Letter

Advanced Analysis of Biosensor Data for SARS-CoV-2 RBD and ACE2 Interactions

Patrik Forssén, Jörgen Samuelsson, Karol Lacki, and Torgny Fornstedt*

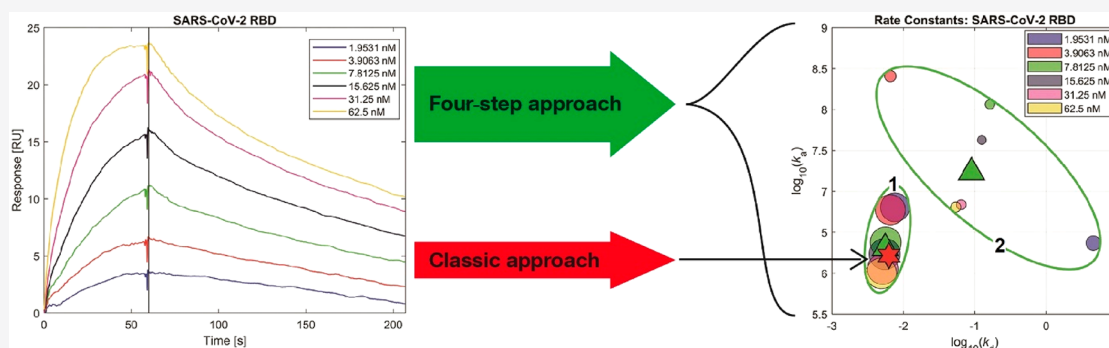Cite This: *Anal. Chem.* 2020, 92, 11520−11524

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The traditional approach for analyzing interaction data from biosensors instruments is based on the simplified assumption that also larger biomolecules interactions are homogeneous. It was recently reported that the human receptor angiotensin-converting enzyme 2 (ACE2) plays a key role for capturing SARS-CoV-2 into the human target body, and binding studies were performed using biosensors techniques based on surface plasmon resonance and bio-layer interferometry. The published affinity constants for the interactions, derived using the traditional approach, described a single interaction between ACE2 and the SARS-CoV-2 receptor binding domain (RBD). We reanalyzed these data sets using our advanced four-step approach based on an adaptive interaction distribution algorithm (AIDA) that accounts for the great complexity of larger biomolecules and gives a two-dimensional distribution of association and dissociation rate constants. Our results showed that in both cases the standard assumption about a single interaction was erroneous, and in one of the cases, the value of the affinity constant K_D differed more than 300% between the reported value and our calculation. This information can prove very useful in providing mechanistic information and insights about the mechanism of interactions between ACE2 and SARS-CoV-2 RBD or similar systems.

After more than two decades of accelerating the technical development of modern biosensor technologies such as surface plasmon resonance (SPR), bio-layer interferometry (BLI), and quartz crystal microbalance (QCM), interaction data are still often analyzed using a single interaction model.[1,2] This model works well when analyzing interactions involving pure receptors and small-size drug candidates in biochemical assays if the kinetics are not heterogonous and slow.[1] However, for detailed characterizations of interactions between more complicated and larger biomolecules, for example, biological drugs and their target receptors in biochemical assays, more advanced data analysis approaches are crucial.[3−5] These approaches should not be considered a replacement for carefully executed experiments but as a way to deal with situations where more complicated interactions are present.

