Biosensors and Bioelectronics 26 (2010) 1692–1695

Contents lists available at ScienceDirect

Biosensors and Bioelectronics

journal homepage: www.elsevier.com/locate/bios

Short communication

Hybridisation thermodynamic parameters allow accurate detection of point mutations with DNA microarrays

Jef Hooyberghs [a,b,c,*], Enrico Carlon [c]

[a] Flemish Institute for Technological Research (VITO), Boeretang 200, B-2400 Mol, Belgium
[b] Hasselt University, Campus Diepenbeek, B-3590 Diepenbeek, Belgium
[c] Institute for Theoretical Physics, KULeuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 14 April 2010
Received in revised form 30 June 2010
Accepted 12 July 2010
Available online 16 August 2010

Keywords:
DNA microarray
Thermodynamics
Nearest neighbour model
Point mutations

ABSTRACT

We consider mixtures of two DNA sequences t and t' differing by a single nucleotide, which are analyzed by an Agilent custom DNA microarray. In particular we focus on the case in which t, the "wild type", is predominantly abundant and t' the "mutant" is at very low concentrations compared to t. We show that by using appropriately designed arrays it is possible to accurately quantify the presence of t' even at low relative concentrations (≈1%). The detection method is based on thermodynamic models of DNA hybridisation and on the analysis of a large number of hybridisation intensities from probes containing one or two mismatches with respect to t and t'.

1. Introduction

This paper introduces a method of detection, identification and quantification of a minority single nucleotide mutant sequence in a sample by means of custom DNA microarrays. Let us consider two DNA oligonucleotide sequences t and t' differing by a single nucleotide, referred to as the wild type and mutant, which are present in a sample at two different concentrations c and c'. Our main interest is in the case $c \gg c'$, i.e. when the wild type is predominantly present in the solution. This type of analysis can be of relevance for the early diagnostics of mutations in specific genomic regions, for instance in HIV (Halvas et al., 2006) or cancer (Prix et al., 2002). We show that using a large number of sequences interrogating both t and t' and inputs from hybridisation thermodynamics, one can detect the presence of a mutant when its abundance is of about 1% of the wild type.

2. Materials and methods

The method presented in this paper relies on previous work (Hooyberghs et al., 2009) aimed at determining hybridisation free energies from microarray data. As in Hooyberghs et al. (2009) we consider hybridisation from a pure target sequence in solution t (thus when only the wild type is present) to probes in the microarray perfectly matching and with one or two mismatches. In total about 10^3 probe sequences are used in the experimental setup, designed starting from the perfect complement of the wild type sequence t as follows (see also Table 1). In the sequence of length N we introduce all possible single mismatches excluding sites at 5 nucleotide distances from the end nucleotide of the sequence, to avoid terminal mismatches. This gives a total of $3(N-10)$ sequences, where the factor 3 counts the three possible mismatching nucleotides. Next we generate all double-mismatch sequences under the constraint that two mismatches cannot be closer than 5 nucleotides, so we obtain $9(N-16)(N-15)/2$ sequences. Since in the experiments $N=30$, one has in total 1006 sequences of which one is the perfect match (PM), 60 carry one mismatch and the remaining 945 contain two mismatches with respect to the wild type sequence. Part of these are listed in Table 1.

The probe sequences are spotted on Agilent custom 8×15 K arrays, with an additional 30-mer poly(A) spacer at the 3′-end to diminish steric hindrance effect from the microarray surface. Each spot is replicated 15 times in the array, for the intensity we take the median over the 15 measured values. The target sequences are synthetic oligonucleotide DNA (purchased from Eurogentec, Seraing, Belgium) terminated with a 20-mer poly(A) spacer at which a Cy3 label dye is attached at its 3′-end. Hybridisations are performed at 65 ° C for 17 h using standard Agilent buffers. The total concentra-

* Corresponding author at: Flemish Institute for Technological Research (VITO), Boeretang 200, B-2400 Mol, Belgium.
E-mail address: jef.hooyberghs@vito.be (J. Hooyberghs).

0956-5663/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.bios.2010.07.039

Table 1
List of target sequences in solution and of probe sequences spotted on a custom microarray as used in the experiments. The solution contains either a pure target t, or a mixture of t and t' at different relative concentrations. The probe sequences are generated starting from the perfect complement of t, labeled as probe 1, and by adding one and two mismatches as explained in the text. Target and probe sequences are given with 5′–3′ orientation.

Target sequences in solution	
t	CAGGGCCTCGTTATCAATGGAGTAGGTTTC
t'	CAGGGCCG CGTTATCAATGGAGTAGGTTTC
Probes at the microarray surface	
1	GAAACCTACTCCATTGATAACGAGGCCCTG
2	GAAACA TACTCCATTGATAACGAGGCCCTG
3	GAAACT TACTCCATTGATAACGAGGCCCTG
4	GAAACG TACTCCATTGATAACGAGGCCCTG
5	GAAACCA ACTCCATTGATAACGAGGCCCTG
6	GAAACCC ACTCCATTGATAACGAGGCCCTG
7	GAAACCG ACTCCATTGATAACGAGGCCCTG
	...
59	GAAACCTACTCCATTGATAACGAGA CCCTG
60	GAAACCTACTCCATTGATAACGAGT CCCTG
61	GAAACCTACTCCATTGATAACGAGC CCCTG
62	GAAACA TACTCA ATTGATAACGAGGCCCTG
63	GAAACA TACTCT ATTGATAACGAGGCCCTG
64	GAAACA TACTCG ATTGATAACGAGGCCCTG
	...
1005	GAAACCTACTCCATTGATC ACGAGT CCCTG
1006	GAAACCTACTCCATTGATC ACGAGC CCCTG

tion $c+c'$ is 50 pM (picomolar), while the ratio c'/c is varied over the experiments.

