Bayesian Estimation of the Active Concentration and Affinity Constants Using Surface Plasmon Resonance Technology

Feng Feng[1]*, **Thomas B. Kepler[1,2]**

1 Department of Microbiology, Boston University School of Medicine, Boston, Massachusetts, 02118, United States of America, **2** Department of Mathematics & Statistics, Boston University, Boston, Massachusetts, 02118, United States of America

* feng.feng@bu.edu

Abstract

Surface plasmon resonance (SPR) has previously been employed to measure the active concentration of analyte in addition to the kinetic rate constants in molecular binding reactions. Those approaches, however, have a few restrictions. In this work, a Bayesian approach is developed to determine both active concentration and affinity constants using SPR technology. With the appropriate prior probabilities on the parameters and a derived likelihood function, a Markov Chain Monte Carlo (MCMC) algorithm is applied to compute the posterior probability densities of both the active concentration and kinetic rate constants based on the collected SPR data. Compared with previous approaches, ours exploits information from the duration of the process in its entirety, including both association and dissociation phases, under partial mass transport conditions; do not depend on calibration data; multiple injections of analyte at varying flow rates are not necessary. Finally the method is validated by analyzing both simulated and experimental datasets. A software package implementing our approach is developed with a user-friendly interface and made freely available.

Introduction

Surface plasmon resonance (SPR) was first applied to the study of molecular binding reactions in biology and chemistry in the 1980s [1–3]. The first commercial SPR instrument was introduced by BIACore in 1991 [4, 5]. Since then, instruments using SPR have gained increasing popularity due to their high sensitivity and simple construction, and have become the accepted standard for measuring the kinetic rate constants for molecular binding interactions. SPR technology has many practical advantages: 1) it allows real-time detection of binding events; 2) no labeling is required; 3) it can generally be applied to binding reactions of many types, such as protein-protein, protein-peptide, protein-DNA, and protein-small molecule; 4) analysis can be carried out on colored or turbid samples without interference from absorption and scattering.

Since SPR measures function—binding between analyte and ligand—it is able to determine active concentration, which is not necessarily identical to the total concentration as measured, for example by optical density. Two approaches have been proposed. The more widely used approach is the "calibration-dependent" method [6–8]. Each quantification run, however, requires a new calibration curve, thereby increasing the time and total material cost. The other approach does not rely on explicit calibration. Christensen [9] developed the mathematical theory and computed the analytical solutions under the partial mass transport conditions for molecular interactions within a two-compartment model. This approach has proven to be useful in applications [8, 10]. The method, however, has a few restrictions. It only exploits the initial binding phase of the sensorgram, requires several injections of analyte at various flow rates for each unknown sample, and assumes low noise levels on response signals in order to accurately determine the rates of change. These restrictions limit the applicability of this method. Sigmundsson *et al.* in 2005 [11] proposed a more general solution for the same process, and derived a complete analytical solution over the entire binding phase [9]. The solution involves Lambert's *W* function, a special function that is not available in many common statistical software packages. Moreover, no user-friendly software implementation has become available as a result of either effort.

In this work, we develop a Bayesian approach to estimating the active concentration and kinetic rate constants from SPR experimental data using the same two-compartment model and quasi-steady state approximation used by Christensen and Sigmundsson [9, 11]. Bayesian statistical inference has been widely used in many biomedical applications and has been proved useful for accounting for multiple sources of uncertainty arising from experimental variation, instrumental noise, etc. [12–16]. In addition, it does not require point estimation or linearization of variations, either of which can lead to inconsistency in the estimation in nonlinear models [17, 18]. Inference of model parameters is given by the posterior density conditional on observed data. Here, we supply prior probabilities on the parameters and derive a likelihood function which together with the SPR data themselves allow us to compute the posterior probability densities on the active concentration and kinetic rate constants. This approach exploits information from the duration of the process in its entirety, including both association and dissociation phases, under partial mass transport conditions. Compared with the two approaches mentioned above, ours does not need multiple injections of analyte at varying flow rates as long as the partial mass transport conditions are met. Finally our methods have been implemented in software and are freely available.

Theory

The theory for molecular binding under conditions of partial mass transport in a two-compartment model was worked out by Christenson [9] and is briefly recapitulated here. If A_0 is the analyte in the bulk flow, A_s is the analyte on the biosensor surface, B is the ligand immobilized on the surface and AB is the analyte-ligand complex, the binding process in the SPR biosensor can be described by the subprocesses,

$$A_0 \quad \underset{k_M}{\overset{k_M}{\rightleftharpoons}} \quad A_S \qquad (1)$$

and

$$A_S + B \quad \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \quad AB \qquad (2)$$

where k_M is the mass transport coefficient, and k_a and k_d are the association and dissociation

rate constants, respectively. Process (1) describes the mass transport of the analyte between the bulk flow (A_0) and surface (A_s); process (2) describes the binding and unbinding of analyte and ligand on the surface matrix.

The mass action equations for these processes are

$$h_{diff}\frac{d[A_S]}{dt} = k_M([A_0] - [A_S]) - k_a[A_S][B] + k_d[AB] \tag{3}$$

and

$$\frac{d[AB]}{dt} = k_a[A_S][B] - k_d[AB] \tag{4}$$

where h_{diff} is the characteristic height of the diffusion layer, given [19] approximately by

$$h_{diff} \approx \sqrt[3]{\frac{Dh^2wl}{F}} \tag{5}$$

D is the analyte diffusion coefficient, F is the bulk flow rate and h, w and l are the height, width and length of the flow chamber of the SPR system, respectively [19]. The diffusion coefficient D can be estimated by Stoke's law and Einstein-Sutherland equation [20] (See details in S1 Appendix).

Under the quasi-steady state condition (where $d[As]/dt = 0$ in Eq (3) is assumed) [21], the analyte concentration at the biosensor surface matrix can be approximated by:

$$[A_s] = \frac{k_M[A_0] + k_d[AB]}{k_M + k_a[B]} \tag{6}$$

which describes the average of the concentration gradient that forms along the surface [21]. Combining Eqs (4) and (6) and rearranging give

$$\frac{d[AB]}{dt} = \frac{k_a[A_0][B] - k_d[AB]}{1 + k_a[B]/k_M} \tag{7}$$

which has been evaluated and demonstrated to be a reasonable approximation for most conditions with the partial mass transport effect [9, 11]. Under experimental conditions, the analyte bulk concentration $[A_0]$ remains constant by a continuous injection of fresh analyte solution at a fixed flow rate. The amount of free ligand $[B]$ decreases with time until steady-state is reached. Furthermore, the density of free ligand can be expressed through the conservation relation as

$$[B] = [B_{\max}] - [AB] \tag{8}$$

where $[B_{max}]$ is the density of total ligand (bound and unbound) on the surface and $[AB]$ is the density of bound complex.

The mass transport coefficient, k_M, is approximately proportional to the cube root of the flow rate [22],

$$k_M = C_{k_M}\sqrt[3]{\frac{D^2F}{h^2wl_2}} \tag{9}$$

and

$$C_{k_M} = 1.47\frac{1 - (l_1/l_2)^{2/3}}{1 - l_1/l_2} \tag{10}$$

where l_1 and l_2 are the lengths to the start and end, respectively, of the detection area from inlet of the flow cell [22]. For a BIAcore system, these values are known. Therefore, if the molecular weight and the bulk flow rate of the analyte are also given, k_M can be obtained. When l_1 and l_2 or other parameters (such as h, w or l) are not easily obtained for a biosensor system other than BIAcore instruments, k_M can still be estimated empirically (see discussions in the Results section).

The response signal, R, output by an SPR sensor is proportional to the amount of complex formed at the biosensor surface with an empirical factor given by

$$R = W_M \, G[AB] \tag{11}$$

where G is the response per mass per area for proteins with an approximate value of 1000 RU·mm^2/ng [23] and W_M is the molecular weight of analyte. Therefore, Eq (7) now can be rewritten as

$$\frac{d[R]}{dt} = \frac{k_a[A_0]([R_{max}] - R) - k_d[R]}{1 + k_a([R_{max}] - R)/k_M} \tag{12}$$

where R_{max} is the maximum value of the response signal when all the immobilized ligands have been fully bound into complexes, and k_M is rescaled to have unit of $RU/M/s$ instead of m/s.