For instance, one of the more advanced approaches is based on single interaction models that consider slow mass transfer, bivalent binding, and conformal changes. These models do not necessarily give a more accurate description of the system studied but will give a better fit because they have more degrees of freedom. Furthermore, there are usually several permutations of these models that fit data equally well, so it is hard to discriminate between them.[6] Another advanced approach, which in the authors' opinions is more robust and consistent, is to consider that the experimental data is the result of multiple interactions. To analyze for multiple interactions, the rate constant distribution (RCD) approach has been proposed which does not require an *a priori* assumption about any specific number of interactions.[7,8] A RCD is a surface with two-dimensional distributions of association and dissociation rate constants where each distribution in this space, represents an important interaction. The RCD approach was recently used by Multia et al.[3] to investigate the interactions between antihuman apoB-100
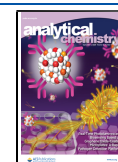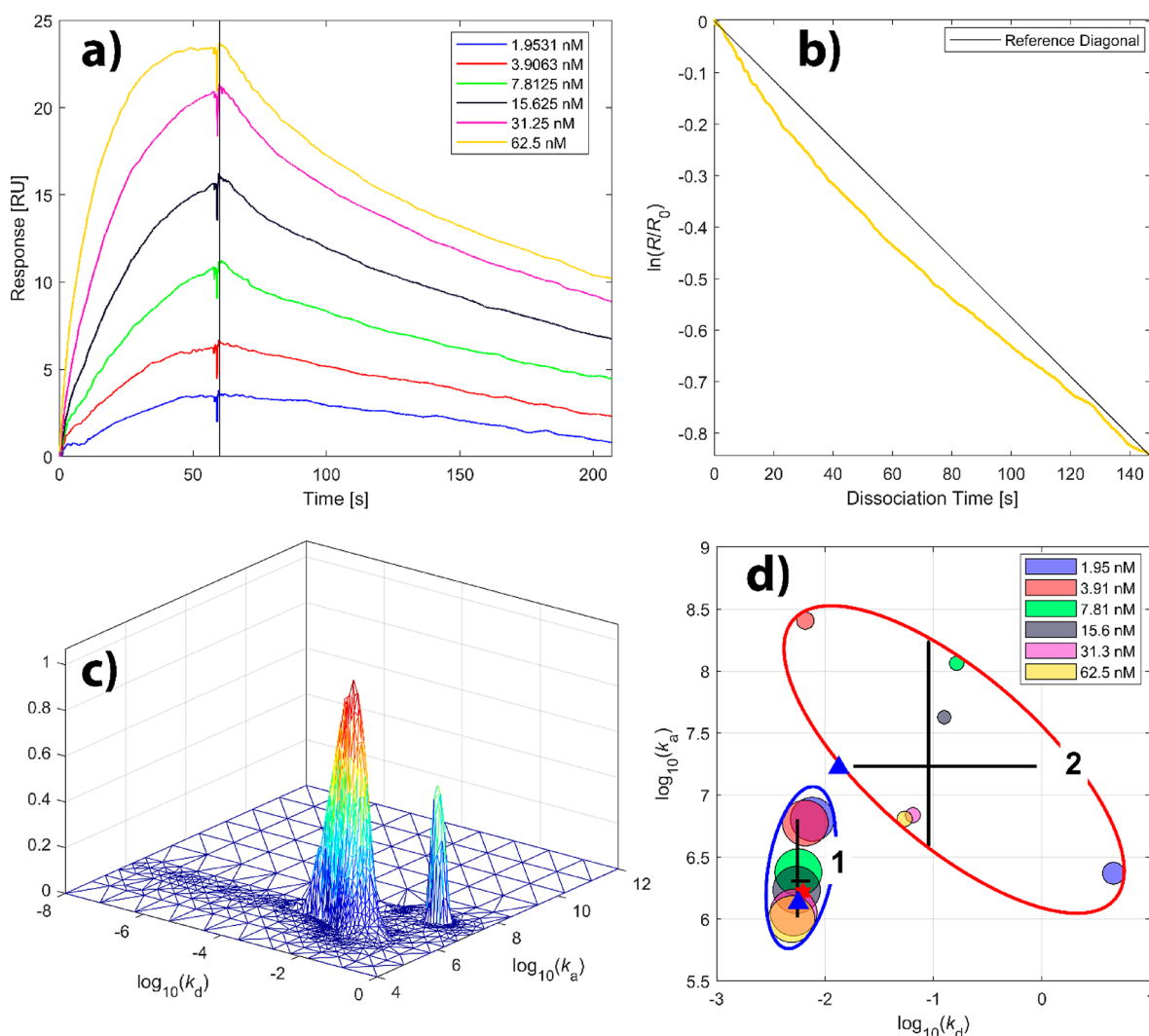
Figure 1. (a) Sensorgrams from the study by Lan et al.[15] using SPR from human ACE2 as immobilized ligand and SARS-CoV-2 RBD as analyte, at different analyte concentration levels; vertical line indicates injection duration. (b) Dissociation graph for the 62.5 nM injection. (c) RCD for a 31.25 nM injection, using the regularization factor $\lambda = 1$. (d) Clustered rate constants estimated from local fitting. The circle areas indicate mean contribution of the two interactions to the total sensorgram response; the crosses indicate the median of the clustered rate constants with corresponding 95% confidence intervals. The red star indicates the rate constants estimated using one interaction overall fitting, and the blue triangles indicate the rate constants estimated using two interactions overall fitting. The corresponding rate constants are presented in Supporting Information Table S1.

monoclonal antibody and lipoproteins. Recently, we developed an improved RCD algorithm, the adaptive interaction distribution algorithm (AIDA), for more refined processing of complex biosensor data.[4].