2.1. Hybridisation to wild type targets

Let us focus first on the case in which the mutant is absent ($c'=0$). The wild type sequence t hybridises with the highest affinity to the PM probe, which results in the highest fluorescence signal from a PM spot. In the presence of mismatches one observes a decrease of the fluorescence intensity with respect to that measured for a PM. From the decay of these intensities one can deduce the corresponding hybridisation free energies ΔG (Hooyberghs et al., 2009). In fact, basic thermodynamics predicts that at low target concentrations c, the measured fluorescent intensities should be distributed as

$$I = Ace^{\Delta G/RT} \tag{1}$$

where A is a scale factor, T the temperature and R the gas constant.

According to the nearest neighbor model (SantaLucia, 1998) the hybridisation free energy ΔG is expressed as a sum of dinucleotide terms. For instance, for a given sequence containing a mismatch, one has:

$$\Delta G\begin{pmatrix} AA\underline{T}GC \\ TT\underline{C}CG \end{pmatrix} = \Delta G\begin{pmatrix} AA \\ TT \end{pmatrix} + \Delta G\begin{pmatrix} A\underline{T} \\ T\underline{C} \end{pmatrix} \Delta G\begin{pmatrix} \underline{T}G \\ \underline{C}C \end{pmatrix} + \Delta G\begin{pmatrix} GC \\ CG \end{pmatrix}. \tag{2}$$

Taking into account all possible dinucleotide combinations and the strand symmetries one has 10 perfect match and 48 mismatch dinucleotide parameters. These were determined from the experimental data as explained in Hooyberghs et al. (2009).

Fig. 1(a) shows a plot of intensities (on a logarithmic scale) vs. free energies differences ($\Delta\Delta G = \Delta G - \Delta G_{PM}$) for which the perfect match free energy is taken as a reference. According to Eq. (1) a plot of the log I vs. ΔG (or equivalently vs. $\Delta\Delta G$) should appear as a straight line with slope $1/RT$. This line is drawn in Fig. 1 (a) as a reference. We note that the experimental data deviate from Eq. (1) in the high intensity scale. The origins of these deviations are non-equilibrium effects which have been discussed in Hooyberghs et al. (2010). They result in a non-linear relationship between intensities and exp $(\Delta G/RT)$:

$$I = Acf\left(e^{\Delta G/RT}\right). \tag{3}$$

The function $f(x)$ is linear ($f(x)\approx x$) at low binding strengths (low $\Delta\Delta G$). A non-equilibrium thermodynamic model describing I vs. ΔG has been presented in Hooyberghs et al. (2010). The function

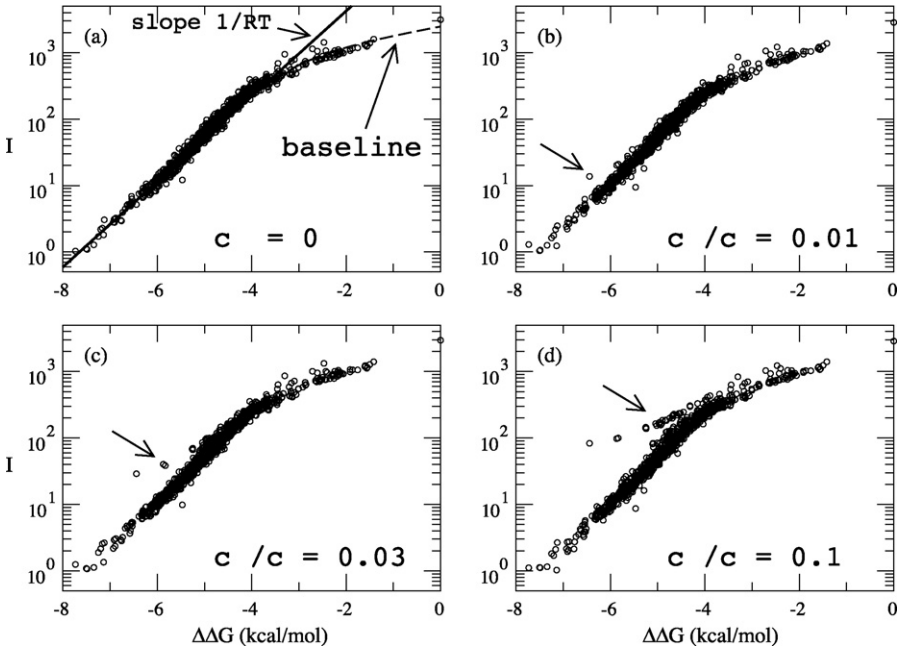

Fig. 1. (a) Plots of intensities vs. $\Delta\Delta G$ for the experiment with a pure wild type sequence ($c'=0$). There are 1006 different intensities in the plot, with up to two mismatches with respect to the target sequence t (see Table 1). The dashed line is the baseline, described by the non-equilibrium model of Hooyberghs et al. (2010). The solid line is a fit to Eq. (1). (b–d) Intensities vs. $\Delta\Delta G$ analogous of that of (a) for different mixtures of t and t' with concentrations c and c' respectively. The total concentration is fixed to $c+c'=50$ pM, while the relative concentration of the mutant with respect to the wild type varies from 0.1% to 10%. The arrows in (b)–(d) indicate the intensities departing from the baseline.