The mass transport limitation for a specific analyte-ligand system is determined by both the amount of free immobilized ligand and the bulk flow rate of the analyte (k_M). A limit coefficient, $k_a[B]/k_M$, can be obtained from Eq (7) to identify conditions for mass transport-limited binding and for kinetic binding (see S1 Appendix for details). Practically, zero and full mass transport limitation are not easily obtained with a given analyte-ligand system, since to approach these extreme conditions it is necessary to use a wide range of ligand concentrations. On the other hand, it is relatively easy to obtain partial mass transportation limitation. In the next section we describe a Bayesian approach to determining the active concentration as well as the rate constants under the partial mass transport conditions of SPR responses.

Bayesian Inference

We use Bayesian methods to obtain the posterior distributions of active concentrations, A_0, association/dissociation rate constants, k_a and k_d, the theoretical maximum response signal, R_{max}, and the initial response level at the beginning of the dissociation phase, R_0. As discussed in the previous section, the analyses are performed on SPR data assuming partial mass transport limitation conditions described by the differential equation Eq (12). The initial-value problem for Eq (12) does not have a closed-form solution, therefore we carry out inference on its numerical solution.

In addition to the process model, we must have a measurement, or statistical, model. The statistical model for the ith SPR sensorgram signal R_i conditional on the parameter vector $\theta \equiv \{A_0, k_a, k_d, \ldots\}$ is

$$R_i = r(t_i, \theta) + \varepsilon_i \tag{13}$$

where the errors ε_i are independent, identically-distributed Gaussian random variables with mean zero and variance σ^2.

$$\varepsilon_i \sim N(0, \sigma^2) \tag{14}$$

The following diffuse prior distributions are chosen [24, 25].

$$k_a \sim N(5 \times 10^{11}, 10^{12}) \tag{15}$$

$$k_d \sim N(50, 10^2) \tag{16}$$

$$R_{\max}, R_0 \sim N(5 \times 10^4, 10^5) \tag{17}$$

$$\sigma^2 \sim InvGamma(0.1, 0.1) \tag{18}$$

The variances for the normal distributions are set so to make the prior distributions sufficiently diffuse/non-informative. Given the amount of observed data in a typical SPR experiments, the posterior distributions of these parameters are insensitive to modest variation in the prior distributions.

For estimation of the posterior density, we use the Markov Chain Monte Carlo (MCMC) method with the Gibbs sampler to generate samples from the posterior distributions of parameters as follows.

- Define the full conditional distributions [26] as

- $\{\pi(\theta_1 \mid \theta_2, \theta_3, ... \theta_p); \pi(\theta_2 \mid \theta_1, \theta_3, ... \theta_p); ..... \pi(\theta_p \mid \theta_1, \theta_2, ... \theta_{p-1})\}$

One cycle of the Gibbs sampler is completed by drawing $\{\theta_k\}_{k=1}^p$ from these distributions successively updating the conditional variables

Algorithm

1. define the initial values, $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, ... \theta_p^{(0)})$;

2. repeat for j = 1.....M

   Generate $\theta_1^{(j+1)}$ from $\pi(\theta_1 \mid \theta_2^{(j)}, \theta_3^{(j)}, ... \theta_p^{(j)})$;

   Generate $\theta_2^{(j+1)}$ from $\pi(\theta_2 \mid \theta_1^{(j+1)}, \theta_3^{(j)}, ... \theta_p^{(j)})$;

   . . .. . .

   Generate $\theta_p^{(j+1)}$ from $\pi(\theta_p \mid \theta_1^{(j+1)}, \theta_2^{(j+1)}, ... \theta_{p-1}^{(j+1)})$;

3. return values $\{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, ... \theta^{(M)}\}$

Following an equilibration period, the Markov chain approaches its stationary distribution and samples from the MCMC closely approximate samples from the posterior parameter distribution. Using these samples, the posterior means and independent 95% Bayesian credible intervals (BCI) from the posterior marginal densities are computed.

Combining the priors and observed data, the full conditional probability can be formulated as

$$k_a : \pi(k_a | k_d, A_0, R_{\max}, R_0, \sigma^2, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{(k_a - 5 \times 10^{11})^2}{2 \times 10^{12}}} \tag{19}$$

$$k_d : \pi(k_d | k_a, A_0, R_{\max}, R_0, \sigma^2, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{(k_d - 50)^2}{2 \times 10^2}} \tag{20}$$

$$A_0 : \pi(A_0|k_a, k_d, R_{\max}, R_0, \sigma^2, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{(A_0 - 50)^2}{2 \times 10^2}} \tag{21}$$

$$R_{\max} : \pi(R_{\max}|k_a, k_d, A_0, R_0, \sigma^2, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{(R_{\max} - 5 \times 10^4)^2}{2 \times 10^5}} \tag{22}$$

$$R_0 : \pi(R_0|k_a, k_d, A_0, R_{\max}, \sigma^2, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{(R_0 - 5 \times 10^4)^2}{2 \times 10^5}} \tag{23}$$

$$\sigma^2 : \pi(\sigma^2|k_a, k_d, A_0, R_{\max}, R_0, R_t, t) \propto \prod_i e^{-\frac{(R_t - r(t,\theta))^2}{2\sigma^2}} \times e^{-\frac{0.1}{2\sigma^2}} \times \sigma^{-n-2.2} \tag{24}$$

Because we are using the numerical solution to a dynamical system initial value problem, we employ the adaptive rejection with Metropolis sampling (ARMS) algorithm [27, 28] with modifications [29], which allows one to draw samples from an arbitrary distribution. The overall inference processes have been summarized in Fig 1 flow chart.

Software Development

The software implementation of the proposed approach was developed in C# for the .NET framework. The executable is available for free distribution. The program has a simple and intuitive graphical interface. Fig 2 shows a screenshot of the interface and illustrates its various components. The input of the program is through text data files. Users can specify the initial values of the parameters through the interface. The results are then written to files with simulated posterior distributions, allowing further analysis such as MCMC diagnostics, hypothesis testing and the fitting of models for biological processes.

Because the proposed approach requires numerical solution of the initial value problem for differential equation Eq (12), the analysis is a computation-intensive task. Special optimization algorithms have been incorporated to make the software efficient, e.g. the forth-order Runge-Kutta algorithm for calculating numerical solution and the Nelder-Mead algorithm for finding maxima and minima of functions.

Experimental Procedures

Reagents

Cabonic anhydrase isozyme II (CAII) (MW = 30kD) from bovine erythrocytes, 4-carboxyben-zenesulfonamide (4-CBS) (MW = 201.2D) and TWEEN 20 were purchased from Sigma (St. Louis, MO). N-(3-dimethylaminopropyl-N'-ethylcarbodiimide (EDC), N-hydroxysuccini-mide (NHS), ethanolamine-HCl and sodium acetate were purchased from Bio-Rad (Hercules, CA).