In late 2019, a pneumonia associated with a coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China,[9,10] and rapidly spread worldwide. Currently, there are no vaccines or any effective specific therapeutic options available for combating the infection. Meanwhile, it is crucial to obtain detailed knowledge of COVID-19 pathogenesis, i.e., the biological mechanisms by which the virus enters and causes the disease in the target hosts.[11] It has been shown that the recently discovered angiotensin-converting enzyme 2 (ACE2), attached to the outer cell membranes of cells in the lungs and in other organs, is the main receptor responsible for SARS-CoV-2 entering the human target body.[12] It was earlier found that ACE2 is also the entry port for the previous coronavirus known as SARS-CoV.[13] Recent studies indicate that SARS-CoV-2 binds more strongly

to ACE2 than does SARS-CoV;[14,15] providing a most interesting starting point for further studies intended to enhance our mechanistic understanding of COVID-19.

Recently, two studies used biosensor assays to determine the interactions between the very complex biomolecules SARS-CoV-2 receptor binding domain (RBD) and ACE2.[15,16] One of the reported findings was that the virus spike proteins have higher affinity to ACE2 than did the previous SARS-CoV.[15] However, biosensors data were analyzed using a simplified model in the software packages that come with the biosensors equipment. The validity of the reported findings therefore needs to be tested using more advanced numerical data processing approaches in order to validate the knowledge gained about COVID-19 pathogenesis.[11]

The aim of this study is to reanalyze recently published interaction data from two selected publications[15,16] using our recently validated four-step approach[4] to see if the results differ.
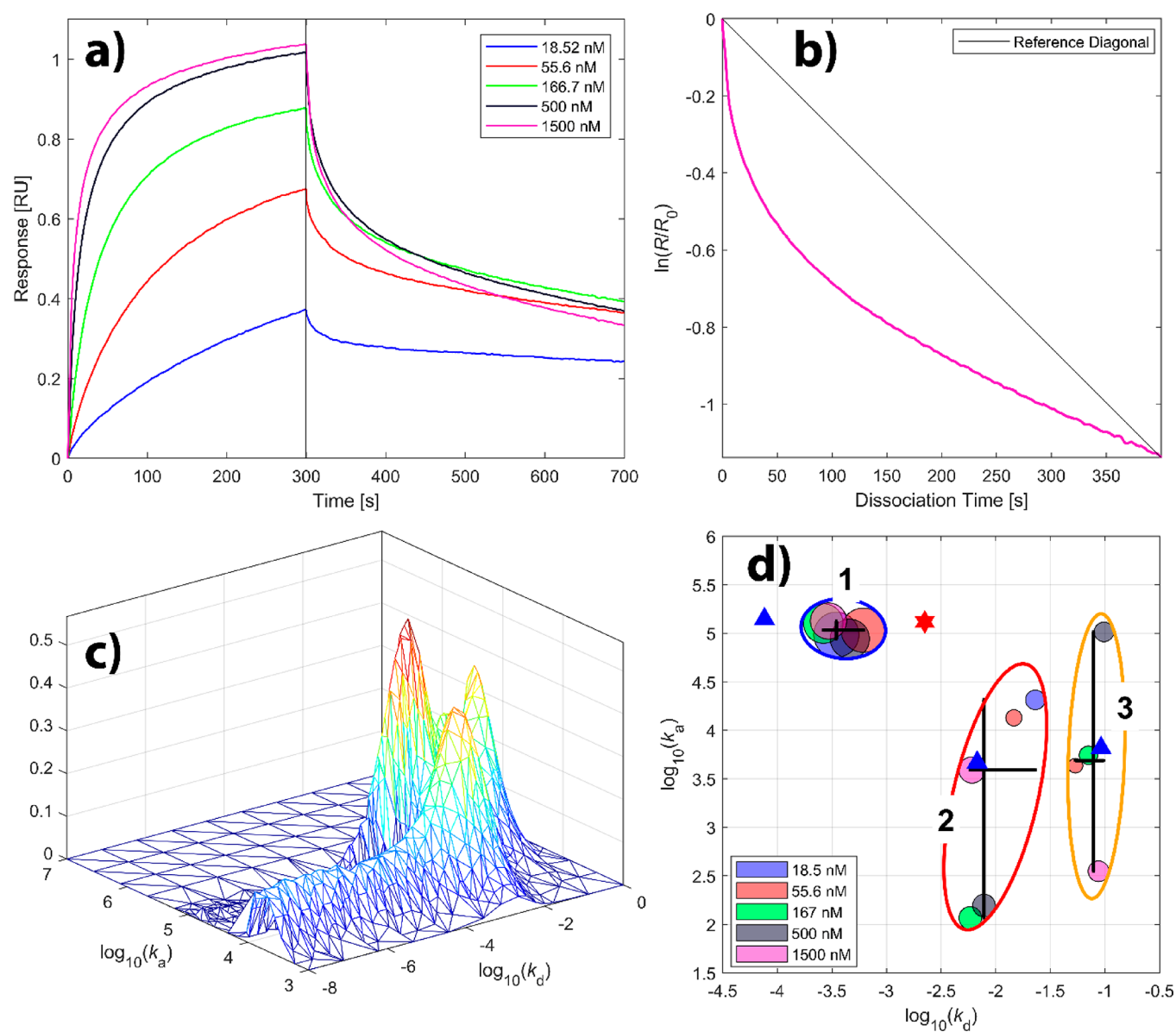
Figure 2. (a) Sensorgrams from the study by Tian et al.[16] using BLI from biotinylated 2019-nCov RBD as the ligand and ACE2 as analyte, at different analyte concentration levels; vertical line indicates injection duration. (b) Dissociation graph for a 1 500 nM injection. (c) RCD for a 1500 nM injection, using the regularization factor $\lambda = 1$. (d) Clustered rate constants estimated from local fitting; one outlier (18.52 nM in the third group) is removed. The circle areas indicate mean contribution of the three interactions to the total sensorgram response; the crosses indicate the median of the clustered rate constants with corresponding 95% confidence intervals. The red star indicates the rate constants estimated using one interaction overall fitting, and the blue triangles indicate the rate constants estimated using three interactions overall fitting. The corresponding rate constants are presented in Supporting Information Table S2.

