A unified approach for high throughput analysis of real-time biomolecular interactions in surface plasmon resonance and fluorescence imaging

Catalin Fetita\textsuperscript{a}, Nicolas François\textsuperscript{a}, Françoise Prêteux\textsuperscript{b} and Hervé Delacroix\textsuperscript{c}

\textsuperscript{a}Dept. ARTEMIS, Institut TELECOM / TELECOM SudParis, MAP5 CNRS UMR 8145
9 rue Charles Fourier, 91000 Evry, France
\textsuperscript{b} Mines ParisTech, Paris, France
\textsuperscript{c} Université Paris Sud, Orsay, France

ABSTRACT

The analysis of real-time biomolecular interactions (observation is performed as the biological interaction occurs) provides information on the formation of target/probe complexes, particularly on their dynamic behaviours. Namely, it allows the determination of the affinity constant, a static value that characterizes the interaction properties, using two dynamic values, the association and dissociation constants. Such dynamic behaviour can be assessed either with surface plasmon resonance (SPR) or fluorescence-based biosensors. The challenging issue is the automatic extraction and analysis of the interaction signal for each spotted probe on the biosensor in a high-throughput framework (hundreds of probes). This paper addresses such issue and develops a unified approach for analyzing the image data provided by the above-mentioned technologies. A mathematical modelling of the image data allowed building-up a virtual biosensor able to simulate biologic experiences related to various possible parameters (level of signal and noise, presence of artefacts, surface functionalization, spotting heterogeneity). Based on such simulation, a generic and automated approach combining 3D mathematical morphology and spatio-temporal classification is proposed for detecting the interacting probes, segmenting the regions of effective signal, and characterizing the associated affinity constants. The developed method has been assessed both qualitatively and quantitatively on simulated and experimental datasets and showed accurate results (maximum error of 7\% for the most difficult cases in terms of noise and surface functionalization).

Keywords: biosensors, real time assays, surface plasmon resonance, fluorescence, biomolecular interactions, hybridization constant, filtering, classification, segmentation, lab-on-chip

1. INTRODUCTION

Recent developments of biosensor technology for biospecific interaction analysis permits the monitoring of molecular reactions in real time both by surface plasmon resonance (SPR) or fluorescence-based approaches. The basis of these approaches resides in the specificity of recognition observed between separate analytes through the formation of stable complexes in which one of the biological elements (an enzyme, receptor, peptide or even a living cell) is integrated onto a solid functional layer (probe) while the other (target) is free to circulate in a liquid medium. The biological reaction is coupled with a physical transducer and converted into a measurable signal. The extent of this signal constitutes the degree of biomolecular interaction per surface unit. The detection method is either direct (without labeling, in SPR approach) or indirect (requiring the target labeling by a fluorophore) in a heterogeneous environment and it allows, for example, the characterization of antibodies and antigens or the detection of particular molecules.

The absence of labeling in the SPR biosensors offers other many advantages. Sample preparation is simplified and allows all the steps of the procedure from the use of biosensor to the biomolecular interaction to be evaluated with the biosensor machine.\textsuperscript{1} A labeling step may potentially influence the properties of the biomolecule\textsuperscript{2} and

\textsuperscript{1} Further author information: (Send correspondence to C.F.)
C.F.: E-mail: catalin.fetita@it-sudparis.eu
may result in some bias in integration. Thus, evaluation of the degree and fidelity of labeling needs to be performed.\textsuperscript{3} If detection of a labeled molecule is used, the measurements will depend drastically on the type of marker used and is totally independent from the molecule of interest. Conversely, since the detection of markers is done on a black background - contrary to the SPR - the photon noise is much lower in this case, which allows a lower threshold of detection and measurement of lower affinities\textsuperscript{4}.

With the SPR methods, the variations of the index of reflection are observed at the level of the metal/dielectric interface and, consequently, the measured signal amplitude depends on the mass of the biological elements. Thus, this system should have better results in the case of large molecule analysis, such as the antibodies, whereas, for the study of small molecules, such as short oligonucleotides, the resolution of the SPR would be lower than a fluorescence-based technology.