Surface plasmon resonance measurements

SPR measurements were carried out on a SensiQ Pioneer biosensor (SensiQ Technologies, Inc., OK). The CHOO2 two dimensional surface chip was used in the assay. The ligand CAII was immobilized via the common amine coupling chemistry. This involved activating the chip surface with freshly made 2mM EDC and 0.5mM NHS (as suggested by the manufacturer), injecting CAII in sodium acetate (pH 5.0) and then blocking excess reactive esters with etha-nolamine. Subsequently, the analyte 4-CBS was injected with a concentration of 50µM at a
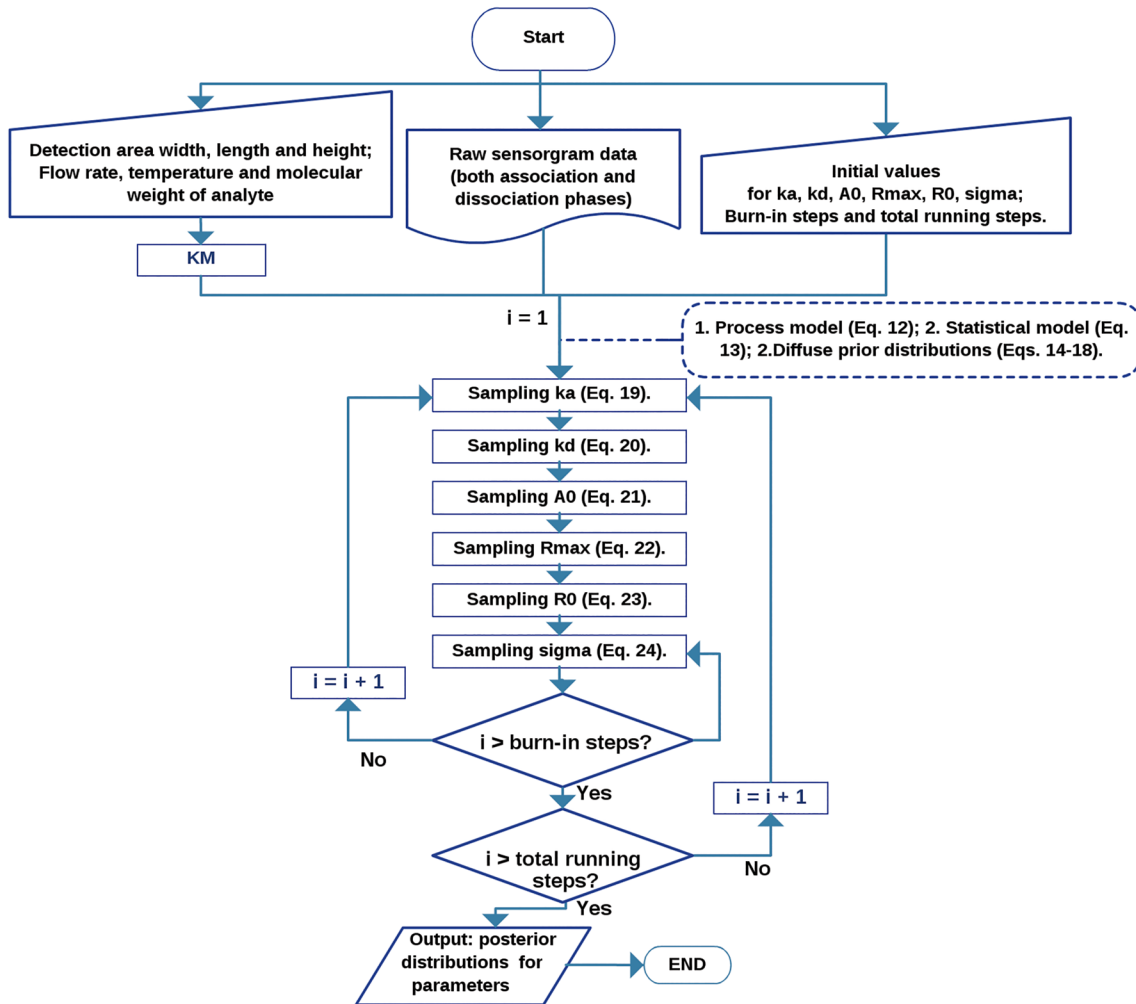
Fig 1. Flow chart summarizing the inference processes in our proposed Bayesian approach.

flow rate of 50μl/min for 60 seconds and followed with a dissociation phase for another 60 seconds. The interaction was carried out in PBS with 0.05% TWEEN 20 (pH 7.0) (PBST) at 25°C.

Data analysis

The SPR response data were first processed in SensiQ Qdat (version 2.2) software for background subtraction, and then were exported as text files for analyses using the software development in this work.

Results

We validated the method using simulated SPR response data, experimental SPR data generated in our laboratory, and publicly available experimental data.

Validation on Simulated Data

The proposed Bayesian approach was validated by analyzing simulated SPR response data. To simulated SPR data, a strategy similar to the one employed by Karlsson were adopted [30]. Eq
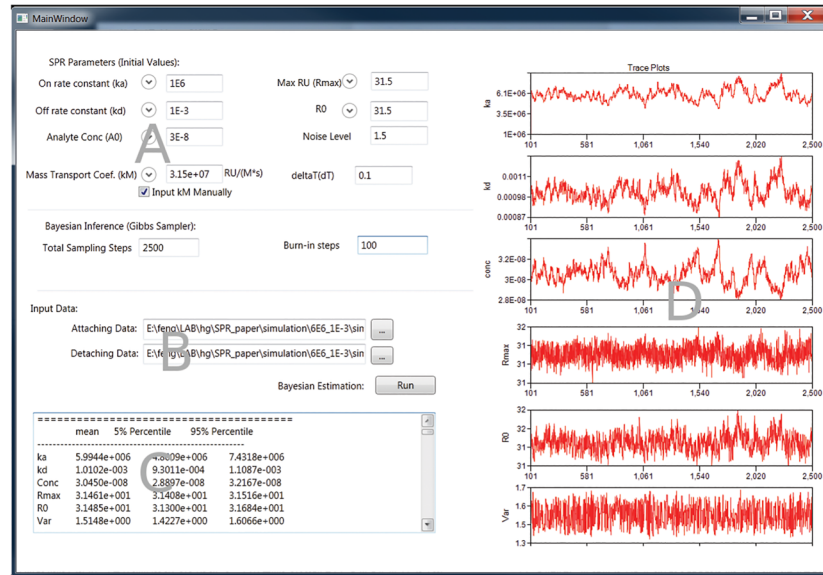
Fig 2. Screen Shot of the implemented software. The implemented software has the interface to allow the user to define the initial values of the parameters (A) and specify the input files (B) as well as controlling the behavior of the Gibbs sampler inference. It also has the interface to show the summary statistics (C) and trace plots of estimations (D).

doi:10.1371/journal.pone.0130812.g002

([12]) was used to simulate interaction data for reactions with $k_M = 3.15 \times 10^7$ RU/M/S. This k_M value is similar to that in the Inogatran/Thrombin system [30]. The concentration of analyte was set to 30nM, the maximum immobilized ligand in terms of the response signal (R_{max}) was 31.5RU, k_a values varied between 1×10^5 and 3×10^8 /M/s, k_d values were adjusted so that the affinity, K_D, was held constant at 1nM. The response curves in Fig 3 represent binding curves with differing kinetic rate constants, but constant K_D. The Gaussian noise at a level of 1.5RUs, which was equivalent to that (1~3RUs) seen on Bio-Rad Proteon XPR36 biosensors but bigger than that on BIACORE and SensiQ systems, was added to the simulated data as specified by Eqs ([13]) and ([14]).

The simulated data were then analyzed using the proposed Bayesian method. The parameter values returned were summarized in Table 1, and the chain trace plots for one set of data ($k_a*R_{max}/k_M = 3.0$) are shown in Fig 4. The proposed approach did not return accurate k_a, k_d or A_0 under the situations of little or no mass transport effect (case i and ii, where k_a*R_{max}/k_M <1), although R_{max} and noise level σ^2 could be determined correctly under the same conditions. As the mass transport effect started to dominate, k_a, k_d and A_0 as well as others (R_{max} and noise level) could all be determined accurately (cases iii~vi, where k_a*R_{max}/k_M was between 1~100). In case of very high mass transport effect conditions (case vii and viii, where k_a*R_{max}/k_M >100), k_a and k_d estimations became inaccurate while the affinity constant K_D and other parameters were still correctly determined. Similar behavior has been reported previously [30].

