varepsilon_i^Y \sim N(0, \sigma_Y^2).$$

Plugging the values for the saturation rates $Y_{i1} \dots Y_{i,10}$, $i = 1, \dots, k$ in the Adair equation,

$$Y = \eta([L]; K) + \varepsilon = \frac{K[L]}{1 + K[L]},$$

the free ligands may be obtained from it, $[L]_{i1} \dots [L]_{i,10}$, $i = 1, \dots, k$. Then, a random error is added to these values according to a Normal distribution with mean zero and constant variance,

$$[L]_{ij}' = [L]_{ij} + \varepsilon_i^L, \quad \varepsilon_i^L \sim N(0, \sigma_L^2), \quad i = 1, \dots, k, \quad j = 1, \dots, 10,$$

obtaining new free ligand concentrations $[L]_{i1}' \dots [L]_{i,10}'$. This process is summarized in Table 1.

These are all the designs obtained:

1. Optimal one-point design for the Adair model, ξ^* .
2. Optimal one-point design for the transformed Adair model, ξ^{*mod} .
3. Suboptimal seven-point design for a harmonic sequence, ξ^{harm} .
4. Suboptimal seven-point design for a geometric sequence, ξ^{geom} .
5. Seven-point uniform design at distances $\frac{1}{7}$, ξ^{unif} .

3.3.1. Example

Table 2 shows these designs. According to the literature, the nominal value of the parameter used for the computations was taken as $K = 0.1$ and the fixed protein concentration as $[P_0] = 11$. The maximum of the information is reached at value $a^* = 12.1663$ for the harmonic sequence, $r^* = 1.35674$ for the geometric and $a^* = 5.52922$ for the uniform. Then, simulations were performed for those designs using variances for the errors $\sigma_{Ymod}^2 = (0.015)^2$ and $\sigma_L^2 = (0.9)^2$, respectively for the transformed model and for the free ligand. The simulation was repeated 10,000 times.

Fig. 2 shows the scatter plot of the simulated points of variable Y versus $[L]$ and the scatter plot of simulated points, using the transformed Adair model, of Y versus $[L_0]$ at the seven points of the harmonic optimal sequence for the initial ligand concentration. Then, the MLE will be obtained and compared (last column of Table 3).

With them the parameter K is estimated for the Adair model and for the transformed model. In Table 3 a comparison of the estimates is made. The estimate obtained with the transformed model, which is the correct one, has less bias than the rest. Also for the one-point designs it is more efficient (less variance). Actually, the most efficient and unbiased estimate is obtained for the optimal design for the transformed model, ξ^{*mod} . From the seven-point designs the most efficient and unbiased estimate was the harmonic sequence design. The last row shows the good approximation of the covariance matrix through the inverse of the information matrix.

All of these agree with the a priori comparisons of the design efficiencies. The efficiencies of the five designs computed above with respect to the actual transformed model are shown in the last row of Table 3. The efficiencies were computed according to the formula

$$eff_{M_{mod}}(\xi) = \frac{\sum_i p_i M_{mod}([L_0]_i; K)}{M_{mod}([L_0]^{*mod}; K)}.$$

Table 1
Scheme of the simulation of k experimental values replicated 10 times.

L_0	Y for each replicate	L for each replicate	$L' = L + \varepsilon$
$[L_0]_1$	$Y_{11} \dots Y_{1,10}$	$[L]_{11} \dots [L]_{1,10}$	$[L']_{11} \dots [L']_{1,10}$
\vdots	\vdots	\vdots	\vdots
$[L_0]_i$	$Y_{i1} \dots Y_{i,10}$	$[L]_{i1} \dots [L]_{i,10}$	$[L']_{i1} \dots [L']_{i,10}$
\vdots	\vdots	\vdots	\vdots
$[L_0]_k$	$Y_{k1} \dots Y_{k,10}$	$[L]_{k1} \dots [L]_{k,10}$	$[L']_{k1} \dots [L']_{k,10}$

The seven-point designs have large efficiencies while the largest efficiency is for the one-point optimal design for the Adair model with a small loss of about 3%. Then, one might conjecture that optimal designs directly computed for the Adair model (on $[L]$) may be still good for the transformed model. There is another interesting result here. The suboptimal designs computed are rather competitive, so one should focus on model fitting and selection and residual diagnostics. The combination of both ideas gives a nice practical tool for dealing with Adair models with more parameters.

