Analyst

Cite this: *Analyst*, 2012, **137**, 3119

www.rsc.org/analyst

PAPER

Biochip data normalization using multifunctional probes

Jérôme Hottin, Julien Moreau, Alain Bellemain and Michael Canva*

Received 26th January 2012, Accepted 12th April 2012
DOI: 10.1039/c2an35120j

Using a biochip with stable probe functionalization and a detection system capable of real time measurements, it is demonstrated that acquired probe–target interaction data are more reproducible in time – on a given probe spot using sequential target runs – than in space, using many probe spot replicates on the biochip in one single parallel target run. To increase the biochip data precision, a normalization method that quantifies and corrects the surface inhomogeneity without the use of complex data post-processing has been developed. This simple and effective method is based on adding a common reactive group to all probes and quantifying the biochip response to a calibration target, thus quantifying the spatial heterogeneity in the biosensor responsiveness. The usefulness of such methodology, which can be easily generalized, is demonstrated in the model case of DNA:DNA interactions, using a surface plasmon resonance imaging system as the dynamical reader. The biochips are based on streptavidin biochemically functionalized gold films onto which biotinylated ssDNA probe sequences, related to cystic fibrosis genotyping, are spotted. This normalization method provides high gain in data precision and allows, in this example, unambiguous genotyping of SNP, including discrimination of the heterozygote case from the two homozygote cases.

Introduction

Since the demonstration of Southern Blot thirty years ago, surface detectors and in particular biochip systems have become a standard tool for molecular interaction studies in research laboratories.[1–3] More and more laboratories and companies offer their own patented biochip manufacture,[4–6] pointing out its capacity to precisely identify and differentiate each biomolecular interaction in fields such as protein–protein,[7,8] protein–DNA,[9] or DNA–DNA[10–12] interactions. However, these systems have not yet really reached accurate routine analysis outside the research field. The main reason is the difficulty of routinely interpreting the biochip raw datasets due to many flaws in the realization processes of the chip and its associated reactants and consumables. In the important case of gene expression using labeled targets, such an issue has been discussed and various innovative techniques have been set up mostly for non-dynamical systems. Biochip dedicated structures, such as co-immobilized probes,[13,14] labeled control protocol[15] as well as numerical data processing,[16,17] present attractive methods to remove variability during chip printing or data analysis dealing with data noise. Although these procedures are well suited in the former case, where typically two targets compete for one probe, they are not always adopted, as for example in the case of genetic diagnosis where, for a single nucleotide polymorphism, one or two ssDNA targets from the patient will compete for two probes differing by only one base. In this case, as in others, it is preferable to identify and quantify the cause of data dispersion and then be able to correct for it. Moreover, to increase diagnosis accuracy it is necessary, in addition to the background normalization, described in previous publication,[15] to drastically reduce the response dispersion between spots in order to quantify minute differences between them. This is the issue we address in this communication by adding to all individual probes a sequence allowing quantification and correction of effective probe density.

A major source of biochip data dispersion is attributed to sensors' surface inhomogeneity.[18–20] Indeed, a given spot, which is a well-defined spatial zone where probes are linked to the surface, is never totally homogeneous in probe concentration. Furthermore replicates of a spot localized in different areas of the biochip will not show the same reactivity. Such variations tend to become even more critical when comparing responses of probes whose spotting cannot be reproduced in a rigorously identical way. No truly satisfactory simple strategies to successfully handle such spot variations are proposed in the literature. This can directly impact the data precision one can obtain from biochip measurements and drastically limit the applications when minute differences are to be quantified and sorted. This is the case for example in the quantification of a mutation's impact on protein functionalities or in the diagnosis of the genotype of patients for known possible mutations, including SNP (Single Nucleotide Polymorphism). In this paper, a new, simple, non-labeled and effective normalization methodology for biochip data dispersion and subsequent correction for assays with low to high number of

Biophotonic Team, Laboratoire Charles Fabry, Institut d'Optique – Graduate School, 2 Avenue Augustin Fresnel, RD128, 91127 Palaiseau, France. E-mail: michael.canva@institutoptique.fr

spot replicates is presented. To illustrate such a methodology, biochip data dedicated to SNP diagnosis related to Cystic Fibrosis (CF) are acquired using a laboratory made Surface Plasmon Resonance Imaging (SPRI) system. Significant improvement in data reproducibility and precision is demonstrated.

Materials

(a) SPRI reader system

A dynamical biochip SPRI reader system[21–23] allowing sequential measurement cycles was used. It monitors the reflected spatial intensity evolutions $\Delta R(x, y, t)$ of a light beam whose reflection is locally modified by the local quantity of the biomaterial present in the vicinity of the biochip surface. This monitoring therefore allows for characterization of the surface interactions between targets and probes bound to the biochip (Fig. 1) – 'x' and 'y' being the spatial coordinates on the biochip and "t" the time. This system, previously described in detail,[9,24] is based on a Kretschmann SPR configuration[25] and uses a TM polarized ($\lambda = 660$ nm, FWHM 50 nm) light source, an afocal system for imaging purpose and a SF11 high index prism ($n \approx 1.78$ at 660 nm). The spatially generated plasmonic images of the biochip are recorded by a digital 12 bit CCD camera system (PCO, PixelFlyQe).