Validation on experimental data from our laboratory

The proposed approach was further tested with experimental data. The enzyme carbonic anhydrase (CAII) and its small molecule inhibitor 4-carboxybenzenesulfonamide (4-CBS) were chosen to be studied on a SensiQ Pioneer biosensor. In the experiment, about 3000RU of CAII was amine coupled to the chip surface as the ligand and then 4-CBS was injected at a flow rate of 50μl/min. This experiment setting resulted in a partial mass-transport limited condition for
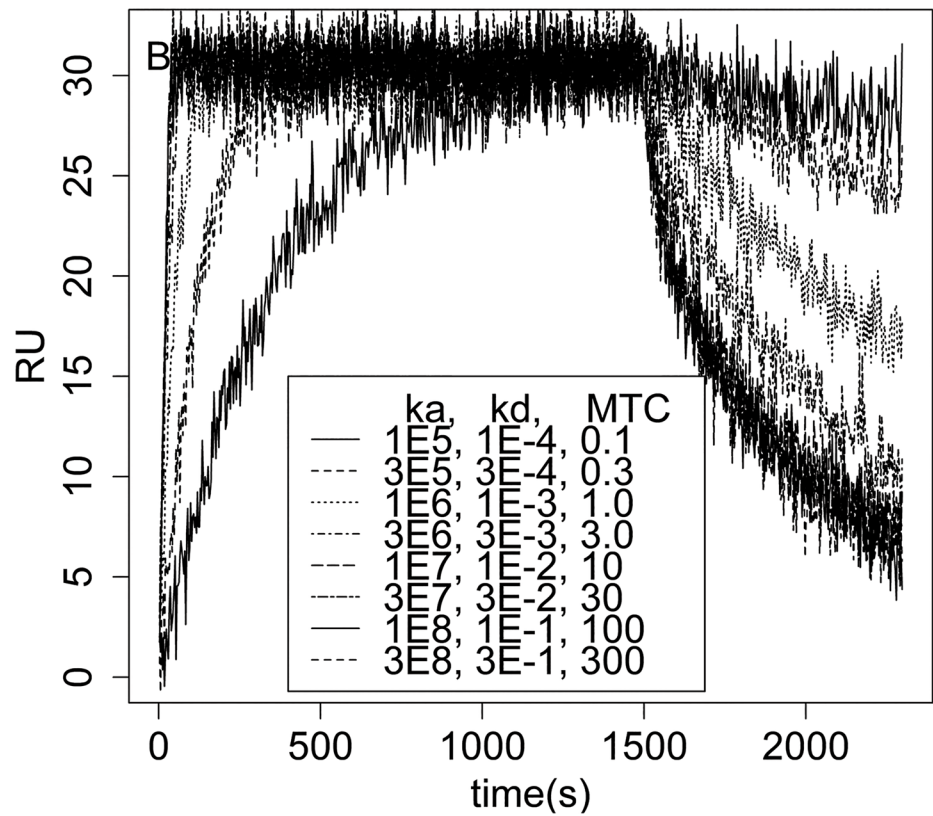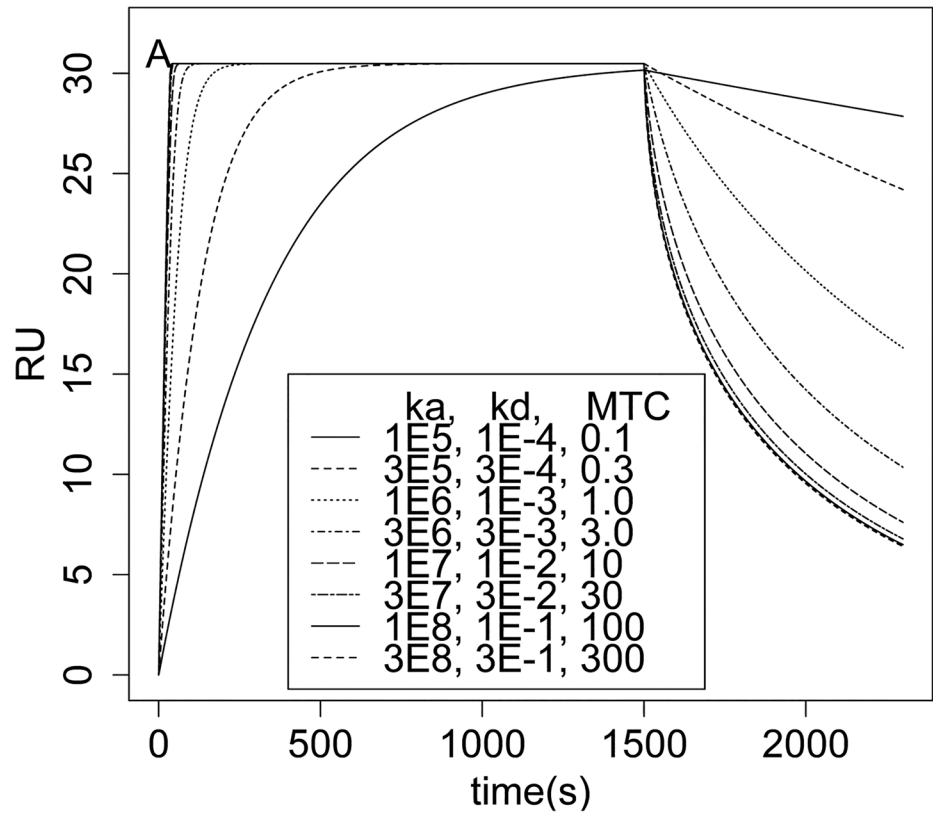
Fig 3. Simulated SPR response data. (A), the SPR response data were simulated according to Eqs (12), (13) and (14) with no measurement error. In simulations R_{max} was 31.5 RU, A_0 was 30nM, k_a values varied between 1×10^5 and 3×10^8/M/s, k_d values were adjusted so that K_D was constant 1nM and k_M was 3.15×10^7 RU/M/s. The mass transport limiting coefficients (MTLC) of these responses varied between 0.1 (little/no mass transport effect) and 300 (high/full mass transport effect). (B), same as in A, but measurement error with a constant standard deviation of 1.5 RUs was added to the simulated data. The level of noise was chosen based on empirical data obtained from the Bio-Rad Proteon XPR36 biosensor in our lab through a variety of antibody-antigen interactions (1~3 RUs in our experiments).

doi:10.1371/journal.pone.0130812.g003

the interaction, and generated small but detectable responses (Fig 5). The data were then exported and analyzed to determine the active concentration and the kinetic constants of the analyte. First, the mass transport coefficient k_M, which primarily depended on the diffusion coefficient, the dimensions of the flow cell and the flow rate, had to be determined. Under the current experimental condition, the diffusion coefficient D of the analyte, 4-CBS (molecular weight = 201.2 Dalton), was calculated to be 4.76×10^{-10} m^2/s [20] (Table A in S1 Appendix). Furthermore, the flow cell dimensions of the SensiQ Pioneer biosensor were given with a width

Table 1. Parameter values estimated by the proposed Bayesian method using the simulated SPR response Data.

	k_a (1/M/s)	k_d (1/s)	A_0 (nM)	R_{max} (RU)	K_D (nM)	k_a*R_{max}/k_M
	Parameters in simulation					
I	1.0×10^5	1.0×10^{-4}	30.0	31.5	1.0	0.1
Ii	3.0×10^5	3.0×10^{-4}	30.0	31.5	1.0	0.3
iii	1.0×10^6	1.0×10^{-3}	30.0	31.5	1.0	1.0
iv	3.0×10^6	3.0×10^{-3}	30.0	31.5	1.0	3.0
V	1.0×10^7	1.0×10^{-2}	30.0	31.5	1.0	10.0
vi	3.0×10^7	3.0×10^{-2}	30.0	31.5	1.0	30.0
vii	1.0×10^8	1.0×10^{-1}	30.0	31.5	1.0	100.0
viii	3.0×10^8	3.0×10^{-1}	30.0	31.5	1.0	300.0
	Parameters estimated by the proposed approach					
I	$1.2 \times 10^5 [1.1 \times 10^5, 1.5 \times 10^5]$	$1.1 \times 10^{-4} [0.8 \times 10^{-4}, 1.3 \times 10^{-4}]$	25.3[20.9, 26.3]	31.0[29.9, 31.9]	0.9[0.63, 1.06]	
Ii	$1.6 \times 10^5 [1.4 \times 10^5, 1.9 \times 10^5]$	$3.1 \times 10^{-4} [3.0 \times 10^{-4}, 3.3 \times 10^{-4}]$	49.8[45.3, 56.5]	31.8[31.6, 32.0]	1.8[1.66, 2.10]	
iii	$0.95 \times 10^6 [0.8 \times 10^6, 1.1 \times 10^6]$	$1.0 \times 10^{-3} [0.9 \times 10^{-3}, 1.0 \times 10^{-3}]$	31.2[27.9, 35.0]	31.5[31.3, 31.7]	1.0[0.89, 1.22]	
iv	$3.3 \times 10^6 [2.9 \times 10^6, 3.8 \times 10^6]$	$3.1 \times 10^{-3} [2.9 \times 10^{-3}, 3.4 \times 10^{-3}]$	28.9[27.3, 30.4]	31.5[31.4, 31.6]	1.1[0.86, 1.20]	
V	$1.1 \times 10^7 [0.9 \times 10^7, 1.4 \times 10^7]$	$1.1 \times 10^{-2} [0.9 \times 10^{-2}, 1.3 \times 10^{-2}]$	29.1[28.2, 30.2]	31.5[31.4, 31.5]	1.0[0.93, 1.03]	
vi	$4.0 \times 10^7 [2.3 \times 10^7, 7.8 \times 10^7]$	$3.9 \times 10^{-2} [2.4 \times 10^{-2}, 7.3 \times 10^{-2}]$	29.6[28.7, 30.5]	31.5[31.5, 31.6]	1.0[0.94, 1.03]	
vii	$0.7 \times 10^8 [0.4 \times 10^8, 1.0 \times 10^8]$	$0.7 \times 10^{-1} [0.5 \times 10^{-1}, 1.0 \times 10^{-1}]$	30.3[29.8, 30.7]	31.5[31.5, 31.6]	1.0[1.00, 1.04]	
viii	$0.6 \times 10^8 [0.2 \times 10^8, 1.1 \times 10^8]$	$0.6 \times 10^{-1} [0.3 \times 10^{-2}, 1.1 \times 10^{-2}]$	30.9[30.1, 31.8]	31.5[31.5, 31.6]	1.1[1.01, 1.10]	