3.4. Robustness with respect to the choice of the nominal value of the parameter

As mentioned above the designs computed in this paper depend on the choice of the nominal value of the parameter of interest K . In this Section we check how the efficiency changes if the nominal value is wrongly chosen.

Let K_0 be the nominal value chosen for computing the designs. This is a fix value during the whole sensitivity analysis since this is the guessed value given by the practitioners, actually 0.1 in the case considered. The true value, say K^* , is unknown, but we assume it will be near K_0 , that is, in a neighborhood of it. Thus, two different optimal designs can be computed, one for K_0 , say ξ_{K_0} , and one for K^* , say ξ_{K^*} . Then, the efficiency of ξ_{K_0} has to be computed with respect to the true one ξ_{K^*} in the following way,

$$\text{eff}_{K^*}(\xi_{K_0}) = \frac{M_{\text{mod}}([L_0]^{*\text{mod}}(K_0); K^*)}{M_{\text{mod}}([L_0]^{*\text{mod}}(K^*); K^*)}$$

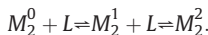
An analytic form of this efficiency can be given

$$\text{eff}_{K^*}(\xi_{K_0}) = \frac{2(2 + [P_0]K^*) \left(K^* + K_0 + [P_0]K^*K_0 - K_0 \sqrt{1 + 2K^*[P_0] + \frac{2K^*}{K_0} + \frac{K^2}{K_0^2}} \right)^2}{(2 + [P_0]K^* - \sqrt{4 + 2K^*[P_0]})^2 (K^{*2} + K_0^2 + 2K^*K_0(1 + [P_0]K_0))}$$

Fig. 3 displays the shape of this function for $K_0 = 0.1$ for different possible true values of K^* . For $K^* > 1$ the efficiency decays below 30%, which is not admissible at all. For $0.0426 < K^* < 0.2488$, the efficiency is always larger than 80%. If $K^* < K_0$, the design is less robust than it is for $K^* > K_0$. Then, it will be better to “underguess” the nominal value, since the big problem is with low possible true values, instead of “overguess” it. This graphic provides a good tool to choose the nominal values needed to design the experiment. Here we have considered the limits to be safe. Beyond those limits [0.0426, 0.2488] there are risk areas we must avoid in practice. This is a general consideration taking into account that the practitioner knows what are the limits of the possible error in choosing the nominal value. For instance, if the practitioner is completely sure that the true value of the parameter is not above 0.2488 or below 0.0426, then we can state that the efficiency of the design provided will be over 80% almost for sure.

4. Optimal designs for the dimer model

The second order Adair model is given by the so called dimer equation for two sites of binding in the macro-molecule. The chemical reaction is then



The association constants for the first and second reactions are K_1 and K_2 respectively

$$K_1 = \frac{[M_2^1]}{[M_2^0][L]} \quad \text{and} \quad K_2 = \frac{[M_2^2]}{[M_2^1][L]}, \tag{10}$$

and the saturation rate is then

$$\eta([L]; \theta) = \frac{[M_n^1] + 2[M_n^2]}{[M_n^0] + [M_n^1] + [M_n^2]} \tag{11}$$

By plugging (10) into (11), the Adair model with two binding sites is obtained,

$$Y = \eta([L]; \theta) + \varepsilon = \frac{K_1[L] + 2K_1K_2[L]^2}{2(1 + K_1[L] + K_1K_2[L]^2)} + \varepsilon,$$

where the parameters to be estimated are $\theta = (K_1, K_2)$ and the random error, ε , is assumed Normal with zero mean and constant variance. It is a non-linear model, and $[L]$ is again the concentration of the free ligand once the equilibrium in the experiment is reached.