Finally, the analysis in real-time (observation is performed as the biological interaction occurs) generates more information on the formation of target/probe complexes, particularly on their dynamic behaviors. It permits the determination of the affinity constant, a static value that characterizes the interaction properties, using two dynamic values, the association and dissociation constants\textsuperscript{5}. In addition, the possibility of undertaking parallel, sequential detection is also an important point since it allows simultaneous observation of the interactions of multiple targets with a variety of probes which are fixed on the surface of the biosensors (high-throughput analysis). This will result in savings in time and will limit the quantity of biological materials that are needed to perform the experiments. Moreover, the individual interactions are measured in parallel under identical conditions with a single machine thus resulting in lower experimental variation and more precise/reliable comparison of results.\textsuperscript{6}

The objective of this study is to provide a unified and automated approach for high-throughput analysis of real-time biomolecular interactions investigated with surface plasmon resonance (SPR) or fluorescence-based biosensors, for clinical lab-on-chip applications. Under the assumption of well calibrated experimental test rigs (specific to each technology), the challenge refers to the ability of characterizing the interactions between injected targets and each spotted biologic entity (probe) on the biosensor, regardless to the biases introduced by the optoelectronic device and unpredictable events (spotting heterogeneity, leakages, chip scratches). We address here the image analysis problem of detecting the probes on the biosensor that interacted with the targets, localizing the homogeneous region within each probe corresponding to the unbiased interaction signal, and characterizing the associated affinity constants.

The study of experimental data collected both by fluorescence and SPR biosensors allowed establishing a set of hypothesis which led to a mathematical model simulating these acquisitions. Based on this model, a virtual biosensor able to simulate biologic experiences related to various possible parameters (level of signal and noise, presence of artefacts, surface functionalization, spotting heterogeneity) was built up. Based on such simulation, a generic and automated approach combining 3D mathematical morphology and spatio-temporal classification is proposed for detecting the interacting probes, segmenting the regions of effective signal, and characterizing the associated affinity constants (§ 2). The developed approach is evaluated both qualitatively and quantitatively on simulated and experimental datasets, showing accurate results (§ 3).

2. MATERIALS AND METHOD

Studies of the real-time biomolecular interaction kinetics on several dynamic experiences performed in SPR and fluorescence have shown that such interactions are characterized by a bi-exponential model corresponding to the association/dissociation phases during the target injection and wash-out, respectively (eq. 1, Figure 1(a)).

\[
R(t) = \begin{cases} 
R_a(t) = R_0, & t < t_0 \\
R_d(t) = R_0 + R_{aM} \left(1 - \exp\left[-k_{obs}(t - t_0)\right]\right), & t \in [t_0, t_d), \\
R_d(t) = (R_d(t_d) - R_{dM}) + R_{dM} \exp\left[-k_{off}(t - t_d)\right], & t \geq t_d,
\end{cases}
\]

where the \( a \) and \( d \) indices refer to the association and dissociation phase, respectively. Such ideal signal is biased in practice due to phenomena associated with the experimental protocol and the optoelectronic transducer: optoelectronic noise, spotting heterogeneities and artefacts (leakages, chip scratches), surface functionalization (other possible signal perturbations like temperature drift or quenching effect in fluorescence are considered to be dealt with during the calibration phase specific to each technology).
Given a SPR or fluorescence experience, the objective is to automatically detect the bi-exponential model parameters of the interaction monitored by each probe on the biosensor, overcoming the above-mentioned biases. In order to develop a generic approach for SPR and dynamic fluorescence image analysis, we have first simulated a virtual biosensor including several probes randomly distributed and having different degrees of spotting heterogeneity, chip scratches and leakages (Figure 1(b)).