The SPR response data were simulated according to Eqs (12), (13) and (14) with $k_M = 3.15 \times 10^7$ as in Fig 1. The parameters used in the simulation were specified in the top section of the table. Then the proposed approach was applied to estimate the parameters, which were listed in the bottom section in the table with the 95% Bayesian credible intervals (in square brackets). The parameters in the simulation ranged from little mass transport limited conditions ($k_a * R_{max}/k_M = 0.1$) to high/full mass transport limited conditions ($k_a * R_{max}/k_M = 300$).
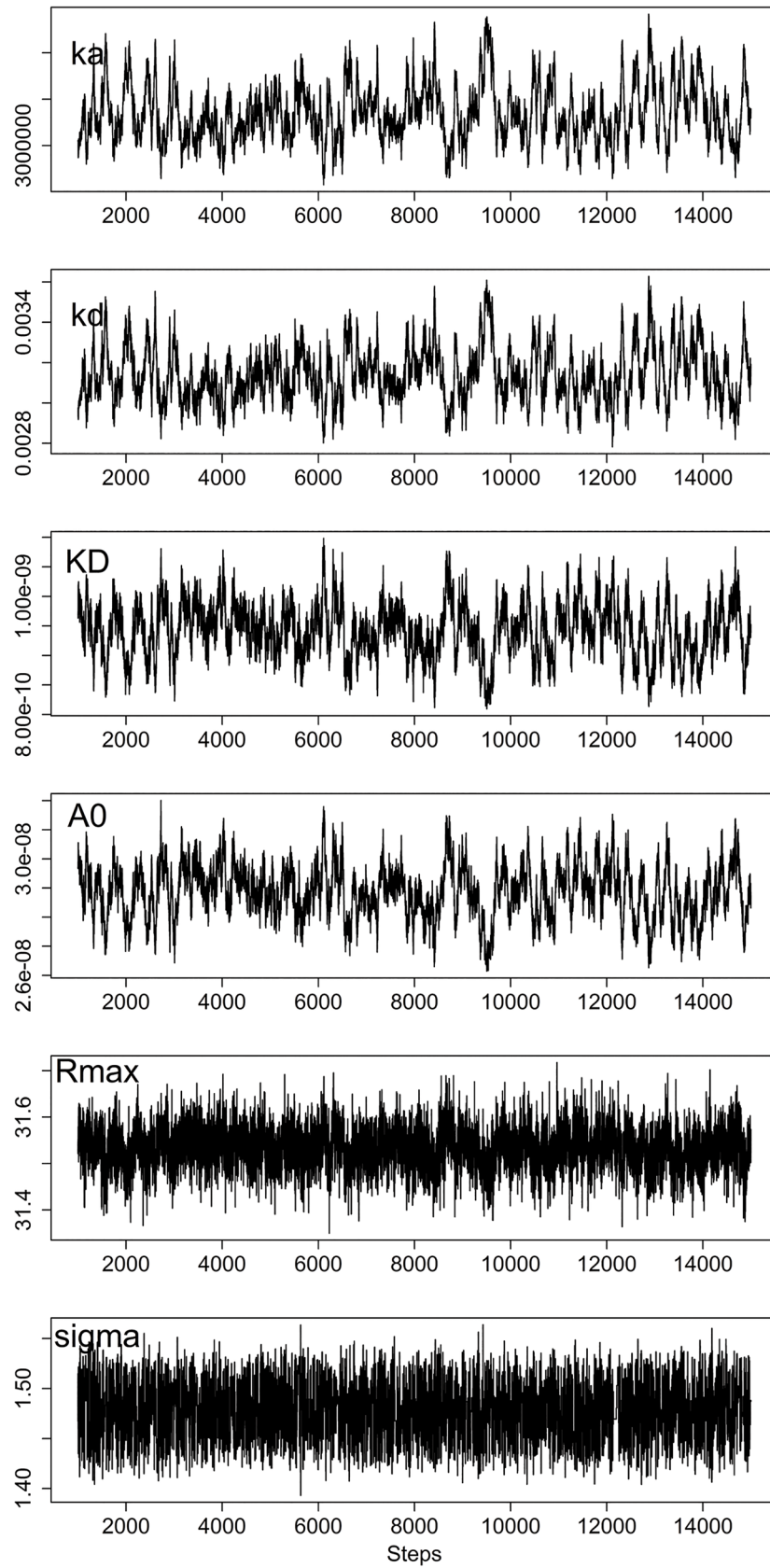
doi:10.1371/journal.pone.0130812.t001

Fig 4. The trace plots for evaluation of the simulated SPR response data. The SPR response data were simulated as in Fig 2 with R_{max} 31.5RU, k_a 3x10^6/M/s, k_d 3x10^{-3}/s, k_M 3.15x10^{7}RU/M/s, A_0 30nM and σ^2 1.5. The data were analyzed as described in the text. The trace plots are shown for six parameters in the analysis of the data set with $k_a * R_{max}/k_M = 3.0$. The returned parameters are summarized in Table 1.