As mentioned above in a dimer model, the protein has two binding sites for the ligand. This binding yields a substantial change in the protein, which is transmitted to the rest of the binding sites. In particular, there is a change of the affinity of the protein to other molecules of the ligand. The cooperativity comes from this change of affinity of one site for the ligand when the saturation increases. Specifically, the binding of a ligand may help (positive cooperativity) or may hinder (negative cooperativity) other binding of other ligands. Positive cooperativity does not mean that the global affinity of the enzyme to the ligand is greater than it is with negative cooperativity. The actual meaning is that a shorter change of the ligand concentration is needed to modify the saturation rate. Fig. 4 shows that in order to go from 10% to 80% saturation, the change of the concentration of the free ligand is much greater for negative than for positive cooperativity.

The FIM for the dimer model, not taking into account the uncertainty of the free ligand concentration, is computed differentiating the mean model with respect to K_1 and K_2 ,

$$f([L]; K_1, K_2) = \begin{pmatrix} \frac{\partial \eta([L]; K_1, K_2)}{\partial K_1} \\ \frac{\partial \eta([L]; K_1, K_2)}{\partial K_2} \end{pmatrix} = \begin{pmatrix} \frac{[L](1 + 2K_2[L])}{2(1 + K_1[L] + K_1K_2[L]^2)^2} \\ \frac{K_1[L]^2(2 + K_1[L])}{2(1 + K_1[L] + K_1K_2[L]^2)^2} \end{pmatrix}$$

The FIM for a general design ξ is then

$$M(\xi; \theta) = \sum_{[L]_i \in \chi} f([L]_i; K_1, K_2) f^T([L]_i; K_1, K_2) \xi([L]_i)$$

In order to compute the D-optimal design we first check two-point designs, $[L]_1, [L]_2 \in \chi$. If the number of support points of the D-optimal design is equal to the number of parameters, then the weights in the design are equal. Thus, weights $p = 1 - p = 1/2$ are assumed in the

Table 2
Optimal and suboptimal designs for $K = 0.1$.

sign	$[L_0]_i$	$[L]_i$
ξ^*	25.5	10
ξ^{mod}	31	12.4
ξ^{harm}	{17.3, 20.2, 24.3, 30.4, 40.5, 60.8, 121.6}	{6.5, 7.7, 9.4, 12.1, 16.8, 26.4, 56.1}
ξ^{geom}	{19.5, 26.4, 35.9, 48.7, 66.1, 89.6, 121.6}	{7.4, 10.4, 14.6, 20.6, 28.9, 40.4, 56.1}
ξ^{unif}	{7.8, 15.7, 23.6, 31.5, 39.4, 47.3, 55.2}	{2.7, 5.8, 9.2, 12.7, 16.3, 20.0, 23.7}

following computations. The determinant of the information matrix for a design of this type is then

$$\det M(\xi; \theta) = \frac{K_1^2 [L]_1^2 ([L]_1 - [L]_2) [L]_2^2 (2 + k_1 [L]_1 + k_1 [L]_2 + 2k_1 k_2 [L]_1 [L]_2)^2}{64 (1 + k_1 [L]_1 + k_1 k_2 [L]_1^2) (1 + k_1 [L]_2 + k_1 k_2 [L]_2^2)},$$

where $[L]_1, [L]_2 \in [0, L]$. Table 4 displays ligand concentrations maximizing this determinant for different nominal values of the parameters. The equivalence theorem (2) was checked numerically for these designs and all of them satisfied it, so they are actually D-optimal. Fig. 5 shows function $f^t([L]; \theta)M^{-1}(\xi; \theta)f([L]; \theta)$ for the three situations considered in Fig. 4

These are some comments about the results in Table 4:

Negative cooperativity Increasing K_1 for fixed values of K_2 , both design points and the determinant decrease.

Hyperbolic Increasing K_1 and K_2 both design points and the determinant decrease. Moreover, for $K'_1 = aK_1$ and $K'_2 = aK_2$, $a > 0$, the optimal points are $[L]'_1 = \frac{[L]_1}{a}$, $[L]'_2 = \frac{[L]_2}{a}$ and

$$\det M([L]'_1, [L]'_2; K'_1, K'_2) = a^{-4} \det M([L]_1, [L]_2; K_1, K_2).$$

Positive cooperativity (sigmoidal) Increasing K_2 for a fixed value of K_1 both design points and the determinant decrease again.