Each homogeneous region has associated a temporal bi-exponential interaction signal of parameters which may vary according to real acquisition ranges. A set of hypothesis were summarized from real biological interaction data and implemented in the virtual biosensor as follows (Figure 2):

- the spatial position of the spots is assumed unknown, as in the case of manual spotting. The expected spot diameter on the biosensor image is supposed known as it can be deduced from the size of the spotting needle and optoelectronic transducer properties;
- the biosensor should include a minimum number of spots to simulate the variability of a high-throughput analysis;
- the resulting image dataset is affected by a zero-mean normal-distributed noise inherent to the optoelectronic device (Figures 2(b), 2(c));
- the absolute temporal signal of each spotted region \( R(t) \), eq. 1, is always greater than the background signal, but this assumption does not hold for the signal amplitude \( R(t) - R_o \). Indeed, in SPR imaging, if the background region is not properly deactivated during the surface functionalization, it may result in a larger amplitude of the background signal than some of the interacting probes (Figure 2(b),2(c)). This case will be taken into account by the analysis approach developed;
- the spotted probes are not assumed homogeneous: several spots include regions of deposition heterogeneity and chip scratches (artefacts) which may occur at any spatial location in the spot. The signal associated with such regions differ from the effective interaction signal and several situations were included in the simulation. In order to discriminate a valid signal inside a spot (and thus validate the measured parameters) the following assumptions are considered: (1) the surface of the valid signal region inside the spot is larger than that of possible heterogeneities, (2) the amplitude of the valid signal is larger than in artefacts; however, the signal value before target injection, \( R_o \) may be larger in artefacts, and (3) the signal inside chip scratches is similar with a deactivated background;

(a) Bi-exponential model corresponding to the association/dissociation phases during the target injection and wash-out.

(b) Virtual biosensor spotting simulation. 0: background; 1-8: probe regions of effective signal; 9-10: leakages; 11-19: heterogeneous regions (biased signal); 20-22: chip scratch.

Figure 1. Interaction model and virtual biosensor setup.
• during spot deposition, leakages may occur (Figure 1(b)), their signal being similar with the one of biochip artefacts.

![Image](image.png)

(a) Biosensor image at a point of the association phase.

(b) Absolute signal associated with each region (top) and amplitude value (bottom) for deactivated background.

(c) Absolute signal associated with each region (top) and amplitude value (bottom) for non deactivated background.

Figure 2. Example of signal simulation for the virtual biosensor.

By varying the signal parameters, eq. 1, for each defined region of the virtual biosensor, several biological interactions can be simulated simultaneously on the same biochip. The purpose of this paper is to develop an automated image analysis approach (automation being a strong requirement for high-throughput investigation) to extract the effective interaction signal inside each spot and to evaluate the accuracy of parameter estimation with respect to the simulations.

The proposed generic analysis approach combines 3D mathematical morphology for spot detection and signal amplitude extraction with spatio-temporal classification for the effective signal identification inside each spot. The procedure performs successively:

1. definition of the biosensor background region allowing to manage possible non-deactivation of the background signal. This step requires the detection of probe locations under the assumption of a manual spotting or a missing spotting file. Such procedure exploits the morphological operator ascension cost applied on the average biosensor image obtained before the target injection, § 2.1;

2. detection of the interacting probes - this step involves the optoelectronic noise filtering and background signal suppression providing the amplitude signals of the spots. The background signal suppression exploits the background region previously detected, which signal is subtracted form the original data via a 3D gray level geodesic reconstruction. The interacting probes are detected by a hysteresis thresholding of the filtered data, § 2.2;

3. surface partitioning of each interacting probe in regions of effective signal versus background/artefacts. The objective is to extract the probe region of homogeneous temporal signal characterizing the interaction (and avoid including regions affected by noise filtering, namely the spot border, and spotting artefacts).
A k-means partitioning-fusion approach exploiting the temporal behaviour of each point in a ROI around each probe is applied here (highly parallelizable), § 2.3;

4. interaction parameter estimation of the average signal in the effective probe region - the bi-exponential model parameters are obtained by non-linear regression.

2.1 Background region detection

In the following, we consider the reader familiar with the definition of the basic binary and grayscale morphological operators, as well as with the operators of grayscale reconstruction by geodesic dilation and erosion. We introduce three additional operators which will be exploited in the detection scheme.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $Y \subseteq \text{supp}(f)$ a nonempty set. The connection cost of $f$ with respect to $Y$, Figure 3(a), is given by:

$$C_f(., Y) = \rho_f^Y(g_Y),$$

with

$$g_Y(x) = \begin{cases} +\infty & \text{if } x \notin Y \\ f(x) & \text{otherwise} \end{cases},$$

where $\rho_f^Y(g)$ denotes the grayscale reconstruction by erosion of $f$ from $g$.