■ THEORY

Calculations and Algorithms. The binding of analyte(s) A to an immobilized ligand L on a biosensor chip is assumed to proceed according to

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

where k_a and k_d are the association and dissociation rate constants, respectively.

The total response of the biosensor system, R_{tot}, at time t for a system with n interactions can then be written

$$R_{tot}(t) = R_{I,i} + \sum_{i=1}^{n} R_{max,i} K(t; C_i, k_{a,i}, k_{d,i}) \tag{2}$$

where $R_{max,i}$ is the maximum analyte binding capacity (in response units) for the ith interaction; $R_{I,i}$ is an optional "bulk effect" parameter for the ith interaction (used to account for the fact that the base response might change during analyte injection); and K is a function that depends on time, the rate constants $k_{a,i}$ and $k_{d,i}$, and the initial analyte concentration C_i (see refs 4, 17, and 18 for more details).

It is possible to show that in the dissociation phase, i.e., when $t > t_{inj}$, where t_{inj} is the injection duration, the natural logarithm of the ratio between $R_{tot}(t)$ and R_0, where $R_{tot}(t)$ is the measured response during the dissociation phase and R_0 is

the measured response at the end of injection phase, is a linear function of time if and only if we have a single interaction. Thus, by plotting the test function $d(t)$, defined by eq 3, against t, which is known as a *dissociation graph*, we can determine if we have one or several interactions in the biosensor system.

$$d(t) = \ln(R_{tot}(t)/R_0), \quad R_0 = R_{tot}(t_{inj}) \qquad (3)$$

If we let the number of interactions n in eq 2 $\rightarrow \infty$, the problem of estimating the rate constants becomes a Fredholm integral equation of the first kind. This is an ill-posed inverse problem, which requires a so-called regularization in order to solve it, and the solution will depend on the type and amount (indicated by the regularization parameter λ) of regularization applied. The solution will be a rate constant distribution (RCD) surface described above (for more details, see refs 4, 17, and 18).

RESULTS AND DISCUSSION

The measured SPR and BLI biosensor data used here were provided by the authors of the original publications.[15,16] The data was analyzed by the four-step approach developed and validated previously[4] involving first (I) estimating the heterogeneity of the interactions using dissociation graphs and second (II) generating RCDs with AIDA. The two first steps are for obtaining a complete census of all possible existing interactions. In step III, we estimate the rate constants by fitting a suitable interaction model to each sensorgram, and in step IV, we cluster the individual rate constants to obtain overall estimates.