Once the D-optimal design is computed in terms of the free ligand, it has to be transformed in terms of the initial ligands that will be tried experimentally. From Eq. (10),

$$[M_2^1] = K_1 [M_2^0] [L] \quad \text{and} \quad [M_2^2] = K_2 [M_2^1] [L] = K_1 K_2 [M_2^0] [L]^2. \quad (12)$$

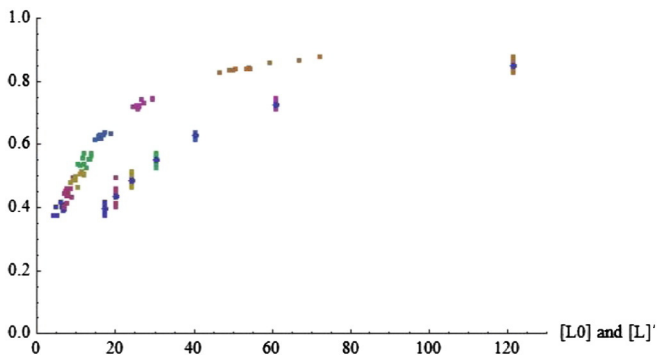


Fig. 2. Scatter plot of simulated points of variable Y versus $[L]'$ (upper spreaded cloud) and scatter plot of simulated points of Y versus $[L_0]$ at the seven points of the harmonic optimal sequence.

Table 3

Results of the simulations for $K = 0.1$ for the Adair model and the transformed model respectively for each statistic. The bottom row shows the efficiencies of the designs computed with respect to the actual transformed model.

	ξ^*	ξ^{mod}	ξ^{harm}	ξ^{geom}	ξ^{unif}	
$E[\hat{K}]$	0.0996	0.0998	0.1000	0.1000	0.0999	Adair
	0.1000	0.1000	0.1000	0.1000	0.1000	Transformed
$10^4 E(\hat{K}) - K $	3.540	1.614	0.4825	0.1516	1.306	Adair
	0.6292	0.4772	0.6608	0.0100	0.0970	Transformed
$10^6 \text{Var}[\hat{K}]$	8.050	5.070	1.211	0.8213	3.424	Adair
	4.611	4.420	5.722	6.057	5.613	Transformed
$10^6 \frac{\sigma_{\text{Adair}}^2 M_{\text{mod}}^{-1}([L_0]; K)}{10}$	4.658	4.535	5.617	6.111	5.617	
Efficiencies	0.968	1	0.807	0.742	0.807	

The initial concentration of a protein with two binding sites is the sum of the concentrations of the free protein, $[M_2^0]$, the protein with one site bound, $[M_2^1]$, and the protein with two sides bound, $[M_2^2]$,

$$[P_0] = [M_2^0] + [M_2^1] + [M_2^2], \quad (13)$$

and plugging Eq. (12) into Eq. (13), then

$$[P_0] = [M_2^0] + K_1 [M_2^0] [L] + K_1 K_2 [M_2^0] [L]^2. \quad (14)$$

The concentration of the initial ligand to be fixed in the experiment will be

$$[L_0] = [L] + [M_2^1] + 2[M_2^2], \quad (15)$$

and plugging Eq. (12) into Eq. (15) then

$$[L_0] = 2[L] + K_1 [M_2^0] [L] + 2K_1 K_2 [M_2^0] [L]^2. \quad (16)$$

The equation system of Eq. (14) and Eq. (16) will be solved numerically obtaining $[M_2^0]$ and $[L_0]$. Thus, the D-optimal design is transformed into

$$\tilde{\xi}^* = \left\{ \begin{matrix} [L_0]_1 & [L_0]_2 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\}.$$

This design is suboptimal since the proper D-optimal design comes from the corresponding transformed Adair model in function of the initial ligand. But, for $n > 1$ an analytic form is not available. The high