Physically, the energy of the incident light beam will be partially transferred to plasmonic surface modes depending on the surface optical characteristics. When biomolecular targets in solution interact and bind to probes fixed on the surface, the change in the local optical index induces a local shift in the plasmon resonance conditions. This shift induces a local increase in reflectivity which is recorded and quantified by the CCD camera. Recorded images are processed using lab made codes that involve automatic segmentation of the spot areas and averaging the reflectivity coefficient R within, associating to each spot, indexed i, a physical $R^i(t)$ dataset. Taking a reference at a time t_0 leads to a useful signal $\Delta R^i(t) = R^i(t) - R^i(t_0)$. Typical sensitivity of a few percent of reflectivity variation per nanometre of the average thickness of the biomolecular film, like for hybridized DNA targets, are obtained.

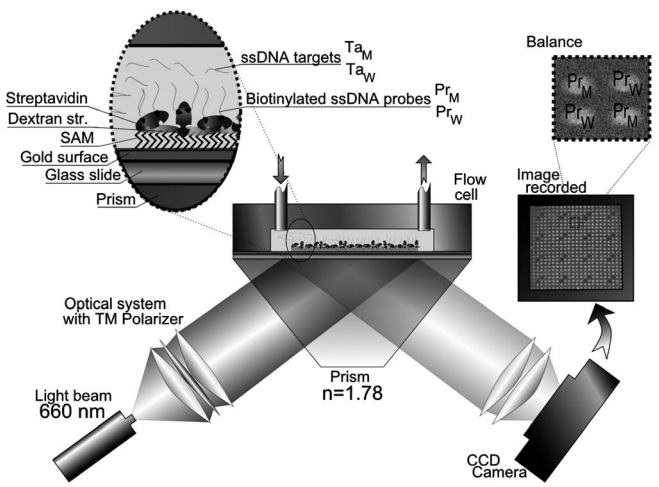

Fig. 1 Scheme of the dynamical SPRI system with a 660 nm TM polarized light beam, a high index coupling prism, a CCD camera, and a dextran functionalized biochip – ssDNA probes were spotted on the biochip as a matrix – ssDNA targets are injected in the flow cell using a peristaltic pump.

A 10 µl flow cell is fixed on the biochip. Solutions are injected using a peristaltic pump (Minipuls, Gilson) with a flow rate of about 70 µl min^{-1} with a void volume of about 120 µl. The temperature of the cell can be set in the 10 to 40 °C range using a Peltier device with a 0.1 °C stability precision.

(b) Biochips and reactants

Biochip plasmonic substrates consisted of a 48 nm gold layer evaporated on a 2 mm SF11 substrate covered by a 2 nm chrome transition layer, chosen for achieving near optimal performances in sensitivity,[26] as evaluated using the previously reported Rouard extended method. A dextran based chemistry was used to functionalize the gold surface,[27] as it offers good performances in terms of specificity, probe binding strength and regenerative capability. Biotin functionalized probes can then be strongly attached to the last layer of streptavidin with a very high affinity ($K_a \approx 10^{13}$ to 10^{15} M^{-1}). 24 × 24 ssDNA spot matrices were prepared, as shown in Fig. 1 (image recorded). The 576 spot areas allowed many replicates to be realized on the same biochip.

The DNA sequences were chosen to serve the clinical purpose of a Cystic Fibrosis (CF) molecular diagnosis. The CF genetic disease is the result of mutation(s) in the Cystic Fibrosis Transmembrane Regulator (CFTR) protein's gene. We have previously shown that these mutations can be detected from exons and introns of interest obtained from patient genes, using SPRI.[24,28] For such a hereditary genetic disease it is necessary to identify three types of patients: those homozygous with two identical genes, either with 2 wild genes referred to herein as "Wild Type" (WT), those with 2 mutated genes referred to herein as "Mutated Type" (MT) and those with two different genes, heterozygous, referred to herein as "Heterozygote Type" (HT) characterized by the presence of both one wild gene and one mutated gene. Such classification into three groups is a classical experimental necessity in all studies of recessive genetic diseases. For this study, we have chosen two oligonucleotide sequence targets related to CFTR. These oligonucleotides are 25 base long ssDNA and are centred on the point mutation location to analyse, referred herein as the M470V polymorphism located at position 1540 in exon 10 of the CFTR gene. Sequences for these two ssDNA targets, respectively referred to as Ta_W (Target with Wild sequence) and Ta_M (Target with Mutated sequence) are: 5′CTCCCATAATCATCATTAGAAGTGA3′ and 5′CTCCCATAATCACCATTAGAAGTGA3′, where 5′ and 3′ indicate the two respective ends of the ssDNA sequence. A third ssDNA target, referred to as the calibration target, Ta_{Cal}, is used for the hereafter described calibration protocol. It is a 20 base ssDNA sequence: 5′CGCGCGCATACCGGTCAAAA3′. The novelty described in this paper lies in the use of multifunctionalized probes made of four, instead of the classical three, different parts (Fig. 2): (1) a sensing sequence, (2) an added calibration sequence and (3) a spacer sequence ending with a (4) biotin linker to ensure efficient binding to the streptavidin functionalized surface. These DNA sensing sequences are 15 bases long with the corresponding sequences perfectly complementary to the targets to analyse and centered on the point mutation to characterize:

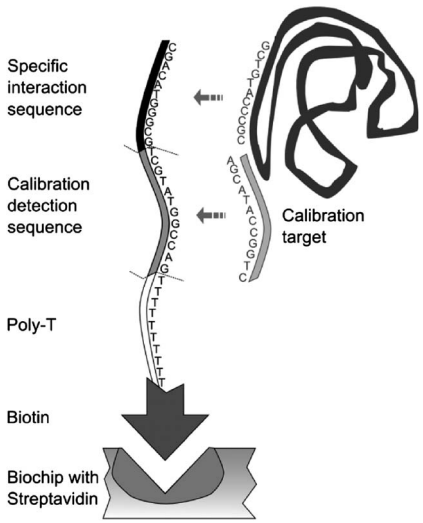

Fig. 2 Multifunctional probes are divided into four parts: first, the specific sequence devoted to the particular sample to characterize – second, a calibration dedicated sequence to allow for measuring of the local quantity of operational probes – third, the spacer sequence – and finally, fourth, a grafting group to link the probe to the surface, a biotin in the case presented.

5′TTCTAATGATGATTA3′ for the wild sequence and 5′TTCTAATGGTGATTA3′ for the mutated sequence. The spacer sequence is a 10 base poly-T sequence. The calibration sequence 5′GACCGGTATGCG3′ is common to all probes on the biochip and is perfectly complementary to the calibration target sequence. The choice of the calibration sequence, probe and target has to be made so that they will interact only between each other and not with any of the other possible probes/targets of interest, as quantified further. Finally, the multifunctionalized probes, referred herein respectively as Pr_W (Probe with Wild sequence) and Pr_M (Probe with Mutated sequence), have the following designs: B-5′TTTTTTTTTT*GACCGGTATGCG*TTCTAATGATGATT3′ and B-5′TTTTTTTTTT*GACCGGTATGCG*TTCTAATGGTG ATT3′, where B is the attached biotin and the calibration part is italicized.

In order to estimate the stability of the different possible duplexes, we evaluated their melting temperature, T_m. For this we used the next nearest neighbour method[29] [http://mfold.rna.albany.edu//?q=DINAMelt/Two-state-melting] with: $T = 25$ °C, $[Na^+] = 0.138$ M, $[Mg^{2+}] = 0$ M, [DNA] = 10 nM for Ta_W and Ta_M and [DNA] = 5 μM for Ta_{Cal}.

Here, we have used a probe/target double-stranded DNA (dsDNA) calibration duplex sequence of 12 bases, rich in GC, which is characterized by a melting temperature of 58.9 °C, indicating that such a calibration duplex will be very stable at our experimental temperature.

The other corresponding melting temperatures were calculated to be the following: around 30 °C for the perfect matches and about 10 °C less for the partial matches. Ta_W : Pr_W 30.3 °C, Ta_M : Pr_M 33.3 °C, Ta_W : Pr_M 21.8 °C and Ta_M : Pr_W 15.4 °C. As previously demonstrated, such temperatures are well suited for efficient signal discrimination[24] in our experimental operating conditions.

To simulate the three possible patient genotypes, different target solutions were used: one, containing only the Ta_W, for the WT, another containing only the Ta_M, for the MT, and a third containing a 1 : 1 mix of both Ta_W and Ta_M for the HT; the total concentration of ssDNA targets was kept constant at 10 nM. All DNA:DNA interactions studied were done sequentially in a phosphate buffer solution (PBS) 1× in concentration.

Normalization methodology

(a) Identification of dispersion causes

Biochips are relatively complex structures, made of stacks of different layers; in our case, chrome, gold, a self-assembled monolayer, the dextran layer and the final streptavidin active layer to which the probes are bound. It is therefore nearly impossible to ensure that all the associated physical parameters such as thickness and density will be strictly homogeneous over the whole sensing area of the biochip (about 1 cm^2). As the plasmonic coupling depends on these parameters, local reflectivity and sensitivity to a change in optical index will depend on the spatial position. Such spatial offset and inhomogeneity in the transducer sensitivity have to be removed in order to obtain consistent results. This "instrumental inhomogeneity correction" is the first discussed cause of data dispersion and is addressed in the next section b.

A second and generally more important source of biochip data dispersion comes from the fact that probes are deposited on the surface in variable quality and quantity. For a given spot, if a fraction of the probes are not functional or in a lower concentration, the response measured by the biochip system will be proportionally lower. This is especially critical in a genetic diagnosis experiment where it is necessary to compare the interaction of a given target with at least two different types of probes in different spots. Small differences in target/probe affinities can be completely masked by a difference in concentration and reactivity of probes in different spots. The smaller the difference one wishes to characterize, the more precisely, reliably and independently of spot quality must the measurements be performed. The spot's efficiency variation must therefore be quantified and corrected. This "biochip reactivity inhomogeneity correction" is the second discussed cause of data dispersion and is addressed in the subsequent section c.