Figure 1a shows the sensorgrams used in reanalyzing the SPR data from Lan et al.[15] Figure 1b shows the dissociation plot for the 62.5 nM experiment, i.e., $\ln(R/R_0)$ versus time, to indicate potential heterogeneity. As discussed in the Theory section, a single interaction would result in a linear dissociation plot, while a heterogeneous interaction would result in concave dissociation plots. In this case, the graph is slightly concave, indicating that the interaction probably is heterogeneous. However, the deviation from linearity is mild, indicating that the contribution of potential secondary interactions to the total observed response is only minor. The root mean square error normalized (RMSEN) for a two-interaction model is 3.5% compared with 5.0% for a one-interaction model. In this case, the secondary interactions account for 13.5% of the overall response (Supporting Information Table S1). The modest improvement in the model fit from using a more complex model indicate that while a one-interaction model is probably sufficient to describe the system it does not provide information about the system heterogeneity, and the system heterogeneity can prove very important when characterizing the system. The system heterogeneity was estimated by calculating RCDs. The RCD for the 31.25 nM experiment (Figure 1c) revealed a major interaction at approximately $\log_{10}(k_{a,1}) \approx 6$, $\log_{10}(k_{d,1}) \approx -2.5$ and a minor interaction at $\log_{10}(k_{a,2}) \approx 7.5$, $\log_{10}(k_{d,2}) \approx -1$. As can been seen from the RCD, the minor interaction is small, but the most suitable model should nevertheless be a two-interaction model. The sensorgrams were therefore fitted individually to a two-interaction model; the fits are shown in Supporting Information Figure S1 together with overall fits for one- and two-interaction models. The clustered estimated rate constants are presented in Figure 1d. The circle areas are proportional to

the interaction's 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 interactions in the RCD (Figure 1c). The estimated rate constants for the minor interaction are much more scattered than those for the major interaction, probably because this interaction contributes very little to the overall response and is therefore very sensitive to system noise. The red star in Figure 1d represents the rate constants determined using a one-interaction overall fit (as the original author did) with R_I adjustment, and the blue triangles represents the rate constants determined using the two-interactions model overall fit with R_I adjustment. To conclude, one interaction accounts for almost 90% of the observed response, and a one-interaction model could therefore be considered a valid model for this system. However, it does not provide information about the system heterogeneity that can be very important.

Figure 2a presents the raw data from Tian et al.[16] To investigate whether the data originates from a single interaction or from more complicated heterogeneous interactions, the dissociation plot, i.e., $\ln(R/R_0)$ versus time, was calculated for the 1500 nM experiment (Figure 2b). The graph is clearly concave, indicating that the data describes more complicated heterogeneous interactions. In the RCD for the 1500 nM experiment, one can identify three different interactions. A three-interaction individual model fit indicates that the major interaction accounts for less than 70% of the overall response (Supporting Information Table S2). In Supporting Information Figure S2, the model fits are presented. In Figure 2d, the clustered estimated rate constants are shown. As in the previous case, the smaller the contribution to the total response is, indicated in Figure 2d by the circle areas, the larger the uncertainty of the determined rate constants becomes. The red star in Figure 2d represents the rate constants determined using the one-interaction overall fit with the R_I adjustment, and the blue triangles represent the rate constants determined using the three-interactions overall fit with the R_I adjustment. What is very important here is that the affinity constant K_D, estimated using an overall fitting strategy with a one-interaction model, and the affinity constant K_D for the interaction making the greatest contribution to the total response, estimated using the four-step approach, differ significantly (Supporting Information Table S2).

CONCLUSIONS

Before presenting our conclusions, it is worth mentioning that when evaluating sensorgrams it is only possible to resolve the interactions that give significant contribution to the response in the studied time interval. For example, very slow interactions in the presence of fast ones cannot be resolved unless the association phase is long (for an example, see Supporting Information Figure S). Ideally, one should reach steady state, i.e., flat sensorgram curve, before ending the association phase. As this can take hours or even days to achieve, very slow interactions are often difficult/impractical to measure, especially using SPR. However, the approach described in this letter can easily detect very slow interactions if the studied association phase is long enough.