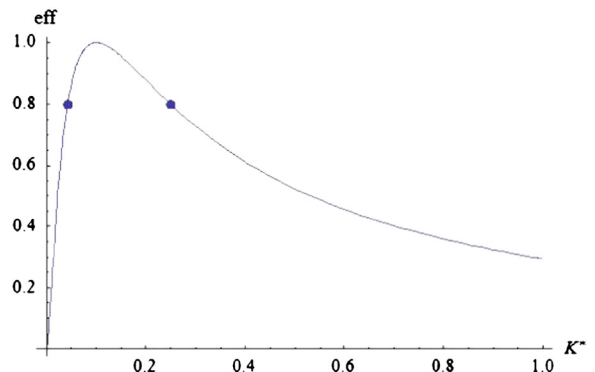


Fig. 3. Relative efficiency of the optimal design for $K_0 = 0.1$ as a function of the possible true values, K^* .

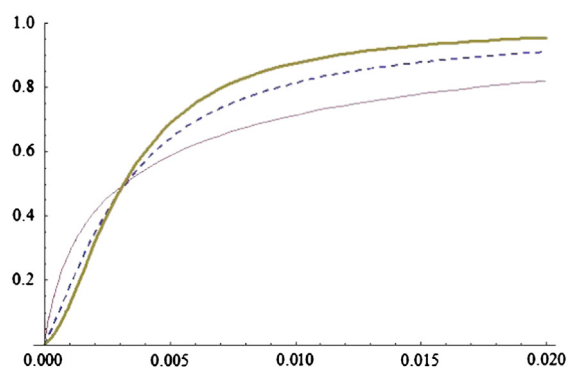


Fig. 4. Saturation rate in function of free ligand (dimer model) for positive cooperativity, $K_1 = 100$, $K_2 = 1000$ (thick line), hyperbolic, $K_1 = 310$, $K_2 = 310$ (dashed line) and negative cooperativity, $K_1 = 1000$, $K_2 = 100$ (thin line).

efficiencies obtained for the monomer somehow justify doing this for the dimer model.

5. Discussion

The Adair model relates two random variables, the saturation rate (response) and the free ligand (explanatory). In practice, this model is fitted using standard procedures, as ordinary least squares, which assume a non-random explanatory variable. Moreover, no optimal experimental design can be used here to improve the estimates because the free ligand is not under the control of the experimenter. On the other hand, this equation is analytically derived from the kinetics theory. Thus, there is a relationship between the free ligand and the initial ligand used by the experimenter. The latter can be controlled and so designed.

In the case of the monomer, with just one parameter, an analytic relationship between the saturation rate and the initial ligand can be derived easily. This can be used as the model to be fitted with two important advantages. On the one hand, the estimates obtained are the right ones, and on the other hand, optimal designs can be obtained to improve the estimation. The universal optimal design is a one-point design. Thus, the parameter can be estimated, but fitting a different model with more parameters will not be possible. Moreover, most experimenters are more satisfied with some more points, e.g. to check whether the postulated model is fine and if not to fit a different one. For that in this paper seven-point designs were forced to follow regular sequences and then they were optimized. This is a compromise between what is optimal and what is convenient in practice. The designs obtained in this way have good efficiencies. A simulation study provided comparisons of the bias and the variance of the estimates. In summary, it may be said that for a reasonable sample size the original Adair model (monomer case) may be used, either for estimating or for obtaining optimal designs on the free ligand, and then transforming it to the initial ligand. The advantage of this is that a fixed concentration of initial ligand produces different concentrations of the free ligand in each replicated

experiment. Notice that minimal variance may only be reached for the transformed model since the experiments are realized at the proper optimal design point $[L_0]^{mod}$. The optimal design for the Adair model cannot be performed exactly at the optimal point $[L]^*$. For the monomer model, three different types of designs were obtained and compared,

1. A one-point design for the Adair model was computed for the free ligand. Then, it was transformed into a one-point design in terms of the initial ligand. Realizing the experiment N times at this point produces N different free ligand concentrations, which is quite convenient for fitting purposes.
2. A one-point design for the transformed Adair model in the function of the initial ligand. This is the proper optimal design, if this model is going to be fitted. But then there is just one point to do it and there is no room for checking a different model with more parameters.
3. Suboptimal designs forced in a sequential way in order to have several different points.

