# Reliable Strategy for Analysis of Complex Biosensor Data

Patrik Forssén,[†] Evgen Multia,[‡] Jörgen Samuelsson,[†] Marie Andersson,[†] Teodor Aastrup,[§] Samuel Altun,[§] Daniel Wallinder,[§] Linus Wallbing,[§] Thanaporn Liangsupree,[‡] Marja-Liisa Riekkola,*,[‡] and Torgny Fornstedt[†]

[†]Department of Engineering and Chemical Sciences, Karlstad University, SE-651 88 Karlstad, Sweden
[‡]Department of Chemistry, P.O. Box 55, FI-00014 University of Helsinki, Finland
[§]Attana AB, Björnäsvägen 21, SE-114 19 Stockholm, Sweden

Ⓢ *Supporting Information*

**ABSTRACT:** When using biosensors, analyte biomolecules of several different concentrations are percolated over a chip with immobilized ligand molecules that form complexes with analytes. However, in many cases of biological interest, e.g., in antibody interactions, complex formation steady-state is not reached. The data measured are so-called sensorgram, one for each analyte concentration, with total complex concentration vs time. Here we present a new four-step strategy for more reliable processing of this complex kinetic binding data and compare it with the standard global fitting procedure. In our strategy, we first calculate a dissociation graph to reveal if there are any heterogeneous interactions. Thereafter, a new numerical algorithm, AIDA, is used to get the number of different complex formation reactions for each analyte concentration level. This information is then used to estimate the corresponding complex formation rate constants by fitting to the measured sensorgram one by one. Finally, all estimated rate constants are plotted and clustered, where each cluster represents a complex formation. Synthetic and experimental data obtained from three different QCM biosensor experimental systems having fast (close to steady-state), moderate, and slow kinetics (far from steady-state) were evaluated using the four-step strategy and standard global fitting. The new strategy allowed us to more reliably estimate the number of different complex formations, especially for cases of complex and slow dissociation kinetics. Moreover, the new strategy proved to be more robust as it enables one to handle system drift, i.e., data from biosensor chips that deteriorate over time.
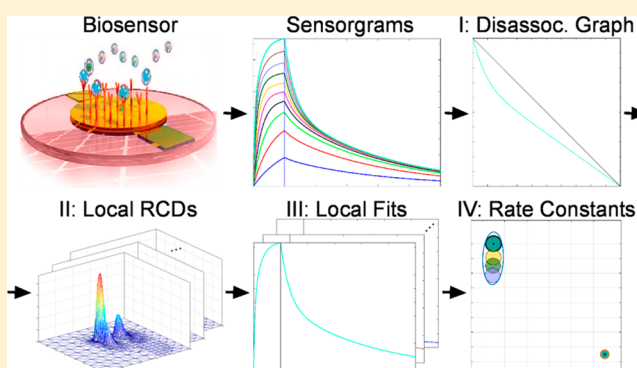
Biological medicines, such as peptides, proteins, mRNA, oligonucleotides, vaccines, and plasmids,[1−4] are a strongly growing area of importance for the pharmaceutical industry.[5,6] Reliable analysis of biomolecular interactions is crucial in order to fulfill modern drug quality assurance criteria, both for traditional small API molecules[7] and for next-generation biological drugs. Therefore, there has been an accelerated technological development of biosensor instruments based on, among others, surface plasmon resonance (SPR) and quartz crystal microbalance (QCM).[8−10]

The interaction data from Bbiosensor instruments is nearly always processed using a simple model fitting procedure assuming just one, or sometimes two, distinct interactions (complex formation reactions) with or without mass transport limitations.[8,11−15] In a global benchmarking study, 150 participants from 20 different countries were provided with aliquots of two binding partners (a 50-kDa Fab and a 60-kDa glutathione S-transferase [GST]-tagged Ag) that form a relative high affinity complex.[11] The participants were asked to determine the rate constants for the interaction. The majority of the participants fitted the data to a simple or mass-transport limited one-to-one model. Only four of the participants fitted to a two-to-one interaction model.[11] The authors speculated that more advanced models were applied to compensate for nonspecific interactions. Other studies[12,13] suggest that experiments should be redone using, e.g., alternative immobilization techniques if one-to-one model does not fit well. Applying models that are more complex is not preferred, and one should "go back and redesign the experiment to improve the quality of the data".[12] In some extreme cases, data that do not fit the model are excluded.[16] Rich et al. state that a more complex model fits always better, "*simply because it fits better does not mean it is correct*",[12] which, of course, is true. The use of a more complex model will result in a better model fit, but it will not necessarily reveal the system's true mechanistic behavior. However, it is also true that if a more complex model fits the data better, it will more accurately describe the system's behavior during the studied time range, even if the more
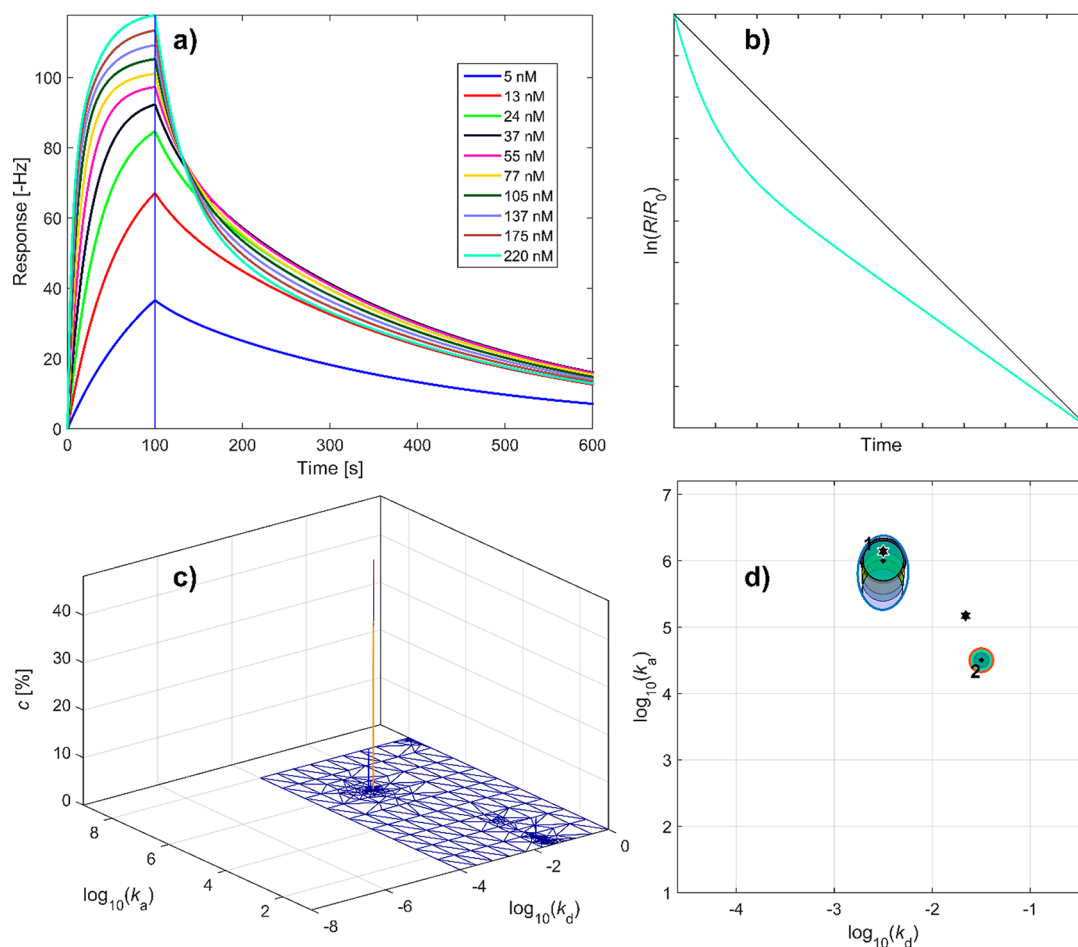
**Figure 1.** (a) Sensorgrams for a deteriorating synthetic system at different analyte concentration levels; vertical line indicates injection duration. (b) Dissociation graph for a 220 nM injection. (c) Rate Constants Distribution (RCD) for a 24 nM injection. (d) Rate constants obtained by fitting a two interactions model to the sensorgrams one by one. In part d, the circled areas are proportional to the relative contributions, the crosses indicate median rate constants, and the stars are the estimated rate constants from global fitting to a two interactions model. See the Supporting Information Table S1 for rate constants.