The sup-constraint connection cost, Figure 3(b), is given by:

$$C_{\text{supc}}^t(f) = \rho_f^Y ([f \bullet D_t] \land g_Y), Y = \partial f,$$

where $f \bullet D_t$ denotes the morphological closing of $f$ by a disk structuring element of radius $t$, $\partial f$ the border of $\text{supp}(f)$ and $\land$ the infimum operand.

Let $f, g : A \subseteq \mathbb{R}^n \rightarrow B \subseteq \mathbb{R}, f \geq g$. The ascension cost of $g$ in $f$, Figure 3(c), is given by:

$$C_{\text{asc}}^f(g) = f - \rho_f^g(g),$$

where $\rho_f^g(g)$ denotes the grayscale reconstruction by dilation of $f$ from $g$.

As mentioned previously, the expected spot diameter is supposed known and will be further denoted by $d_s$. Two additional parameters are defined as follows:

$$r_1 = \frac{3}{8}d_s, \quad r_2 = \frac{r_1}{2}.$$
The purpose to detect a (sub)set of the biochip background is equivalent to detecting an (over)set of the spots. According to the hypotheses previously formulated on the spots and background signals, this task can be achieved by designing a contrast-marker module (CM) combining the ascension cost and a morphological opening; the spot position is then specified by local maxima detection, Figure 4.

![Figure 4. Contrast-marker module. LM denotes local maxima, \( \wedge \) the \textit{infimum} operand, and \( \circ \) the morphological opening.]

The contrast-marker module is applied on an average image, \( I_{\text{ave}} \), of the biosensor signal obtained before the target injection. The reference subset (here denoted by R) is necessarily a subset of the background. A possible estimate could be a border image, \( I_{\text{border}} \), which has the maximum grayscale value for all pixels in a small distance (here 5 units) from the image border, and zero elsewhere. In order to include in the background region possible high contrast zones corresponding to spotting leakages, a better estimate will be defined, \( I_{\text{msk}} \), by using the scheme in Figure 5. Figure 6 shows the first estimate of the border image and few intermediate results of the scheme in Figure 5.

![Figure 5. Reference mask definition for background detection. \( \oplus \) denotes morphological dilation and \( \bar{f} \) the negative of \( f \).]

The spotted probes on the biochip surface can now be more accurately detected by using the first background estimate \( I_{\text{msk}} \) and the contrast-marker module as shown in Figure 7. The contrast image \( I_{f2} \) is then reconstructed by grayscale dilation based on the local maxima extracted from each spot, which results in a better suppression of the background. A regularisation of the probes surface is achieved by applying the connection cost operator with respect to the image border, \( C_f(., \partial f) \), which “levels” the low-contrast artefact regions inside the spots. The resulting image, \( I_{cc} \) does not include anymore large leakages around the spots, only small ones may subsist.

![Figure 6. Several steps of the biochip background first estimate according to the scheme of Figure 5. The background subset is shown in white.](a) \( I_{f1} \). (b) \( I_{\text{mx1}} \). (c) \( I_{\text{msk}} \).]
These latter will be removed in the next step or during the final classification stage (§ 2.3). Figure 8 illustrate the intermediate images produced by the scheme of Figure 7.

Finally, a better delineation of the spots is achieved by applying the sup-constraint connection cost in a multiresolution scheme (Figure 9) to the filtered spots image, $I_{cc}$, which allows removing the leakages of small spatial range and of quasi-homogeneous signal level. A first filtering by $C_{cc1}$ allows leveling the “valleys” of size smaller than $r_1$ ($I_{cc1}$, Figure 10), while a second one by $C_{cc2}$ selects all the spots and preserves intact the homogeneous leakage regions around them ($I_{cc2}$, Figure 10). The binarisation of the difference segments the spot subset, $I_{seg}$, the missing part corresponding to the scratches on probes border. The searched spot set $I_{spots}$ is obtained by a closing-opening filtering of sizes $r_2, r_1$. The background (sub-)mask $I_{bkg}$ is obtained by the erosion of the $I_{spots}$ complement in order to guarantee an empty intersection between $I_{bkg}$ and the set of spots.