(b) Instrumental inhomogeneity correction

To correct the spatial dispersion in reflectivity sensitivity, two solutions with different reflective optical indexes are injected sequentially in the flow cell. Specifically, we have used a regular PBS 1× buffer, followed by a PBS 1.25× solution; the corresponding optical index step, $n_{PBS\ 1.25\times} - n_{PBS\ 1\times}$ is about 5×10^{-4} R.I.U. (Refractive Index Unit) at our experimental wavelength, $\lambda = 660$ nm. Each solution is injected for 5 minutes during which the reflectivity is monitored and considered having reached homogeneous equilibrium both temporally and spatially. This cycle of injection is referred to as "SI" (Step of Index). The reflectivity variation between the PBS 1× and the PBS 1.25× circulation for the spot "i" is therefore denoted $\Delta R^i(SI)$ (Fig. 3). Theoretically with a perfect surface, every $\Delta R^i(SI)$ should be equal. The fact that different values of $\Delta R^i(SI)$ are measured is the direct consequence of the surface sensitivity

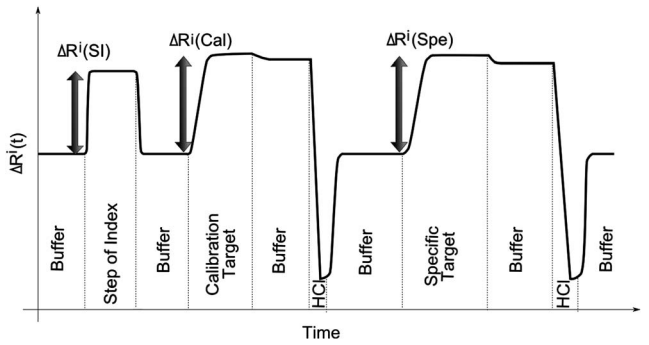

Fig. 3 Reflectivity signal *versus* time, $\Delta R^i(t)$. A complete characterization cycle corresponds to a series of product injections whose names are written above the time line. Specifically, the buffer is a PBS 1× solution; the solution used as the step index is a PBS 1.25× solution. The ΔR^i values correspond to the differences between the level corresponding to the buffer circulation, and respectively those of the step index solution (SI), the calibration target circulation (Cal) and the specific target circulation (Spe).

inhomogeneity. A correction factor C_{SI} is defined for every spot "i" and used to correct all following reflectivity variations:

$$C^i_{SI} = \frac{\langle \Delta R^i(SI) \rangle}{\Delta R^i(SI)}$$

$$\Delta R^i_{CorSI}(t) = \Delta R^i(t) \times C^i_{SI}$$

where $\langle \Delta R^i(SI) \rangle$ is the centre value of the Gaussian fit, or in the absence of enough significant data, the mean value, of all $\Delta R^i(SI)$ values.

(c) Biochip reactivity inhomogeneity correction

The second and major cause of data dispersion comes from the inhomogeneity in the reactive probe effective concentration. To quantify this data dispersion source independently from the instrumental one, one should use the previously SI corrected reflectivity variations. In order to quantify the level of active groups in each spot of the surface, a calibration target perfectly matched to a common calibration sequence present on all probes is used. Like in the previous correction protocol, different solutions are injected sequentially into the flow cell. First the PBS 1× buffer solution is injected during 5 minutes, setting the reference levels for all spots. Subsequently, a concentrated solution of the calibration target, in our case 5 μM, in PBS 1× circulates during 10 minutes, saturating the corresponding active probe sites. The buffer PBS 1× solution is again injected during 5 minutes and finally de-hybridization of the calibration target is achieved with the injection of HCl (11 mM) during 1 minute. Note that if a surface functionalization chemistry with no regenerative capability is used, this last injection can be avoided in the correction protocol as the hybridization of the calibration target does not hinder the sensing sequence above. The SI corrected reflectivity variation between the PBS 1× circulation and the calibration solution circulation for the spot "i" is noted $\Delta R^i_{CorSI}(Cal)$. Dispersion in the values of $\Delta R^i(Cal)$ is the direct consequence of the probe density and reactivity inhomogeneity. For every spot "i" a second correction factor C^i_{Cal} is defined:

$$C^i_{Cal} = \frac{\langle \Delta R^i_{CorSI}(Cal) \rangle}{\Delta R^i_{CorSI}(Cal)}$$

where $\langle \Delta R^i_{CorSI}(Cal) \rangle$ is the centre value of the Gaussian fit (or in the absence of enough data values, the mean value) of all $\Delta R^i_{CorSI}(Cal)$ values. Finally all subsequent measurements on the biochips for a given spot "i" will be corrected as follows:

$$\Delta R^i_{Cor}(t) = \Delta R^i_{CorSI}(t) \times C^i_{Cal}$$
$$\Delta R^i_{Cor}(t) = \left[\Delta R^i(t) \times C^i_{SI}\right] \times C^i_{Cal}$$