complex model does not reflect the system's true mechanistic behavior it will be a very accurate model of it. As biological systems are indeed complex, it is very risky to use only the simplest kinetic model regardless of how well it fits the data. Regardless of the model, if the model fit is poor it does not describe the system accurately.

If the complex formation reactions are sufficiently fast to reach steady-state, one way to accurately estimate the number of interactions is to use a three-step approach based on an algorithm called Adsorption Energy Distribution (AED) calculations.[17,18] This approach has been used for adsorption studies in several publications, both for SPR[19] and QCM[20,21] biosensors. However, AED has limited use for biosensor data since steady state, which is rarely reached in most biological biosensor applications of interest, is required. Therefore, a numerical tool similar to AED, called Rate Constant Distribution (RCD) or Interaction Maps,[22−24] has been developed for biosensor data. Using RCD calculations, one gets the number of interactions needed to describe the system, and their associated rate constants, as peaks on a surface. However, RCD calculations are not usually part of biosensor instruments software, and the implementations that do exist take up to days to calculate a single RCD.[23] Our new Adaptive Interaction Distribution Algorithm (AIDA)[25] does not only

give higher resolution since it adapts to the solution, but it is also fast, the calculations take only a few seconds, and thereby allows RCDs to be used as a standard tool.

The aim of this study is to introduce and evaluate a new four-step strategy, based on our new numerical algorithm AIDA, for more reliable analysis of interaction data. Verification will be done using both synthetic and experimental data obtained from three different experimental systems exhibiting fast to slow interactions. The new strategy will also be compared to the conventional method, i.e., to simultaneously fit a one or two-site kinetic model to all sensorgrams, so-called global fitting. This four-step strategy can also be applied to the clarification of interactions between biomolecules and cell surfaces, present in a very complex and biologically relevant environment where a more complex fitting model is often needed.

## ■ THEORETICAL SECTION

For a more detailed description of the theory, see the Supporting Information. Here we will assume that we have a biosensor system where the kinetics is described by an "n-to-one" model, i.e., we have $n$ interactions of the following type,

$$[A_i] + [L] \underset{k_{d,i}}{\overset{k_{a,i}}{\rightleftharpoons}} [A_iL] \tag{1}$$

in the system. Here $[A_i]$ is an injected analyte, $[L]$ is a ligand immobilized on the biosensor chip surface, $[A_iL]$ is a complex formed between the analyte and the ligand, and this reaction has the corresponding association and disassociation rate constants, $k_{a,i}$ and $k_{d,i}$, respectively.

If we have $n$ different interactions in the system of the type in eq 1, the total response of the system will be the sum of the individual responses of the interactions. Here we define the contribution $c_i$ of interaction $i$ to the total response as the mean of the interaction's relative contribution to the association and dissociation phase.

In a practical situation, we want to estimate the rate constants $k_{a,i}$, $k_{d,i}$ from experimental sensorgram data. This is usually done by fitting all experimental sensorgrams simultaneously to a model with one, or perhaps two, interaction(s) of the type in eq 1. Assuming the "$n$-to-one" kinetic model in eq 1, a useful simple tool that indicates if one has one or more interactions in the biosensor system is the *dissociation graph*. Here $\ln[R(t)/R_0]$ is plotted against $t$ for the dissociation phase, $R(t)$ is the response at time $t$, and $R_0$ is the response at the start of the dissociation phase. If the resulting curve is near the top-left to bottom-right diagonal, then there is only one interaction in the system, but if it is convex, then there are at least two different interactions in the system. Note that only the *shape* of the graph is important, so for the clarity of presentation, axis numbering is omitted here, see for example Figure 1b.

It is well-known that the assumption of a "$n$-to-one" kinetic model might not hold in some cases. For example, Ogi et al. have studied the effect of flow rate on determined rate constants between human IgG and protein A.[26] They found that an increase of the flow rate 20 times lead to an increase, in both the association and dissociation rate constant, with a factor of three. However, our proposed strategy can be used as an accurate representation of a wide range of different systems, regardless of the true underlying interactions. In a similar way as any measured signal can be represented with a Fourier series expansion, whatever the true underlying process(es) that generates the signal are.

**Rate Constant Distributions.** If we let the number of interactions in the system $\rightarrow \infty$, assuming interactions of the type in eq 1, the total response of the system can be modeled by an inhomogeneous Fredholm Integral Equation of the first kind. One can estimate a so-called Rate Constant Distribution (RCD) by solving the corresponding inverse Integral Equation problem using an algorithm called AIDA, developed by us.[25] In an RCD we have the interaction contribution vs the rate constants $k_a$ and $k_d$, see for example Figure 1. An RCD consists of a number of discrete distributions, and the mode of these distributions can be viewed as estimates of the biosensor systems rate constants.