Note that the extraction of the biosensor background mask relies on the detection of the spotted probes, prior to the target injection. Such detection do not provide any information about the biological interaction (i.e. which probes will interact with the targets) and is not required if a spotting robot is used and records the spot position.
Figure 9. Biosensor background definition.

Figure 10. Several steps of the biochip background estimate according to the scheme of Figure 9. The background subset is shown in white.
2.2 Detection of the interacting probes

In order to detect which probes have interacted with the targets, the information exploited is the amplitude of
the interaction signal. In this respect, the optoelectronic noise has to be filtered out and the background signal
suppressed, especially in the case of non-deactivated biochip surface.

The noise filtering uses a spatio-temporal Gaussian smoothing of small kernel size (with respect to the spot
diameter and temporal frame rate). This will result in a trade-off between noise suppression and the induced
signal bias on the probes border due to the blurring effect. Figures 12(a), 12(d) show the result of the filtered
sequence further denoted by \( V_{ini} \).

The background signal is suppressed by using the ascension cost operator, Figure 11, applied in a 3D (spatio-
temporal) framework to the whole interaction sequence \( V_{ini} \) with respect to a 3D mask of the background, \( V_{bkg} \),
which is obtained by replicating the \( I_{bkg} \) image throughout the sequence, Figures 12(b), 12(c). The amplitude
signal is then obtained by subtracting a reference image from the whole sequence, which represent the average
signal before target injection, Figures 12(c), 12(f). The biosensor regions which have interacted with the targets
are selected via a hysteresis thresholding performed with respect to the noise level estimated from the original
data, Figures 11, 13(a). The potential noise in the resulting image, \( I_{actspot} \), is suppressed by a morphological
opening exploiting the expected size of the spots, which provides the interacting probes, \( I_{inter} \). Figures 11, 13(b).
At this stage an affinity analysis can be performed with respect to the maximum amplitude of the signals over
the biosensor in order to define strong and weak interactions. An example shown in Figure 13(c) defines strong
interactions as those of signal amplitude greater than the half-maximum signal amplitudes over the biosensor.
In the following we shall however preserve all the interacting spots for the affinity constants analysis.

![Figure 11. Synoptic scheme for detection of the interacting probes.](image-url)
2.3 Effective signal extraction and interaction constants computation

The effective signal enclosed by each interacting probe is extracted individually by analyzing a region of interest (ROI) defined around each spot of $I_{\text{inter}}$ (highly parallelisable procedure). For a homogeneous behavior of the
temporal interaction signal at the spot surface, the effective signal can be computed as the average of the signals of all points of the spot. The presence of spotting heterogeneities and also the noise filtering which smoothes the spots border will introduce biases in the signal estimation. Thus, the effective signal should be computed only from the spot points with similar interaction behavior. In order to extract such points, we have developed a classification procedure of the regions in the spot ROI by means of a partitioning-fusion approach.

The partitioning approach relies on a k-means classification of the point signals in the ROI, implementing the algorithm of Hartigan-Wong\textsuperscript{8} where the class initialisation is performed as in.\textsuperscript{9} The key issue is the choice of the number $K$ of classes. Due to the above-mentioned considerations, it is required that $K > 2$, even for homogeneous probe spotting.

After the k-means partitioning, the fusion approach will group back together the ROI classes obtained, which average signal amplitude is 95\% similar. The class corresponding to the effective signal is selected according to its signal amplitude and its surface on biochip. Figure 14 shows the class of effective signal obtained for the simulations in Figure 2 with $K = 7$.

For each interacting probe, the parameters of the bi-exponential model eq. 1 are estimated from the effective signal by means of least squares nonlinear regression.\textsuperscript{10} Figure 2 shows the estimated effective signals superimposed on the ideal signals injected in the simulation (before adding the optoelectronic noise). Note that, in Figure 14(c), the inclusion of region \#14 into the effective signal of spot \#3 (Figure 1(b)) leads to a bias in signal parameter estimation (Figure 14(d) - third curve from top to bottom).