doi:10.1371/journal.pone.0130812.g004

0.635mm, a height 0.05mm and a length 3.0mm by the manufacturer. Then with a flow rate of 50μl/min, k_M was determined to be 1.03x10^7 RU/(Ms) according to Eqs (9) and (10) (Table A in S1 Appendix). Finally, the kinetic parameters and the active analyte concentration were estimated by the proposed Bayesian approach and shown in Table 2. The original sensorgram and the fitted values were overlaid in Fig 5, which showed a very good fit. The trace plots of each parameter in the analysis were shown in Fig A in S1 Appendix.
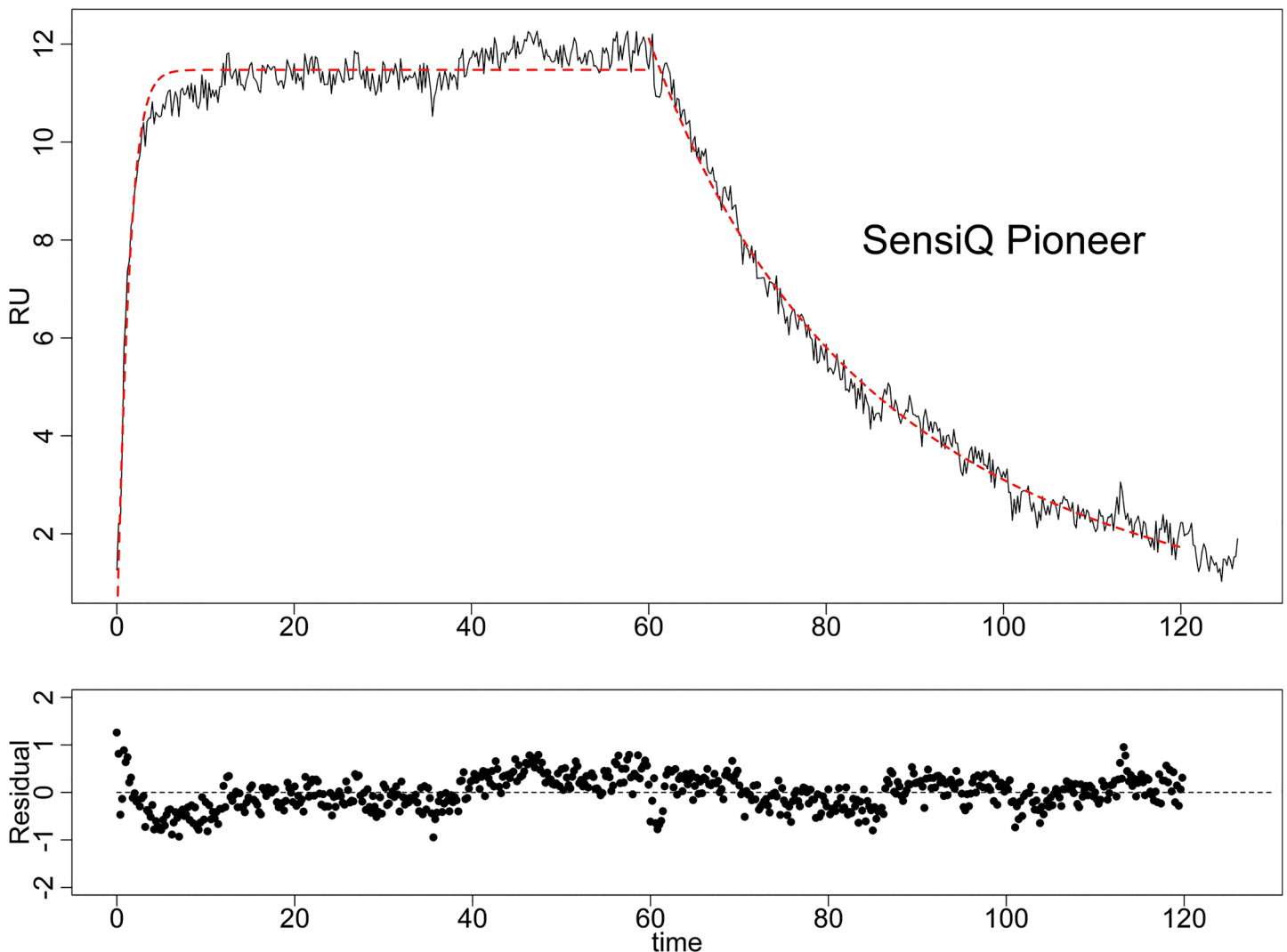

Fig 5. The kinetic analysis of the emzyme CAII and its small molecule inhibitor 4-CBS in the SensiQ Pioneer biosensor. CAII was immobilized by the amine coupling chemistry at a level of about 3000RU, and then 50μM 4-CBS was injected at a flow rate of 50μl/min at 25˚C. The generated sensorgram data were exported and analyzed by the proposed Bayesian approach. The diffusion coefficient and mass transport coefficient were first derived, and then the kinetic parameters as well as the active concentration were determined. The fitted values of the responses were overlaid with the original sensorgram (the red dashed line) on the top figure, and the residuals were showed on the bottom.

doi:10.1371/journal.pone.0130812.g005

The active concentration A_0 of the analyte was determined in our analysis to be 2.0 ± 0.22 μM, which was about 4% of the total concentration (50μM of the analyte). This percentage of the active concentration in the total concentration was similar to that of the active phosphorylated GST-CEACAM1-4L protein determined by both SPR and other radioactivity based method [11]. Moreover, our analysis showed that the dissociation rate constant k_d was $(4.23 ± 0.12) \times 10^{-2} \, s^{-1}$, which is consistent with what has been reported in the literature $((3.65 ± 0.06) \times 10^{-2} \, s^{-1})$ [31]. On the other hand, the estimated association rate constant k_a, $(4.67 ± 0.66) \times 10^5 \, M^{-1}s^{-1}$, was about ten times higher than what was reported $((4.8 ± 0.2) \times 10^4 \, M^{-1}s^{-1})$ in an analysis that did not take into account the difference between active concentration and total concentration [31]. Thus, the discrepancy is not a surprise since the active concentration was more than ten times lower than the total concentration. When we used the total concentration of analyte (50μM) rather than the active concentration, we get a value for k_a $(1.4 \times 10^4 \, M^{-1}s^{-1})$ that was similar to the previously-reported value. We argue that it is as more appropriate to use the active concentration under these circumstances, and that this comparison illustrates the importance of doing so.

Validation on publicly available data

In order to further validate the proposed approach, we analyzed two publicly available experimental data sets generated using other popular biosensors. The first dataset was from Castonguay et al.[32], in which the bone morphogenetic protein 10 (BMP10) interacted with immobilized human endoglin extracellular domain-Fc chimera (hEng^{ECD}-FC) in a Biacore 3000 system. The second dataset was generated on a Proteon XPR36 biosensor from the work of Abdiche et al.[33], in which human calcitonin gene-related peptide-α (CGRPα) reacted with immobilized 4901 IgG. These two datasets were selected because they were generated under partial mass transport-limited conditions and the molecular weights of the analytes in them were known. With the latter information, the diffusion coefficient and then the mass transport coefficient could be derived from Eqs (5), (9) and (10) (also see details in S1 Appendix about the diffusion coefficient D). Therefore, the proposed approach could be applied to estimate the active concentration of analyte A_0 and rate constants k_a and k_d. The results are listed in Table 2, together with parameter estimates from the original reference papers for comparison.

In the first case, the SPR response data between BMP10 and hEng^{ECD}-hFC were generated on a Biacore 3000 system, in which the detection area of the flow chamber has been very well documented. The parameters, such as the width, length and height of the measured SPR, are known [9, 11] (Table A in S1 Appendix). The diffusion coefficient was approximated using the available molecular weight of the analyte BMP10 (see details in S1 Appendix). The mass transport coefficient k_M was thus determined to be approximately 8×10^8 RU/M/s (Table A in S1 Appendix) and used to estimate the active analyte concentration and the kinetic rate constants (Table 2 and Fig B in S1 Appendix). The fitted curve was overlaid with the raw sensorgrams in Fig 6A. In the second dataset, the interaction between hCGRPα and 4901 IgG was measured on a Proteon XPR36 system. In this case, the parameters for the detection flow chamber were not available. Similar parameters as in the Biacore systems, however, can still be assumed to approximate k_M; this has been verified empirically (data not shown). As shown in Table A in S1 Appendix, the mass transport coefficient k_M between analyte hCGRPα and immobilized 4901 IgG was determined to be about 8.5×10^7 RU/M/s. Finally, the active concentration and rate constants were estimated (Table 2 and Fig C in S1 Appendix) and the fitted curved were showed in Fig 6B. Overall, the fitted SPR responses based on the estimated parameters matched the observed sensorgrams very well in both cases.