Calculation of RCDs using AIDA[25] is considerably faster than fitting to the experimental data, and here we do not have to select the number of interactions in the system. However, RCDs should be used with caution, as they are the solution of an ill-posed and ill-conditioned inverse problem, and the solution depends heavily on the amount and type of regularization applied. One generally needs to use several sensorgrams when calculating an RCD in order to get reliable results and a good advice is to check the results by also doing model fitting. When using several sensorgrams, and proper regularization, one can get a good estimate of number of interactions and their corresponding rate constants from the peak maxima. These can, for example, be used as input to the

fitting algorithm, but one should never try to draw any conclusions about the system from the peak shapes.

In the proposed strategy, we will calculate RCDs separately for each sensorgram, and the resulting RCDs can be used as a guide to decide the number of interactions and their rate constants.

## ■ EXPERIMENTAL SECTION

**Chemicals and Materials.** Trastuzumab ($M_w$ 145531.5 g/mol) (Roche, Basel, Switzerland), purified recombinant human epidermal growth factor receptor 2 (HER2) (R&D Systems Europe Ltd., Abingdon, United Kingdom), parathyroid hormone (1-34) (PTH) ($M_w$ 4117.75 g/mol) (Tocris Bioscience, Bristol, U.K.), and Human PTH1 receptor with a His-tag (NovoPro, Shanghai, China) was used. Zeba spin desalting column from Thermo Fisher Scientific, was used to for the buffer exchange of trastuzumab. Other chemicals were purchased from Sigma-Aldrich unless noted otherwise. Hydrochloric acid (0.1−1.0 M) and sodium hydroxide (0.1−1.0 M) were obtained from Oy FF-Chemicals (Yli-Ii, Finland). Phosphoric acid was from Merck (Darmstadt, Germany). Ammonia (25%) was purchased from Riedel-de Haën (Seelze, Germany). NaCl was from Fisher Chemicals (Loughborough, U.K.). Ethylenediaminetetraacetic acid disodium salt dihydrate ($Na_2EDTA$), bovine serum albumin (BSA), sodium citrate trihydrate, and sodium azide were from Sigma Chemicals Co. (St. Louis, MO). Gentamycin sulfate (BioReagent, suitable for cell culture) was from Lonza (Basel, Switzerland). Sulfo-N-hydroxy-succinimide (S-NHS, 98.0%) and 1-ethyl-3-(3-(dimethylamino)propyl) carbodiimide hydrochloride (EDC-HCl, 98.5%) were purchased from Aladin Chemical Reagent Company (Shanghai, China). The water used in the experiments was purified in a Milli-Q system (Millipore AB, Billerica, MA) and a Millipore water purification system (Millipore, Molsheim, France). Intermediate-Density Lipoprotein (IDL)-Very Low-Density Lipoprotein (VLDL) and antihuman apoB-100 (anti-apoB-100) monoclonal antibody (Mab) were produced according to ref 27. Molar concentrations of IDL-VLDL were calculated using the molecular weight of apoB-100 (512000 g/mol).

The Attana Cell 200 (Attana AB, Stockholm, Sweden) QCM Biosensor was used for the trastuzumab-HER2 and PTH-PTH1R experimental analyses. The Attana A100 instrument (Attana AB, Stockholm, Sweden) was used for the IDL-VLDL-anti-apoB-100 Mab experimental analyses.

The fundamental frequency of the crystals in this study was around 10 MHz in thickness shear mode. The immobilized protein constitutes a rigid layer and is suitable for the following kinetic measurement. Thus, the Sauerbrey equation[28] should give a good approximation for correlating sensor response to adsorbed amount of solute; in other words, the sensor signal is proportional to the adsorbed amount of solute on the chip.

**Sensor Procedures.** *Synthetic Assay.* For the synthetic data we used a system with two interactions and rate constants $\log_{10}(k_{a,1}) = 6.00$ (M s)$^{-1}$, $\log_{10}(k_{d,1}) = -2.50$ s$^{-1}$, $\log_{10}(k_{a,2}) = 4.50$ (M s)$^{-1}$, $\log_{10}(k_{d,2}) = -1.50$ s$^{-1}$ with $R_{max,1} = 100.0$ RU, $R_{max,2} = 500.0$ RU. We made 10 injections with concentrations $C = 5, 13, 24, 37, 55, 77, 105, 137, 175$, and 220 nM, and the injection time was 100 s and the run time was 600 s. To simulate a deteriorating system, $R_{max}$ was set to 95% of the value in the previous injection.

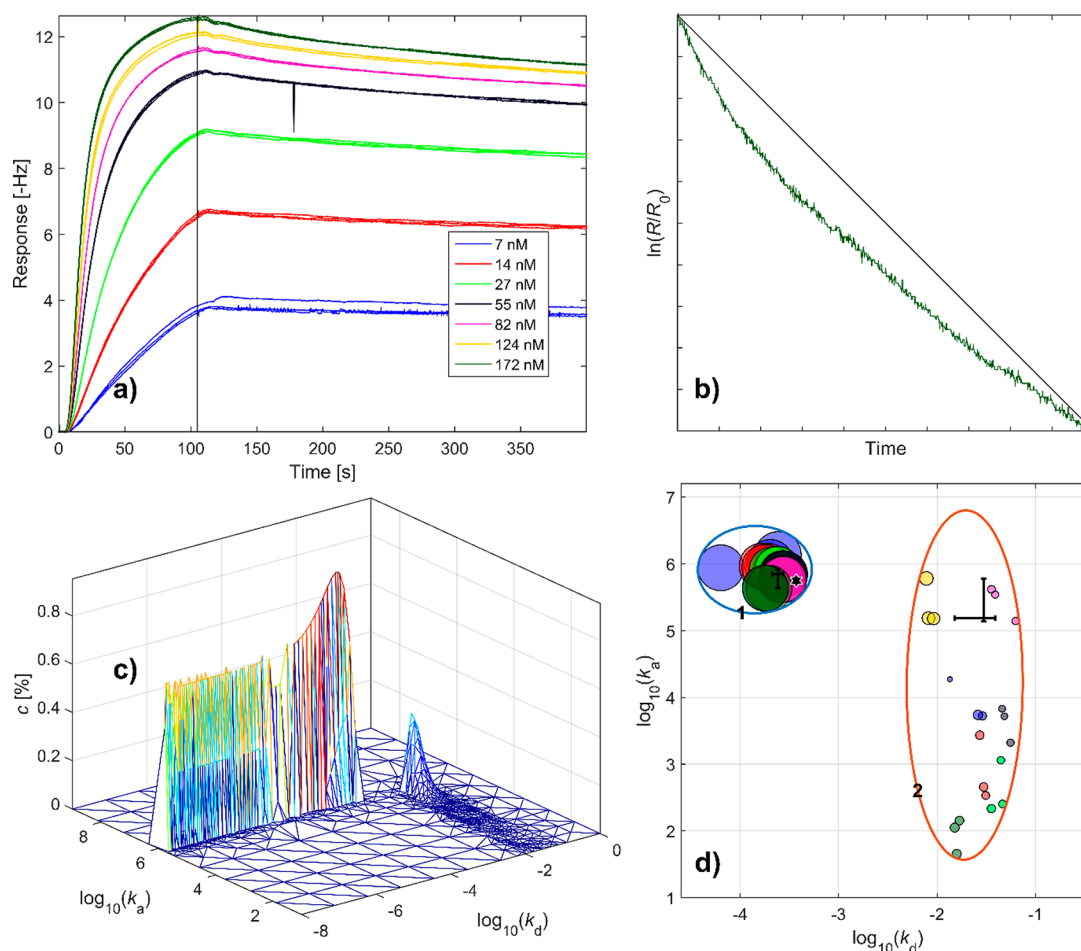*HER2 Assay.* HER2 was amine coupled to a LNB-carboxyl chip using the manufactures' amine coupling kit and protocol