![Figure 14. Effective signal extraction from the interacting probes of Figure 2 and parameter estimation (red) with respect to the simulated ground truth (black). Left: deactivated background, right: non-deactivated background.](image)

3. RESULTS AND DISCUSSION

The developed methodology has been assessed on simulated data of various parameters for eq. 1 (strong/weak affinity), and different levels of noise, for both deactivated and non-deactivated background, Figures 15, 16, and on few experimental data\textsuperscript{11} by comparison with the expected affinity constants known from literature or obtained via immunoassay tests (ELISA).
Figure 15. Different signal simulations (ideal and with added optoelectronic noise $\mathcal{N}(0, \sigma)$) for the biosensor regions of Figure 1(b) considering a deactivated background. Simulations of regions #1, #2, #3, #4 and #8 (Figure 1(b)) correspond to strong affinity interactions, #5 and #6 to weak and #7 to very weak affinity interactions respectively. The bottom row corresponds to a higher sensibility of the optoelectronic transducer (higher signal and noise amplitude).
Figure 16. Different signal simulations (ideal and with added optoelectronic noise $N(0, \sigma)$) for the biosensor regions of Figure 1(b) considering a non-deactivated background. Simulations of regions #1, #2, #3, #4 and #8 (Figure 1(b)) correspond to strong affinity interactions, #5 and #6 to weak and #7 to very weak affinity interactions respectively. The bottom row corresponds to a higher sensibility of the optoelectronic transducer (higher signal and noise amplitude).

The use of simulated data allowed to find the method parameter range for optimal results and to formulate recommendations for the analysis. Thus, the value of the unique parameter of the method, the number of classes $K$, was found to be appropriate in a range $K \in [7, 10]$. Large values of $K$ will induce the inclusion of biased spot border regions into the effective signal for non-deactivated background. This behaviour may be also encountered for the defined $K$ range when the non-deactivated background signal amplitude is close to the one of the effective signal (Figure 17). This suggests that background deactivation should be considered as a requirement for the interaction data analysis. Note also that for high noise levels and very low interaction affinity (e.g. zone #7 of biosensor which has a behavior similar with the background - simulation $S_6$, Figure 15(h)) some border regions are included in and will bias the effective signal, Figure 17(f). Nevertheless, this is an extreme situation included here for the purpose of the analysis, but such low affinity interaction is unlikely to be considered for further biological interpretation.
The simulated data analysis showed that for strong affinity interactions and background deactivation, the interaction parameter estimation error was less than 3%, whereas for weaker affinity interactions such error climbs up to 5.6% and respectively to 6.5% for background non-deactivation.\footnote{11}

The developed method was also applied to few experimental datasets for a qualitative evaluation. Figure 18 illustrates the result obtained on a manual spotting experience in SPR involving several probe types and successive injections of different targets, showing problems of leakages, chip scratches and non-deactivated background. Figure 19 illustrates an example of interaction detection in dynamic fluorescence.

To sum up, the proposed approach, fully-automatic, responds to the requirements of high-throughput analysis of dynamic biological interactions but it can also be used for applications targeting only the detection of an interaction (\textit{e.g.} presence of pathogens, mutations) cf. \S\ 2.2. Note also that the methodology developed in \S\ 2.1 can be applied for the analysis of static (final point) interactions ($I_{spots}$).
Figure 18. Example of interaction analysis in SPR: biosensor with three types of biologic entities manually spotted. Interaction detection and hybridization curves obtained during a first injection of against-mouse specific antibodies, top (only the mouse specific antibodies responded), followed by a washing step and a second injection of p53 protein, bottom (only the p53 antibodies responded). Non-specific interactions occurring on the background or at some spots surface were correctly bypassed by the developed approach. White - strong affinity, grey - low affinity interaction regions.

Figure 19. Example of interaction analysis in fluorescence. White - strong affinity, grey - low affinity interaction regions.
Conclusion

High-throughput quantitative analysis of biomolecular interactions via confirmed (fluorescence) or emerging (SPR) nanotechnologies attempt to breach the monopoly of the classic immunoassays. This requires the development of appropriate technology platforms that provide reliable, ultra-sensitive, rapid, quantitative, low-cost and multiplexed identification of biomarkers. This paper develops a unified methodology for automated quantitative analysis of image data provided by the SPR and dynamic fluorescence technologies, showing good accuracy in estimating the interaction parameters.

REFERENCES