For the dimer model, and greater order, an analytic expression in function of the initial ligand cannot be obtained. Taking into account, that the monomer model shows good efficiencies for optimal designs with the original Adair model, we have computed optimal designs numerically for it expecting they are also adequate for the dimer model.

If the error of the free ligand is small or just the prediction of the response variable is the aim of the inferences, then the original Adair model can be considered to compute optimal designs. If the interest is in estimating the parameter and the error is not negligible a better approach to the problem should be considered.

Because the models considered are nonlinear, the optimal experimental designs depend on the values of the parameters to be estimated. In this paper (guessed) nominal values are provided for that. Thus, a robustness analysis was performed to check how sensitive the designs are versus wrong choices of the nominal values of the parameters. A reference is provided in this way in order to know how risky the choice of the nominal value is.

The reason of choosing the Gaussian distribution in a first approach is that in the literature LSE are usually computed, but a more appropriate distribution could be looked for. Since the response variable is a proportion, a Beta distribution may better fit this problem. This is deserved for further work using real data.

Another important point is that the model is considered homoscedastic in practice. But we have noticed that there is potential heteroscedasticity. This is a future point to be explored, both from the estimation and from the experimental design point of view.

Throughout this paper, the protein concentration was fixed. Although this is justified in practice in many situations, this concentration can also be designed optimally. This means an important degree of difficulty is included in the problem, but it is still affordable numerically.

In practice, it is common to have two types of proteins to which the ligand may be bounded. This leads to computing optimal designs for mixtures of distributions, which has not been considered much in the literature.

Conflict of interest

There is no conflict of interest.

Acknowledgments

This work has been supported by the Ministerio de Educación y Ciencia and Fondos FEDER MTM2010-20774-C03-01 and Junta de Comunidades de Castilla la Mancha PEI10-0291-1850. The authors are thankful for the helpful comments of Dr. J. Burguillo and Prof. Mucientes about the chemical problem. We thank the referees and the AE for all the suggestions and good, clear and constructive words. We want to thank one of the referees in particular for correcting the grammatical errors. Not being native English speakers we appreciate this very much.

Table 4
Optimal designs for different nominal values of the parameters.

	K_1	K_2	$[L]_1$	$[L]_2$	$detM([L]_i; \theta)$
Negative cooperativity	250	100	$2.64 \cdot 10^{-3}$	$1.50 \cdot 10^{-2}$	$1.25 \cdot 10^{-13}$
	500	100	$1.54 \cdot 10^{-3}$	$1.27 \cdot 10^{-2}$	$3.17 \cdot 10^{-14}$
	750	100	$1.17 \cdot 10^{-3}$	$1.13 \cdot 10^{-2}$	$1.38 \cdot 10^{-14}$
	1000	100	$9.27 \cdot 10^{-4}$	$1.07 \cdot 10^{-2}$	$7.58 \cdot 10^{-15}$
Hyperbolic	100	100	$4.73 \cdot 10^{-3}$	$2.11 \cdot 10^{-2}$	$6.66 \cdot 10^{-13}$
	200	200	$2.36 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$4.16 \cdot 10^{-14}$
	300	300	$1.57 \cdot 10^{-3}$	$7.03 \cdot 10^{-3}$	$8.22 \cdot 10^{-15}$
Positive cooperativity (sigmoidal)	100	250	$3.22 \cdot 10^{-3}$	$1.23 \cdot 10^{-2}$	$7.57 \cdot 10^{-14}$
	100	500	$2.36 \cdot 10^{-3}$	$8.44 \cdot 10^{-3}$	$1.31 \cdot 10^{-14}$
	100	750	$1.96 \cdot 10^{-3}$	$6.78 \cdot 10^{-3}$	$4.53 \cdot 10^{-15}$
	100	1000	$1.71 \cdot 10^{-3}$	$5.82 \cdot 10^{-3}$	$2.10 \cdot 10^{-15}$

