10th International IFAC Symposium on
Computer Applications in Biotechnology
Preprints Vol.1, June 4-6, 2007, Cancún, Mexico

OPTIMIZATION OF EXPERIMENTS FOR IMPROVED ESTIMATION OF PROTEIN INTERACTION PARAMETERS

Lyne Woodward, Gregory De Crescenzo, Bala Srinivasan

Department of Chemical Engineering,
École Polytechnique Montréal,
Montréal, Canada, H3C 3A7

Abstract: The emergence of surface plasmon resonance-based biosensors has facilitated the identification of kinetic parameters for macromolecular interactions. Normally, these parameters are computed from experiments with arbitrarily chosen periods of protein and buffer injections, and varying protein concentrations. Instead of choosing the above mentioned variables arbitrarily, in this paper, an optimization approach is used to determine them so as to reduce the experimentation time, while treating the required confidence level as constraints. It is shown using experimental data that the desired confidence is reached with a much shorter experiment, compared to the standard set of experiments typically performed.

Keywords: Optimization, Identification, Confidence level, Macromolecular interactions, Biophysical characterization, Kinetics, Surface plasmon resonance.

1. INTRODUCTION

Identification of kinetic parameters for macromolecular interactions is a key issue in biochemistry, protein engineering and rational drug design. The confidence in these kinetic parameters is key during early and late stages of in vitro drug development (Stenlund *et al.*, 2006) as well as in the growing field of quality assessment of recombinant proteins used as therapeutics (Kikuchi *et al.*, 2005).

The above mentioned identification requires appropriate measurements, for which surface plasmon resonance (SPR) biosensors have emerged as a new approach during the last decade (Myszka, 1997). This approach for characterizing macromolecular interactions gives not only an opportunity to quantify the interactions at equilibrium, but also to measure their kinetics. Among the SPR biosensors available in the market, the Biacore instruments (BIACORE Inc.) are widely recognized to provide the state-of-the-art technology for kinetic characterization (Myszka, 2004).

The rapid expansion of the SPR technology has been accompanied by the development of numerical approaches to analyze the recorded data sets (Myszka, 1997; De Crescenzo *et al.*, 2001). Though it is relatively easy to collect Biacore data, the precise determination of kinetic constants still depends on experimenter's know-how and is based on trial-and-error approaches. More importantly, to assign confidence levels to these parameters is a tricky issue, since the variation in the estimated parameters arises from two sources: the measurement noise and the process noise (variation in biological behavior). In this paper, only the error in the parameters caused by the measurement error is addressed and the goal herein is to design the experiments so as to have a desired confidence in the parameters.

It is shown in this paper that optimization at the level of experimental planning can indeed help in having shorter experiments that give the desired confidence in the estimated parameters. Also, the total consumption of the material can be reduced. The optimization problem defined here is to find the shortest experiment that leads to the given confidence levels on the estimated parameters.

The literature on optimization of experiments is vast and has used certain optimality criteria derived from the Hessian of the identification problem (Pukelsheim, 1993), also referred to as the Fisher information matrix (Walter and Pronzato, 1985). Typically, constraints are not taken into account in the problem formulation. The novelty of this paper is the use of an experimentally relevant criterion such as the total time as the objective function, while taking into account the confidence level of the identified parameters as constraints. The confidence levels are indeed calculated using the Hessian resulting from the identification step.

The paper is organized as follows. Section 2 presents the problem definition and includes all the identification tools used in this project. Section 3 defines the optimization problem for experiment planning while Section 4 shows the improvement obtained using the optimization by comparing the optimized experiments with the non-optimized ones. Finally the conclusions are presented in Section 5.

2. IDENTIFICATION OF KINETIC PARAMETERS

2.1 Surface plasmon resonance experiment

In a typical Biacore experiment, one of the species under study is immobilized on the sensor chip surface. A solution containing the other binding partner is then injected at a defined concentration in a continuous fashion over the sensor chip surface and the mass accumulation resulting from the interaction is recorded in real-time in arbitrary Resonance Units (RU). The signal recorded in RU is proportional to the change in mass that occurs at the sensor chip surface (Fig 1 for example). After a pre-defined period of time, the protein solution is replaced by a continuous injection of a buffer and the complexes formed at the biosensor surface starts dissociating. The RU measurements are obtained in real-time both during the association and dissociation periods. Since the signal is recorded in real time, kinetic information can be obtained.