**Figure 2.** (a) Sensorgrams for a trastuzumab-HER2 system at different analyte concentration levels. (b) Dissociation graph for a 172 nM injection. (c) Rate Constants Distribution (RCD) for a 14 nM injection. (d) Rate constants obtained by fitting a two interactions model to the sensorgrams one by one. See Figure 1 for more details and the Supporting Information Table S2 for rate constants.

(Attana AB, Stockholm, Sweden). In brief, 10 mM HEPES, 150 mM NaCl, 0.005% Tween-20, pH 7.4, was used as immobilization buffer. Carboxyl groups on the chip surface were activated by a 50 $\mu$L injection of 0.2 M EDC and 0.05 M S-NHS. A volume of 50 $\mu$L of HER2 (5 $\mu$g/mL diluted in 10 mM acetic acid, pH 4.5) was injected over the activated surface at a flow rate of 10 $\mu$L/min. The frequency change was typically between 20 and 30 Hz. Finally, free carboxyl groups were blocked by a 50 $\mu$L injection of 1 M ethanolamine pH 8.5.

The adsorption data for trastuzumab at 35 °C were acquired by 35 $\mu$L of trastuzumab (1, 2, 4, 8, 12, 18, and 25 $\mu$g/mL) at a flow rate of 20 $\mu$L/min. Following dissociation for 300 s the surface was regenerated with a 6 $\mu$L injection of 10 mM glycine pH 2.0 (Acros Organics, Fisher Scientific) and re-equilibrated with running buffer (phosphate buffered saline (PBS), pH 7.4, Giboc, Fisher Scientific). Blank injections were performed prior to every analyte injection, see Figure S1 in the Supporting Information, and after each analyte injections, the loops were washed with 100 mM acetic acid and repeated washes of running buffer.

*Anti-ApoB-100 Assay.* Anti-apoB-100 Mab was immobilized onto LNB-carboxyl chip (Attana AB, Stockholm, Sweden) using amine coupling according to the protocol described in a previous study.[27] In brief, the chip was prewetted *ex situ* with 20 $\mu$L of Milli-Q and left to stabilize in the instrument. The immobilization was performed at the flow rate 10 $\mu$L/min

under 25 °C with 10 mM HEPES, 150 mM NaCl, 0.005% Tween 20, pH 7.4, as the immobilization buffer. The surface was activated once by injecting 50 $\mu$L of freshly mixed 0.4 M EDC and 0.1 M S-NHS (1:1, v/v). Followed by two injections (50 $\mu$L each) of anti-apoB-100 Mab (100 $\mu$g/mL in phosphate buffer; pH 6.4). The frequency change was typically between 30 and 60 Hz. Deactivation was done with two injections (50 $\mu$L) of 1 M ethanolamine solution (pH 9.0).

The adsorption data of IDL-VLDL at 25 °C were acquired by 35 $\mu$L IDL-VLDL injections (3, 5, 10, 15, 20, 30, 35, 40, 50, 60, 70, 80, 90, 100, and 110 $\mu$g/mL) on the amine coupled chip at the flow rate of 50 $\mu$L/min using phosphate buffer (pH 7.4) as the running buffer. Blank injections were performed prior every analyte injection, see Figure S2 in the Supporting Information. Following dissociation for 500 s, the surface was regenerated with a 0.28 M ammonium hydroxide (pH 11.5) and re-equilibrated with running buffer. Running buffer was prepared as described in ref 27.

*PTH1R Assay.* PTH1R was amine coupled to a LNB-carboxyl chip using the manufacturer's amine coupling kit and protocol (Attana AB, Stockholm, Sweden). The coupling was done according to the description for the HER2 assay above. With the exception that 50 $\mu$L of PTH1R (10 $\mu$g/mL in 10 mM acetic acid, pH 4.5) was injected over the activated surface at a flow rate of 10 $\mu$L/min. The frequency change was typically
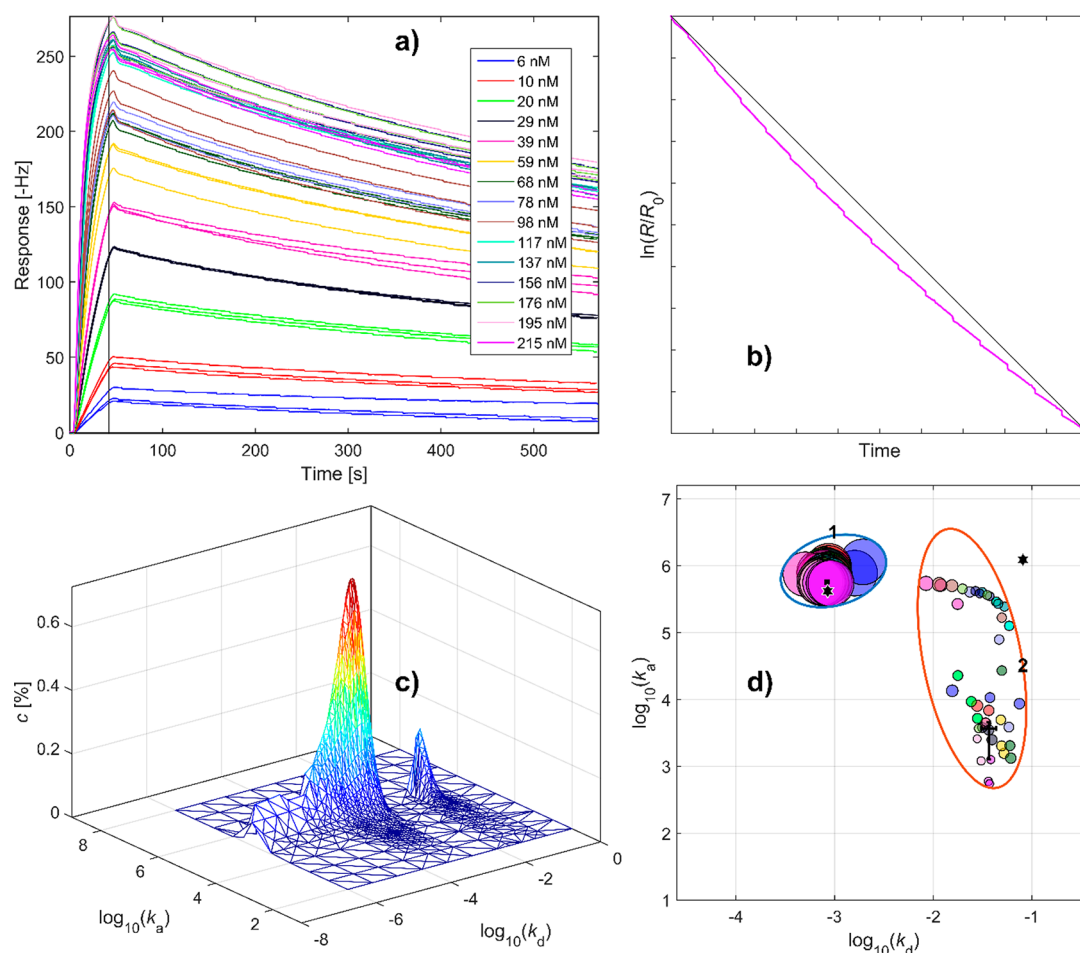
**Figure 3.** (a) Sensorgrams for an IDL-VLDL-anti-apoB-100 system at different analyte concentration levels. (b) Dissociation graph for a 215 nM injection. (c) Rate Constants Distribution (RCD) for a 20 nM injection. (d) Rate constants obtained by fitting a two interactions model to the sensorgrams one by one. See Figure 1 for more details and the Supporting Information Table S3 for rate constants.