Table 2. Parameter values estimated by the proposed Bayesian approach using the experimental SPR response data.

	k_a (1/M/s)	k_d (1/s)	A_0 (nM)	K_D (nM)
Dataset #1 (SensiQ Pioneer)				
in literature [31]	4.8×10^4	3.65×10^{-2}	-	760
by the proposed approach	$4.67 \times 10^5 [3.66 \times 10^5, 5.84 \times 10^5]$	$4.2 \times 10^{-2} [4.04 \times 10^{-3}, 4.45 \times 10^{-3}]$	2020[1705, 2415]	92.1[76, 110]
Dataset #2 (Biacore 3000)				
in the reference [32]	5.47×10^7	3.9×10^{-3}	1.25	0.0713
by the proposed approach	$2.30 \times 10^8 [2.16 \times 10^8, 2.44 \times 10^8]$	$4.4 \times 10^{-3} [4.30 \times 10^{-3}, 4.65 \times 10^{-3}]$	0.612[0.61, 0.62]	0.01946[0.0191, 0.0199]
Dataset #3 (Proteon XPR36)				
in the reference [33]	1.15×10^5	6.71×10^{-4}	300	5.780
by the proposed approach	$3.6 \times 10^7 [2.7 \times 10^7, 4.7 \times 10^7]$	$8.2 \times 10^{-4} [7.5 \times 10^{-4}, 9.2 \times 10^{-4}]$	13.73[13.4, 14.1]	0.023[0.018, 0.030]

Three tests were carried out. In the first experiment, the interaction data between CAII and 4-CBS were studied on a SensiQ Pioneer biosensor under a partial mass-transport limited condition. The data were analyzed and the resulted kinetic parameters and active concentration. In the other two tests, two datasets were collected from the literature. Dataset #1 was from the work done by Castonguay et al.[32], in which BMP10 interacted with immobilized human Eng^ECD-Fc in a Biacore 3000 system. Dataset #2 was from a different work done by Abdiche et al.[33], in which human CGRPα bound immobilized 4901 IgG in a Proteon XPR36 biosensor. The two datasets were then analyzed by the proposed approach to estimate the active concentration of analyte A_0, rate constants k_a and k_d, and the dissociation constant K_D. The returned parameters as well as the parameters estimated in the original reference papers were listed. The 95% Bayesian credible intervals for the estimation were also provided in the square brackets. The biosensor used to collect the experimental data was indicated in the parentheses.

doi:10.1371/journal.pone.0130812.t002

As shown in Table 2, our method obtained similar dissociation rate constant k_d but different association rate constant k_a compared with the ones estimated in the original reference papers in both tests. Again, this was reasonable, since k_d estimation didn't depend on the actual analyte concentration and can be determined correctly without knowing it. On the other hand, k_a estimation relied on the accurate information of the analyte concentration, more precisely the active concentration of the analyte. In the original papers, the total analyte concentrations were taken as a known input in the estimation. These values, however, were different from the active concentrations of analyte. In the first dataset, the active concentration was determined to be about 50% of the total concentration, and the difference between the two concentrations resulted in a higher and probably more accurate estimation of association constant k_a. In the second dataset, the difference in k_a estimation came in not only due to the difference between the active concentration in our approach and the total concentration in the original, but also due to the way in which the data analysis was carried out in the original paper. In the reference paper, a Langmuir model that assumed no mass transport effect was used. Under these conditions, one expects to underestimate the kinetic rate constants. The proposed Bayesian method assumed a model with non-negligible mass transport and was thus more likely to produce more accurate estimates.

In conclusion, the proposed Bayesian approach in this work, assuming a partial mass transport-limited model and exploiting the activate analyte concentration, led to more accurate estimation of kinetic rate constants than was obtained using other methods

Software Availability

Project name: SPR_MCMC; Projection home page: https://sourceforge.net/projects/sprmcmc/; Operating systems: Windows; Programming Language: C#; License: Free for academic use.
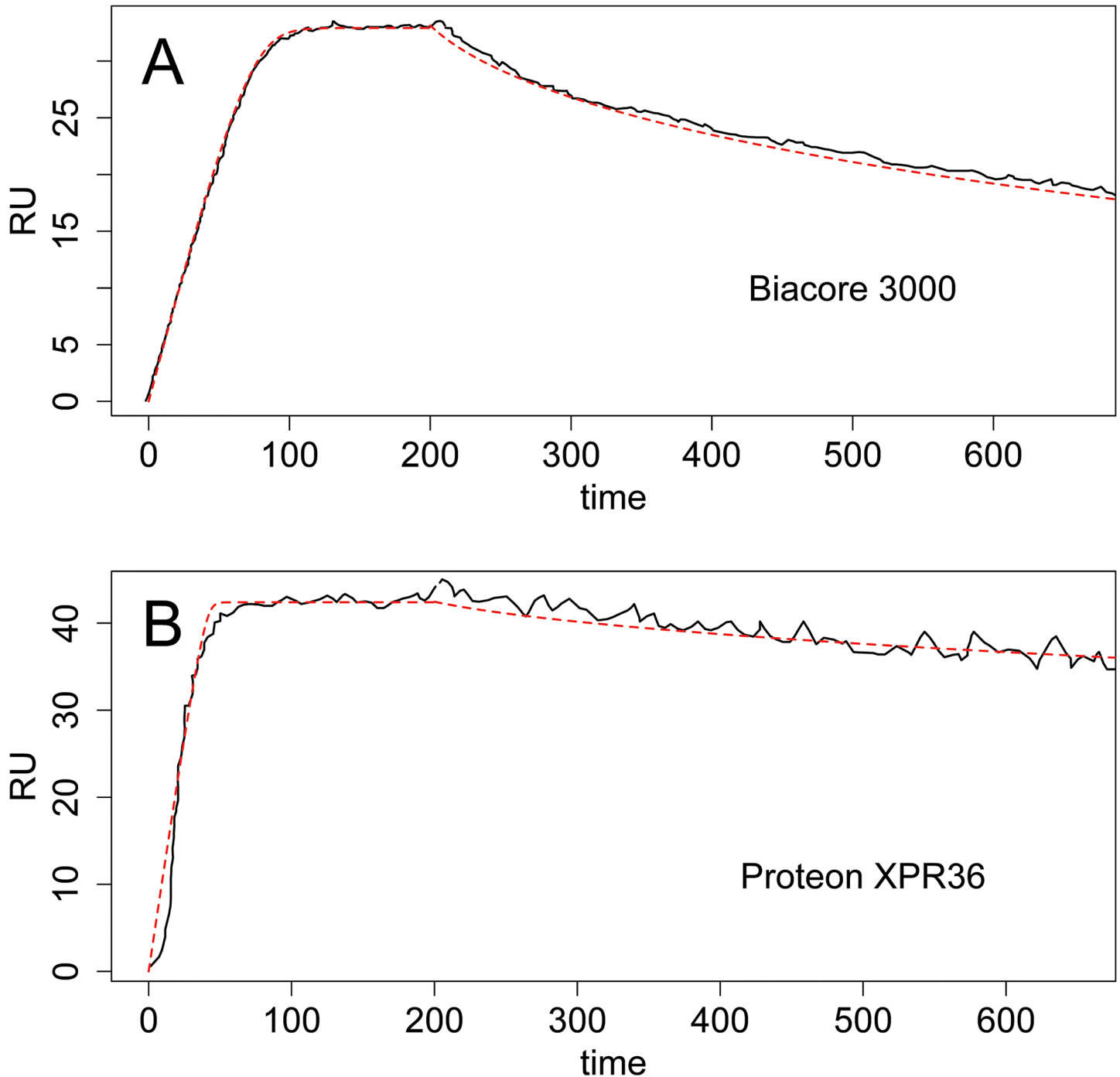
Fig 6. The kinetic analysis of BMP10 with hEnd^ECD-hFC in BiaCore 3000 and hCGRPα with 4901 IgG in Proteon XPR36 by the proposed Bayesian approach. Two sets of data were extracted from literature. (A), the interaction between BMP10 and immobilized hEngECD-hFc was measured in a Biacore 3000 system [32]. (B), the analyte human CGRPα and immobilized 4901 IgG were studied in a Proteon XPR36 biosensor [33]. The values of mass transport coefficient k_M were derived according to Eqs (9) and (10) for both cases. The two datasets were then analyzed by the proposed approach to estimate the active concentration of analyte A_0, rate constants k_a and k_d, and the dissociation constant K_D. The returned parameters were listed in Table 2. The raw sensorgrams (solid lines) and fitted curves (dashed lines) were overlaid.

doi:10.1371/journal.pone.0130812.g006

Supporting Information

S1 Appendix. The supporting information for the Bayesian inference approach to estimate the active concentration and affinity constants.
(DOCX)

Author Contributions

Conceived and designed the experiments: TBK FF. Performed the experiments: FF. Analyzed the data: TBK FF. Contributed reagents/materials/analysis tools: TBK FF. Wrote the paper: TBK FF.

References

1. Bendow B, Lengeler B. Electronic structure of noble metals and polariton-mediated light scattering: Springer-Verlag; 1978.