If necessary, the complexes remaining at the biosensor surface are totally dissociated by injecting a regeneration solution. Once the surface is regenerated, the injection cycle is repeated over the same surface. The protein concentration of the solution to be injected is varied. This leads to the recording of a set of curves, depicting the interactions between the protein and its binding partner that is attached to the sensor chip surface.

2.2 Modeling of the interaction

The system under study corresponds to continuous injections of protein (the analyte, denoted A) at defined concentrations, over a surface on which its binding partner (the ligand, denoted B) had been previously coupled. This results in the formation of a non-covalent complex AB that can, in turn, dissociate. The interaction is depicted by the following scheme:

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} AB \tag{1}$$

where k_a and k_d are the rate constants of the interaction in $M^{-1}s^{-1}$ and s^{-1} respectively. The above interactions are monitored by using the resonance signal of the resulting product AB. The following model is used to characterize this interaction :

$$\dot{R}_{AB} = k_a C_A (R_{max} - R_{AB}) - k_d R_{AB}, \quad R_{AB}(0) = 0 \tag{2}$$

$$R = \begin{cases} R_{AB} + R_A & \text{if } C_A \neq 0 \\ R_{AB} & \text{if } C_A = 0 \end{cases} \tag{3}$$

where C_A is the concentration of the free analyte A in M, R_{max} the total ligand concentration in RU, R_{AB} the concentration of AB complex in RU, R_A the resonance signal resulting from the refraction of analyte A in RU and R is the resulting recorded signal in RU. Note that a simple Langmuirian kinetics is assumed with the rate of the forward interaction being $k_a C_A C_B$, with $C_B \propto R_{max} - R_{AB}$.

2.3 Parameter identification problem

In order to identify the parameters k_a and k_d, a series of experiments with varying concentrations of analyte A, C_A are performed. Each experiment consists of an on-period (protein injection) t_{on} and an off-period (buffer injection) t_{off}, both expressed in seconds. From the obtained data, the following least-squared identification algorithm is utilized to compute the required parameters.

$$\min_{k_a,k_d,R_{max},R_{A_i}} S \tag{4}$$
$$S = \sum_{i=1}^{N} \frac{1}{T^i} \int_0^{T^i} (R^i_{meas}(t) - \hat{R}^i(t))^2 dt$$
$$s.t. \quad \hat{R}^i \text{ obtained from (2), (3)}$$

where N is the number of experiments, $T^i = t^i_{on} + t^i_{off}$, the duration of the i$^{\text{th}}$ experiment, $R^i_{meas}(t)$ the measurements, and $\hat{R}^i$ the prediction from the model. Note that the parameters R_{max} and R_{A_i} the resonance signal of analyte A in the i$^{\text{th}}$ experiment should also be included in the identification procedure.

2.4 Confidence intervals

One of the important aspects while performing identification is the confidence one has in the parameters. This primarily depends on the measurement noise, but also on the experimental planning. To analyze this, consider the general case, where the model parameters θ are obtained by minimizing a least-squared criterion

$$\min_{\theta} S(\theta) = \frac{1}{T} \int_0^T \frac{1}{2} \|y_{meas} - \hat{y}(\theta)\|^2 dt \tag{5}$$

where y_{meas} are the measurements and $\hat{y}$ the model prediction. The best estimate $\hat{\theta}$ of the parameters θ found by minimizing $S(\theta)$ respects the necessary condition of optimality:

$$\frac{\partial S}{\partial \theta} = 0 \Rightarrow$$
$$\frac{1}{T} \int_0^T (y_{meas} - \hat{y}(\theta))^T \left(\frac{\partial \hat{y}}{\partial \theta} \right) \Big|_{\hat{\theta}} = 0 \tag{6}$$

Let y^* be the real data without noise, ϵ be the noise signal and θ^* the real values of the parameters. Then, we have :

$$y_{meas} = y^* + \epsilon \tag{7}$$
$$\frac{1}{T} \int_0^T (y^* - \hat{y}(\theta^*))^T \left(\frac{\partial \hat{y}}{\partial \theta} \right) \Big|_{\theta^*} dt = 0 \tag{8}$$