between 20 and 30 Hz. Finally, free carboxyl groups were blocked by a 50 $\mu$L injection of 1 M ethanolamine pH 8.5.

The adsorption of PTH at 20 °C were acquired by 35 $\mu$L of PTH injections (4, 5, 6, 8, 10, 12, 14, 17, and 20 $\mu$g/mL) at a flow rate of 25 $\mu$L/min, following dissociation for 510 s. Blank injections were performed prior to every analyte injection, see Figure S3 in the Supporting Information. As a running buffer, 10 mM HEPES with 150 mM NaCl and 0.005% Tween-20, pH 7.4 was used.

*Calculations.* The time when the injection begins, $t_0$, was estimated directly from the measured sensorgrams individually by an algorithm based on fitting concavity constrained Hermite polynomials to the association phase. The RCDs were calculated using adaptive Finite Element triangulation.[25] For fitting to the kinetic model, we used a combination of a global Genetic Algorithm and a nonlinear least-squares solver. The Genetic Algorithm was the MATLAB,[29] Global Optimization Toolbox function "ga" with the fitness function set to the residual norm, the population size set to 5000, and the number of generations set to 100. The Genetic Algorithm was run in parallel on a computer cluster with 20 calculation cores (Intel Core i7-3770S 3.10 GHz). The nonlinear least-squares solver was the trust region reflective algorithm,[30] "lsqnonlin" in MATLAB,[29] and Optimization Toolbox with the initial guess set to the parameters found by the Genetic Algorithm.

## RESULTS AND DISCUSSION

A new four-step strategy is used here to analyze the sensor data. The new strategy is based on a recently published improved algorithm[25] for fast and reliable calculations of Rate Constant Distributions (RCDs) from sensorgrams. An RCD is a two-dimensional distribution of association and dissociation rate constants, which will display heterogeneous data as a map where each local distribution observed in the map corresponds to one interaction, see Figure 1c.

The four-step strategy will be used, evaluated, and compared with the conventional approach, i.e., global fitting of all sensorgrams simultaneously.[12,13] The evaluation will be carried out on synthetic (see Figure 1a) and on three experimental biosensor (QCM) data sets representing different kinetics (see Figures 2−4).

**Proposed Strategy for Rate Constant Estimation.** The suggested four-step strategy for more reliable processing of complex binding kinetics consists of the following four steps (see also Figure 1):

(1) Plot a dissociation graph (cf. Figure 1b) to get the first impression of the data under investigation. If the plot is linear, there is just one interaction and one can move directly to step (3), otherwise continue to step (2).

(2) Calculate RCDs for each sensorgram separately (cf. Figure 1c) to estimate the number of interactions.
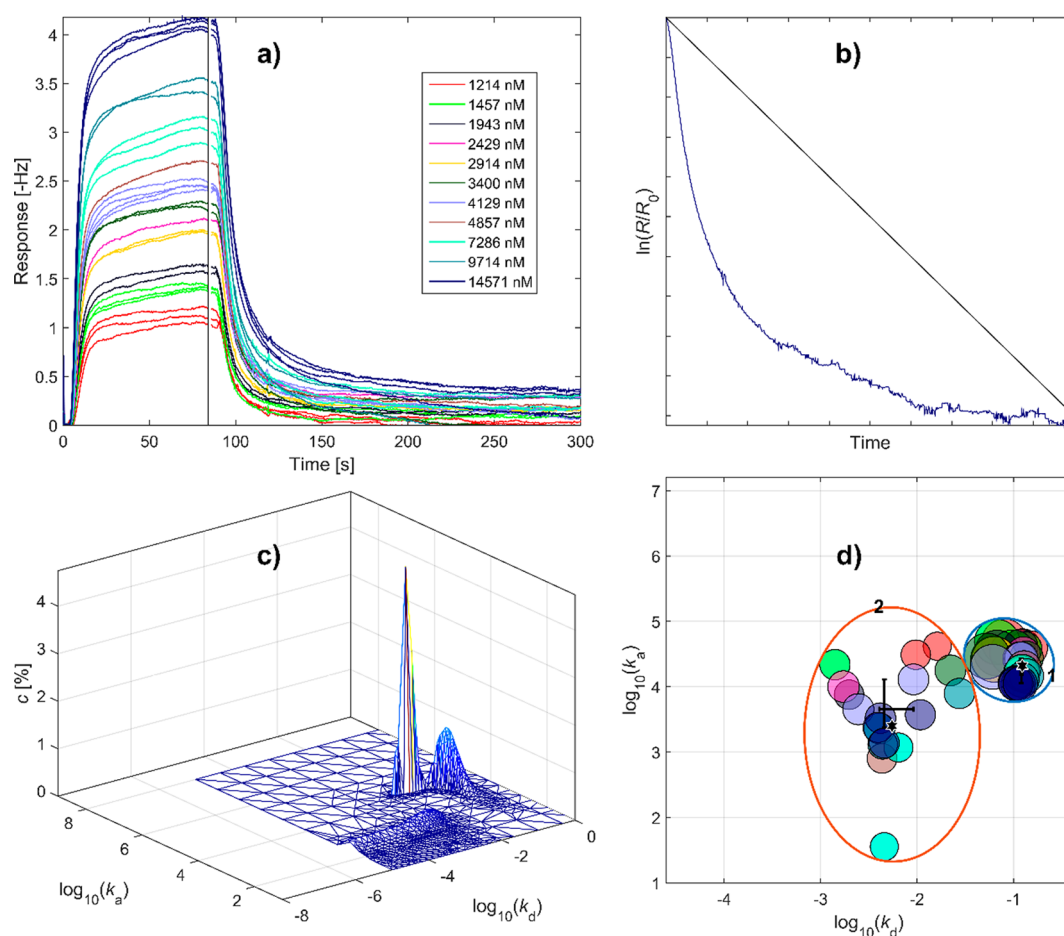
**Figure 4.** (a) Sensorgrams for a PTH-PTH1R system at different analyte concentration levels. (b) Dissociation graph for a 7 286 nM injection. (c) Rate Constants Distribution (RCD) for a 4129 nM injection. (d) Rate constants obtained by fitting a two interactions model to the sensorgrams one by one. See Figure 1 for more details and the Supporting Information Table S4 for rate constants.