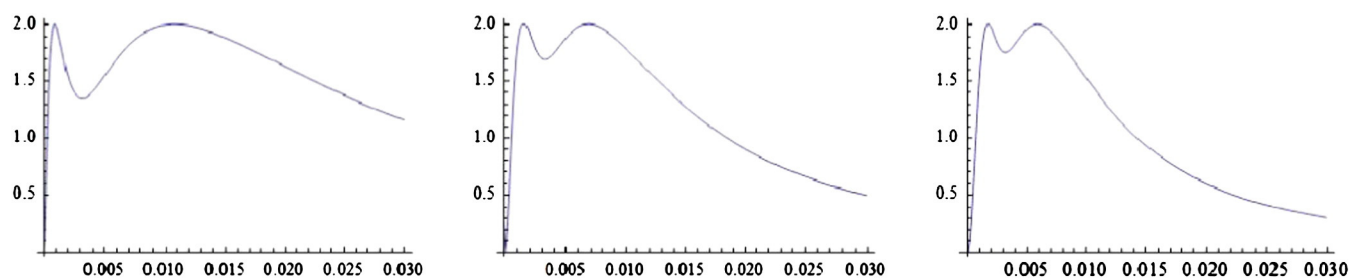


Fig. 5. Sensitivity functions showing that the equivalence theorem is satisfied for negative cooperativity, $K_1 = 1000, K_2 = 100$ (left); hyperbolic, $K_1 = 310, K_2 = 310$ (center) and positive cooperativity, $K_1 = 100, K_2 = 1000$ (right).

References

- [1] G.S. Adair, The hemoglobin system: VI. The oxygen dissociation curve of hemoglobin, *J. Biol. Chem.* 63 (1925) 529–545.
- [2] A.C. Atkinson, A.N. Donev, R.D. Tobias, *Optimum Experimental Designs, with SAS*, Oxford University Press, 2007.
- [3] H. Chernoff, Locally optimal designs for estimating parameters, *Ann. Math. Stat.* 24 (1953) 586–602.
- [4] H. Dette, V.B. Melas, A. Pepelyshev, N.S. Strigul, Design of experiments in the Monod model, *J. R. Stat. Soc. Ser. B* 65 (2003) (2005) 725–742.
- [5] V.V. Fedorov, *Theory of Optimal Experiments*, Academic press, New York, 1972.
- [6] D. Freifelder, *Principles of Physical Chemistry with Applications to the Biological Sciences*, Boston Portola Valley, 1984. (Chapter 16).
- [7] M. Imhof, J. López-Fidalgo, W.K. Wong, Efficiencies of optimal approximate designs for small samples, *Statistica Neerlandica* 55 (2001) 301–315.
- [8] J. Kiefer, Optimum experimental designs, *J. R. Stat. Soc. Ser. B* 21 (1959) 272–319.
- [9] J. Kiefer, J. Wolfowitz, The equivalence of two extremum problems, *Can. J. Math.* 12 (1960) 849–879.
- [10] J. López-Fidalgo, I. Ortiz-Rodríguez, W.K. Wong, Design issues for population growth models, *J. Appl. Stat.* 38 (2011) 501–512.
- [11] J. López-Fidalgo, W.K. Wong, Design issues for the Michaelis–Menten model, *J. Theor. Biol.* 215 (2002) 1–11.
- [12] A.E. Martell, R.J. Motekaitis, *Determination and Use of Stability Constants*, 2nd edition Wiley, 1992.
- [13] L.J. Rodríguez Aragón, J. López Fidalgo, Optimal designs for the Arrhenius equation, *Chemom. Intell. Lab. Syst.* 77 (2005) 131–138.
- [14] L.J. Rodríguez Aragón, J.T. López Fidalgo, T-, D- and c- Optimum Designs for BET and GAB adsorption isotherms, *Chemom. Intell. Lab. Syst.* 89 (2007) 36–44.
- [15] S.Y. Tetin, T. Hazlett, Optical spectroscopy in studies of antibody–hapten interactions, *Methods* 20 (2000).
- [16] P. Whittle, Some general points in the theory of optimal experimental design, *J. R. Stat. Soc. Ser. B* 35 (1973) 123–130.
- [17] H.P. Wynn, The sequential generation of D-optimum experimental designs, *Ann. Math. Stat.* 41 (1970) 1655–1664.