Considering the Taylor series approximation of (6) and using (7) leads to:

$$\frac{1}{T} \int_0^T (y^* - \hat{y}(\theta^*)^T \left(\frac{\partial \hat{y}}{\partial \theta} \right) \Big|_{\theta^*} dt$$
$$- \frac{1}{T} \int_0^T \left(\left(\frac{\partial \hat{y}}{\partial \theta} \right) \Big|_{\theta^*} \tilde{\theta}^T - \epsilon \right)^T \left(\frac{\partial \hat{y}}{\partial \theta} \right) \Big|_{\theta^*} dt = 0 \tag{9}$$

where the variation of $\hat{\theta}$ around θ^* is expressed by $\tilde{\theta} = \hat{\theta} - \theta^*$. Using (8) in (9) gives :

$$\tilde{\theta} = H^{-1} \left(\frac{1}{T} \int_0^T \left(\frac{\partial \hat{y}}{\partial \theta} \right)^T \epsilon \, dt \right) \tag{10}$$
$$H = \left(\frac{1}{T} \int_0^T \left(\frac{\partial \hat{y}}{\partial \theta} \right)^T \frac{\partial \hat{y}}{\partial \theta} \, dt \right) \tag{11}$$

where H is an approximation of the Hessian of the minimization problem (5) and is also referred to as the Fisher information matrix in literature (Walter and Pronzato, 1985).

The variance of $\tilde{\theta}$ can be expressed by :

$$E(\tilde{\theta} \, \tilde{\theta}^T) = H^{-1} \phi^T E(\epsilon \epsilon^T) \phi H^{-1} \tag{12}$$
$$\phi = \left(\frac{1}{T} \int_0^T \frac{\partial \hat{y}}{\partial \theta} \, dt \right) \tag{13}$$

If the variance of the noise measurement is the diagonal matrix, $E(\epsilon \epsilon^T) = \sigma^2 I$, then (12) becomes:

$$E(\tilde{\theta} \, \tilde{\theta}^T) = H^{-1} \sigma^2 \tag{14}$$

where σ^2 can be estimated by the optimal value of S in 5. If the number of measurements is large enough, the dispersion of $\hat{\theta}$ around the true values θ^* follow a normal distribution such as :

$$(\hat{\theta} - \theta^*) \approx N_p(0, \sigma^2 H^{-1}) \tag{15}$$

So, the dispersion around θ^* is normally distributed, and the $100(1-\alpha)$ confidence interval for θ^* can be approximated by :

$$\theta^* : (\theta^* - \hat{\theta}) H (\theta^* - \hat{\theta}) \leq 2p\sigma^2 F^{\alpha}_{p,n-p} = c^2 \tag{16}$$

where p is the number of parameters in the model, n the number of samples, and F the cumulative distribution function (Seber and Wild, 1989). The confidence intervals of θ_i can be evaluated as :

$$\xi_i = \frac{|\hat{\theta}_i - \theta^*_i|}{\hat{\theta}_i} = \frac{c\sqrt{(H^{-1})_{ii}}}{\hat{\theta}_i} \tag{17}$$

2.5 Calculation of the Hessian

Thus, for calculating the confidence intervals, it can be seen that obtaining the Hessian is crucial. The calculation of the Hessian for the optimization problem (5) where the dynamic system (2) acts as a constraint will be studied next.

Applying (11) to the system (2) leads to the following expression of the Hessian:

$$H = \frac{1}{T} \int_0^T \left(\frac{\partial R}{\partial \theta} \right)^T \frac{\partial R}{\partial \theta} dt \tag{18}$$

where :

$$\theta \equiv \begin{bmatrix} k_a \\ k_d \\ R_{max} \\ R_A \end{bmatrix}$$

For the system (2)(3), let x_i be used to denote the elements of the vector $\frac{\partial R}{\partial \theta}$:

$$\frac{\partial R}{\partial k_a} = \frac{\partial R_{AB}}{\partial k_a} \equiv x_1 \quad (19)$$

$$\frac{\partial R}{\partial k_d} = \frac{\partial R_{AB}}{\partial k_d} \equiv x_2 \quad (20)$$

$$\frac{\partial R}{\partial R_{max}} = \frac{\partial R_{AB}}{\partial R_{max}} \equiv x_3 \quad (21)$$

$$\frac{\partial R}{\partial R_A} = \begin{cases} 1 & \text{if } C_A \neq 0 \\ 0 & \text{if } C_A = 0 \end{cases} \quad (22)$$