Experimental results

(a) Robustness of data normalization

To demonstrate the robustness of the bio-calibration protocol, 6 runs of calibration cycles were realized in a row. The aim is to demonstrate that the measurement of $\Delta R^i_{CorSI}(Cal)$ for a given spot "i" is reproducible and depends only on the intrinsic characteristics of the spot "i". It also shows that there is no need to run a calibration cycle before every measurement as a single calibration run at the beginning is enough, provided the chemical functionalization of the biochip is stable enough. Fig. 4 left illustrates, for a set of different spots, the five $\Delta R^i_{CorSI}(Cal)$ values obtained in runs number 1 to number 5 as a function of the initial $\Delta R^i_{CorSI}(Cal)$ measured at run number 0. The spread of data on the x axis is a visual representation of the data dispersion when measuring one given interaction using spots that are theoretically supposed to be identically functionalized, but which experimentally cannot be rigorously identical. Run 0 data have an average value of 1.08%, a median of 1.16% and a standard deviation of 0.39%. The following runs are characterized by similar values, as shown by the data dispersion on the y axis – or illustrated on Fig. 4 right. However, it can be seen that these data points for all six runs are correlated as the representation does not center within a circle, but are aligned along the diagonal. This can be quantified by the dispersion of $\Delta R^i_{CorSI}(Cal)$ values when looking at sequential target injection over the same spot for which the standard deviation is only 0.07%, which corresponds

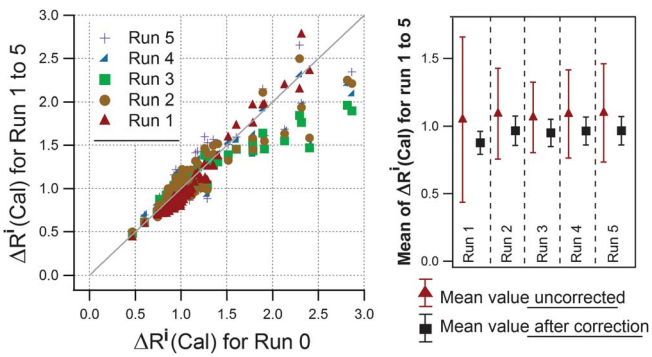

Fig. 4 (Left) Set of $\Delta R^i(Cal)$ values collected on all probe spots for five consecutive runs, plotted from 1 to 5, *versus* $\Delta R^i(Cal)$ values for a preliminary run 0. Data are clearly positioned along the diagonal, demonstrating that initial dispersion is reproducible from run to run and that each spot measurement series is in fact characterized by a much smaller dispersion – (right) comparison of raw and calibration corrected data quantified as mean values of sets of $\Delta R^i(Cal)$ for each run.

to the experimental noise of our system in our operating conditions. Dispersion is, in the typical case presented here, at least five times lower through reproducing the same probe:target interaction characterization using another target injection on the same probe spot (temporal reproducibility dispersion) than using another spot of the same probe with parallel target injection (spatial reproducibility dispersion).

Applying the C^i_{Cal} correcting factor obtained from run 0 for each of the following calibration injection impacts equivalently on the average signal and standard deviation values. Before correction, average signal values are: 1.04%, 1.09%, 1.06%, 1.08% and 1.09% and standard deviation values are: 0.61%, 0.33%, 0.26%, 0.32%, 0.36% and after correction, average signal values are: 0.87%, 0.96%, 0.95%, 0.96% and 0.96% and standard deviation values are: 0.08%, 0.10%, 0.10%, 0.10% and 0.10% respectively for the 1st, 2nd, 3rd, 4th and 5th ΔR^i(Spe) values of injections of runs 1 to 5 (Fig. 5 – left) – standard deviations of the corrected signals are similar to those mentioned above for a sequential target injection over the same spot.

(b) Application to a SNP diagnosis

To demonstrate the impact of the calibration protocol on the precision and reliability of results obtained in a typical diagnosis experiment, we have applied it to the model case of SNP diagnosis. The biochip was functionalized in a 24 × 24 spot matrix, made of "identical" sub-matrices of 4 spots with 2 spots of wild sequences, Pr_W, and 2 spots of mutated sequences, Pr_M. This provides a large number of replicates (144 sub-matrices) in order to extract statistical results from this experiment. Three successive full injection cycles were realized (Fig. 3) consisting of an instrumental correction cycle made of the step of index measurement as well as the biochip correction cycle using injection of the calibration target, followed by the circulation of the targets of interest and a final regeneration cycle (11 mM HCl – 1 minute). As reported in the Materials section, the first target solution of interest contained only the Ta_W sequences (WT genotype), the second one only the Ta_M target (MT genotype) and the last cycle, an equal mix of Ta_W and Ta_M (HT genotype). For these three cycles, the signal of a given sub-matrix is defined as the mean difference between the reflectivity variation $\Delta R^i(t)$ of the Pr_M probes and the $\Delta R^i(t)$ of the Pr_W probes, denoted as $\Delta R^i_{(W-M)}(t)$. As these targets are in very low concentration, like for a real patient product extract, the saturation regime is never reached. In fact $\Delta R^i_{(W-M)}(t)$ is almost linear as it only reports on the initial stage of the target:probe binding in terms of surface coverage. A positive slope for $\Delta R^i_{(W-M)}(t)$ means that the target has a higher affinity for the Pr_W probe than for the Pr_M probe which is the signature of a WT genotype. On the contrary, a negative slope will show a MT genotype. Finally, an intermediate regime, not necessarily a horizontal slope, would indicate that the target solution contains both Ta_M and Ta_W targets as is the case for the heterozygous genotype, HT case.