Both experimental data sets reanalyzed here should describe the same, or very similar, interactions, but the data were generated using different biosensor technologies. Theoretically, the technology used should not affect the interaction

mechanism. However, although the estimated affinity constant K_D is similar in both systems ($10^{-8.62}$ and $10^{-8.44}$ M, respectively) the estimated rate constants k_a and k_d differ significantly (Supporting Information Tables S1 and S2). There could be many reasons for this. For example, the ligand and analyte might not actually be the same in both systems and that they are also switched; i.e., in the Lan et al. study,[15] ACE2 was immobilized, whereas in the Tian et al.[16] study, SARS-CoV-2 RBD was immobilized instead.

However, investigating the reason for the differences between the two studied biosensor systems was not our intention with this work nor was it to speculate on the origin of any potential secondary interactions. We leave the interpreting of the result to the more capable scientists in pathogenesis who have great knowledge of the interaction systems in question.

Our intention has been to demonstrate the importance of the deeper analytical analysis of biosensors data, and we have shown that such an analysis could give a more adequate representation of reality. In contrast, the simplified approaches at best provided an oversimplification and at worst completely erroneous results. Wrong data interpretation may result in wrong conclusions with potentially very serious consequences, for example, regarding the SARS-CoV-2 binding mechanism.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c02475.

Figure S1: Model fits to the Lan et al. data (ref 15). Table S1: Model fit parameters to the Lan et al (ref 15). data. Figure S2: Model fits to the Tian et al. data (ref 16). Table S2: Model fit parameters to the Tian et al. data (ref 16). Figure S3: Slow association example. (PDF)

AUTHOR INFORMATION

Corresponding Author

Torgny Fornstedt − Department of Engineering and Chemical Sciences, Karlstad University, SE-651 88 Karlstad, Sweden; orcid.org/0000-0002-7123-2066; Email: Torgny.Fornstedt@kau.se

Authors

Patrik Forssén − Department of Engineering and Chemical Sciences, Karlstad University, SE-651 88 Karlstad, Sweden

Jörgen Samuelsson − Department of Engineering and Chemical Sciences, Karlstad University, SE-651 88 Karlstad, Sweden

Karol Lacki − Department of Engineering and Chemical Sciences, Karlstad University, SE-651 88 Karlstad, Sweden

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.analchem.0c02475

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Rich, R. L.; Papalia, G. A.; Flynn, P. J.; et al. *Anal. Biochem.* **2009**, *386* (2), 194−216.

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

(3) Multia, E.; Sirén, H.; Andersson, K.; et al. *Anal. Biochem.* **2017**, *518*, 25−34.

(4) Forssén, P.; Multia, E.; Samuelsson, J.; et al. *Anal. Chem.* **2018**, *90* (8), 5366−5374.

(5) Liangsupree, T.; Multia, E.; Metso, J.; et al. *Sci. Rep.* **2019**, *9* (1), 1−10.

(6) Karlsson, R.; Fält, A. *J. Immunol. Methods* **1997**, *200* (1−2), 121−133.

(7) Svitel, J.; Balbo, A.; Mariuzza, R. A.; et al. *Biophys. J.* **2003**, *84* (6), 4062−4077.

(8) Zhao, H.; Boyd, L. F.; Schuck, P. *Curr. Protoc. Protein Sci.* **2017**, *88* (1), 20.2.1−20.2.25.

(9) Chen, N.; Zhou, M.; Dong, X.; et al. *Lancet* **2020**, *395* (10223), 507−513.

(10) Zhou, P.; Yang, X.-L.; Wang, X.-G.; et al. *Nature* **2020**, *579* (7798), 270−273.

(11) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. *Cell* **2020**, *181* (2), 281−292.e6.

(12) Wrapp, D.; Wang, N.; Corbett, K. S.; et al. *Science* **2020**, *367* (6483), 1260−1263.

(13) Kuba, K.; Imai, Y.; Rao, S.; et al. *J. Mol. Med.* **2006**, *84* (10), 814−820.

(14) Carter, L. J.; Garner, L. V.; Smoot, J. W.; et al. *ACS Cent. Sci.* **2020**, *6* (5), 591−605.

(15) Lan, J.; Ge, J.; Yu, J. *Nature* **2020**, *581*, 215.

(16) Tian, X.; Li, C.; Huang, A.; et al. *Emerging Microbes Infect.* **2020**, *9* (1), 382−385.

(17) Zhang, Y.; Forssen, P.; Fornstedt, T.; Gulliksson, M.; Dai, X. *Inverse Probl. Sci. Eng.* **2018**, *26*, 1−26.

(18) Zhang, Y.; Yao, Z.; Forssén, P.; et al. *Ann. Appl. Stat.* **2019**, *13* (4), 2011−2042.