Taking the derivative of (2) with respect to the three first parameters gives the following system of differential equations:

$$\dot{x}_1 = C_A(R_{max} - R_{AB}) - (k_a C_A + k_d)x_1 \quad (23)$$

$$\dot{x}_2 = -R_{AB} - (k_a C_A + k_d)x_2 \quad (24)$$

$$\dot{x}_3 = k_a C_A - (k_a C_A + k_d)x_3 \quad (25)$$

Thus, the Hessian is obtained by solving (23)-(25)along with (2)-(3).

3. OPTIMIZATION OF EXPERIMENTS

The main objective of the paper is to find the shortest experiment that will provide a desired confidence level in the estimated values of the interaction rate constants k_a and k_d.

Optimum design of experiments have been studied in the past using various techniques. Many optimality criteria have been used to define the optimization problem in the literature: the A-optimality (Montgomery, 2005), the E-optimality (Lorenzen and Anderson, 1993) and the D-optimality (Pronzato and Walter, 1989). Let the p eigenvalues of the inverse of the Hessian be denoted by λ_i. The A-optimality criterion minimizes the average variance, i.e. it maximizes the trace of the inverse of the Hessian ,i.e. $\max(\sum_{i=1}^{p} \lambda_i)$. The E-optimality criterion minimizes the maximum variance of the parameters, i.e. maximizes the minimum eigenvalue of the inverse of the Hessian, i.e. $\max(\min \lambda_i)$. The D-optimality criterion minimizes the volume of the space in which the estimated parameters could lie. This corresponds to maximizing the determinant of the inverse of the Hessian, i.e. $\max(\prod_{i=1}^{p} \lambda_i)$.

In this paper, the novelty lies in the formulation of optimization problem, which is motivated by the user requirements rather than a variance criterion as discussed before. The problem is to minimize the experimentation time subject to the desired confidence levels acting as constraints. In addition physical limitations are also taken into consideration. The optimization problem is defined as follows :

$$\min_{t_{on}, t_{off}, C_A} (t_{on} + t_{off}) \quad (26)$$

$$\begin{aligned} s.t. \quad \xi_{k_a} &\leq \beta_a \\ \xi_{k_d} &\leq \beta_d \\ C_A t_{on} &\leq \eta_{max} \\ C_A &\leq C_{max} \\ t_{on} + t_{off} &\leq T_{max} \end{aligned}$$

where t_{on} and t_{off} are the association and the dissociation periods of the experiment respectively, C_A the concentration of analyte to inject, β_a and β_d are the desired level of confidence in estimated parameters k_a and k_d, respectively, η_{max} the maximum amount of analyte allowed to be used and T_{max} the maximum total time allowed for the experiment. The confidence levels ξ_{k_a} and ξ_{k_d} are computed using (17).

As with any experimental design, the value of k_a, k_d needs to be known to choose the optimum experiment. This however is not possible (if we had known the true values, there is no reason to perform an experiment !) and so, the optimum design is based on a previous non-optimized experiment.

4. EXPERIMENTAL RESULTS

4.1 Materials and methods

All the experiments presented in this paper have been performed using the Biacore 3000 biosensor instrument with sensor chips (CM5). All chemicals were purchased from Biacore Inc. and Fisher Scientific Ltd (ON). All SPR experiments were carried out at 25^oC with PBS (Phosphare Buffer Saline) as the running buffer. In order to demonstrate the adequacy of the proposed approach, the interaction between a protein and a specific monoclonal antibody that had been developed for diagnostic and therapeutic use (the analyte), were investigated. More specifically, recombinantly-expressed and purified Prostate-Specific Membrane Antigen (PSMA, 50 000 Da) has been immobilized on the sensor chip (120 RUs) according to standard amine protocols described in (De Crescenzo et al. 2001). Purified anti-PSMA mouse monoclonal IgG antibody (clone ♯17G1, from Dr. S. Moffett, Proscan RX Pharma Inc., Montreal, QC, Canada) was used as the analyte (the species which is injected over sensorchip surface).