2. Flanagan MT, Pantell RH. Surface plasmon resonance and immunosensors. Electronics Letters. 1984; 20(23):968–70. doi: 10.1049/el:19840660

3. Liedberg B, Nylander C, Lunström I. Surface plasmon resonance for gas detection and biosensing. Sensors and Actuators. 1983; 4(0):299–304. http://dx.doi.org/10.1016/0250-6874(83)85036-7.

4. Jönsson U, Fägerstam L, Ivarsson B, Johnsson B, Karlsson R, Lundh K, et al. Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. BioTechniques. 1991; 11(5):620–7. PMID: 1804254

5. Karlsson R, Michaelsson A, Mattsson L. Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. J Immunol Methods. 1991; 145(1–2):229–40. Epub 1991/12/15. PMID: 1765656.

6. Chavane N, Jacquemart R, Hoemann CD, Jolicoeur M, De Crescenzo G. At-line quantification of bioactive antibody in bioreactor by surface plasmon resonance using epitope detection. Anal Biochem. 2008; 378(2):158–65. Epub 2008/05/07. doi: 10.1016/j.ab.2008.04.019 PMID: 18457651.

7. Kikuchi Y, Uno S, Nanami M, Yoshimura Y, Iida S, Fukushima N, et al. Determination of concentration and binding affinity of antibody fragments by use of surface plasmon resonance. Journal of bioscience and bioengineering. 2005; 100(3):311–7. Epub 2005/10/26. doi: 10.1263/jbb.100.311 PMID: 16243282.

8. Richalet-Secordel PM, Rauffer-Bruyere N, Christensen LL, Ofenloch-Haehnle B, Seidel C, Van Regenmortel MH. Concentration measurement of unpurified proteins using biosensor technology under conditions of partial mass transport limitation. Anal Biochem. 1997; 249(2):165–73. Epub 1997/07/01. doi: 10.1006/abio.1997.2183 PMID: 9212868.

9. Christensen LL. Theoretical analysis of protein concentration determination using biosensor technology under conditions of partial mass transport limitation. Anal Biochem. 1997; 249(2):153–64. Epub 1997/07/01. doi: 10.1006/abio.1997.2182 PMID: 9212867.

10. Mehand MS, Srinivasan B, Crescenzo GD. Estimation of analyte concentration by surface plasmon resonance-based biosensing using parameter identification techniques. Anal Biochem. 2011; 419(2):140–4. http://dx.doi.org/10.1016/j.ab.2011.08.051. doi: 10.1016/j.ab.2011.08.051 PMID: 21945965

11. Sigmundsson K, Masson G, Rice R, Beauchemin N, Obrink B. Determination of active concentrations and association and dissociation rate constants of interacting biomolecules: an analytical solution to the theory for kinetic and mass transport limitations in biosensor technology and its experimental verification. Biochemistry. 2002; 41(26):8263–76. Epub 2002/06/26. PMID: 12081475.

12. Conlon EM, Song JJ, Liu A. Bayesian meta-analysis models for microarray data: a comparative study. BMC Bioinformatics. 2007; 8:80. Epub 2007/03/09. 1471-2105-8-80 [pii] doi: 10.1186/1471-2105-8-80 PMID: 17343745.

13. Frigessi A, van de Wiel MA, Holden M, Svendsrud DH, Glad IK, Lyng H. Genome-wide estimation of transcript concentrations from spotted cDNA microarray data. Nucleic acids research. 2005; 33(17): e143. Epub 2005/10/06. 33/17/e143 [pii] doi: 10.1093/nar/gni141 PMID: 16204447.

14. Gelman A, Chew GL, Shnaidman M. Bayesian analysis of serial dilution assays. Biometrics. 2004; 60 (2):407–17. Epub 2004/06/08. doi: 10.1111/j.0006-341X.2004.00185.x BIOM185 [pii]. PMID: 15180666.

15. Sivaganesan M, Seifring S, Varma M, Haugland RA, Shanks OC. A Bayesian method for calculating real-time quantitative PCR calibration curves using absolute plasmid DNA standards. BMC

Bioinformatics. 2008; 9:120. Epub 2008/02/27. 1471-2105-9-120 [pii] doi: 10.1186/1471-2105-9-120 PMID: 18298858.

16. Feng F, Sales AP, Kepler TB. A Bayesian approach for estimating calibration curves and unknown concentrations in immunoassays. Bioinformatics. 2011; 27(5):707–12. doi: 10.1093/bioinformatics/btq686 PMID: 21149344

17. Davidian M, Giltinan DM. Nonlinear Models for Repeated Measurement Data. London: Chapman and hall; 1995.

18. Dellaportas P, Stephens DA. Bayesian Analysis of Errors-in-Variables Regression Models. Biometrics. 1995; 51(3):1085–95.

19. Christensen LLH. Theoretical Analysis of Protein Concentration Determination Using Biosensor Technology under Conditions of Partial Mass Transport Limitation. Analytical Biochemistry. 1997; 249 (2):153–64. http://dx.doi.org/10.1006/abio.1997.2182. PMID: 9212867

20. Knowles PF. Biophysical chemistry: Part II 'Techniques for the study of biological structure and function': By CR Cantor and PR Schimmel. pp 503. WH Freeman and Co, Oxford. 1980. £20.70/£10.60 (paperback) ISBN 0-7167-1189-3/0-7167-1190-7 (paperback). Biochemical Education. 1981;9(4):157-. doi: 10.1016/0307-4412(81)90143-6

21. Segel LA, Slemrod M. The quasi-steady state assumption: a case study in perturbation. SIAM Rev. 1989; 31(3):446–77. doi: 10.1137/1031091

22. Sjoelander S, Urbaniczky C. Integrated fluid handling system for biomolecular interaction analysis. Analytical Chemistry. 1991; 63(20):2338–45. doi: 10.1021/ac00020a025 PMID: 1759716

23. Stenberg E, Persson B, Roos H, Urbaniczky C. Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. Journal of Colloid and Interface Science. 1991; 143(2):513–26. http://dx.doi.org/10.1016/0021-9797(91)90284-F.

24. Box GEP, Tiao GC. Bayesian Inference in Statistical Analysis. New York: Jonh Wiley and Sons; 1973 reprinted 1992. 580 p.

25. Kass RE, Wasserman L. The Selection of Prior Distributions by Formal Rules. Journal of the American Statistical Association. 1996; 91(435):1343–70. doi: 10.1080/01621459.1996.10477003

26. Gilks WR, editor. Full Conditional Distributions. London, UK: Chapman & Hall; 1996.

27. Gilks WR, Best NG, Tan KKC. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. Journal of the Royal Statistical Society Series C (Applied Statistics). 1995; 44(4):455–72. doi: 10.2307/2986138

28. Gilks WR, Wild P. Adaptive Rejection Sampling for Gibbs Sampling. Journal of the Royal Statistical Society Series C (Applied Statistics). 1992; 41(2):337–48. doi: 10.2307/2347565

29. Martino L, Read J, Luengo D. Improved Adaptive Rejection Metropolis Sampling Algorithms. ARXIV. 2012;arXiv:1205.5494.

30. Karlsson R. Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. Journal of Molecular Recognition. 1999; 12(5):285–92. doi: 10.1002/(sici) 1099-1352(199909/10)12:5<285::aid-jmr469>3.0.co;2-y PMID: 10556876

31. Day YS, Baird CL, Rich RL, Myszka DG. Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. Protein science: a publication of the Protein Society. 2002; 11(5):1017–25. Epub 2002/04/23. doi: 10.1110/ps.4330102 PMID: 11967359; PubMed Central PMCID: PMC2373566.

32. Castonguay R, Werner ED, Matthews RG, Presman E, Mulivor AW, Solban N, et al. Soluble endoglin specifically binds bone morphogenetic proteins 9 and 10 via its orphan domain, inhibits blood vessel formation, and suppresses tumor growth. J Biol Chem. 2011; 286(34):30034–46. PMID: 21737454. doi: 10.1074/jbc.M111.260133

33. Abdiche Y, Malashock D, Pinkerton A, Pons J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. Anal Biochem. 2008; 377(2):209–17. http://dx.doi.org/10.1016/j.ab.2008.03.035. doi: 10.1016/j.ab.2008.03.035 PMID: 18405656