(3) Do fitting to the sensorgrams one by one using the number of interactions determined in step (2).

(4) Plot the determined rate constants, remove outliers, and do clustering of all rate constants estimated from step (3), (cf. Figure 1d).

The simplest way to cluster the data is to set manually rough cluster center points and then assign the rate constant pairs based on their distance to the cluster center points. Another strategy is to use statistical clustering tools such as hierarchical clustering or $k$-means.[31] It should be noted that in this strategy, in contrast to the conventional global fitting method, there is no need to exclude sensorgrams for potentially failed injections in the analysis. Those can be easily detected as outliers in step (4) and removed at that stage.

In the following subsections, we will evaluate the four-step strategy on synthetic data sets (Figure 1) as well as three different biosensor (QCM) assays representing slow kinetics (Figure 2), medium-slow kinetics (Figure 3), and fast kinetics (Figure 4) systems.

The Supporting Information contains complementary information, such as model fits and residual plots, obtained by the new strategy and the conventional global fitting approach.

**Synthetic Data.** In this section we are going to analyze the new strategy using synthetic two interactions data with rate parameters $\log_{10}(k_{a,1}) = 6.00$ $(M\ s)^{-1}$, $\log_{10}(k_{d,1}) = -2.50$ $s^{-1}$, $\log_{10}(k_{a,2}) = 4.50$ $(M\ s)^{-1}$, $\log_{10}(k_{d,2}) = -1.50$ $s^{-1}$, and $R_{max,1} =$

100.0 RU, $R_{max,2} = 500.0$ RU. The new strategy is compared to conventional global fitting using both ideal synthetic data and synthetic data for a deteriorating system.

First, we test an ideal system, the sensorgrams at different analyte concentrations can be found in the Supporting Information Figure S4a. The corresponding dissociation graph in Figure S4b indicates that at least two interactions are needed to describe the system. In Figure S4c the RCDs demonstrate that we have two interactions, and Figure S4d shows the clustered rate constants estimated by fitting to the sensorgrams one by one using a two interactions kinetic model. The stars in Figure S4d are the estimated rate constants using a conventional global fit routine and the crosses indicate median rate constants for the clusters. In this case they coincide, meaning that global fitting and the new strategy give the same results.

It is not uncommon for experimental systems to slowly deteriorate, i.e., lose signal. There might be several reasons for this, e.g., some ligand molecules are destroyed or washed out in the regeneration step of the biosensor chip. This is simulated in Figure 1a by using the same system as above (vertical line indicates end of injection) but setting $R_{max}$ to 95% of the value in the previous injection. A common reaction would be to consider the experiment failed and discard the data set, then redo the experiment with a fresh chip, new buffers etc.; a time-consuming and costly procedure without any guarantee that the data will improve. However, here we will try to process the data

using our four-step strategy. In Figure 1d the estimated clustered rate constants from the new strategy, median values indicated by crosses are seen together with the result from conventional global fitting, indicated by stars. In this case, the new strategy provided excellent results but the conventional global fitting gave large errors, especially for the second association rate constant where the correct value is 4.50 but global fitting estimated it to 6.17, see also Figure S5 and Table S1. This is because the conventional global fitting method does not handle deteriorating systems well. The problem was solved using our new strategy, where fitting to the sensorgrams is done one by one.

**High Affinity Complex.** The interaction between human epidermal growth factor receptor 2 (HER2) and trastuzumab that forms a high affinity complex was studied at 35 °C. Trastuzumab, sold under the brand name Herceptin among others, is a monoclonal antibody used to treat breast cancer that is HER2 receptor positive.

The measured sensorgrams are presented in Figure 2a. From the sensorgrams it is clear that the dissociation (i.e., after the vertical line in the figure) was very slow. However, it is hard to deduce directly from the sensorgrams if the data originates from a single interaction or many interactions. The dissociation graph can be used to indicate potential heterogeneity, see Figure 2b for the 172 nM trastuzumab experiment. The graph is only slightly convex indicating that the interaction could be heterogeneous. Using the new strategy, we calculated the RCDs for all sensorgrams to estimate how many different interactions are needed to accurately describe the total response for the system, see Figure 2c for the RCD for a 14 nM injection. From the RCDs we can identify a major interaction at $\log_{10}(k_a) \approx 5.5$, $\log_{10}(k_d) \approx -4$ as well as a minor interaction at $\log_{10}(k_a) \approx 3$, $\log_{10}(k_d) \approx -1.5$. Because of this observation, the sensorgrams were fitted to a two-site model one by one. In Figure 2d the clustered estimated rate constants are presented. The sizes of the circle areas are proportional to the interactions contribution to the total response, i.e., the larger the circle area the more the interaction contributes to the total response. Inspecting the estimated rate constants, we can identify two different clusters, corresponding to the major and minor interaction in the RCD, see Figure 2c. The estimated rate constants for the minor interactions are much more scattered compared to the major interactions. Minor interaction contributes very little to the overall response and is therefore very sensitive to system noise.

We compared simultaneously the result above with conventional global fitting to all sensorgrams; using one- and two-interaction model see Figure S6 and Table S2, respectively. As expected, we see that the fit becomes better if a two interactions model is used. However, this does not necessarily prove it to be more appropriate as the estimated dissociation rate constant for the first interaction is unrealistically small ($\log_{10}k_d = -21.2$) and has extremely large error bounds. Therefore, the conclusion from global fitting is that it is probably more appropriate to use a single interaction model. For this interaction we get similar results from the new strategy and global fitting, $\log_{10} K_D = -9.42$ and $\log_{10}K_D = -9.18$ respectively.