All the kinetic experiments were carried out in duplicate at a flow rate of 50 μ L/min to prevent non-biological artifacts. Injections of antibody were performed using the values of concentration, t_{on} and t_{off} specified later. In between the injections, regeneration of the sensor chip surface (i.e., elution of antibody bound to antigen) was

accomplished by 2 pulses of Glycine (10 mM, 20 s, pH 3.0). Similar injections were performed over a mock surface to correct antigen-antibody data (De Crescenzo *et al.*, 2001) prior to analysis.

4.2 Non-optimized reference experiments

As a point of comparison, experiments were done using 5 different levels of analyte concentration: 18.75 nM, 37.5 nM, 75 nM, 150 nM and 300 nM. Analytes were injected for 4 minutes ($t_{on} = 240$ s) and the dissociation period allowed was 6 minutes ($t_{off} = 360$ s). In addition, an experiment where only the buffer is passed was also performed, using which the baseline is eliminated. Also, each experiment was done in duplicate and an average of each pair of experiments was utilized. Fig 1 shows the resonance signal of these experiments.

The next challenge is to calculate the variance of the measurement error, σ^2. Note that the error between the model and the real data is the result of two different sources: the measurement noise and the process noise (variation in biological behavior). In this project, only the error in the parameters caused by measurement error is addressed. The variation in the parameters due to biological reasons is supposed to be compensated by performing the experiments in duplicate, a regular practice for biological systems.

To isolate the measurement error, each experiment was identified individually and then the difference between the measured and the predicted values of the resonances was used as an indicative of the measurement error. Then, an average of this indicative over the five different experiments was obtained. Here, a noise standard deviation of 0.3 RU was computed. This value has been used for all calculations of the confidence levels.

4.3 Optimized experiment

The main goal of this paper is to show that optimizing the profile can lead to shorter experiments to reach the same or even a better confidence in the estimated parameters. As explained before, an initial value of the parameters is required to perform the optimization. For this, a first experiment is performed with $C_A = 75$ nM, $t_{on} = 240$ s, and $t_{off} = 360$ s, as in the previous case. Next, the optimization problem (26) was solved using the following parameters values: β_a : 1%, β_d : 9%, η_{max} : $1.239e^{-4}$ m, C_{max} : 300nM, and T_{max} : 2400 s.

The last three values were obtained from the reference experiments presented in Fig. 1, such that the maximum amount of analyte, the maximum concentration of the analyte and the total time have to be less or equal to those used during the reference experiments.

The obtained optimal profile was to inject the analyte at its maximum concentration of 300 nM for a period of 300 seconds and to let the dissociation for 275 seconds. Note that the association time is longer to reach higher resonance values. The dissociation time needed to get the same confidence in k_d is then reduced. A margin was provided during the experiments (using periods of $330s$ and $310s$ respectively) in order to assure that the confidence levels were reached. The resonance signals of the initial 75 nM experiment and the optimized 300 nM experiment are shows in Fig 2.

4.4 Comparison

The parameter values and the confidence levels obtained for both the reference and optimized experiments are presented in Table 1. In this table, the total time is calculated taking into account the time for the buffer signals as well as the duplicate experiment. The estimated values obtained for the rate constants were very close to the ones obtained by the software supplied by the manufacturer of the equipment (Biacore).

Table 1. Results of the identification for reference and optimal experiments

	Reference profile	Optimal profile
$\hat{k}_a$ $(M^{-1}s^{-1})$	1.07×10^4	0.88×10^4
$\xi_a(\%)$	0.8	1
$\hat{k}_d(s^{-1})$	0.65×10^{-4}	1.2×10^{-4}
$\xi_d(\%)$	9.1	7.4
Qty of analyte(m)	2.79×10^{-4}	2.34×10^{-4}
Total time (min)	120	73

It can be clearly seen that optimizing the second experiment leads to better results in many ways. The total time is quite shorter and the confidence is better. The amount of analyte to reach that confidence is even lower than for the reference experiments.