In Fig. 5, the three sets of data corresponding to the three different target type injections, $\Delta R^i_{(W-M)}(t)$, have been superimposed for all 144 sub-matrices – on the top graph, without the normalization and, on the bottom graph, using the calibration correction. The effectiveness of the methodology is evidenced by the discrimination into three groups. To quantify the reduction of data dispersion after normalization, standard deviations were calculated. On this typical dataset, without correction, discrimination can possibly reach, in the best of cases, 1σ (68% of results statistically in the correct patient's type). With the bio-calibration correction, discrimination increases from 1σ to 3σ (>99% of results statistically in the correct patient's type). Using data segmentation in Fig. 5, one can see that genotyping cannot be unambiguously achieved prior to data normalization (top) and that 100% correct genotyping is clearly achievable once performed (bottom figure).

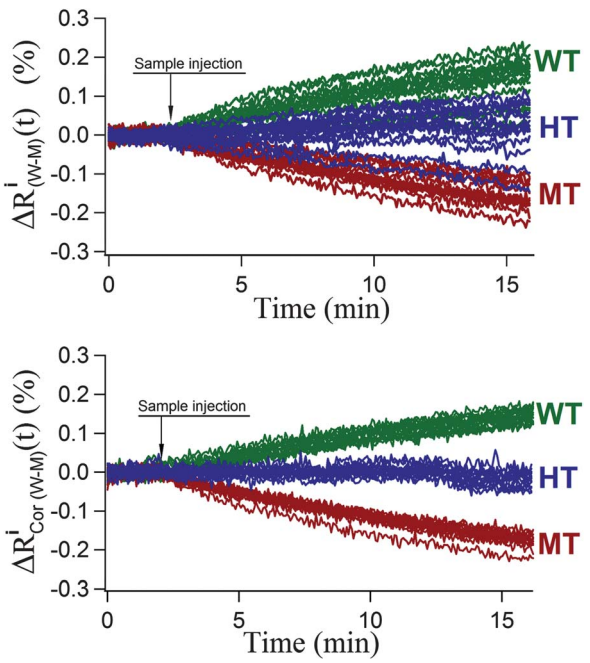

Fig. 5 Superimposition of $\Delta R^i_{(W-M)}(t)$ for 3 sample's injections onto the biochip: WT, MT and HT data. (Top) Data resulting from HT sample injection are, as is often the case, not separated from the two other sample's types. (Bottom) Same set of data using the calibration correction. The unambiguous genetic diagnostic, for the three possible genotypes, is clearly possible within minutes.

Discussion

Biochips are generally used as high or medium throughput systems in order to obtain a large quantity of bio-affinity data in parallel. For reliability constraint, one probe is not spotted only once but in duplicate, triplicate or even more spots, depending on the manufacturers/users. The trade-off is essentially between the reliability and information density and cost, taking into account the specific detection requirements to which the biochip is dedicated. In the example of the CF genetic diagnosis with our SPRI system, to be able to scan 65 mutation locations in the patient gene (allowing for ~95% of the ill Caucasian population, instead of the 85% achieved using only the 35 sequences classically used in commercial kits), probes can be replicated 6 times and thus distributed in 3 sub-matrices.[24] The data in Fig. 5 illustrate the difficulty to identify, without correction, the real type of the patient for a given SNP using only three $\Delta R^i_{(W-M)}(t)$ curves chosen at random in each dataset. Fig. 5 illustrates that, using the normalization, unambiguous genotype discriminations can be

achieved in a few minutes, clearly less than 10 minutes, without requiring extensive replication of spots.

One should also note that, to minimize the experimental time, only the "bio"-calibration step injection is required and not the optical step index one. Indeed, the instrumental pre-calibration using step index allows separation of the inhomogeneities due to the system sensitivity and those due to the biochip reactivity; if only the calibration is performed, its numerical corrections will take both sources of inhomogeneity into account and allow identical correction.

$$\Delta R^i_{\mathrm{Cor}}(t) = \Delta R^i(t) \times C^i_{\mathrm{Eff\ Calib}}$$

$$C^i_{\mathrm{Eff\ Calib}} = \frac{\langle \Delta R^i(\mathrm{Cal}) \rangle}{\Delta R^i(\mathrm{Cal})}$$

$$C^i_{\mathrm{Eff\ Calib}} = C^i_{\mathrm{Cal}} \times C^i_{\mathrm{SI}}$$

The effective calibration coefficient, $C^i_{\mathrm{Eff\ Calib}}$, can quite clearly be quantified only from the $\Delta R^i(\mathrm{Cal})$ data, and thus quantify the dispersion caused by both sources, instrumental and biochip reactivity.