The advantage of using the new strategy was that we could observe that at least two interactions are necessary to give accurate description for the system. We also successfully resolved two different interactions using the new strategy, whereas the conventional global fitting was only able to resolve one, since the two interactions model gave an unrealistic dissociation constant.

**Medium High Affinity Complex.** The interaction of VLDL and IDL with anti-apoB-100 monoclonal antibody at 25 °C was investigated. VLDLs and IDLs are involved in conversion to Low-Density Lipoprotein (LDL),[32] that is responsible for initiation of atherosclerosis.[33] Since these lipoproteins contain apolipoprotein B-100 (apoB-100), anti-apoB-100 Mab can be used to study IDL-VLDL mixture interactions.

From the sensorgrams of Figure 3a it is clear that the dissociation (after the vertical line in the figure) was moderately slow. To analyze the heterogeneity, dissociation graphs in Figure 3b for the 215 nM IDL-VLDL-anti-apoB-100 experiment, was plotted. The graphs are slightly convex, indicating that the interaction could be heterogeneous. From the RCD for the 20 nM injection (Figure 3c) we can identify two different interactions both with $\log_{10}(k_a) \approx 5.5$ and with $\log_{10}(k_d) \approx -4$, $\log_{10}(k_d) \approx -1.5$. We therefore fitted a two-site model to all sensorgrams one by one. From the estimated rate constants, seen in Figure 3d, two distinct clusters can be identified. The first at $\log_{10}(k_d) \approx -3.1$, $\log_{10}(k_a) \approx 5.8$, with a small cluster distribution, and the second at $\log_{10}(k_d) \approx -1.4$, $\log_{10}(k_a) \approx 3.6$, with a broader cluster distribution.

By inspecting the data, we see that the experimental system is deteriorating in a similar way as the synthetic data in Figure 1a. We can therefore suspect that global fitting might give erroneous rate constant estimation. To compare the result with the conventional way of analyzing the data, all the sensor data was simultaneously fitted to one- and two- interactions model, see Figure S7 and Table S3. Here the results are almost the same as for the new strategy for the first interaction. However, if a two interactions model was used, the results for the second interactions deviate strongly, giving $\log_{10}(k_a)$ 3.57 with the new strategy and 6.09 using global fitting, see Table S3. This is because the conventional global fitting procedure is not able to handle deteriorating systems.

Analyzing the global fitting result is tricky in this case. First, the dissociation graph is only slightly convex indicating that a linear model could probably describe the data rather well. Second, the improvement in model fit from one- to two-interactions model is modest. However, the second site contribution to the total response is non-negligible, about 7%, so probably a two-site model is more appropriate.

The advantage of using the new strategy in this case is that we could, from the local RCD, conclude that two interactions are needed to describe the system. In addition, there were no problem with the deteriorating system as was the case with global fitting. Moreover, the interpretation of the global fitting results was not as straightforward. From the results it can be easily concluded that the system fits one interaction model, since only minor improvement was achieved with the two interaction model.

**Low Affinity Complex.** Parathyroid hormone (PTH)-Parathyroid Hormone Receptor (PTH1R) interaction at 20 °C was investigated. PTH is a hormone secreted by the parathyroid glands (84 amino acid long). PTH is important in bone remodeling by regulating the blood calcium cation concentration.[34] In this study, synthetic PTH and PTH1R containing only the first 34 amino acids was used. However, the *in vivo* and *in vitro* activities of synthetic PTH compared to natural one have been shown to be similar.[35]

From the sensorgrams of Figure 4a it is evident that the dissociation (after the vertical line in the figure) is fast. Observing the association phase, we see that we probably have

a fast association and a slow association, indicating that the interaction is heterogeneous. To investigate this in more detail, we plotted the dissociation graph, see Figure 4b for the 7286 nM PTH injection. The graph is strongly convex, indicating clearly that the interaction is heterogeneous. In Figure 4c, the RCD for the 4129 nM PTH experiment demonstrates two distributions with the same $\log_{10}(k_a) \approx 4$ and with $\log_{10}(k_d) \approx -2$, $\log_{10}(k_d) \approx -1$. We therefore fitted a two-site model to all sensorgrams one by one. From Figure 4d, the estimated rate constants can be seen and two distinct clusters could be identified. The first at $\log_{10}(k_d) \approx -0.9$, $\log_{10}(k_a) \approx 4.2$, with a small cluster distribution, and the second at $\log_{10}(k_d) \approx -2.3$, $\log_{10}(k_a) \approx 3.7$, with a broader cluster distribution.

To compare the data achieved with the conventional approach, the sensorgrams were simultaneously fitted to one- and two- interactions model, see Table S4. In this case, the data fitted much better to two- interactions model (RMSEN 7.3%) compared to one- interaction model (RMSEN 14.1%). The errors in the predicted rate constants for the two-interactions model are also small indicating that a good fit was found.

In this case, the new strategy (crosses in Figure 4d) and global fitting (stars in Figure 4d) gave more or less the same predicted rate constants and equilibrium constant. However, although global fitting worked in this case, the new strategy is more robust and can handle cases where global fitting fails. Thus, from a practical point of view the new strategy is preferred.

## CONCLUSIONS

It is important that the recent technological development in the biosensor area is accompanied by a corresponding development in the processing of the kinetic data. To meet this requirement, we suggest the new four-step strategy for the reliable estimation of the rate constants from sensorgram data.

The four-step strategy is based on our new fast Adaptive Interaction Distribution Algorithm (AIDA). The first step is to plot dissociation graphs in order to get an indication of the heterogeneity of the data, i.e., if one or several interactions are present in the system. In the second step, whenever needed, the Rate Constant Distributions for each sensorgram are calculated with AIDA[25] in order to estimate the number of interactions for an accurate representation of the system. In the third step, fitting is done to the sensorgrams, one by one, using a kinetic model with the number of interactions estimated in the second step. In the fourth and final step, the rate constants determined are clustered, and each cluster describes an interaction present in the system.

The new four-step strategy was evaluated and compared with the conventional method, i.e., global fitting to all sensorgrams simultaneously.[12,13] The evaluation was first done using synthetic data, where the true rate constants were known. In order to test the ability of the new strategy to handle experimental noise and other experimental disturbances, the strategy was tested on three experimental biosensor (QCM) assays with slow, moderate, and fast kinetics.

The new four-step strategy was found to be particularly useful when processing complex data and data from deteriorating systems, whereas for simple and stable systems, conventional global fitting worked as well. However, since it is difficult to predict when a system is simple enough to suit to the conventional approach, we recommend that the new four-step approach is preferred. This four-step approach is applicable for any instrumental platform, not only for QCM based biosensors but also for SPR biosensors.