5. CONCLUSION

The results shown in the present paper clearly demonstrate the multi-fold advantages of using optimization in the planning of experiments. From the point of view of biological system identification, this paper shows that one could be to save time and analyte consumption, while obtaining higher confidence. From the point of view of optimizing experiments, instead of performing an unconstrained optimization using a criteria based on the Fisher information matrix, it is interesting to see how all the user specifications could be

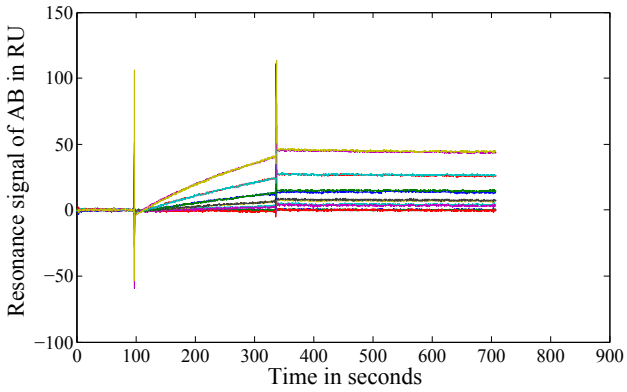

Fig. 1. Resonance signals from the reference experiments - 10 experiments without any optimization

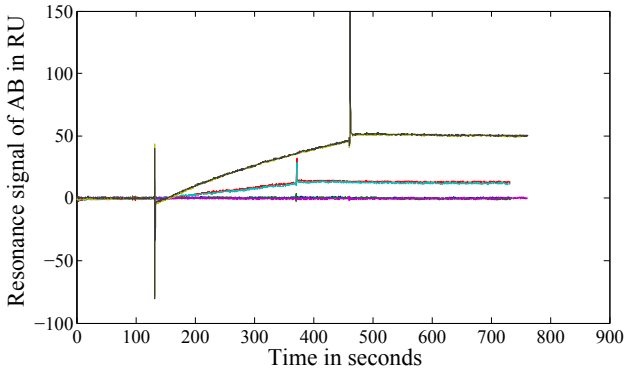

Fig. 2. Resonance signal resulting from a first non-optimized experiment with 75nM injection and a second optimized profile with 300nM injection

effectively included in a constrained optimization framework suitable for experimental planning.

REFERENCES

De Crescenzo, G., S. Grothe, M. Tsang, J. Zwaagstra and M.D. O'Connor-McCourt (2001). Real-time monitoring of the interactions of transforming growth factor- beta (TGF-beta) isoforms with latency-associated protein and the ectodomains of the TGF-beta type ii and iii receptors reveals different kinetic models and stoichiometries of binding. *Journal of Biological Chemistry* **276**(1), 29632–29643.

Kikuchi, Y., S. Uno, M. Nanami, Y. Yoshimura, S. Iida, N. Fukushima and M. Tsuchiya (2005). Determination of concentration and binding affinity of antibody fragments by use of surface plasmon resonance. *Journal of Bioscience and Bioengineering* **100**(3), 311–317.

Lorenzen, T.J. and V.L. Anderson (1993). *Design of experiments: A no-name approach*. Marcel Dekker. New York.

Montgomery, D.C. (2005). *Design and analysis of experiments*. John Wiley and sons, inc.

Myszka, D.G. (1997). Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Current Opinion in Biotechnology* **8**(1), 50–57.

Myszka, D.G. (2004). Analysis of small-molecule interactions using biacore s51 technology. *Analytical Biochemistry* **329**(2), 316–323.

Pronzato, L. and E. Walter (1989). Experiment design in a bounded-error context: comparison with D-optimality. *Automatica* **25**(3), 383–391.

Pukelsheim, F. (1993). *Optimal design of experiments*. Wiley series in probability and mathematical statistics.

Seber, G.A.F. and C.J. Wild (1989). *Nonlinear regression*. Wiley series in probability and mathematical statistics.

Stenlund, P., A. Frostell-Karlsson and OP. Karlsson (2006). Studies of small molecule interactions with protein phosphatases using biosensor technology. *Analytical Biochemistry* **353**(2), 217–225.

Walter, E. and L. Pronzato (1985). How to design experiments that are robust to parameter uncertainty. Vol. 7 of *IFAC Proc. Ser.*. pp. 921–926.