Finally, one should note that this methodology not only corrects the inhomogeneity in spot functionalization but also quantifies and corrects inhomogeneity in probe concentrations. These two points clearly release the constraint and requirements in biochip realization therefore both allowing the reliable use of lab-made biochips and the lower cost of manufactured biochip production. The overall significant gain in routinely achievable precision opens the way to a wider use of such dynamical biochip systems. The extra-imposed requirement consists in choosing the proper calibration sequences that must not interact with other 6 sequences used in the biochip, including possible secondary structure issues.

Conclusions

In summary, a very efficient normalization method for biochip data gathering is demonstrated. It is based on the use of a common calibration target that can bind under identical conditions to all probes allowing the quantification of their relative effective spatial surface concentration. This allows, by a simple multiplication, to correct the effective probe concentration inhomogeneity, the usual major cause of data relative unreliability. Such a methodology significantly improves data reliability and precision, especially when comparing signals of the same nature within a given biochip. Furthermore, this technique diminishes the constraints on good-quality biochip production. To multifunctionalize the probes in the general field of biomolecular interactions, an ssDNA calibration sequence can be conveniently used whenever possible, for ssDNA type probes as illustrated in this communication, or possibly in a more general case for protein type probes using the so-called PDNAs.[30]

Such a patented methodology[31] could be generalized to different systems. As it is not specific to the reading system, it could be used identically in any dynamic reading system allowing sequential injections and readings of different target solutions. In the case of the end point chromophore label detection technology, one could advantageously spare a wavelength that would be attributed to a labelled calibration target in order to obtain such gain in precision.

Such methodology, and its precision gain applied to DNA:DNA interactions, could enable biomolecular biochip analysis to be effectively applied to routine analysis, especially in laboratories and hospitals.

Acknowledgements

This work was partially funded by the Ministère de la Recherche, CNRS, INSERM and the VLM association (Vaincre la Mucoviscidose/Defeat Cystic Fibrosis). J. Hottin was funded by GenOptics (www.genoptics-spr.com), acquired since then by Horiba Jobin-Yvon. The authors also thank Gisèle Roger from IOGS for valuable help in the experimental part as well as Michel Goossens *et al.* from Hospital Henri Mondor for strong collaboration in the cystic fibrosis medical application issue. CNRS, IOGS is a member of the European Network of Excellence, Photonic for Life (P4L).

References

1 J. Brookman, B. Nelson and R. Corn, Surface plasmon resonance imaging measurements of ultrathin organic films, *Annu. Rev. Phys. Chem.*, 2000, **51**, 41–63.
2 J. R. Epstein, I. Biran and D. R. Walt, Fluorescence-based nucleic acid detection and microarrays, *Anal. Chim. Acta*, 2002, **469**, 3–36.
3 M. Piliarik, L. Parova and J. Homola, High-throughput SPR sensor for food safety, *Biosens. Bioelectron.*, 2009, **24**, 1399–1404.
4 M. Campàs and I. Katakis, DNA biochip arraying, detection and amplification strategies, *Trends Anal. Chem.*, 2004, **23**, 46–62.
5 K. Okumura, M. Mihara and M. Yoshioka, *Biochip and biochip kit, and method of producing the same and method of using the same*, European Patent Application EP1619500, 2006.
6 T. Urisu and K. Suzui, *Biochip producing method, biochip, biochip analyzing device, biochip analyzing method*, European Patent Application EP1742054, 2007.
7 L. Grosjean, B. Cherif, E. Mercey, A. Roget, Y. Levy, P. N. Marche, M. B. Villiers and T. Livache, A polypyrrole protein microarray for antibody–antigen interaction studies using a label-free detection process, *Anal. Biochem.*, 2005, **347**, 193–200.
8 G. Krishnamoorthy, E. T. Carlen, J. G. Bomer, D. Wijnperlé, H. L. deBoer, A. van den Berg and R. B. M. Schasfoort, Electrokinetic label-free screening chip: a marriage of multiplexing and high throughput analysis using plasmon resonance imaging, *Lab-on-a-Chip*, 2010, **10**, 986–990.
9 E. Maillart, K. Brengel-Pesce, D. Capela, A. Roget, T. Livache, M. Canva, Y. Levy and T. Soussi, Versatile analysis of multiple macromolecular interactions by SPR imaging: application to p53 and DNA interaction, *Oncogene*, 2004, **23**(32), 5543–5550.
10 C. Ananthanawata, V. P. Hovenb, T. Vilaivanb and X. Suc, Surface plasmon resonance study of PNA interactions with double-stranded DNA, *Biosens. Bioelectron.*, 2011, **26**, 1918–1923.
11 M. G. Manera, J. Spadavecchia, A. Leone, F. Quaranta, R. Rella, D. Dell'atti, M. Minunni, M. Mascini and P. Siciliano, Surface plasmon resonance imaging technique for nucleic acid detection, *Sens. Actuators, B*, 2008, **130**, 82–87.
12 H. J. Lee, D. Nedelkov and R. M. Corn, Surface plasmon resonance imaging measurements of antibody arrays for the multiplexed detection of low molecular weight protein biomarkers, *Anal. Chem.*, 2006, **78**, 6504–6510.
13 G. Peterson, J. Bai and S. Narayanan, A co-printed oligomer to enhance reliability of spotted microarrays, *J. Microbiol. Methods*, 2009, **77**, 261–266.
14 B. C. Yin, H. Li and B. C. Ye, A dual-probe hybridization method for reducing variability in single nucleotide polymorphism analysis with oligonucleotide microarrays, *Anal. Biochem.*, 2008, **383**, 270–278.
15 J. Ritari, L. Paulin, J. Hultman and P. Auvinen, Application of hybridization control probe to increase accuracy on ligation