## ASSOCIATED CONTENT

**Ⓢ Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.anal-chem.8b00504.

> Extended theory for biosensors; additional figures for blank injections; additional figures for the synthetic data, analysis of a "perfect" synthetic system, and a table with the estimated rate constants; additional figures for the high affinity complex and a table with the estimated rate constants; additional figures for the medium high affinity complex and a table with the estimated rate constants; and additional figures for the low affinity complex and a table with the estimated rate constants (PDF)

## AUTHOR INFORMATION

**Corresponding Author**
*E-mail: marja-liisa.riekkola@helsinki.fi.

**ORCID** Ⓞ
Evgen Multia: 0000-0002-8717-8842
Marja-Liisa Riekkola: 0000-0001-7519-9892

**Notes**
The authors declare no competing financial interest.

## REFERENCES

(1) Guiochon, G.; Beaver, L. A. *J. Chromatogr. A* **2011**, *1218* (49), 8836−8858.
(2) Mant, C. T.; Hodges, R. S. *J. Chromatogr. A* **2012**, *1230*, 30−40.
(3) *Handbook of Analysis of Oligonucleotides and Related Products*; Bonilla, J. V., Srivatsa, S., Eds.; CRC Press: Boca Raton, FL, 2011.
(4) Tavernier, G.; Andries, O.; Demeester, J.; Sanders, N. N.; De Smedt, S. C.; Rejman, J. *J. Controlled Release* **2011**, *150* (3), 238−247.
(5) Bunnage, M. E.; Gilbert, A. M.; Jones, L. H.; Hett, E. C. *Nat. Chem. Biol.* **2015**, *11* (6), 368−372.
(6) Sasaki, K.; Takahashi, N.; Satoh, M.; Yamasaki, M.; Minamino, N. *J. Proteome Res.* **2010**, *9* (10), 5047−5052.
(7) Åsberg, D.; Nilsson, M.; Olsson, S.; Samuelsson, J.; Svensson, O.; Klick, S.; Ennis, J.; Butterworth, P.; Watt, D.; Iliadou, S.; et al. *J. Pharm. Biomed. Anal.* **2016**, *129*, 273−281.
(8) *Handbook of Surface Plasmon Resonance*; Schasfoort, R. B. M., Tudos, A. J., Eds.; RSC Pub: Cambridge, U.K., 2008.

(9) *Protein Interactions: Biophysical Approaches for the Study of Complex Reversible Systems*; Protein Reviews; Schuck, P., Ed.; Springer: New York, NY, 2007.

(10) Nirschl, M.; Reuter, F.; Vörös, J. *Biosensors* **2011**, *1* (4), 70−92.

(11) Rich, R. L.; et al. *Anal. Biochem.* **2009**, *386* (2), 194−216.

(12) Rich, R. L.; Myszka, D. G. *J. Mol. Recognit.* **2008**, *21* (6), 355−400.

(13) Singh, P. *Surface Plasmon Resonance*; Nanotechnology Science and Technology; Novinka: New York, 2014.

(14) Yang, D.; Singh, A.; Wu, H.; Kroe-Barrett, R. *Anal. Biochem.* **2016**, *508* (Supplement C), 78−96.

(15) Sun, Y.-S.; Landry, J. P.; Zhu, X. D. *Instrum. Sci. Technol.* **2017**, *45* (5), 486−505.

(16) Ylera, F.; Harth, S.; Waldherr, D.; Frisch, C.; Knappik, A. *Anal. Biochem.* **2013**, *441* (2), 208−213.

(17) Samuelsson, J.; Arnell, R.; Fornstedt, T. *J. Sep. Sci.* **2009**, *32* (10), 1491−1506.

(18) Stanley, B. J.; Guiochon, G. *J. Phys. Chem.* **1993**, *97* (30), 8098−8104.

(19) Sandblad, P.; Arnell, R.; Samuelsson, J.; Fornstedt, T. *Anal. Chem.* **2009**, *81* (9), 3551−3559.

(20) Agmo Hernández, V.; Samuelsson, J.; Forssén, P.; Fornstedt, T. *J. Chromatogr. A* **2013**, *1317*, 22−31.

(21) Lipponen, K.; Stege, P. W.; Cilpa, G.; Samuelsson, J.; Fornstedt, T.; Riekkola, M.-L. *Anal. Chem.* **2011**, *83* (15), 6040−6046.

(22) Svitel, J.; Balbo, A.; Mariuzza, R. A.; Gonzales, N. R.; Schuck, P. *Biophys. J.* **2003**, *84* (6), 4062−4077.

(23) Altschuh, D.; Björkelund, H.; Strandgård, J.; Choulier, L.; Malmqvist, M.; Andersson, K. *Biochem. Biophys. Res. Commun.* **2012**, *428* (1), 74−79.

(24) Magjarevic, R.; Schuck, P. Fredholm Integral Equations in Biophysical Data Analysis. In *26th Southern Biomedical Engineering Conference SBEC 2010*, College Park, MD, April 30−May 2, 2010; Herold, K. E., Vossoughi, J., Bentley, W. E., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, Germany, 2010; Vol. 32, pp 340−343.

(25) Zhang, Y.; Forssén, P.; Fornstedt, T.; Gulliksson, M.; Dai, X. *Inverse Probl. Sci. Eng.* **2017**, 1−26.

(26) Ogi, H.; Fukunishi, Y.; Omori, T.; Hatanaka, K.; Hirao, M.; Nishiyama, M. *Anal. Chem.* **2008**, *80* (14), 5494−5500.

(27) Multia, E.; Sirén, H.; Andersson, K.; Samuelsson, J.; Forssén, P.; Fornstedt, T.; Öörni, K.; Jauhiainen, M.; Riekkola, M.-L. *Anal. Biochem.* **2017**, *518*, 25−34.

(28) Sauerbrey, G. *Eur. Phys. J. A* **1959**, *155*, 206−222.

(29) *MATLAB*; The MathWorks Inc.: Natick, MA, 2015.

(30) Coleman, T. F.; Li, Y. *SIAM J. Optim* **1996**, *6* (2), 418−445.

(31) Clustering Methods. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: New York, 2005; pp 321−352.

(32) German, J. B.; Smilowitz, J. T.; Zivkovic, A. M. *Curr. Opin. Colloid Interface Sci.* **2006**, *11* (2−3), 171−183.

(33) Libby, P.; Ridker, P. M.; Hansson, G. K. *Nature* **2011**, *473* (7347), 317−325.

(34) Gardella, T. J.; Jüppner, H. *Trends Endocrinol. Metab.* **2001**, *12* (5), 210−217.

(35) Potts, J. T.; Gardella, T. J. *Ann. N. Y. Acad. Sci.* **2007**, *1117*, 196−208.