detection or minisequencing diagnostic microarrays, *BMC Res. Notes*, 2009, **2**, 249.
16 H. Xiong, D. Zhang, C. J. Martyniuk, V. L. Trudeau and X. Xia, Using generalized procrustes analysis (GPA) for normalization of cDNA microarray data, *BMC Bioinf.*, 2008, **9**, 25.
17 M. Hulsma, A. Mentink, E. P. van Someren, K. J. Dechering, J. de Boer and M. J. T. Reinders, Delineation of amplification, hybridization and location effects in microarray data yields better-quality normalization, *BMC Bioinf.*, 2010, **11**, 156.
18 A. W. Peterson, R. J. Heaton and R. M. Georgiadis, The effect of surface probe density on DNA hybridization, *Nucleic Acids Res.*, 2001, **29**(24), 5163–5168.
19 E. Huang, M. Satjapipat, S. Han and F. Zhou, Surface structure and coverage of an oligonucleotides probe tethered onto a gold substrate and its hybridization efficiency for polynucleotide target, *Langmuir*, 2001, **17**, 1215–1224.
20 O. G. Somova, O. V. Moiseeva and V. E. Barsky, *A method for immobilization of hydrogels over unmodified polymer materials, a biochip based on unmodified polymer materials, and a method for manufacturing thereof*, UK Patent Application GB2435473, 2007.
21 B. Rothenhausler and W. Knoll, Surface-plasmon microscopy, *Nature*, 1988, **332**, 615–617.
22 J. Homola, S. S. Yee and G. Gauglitz, Surface plasmon resonance sensors: review, *Sens. Actuators, B*, 1999, **54**(1), 3–15.
23 G. Krishnamoorthy, E. T. Carlen, A. van den Berg and R. B. M. Schasfoort, Surface plasmon resonance imaging based multiplex biosensor: integration of biomolecular screening, detection and kinetics estimation, *Sens. Actuators, B*, 2010, **148**, 511–521.
24 J. Hottin, J. Moreau, G. Roger, J. Spadavecchia, M. C. Millot, M. Goossens and M. Canva, Plasmonic DNA: towards genetic diagnosis chips, *Plasmonics*, 2007, **2**(4), 201–215.
25 E. Kretschmann and H. Raether, Radiative decay of non-radiative surface plasmons excited by light, *Naturforscher*, 1968, **23A**, 2135–2136.
26 P. Lecaruyer, M. Canva and J. Rolland, Metallic film optimization in a surface plasmon resonance biosensor by the extended Rouard method, *Appl. Opt.*, 2007, **46**, 2361–2369.
27 I. Mannelli, L. Lecerf, M. Guerrouache, M. Goossens, M. C. Millot and M. Canva, DNA immobilization procedures for surface plasmon resonance imaging (SPRI) based microarray systems, *Biosens. Bioelectron.*, 2007, **22**, 803–809.
28 I. Mannelli, V. Courtois, P. Lecaruyer, G. Roger, M. C. Millot, M. Goossens and M. Canva, Surface plasmon resonance imaging (SPRI) system and real-time monitoring of DNA biochip for human mutation diagnosis of DNA amplified samples, *Sens. Actuators, B*, 2006, **119**, 583–591.
29 J. SantaLucia, Jr, H. T. Allawi and P. A. Seneviratne, Improved nearest-neighbour parameters for predicting DNA duplex stability, *Biochemistry*, 1996, **35**, 3555–3562.
30 A. Laisne, M. Ewald, T. Ando, E. Lesniewska and D. Pompon, Self-assembly properties and dynamics of synthetic proteo-nucleic building blocks in solution and on surfaces, *Bioconjugate Chem.*, 2011, **22**(9), 1824–1834.
31 CNRS and M. Canva, *Biochip self-calibration process*, World patent application WO068725Brevet FR no. 05 12 594, December 13 2005, international extension PCT/EP2006/069683 December 13 2006, 2007.