arXiv:q-bio.BM/0411014 v1 3 Nov 2004

Hybridization Isotherms of DNA Microarrays and the Quantification of Mutation Studies

Avraham Halperin,[1] Arnaud Buhot,[1] and Ekaterina B. Zhulina[2]

[1] *UMR 5819 (UJF, CNRS, CEA) DRFMC/SI3M, CEA Grenoble, 17 rue des Martyrs, 38054 Grenoble cedex 9, France*

[2] *Institute of Macromolecular Compounds of the Russian Academy of Sciences, 199004 St Petersburg, Bolshoy prospect 31, Russia*

Background: Diagnostic DNA arrays for detection of point mutations as markers for cancer usually function in the presence of a large excess of wild type DNA. This excess can give rise to false positives due to competitive hybridization of the wild type target at the mutation spot. The analysis of the DNA array data is typically qualitative aiming to establish the presence or absence of a particular point mutation. Our theoretical approach yields methods for quantifying the analysis so as to obtain the ratio of concentrations of mutated and wild type DNA.

Method: The theory is formulated in terms of the hybridization isotherms relating the hybridization fraction at the spot to the composition of the sample solutions at thermodynamic equilibrium. It focuses on samples containing an excess of single stranded DNA and on DNA arrays with low surface density of probes. The hybridization equilibrium constants can be obtained by the nearest neighbor method.

Results: Two approaches allow us to obtain quantitative results from the DNA array data. In one the signal of the mutation spot is compared with that of the wild type spot. The implementation requires knowledge of the saturation intensity of the two spots. The second approach requires comparison of the intensity of the mutation spot at two different temperatures. In this case knowledge of the saturation signal is not always necessary.

Conclusions: DNA arrays can be used to obtain quantitative results on the concentration ratio of mutated DNA to wild type DNA in studies of somatic point mutations.

I. INTRODUCTION

Cancers are attributed to accumulation of somatic mutations [1]. In turn, the mutated DNA can provide molecular markers of diagnostic utility [2]. Among them are point mutations, involving a change in a single base pair in the DNA, such as those occurring in the p53 and K-ras genes [3, 4, 5]. The detection of such point mutations is useful in screening for cancers [6] as well as typing the cancer in order to optimize the treatment protocol [7]. DNA microarrays [8, 9], "DNA chips", are among the analytical techniques of proven potential to this end [6, 10]. The detection of somatic point mutations is hampered by the presence of an excess of wild type DNA. This favors hybridization of the wild type single stranded DNA (ssDNA) with mismatched sequences resulting in false positives. A similar problem occurs in analyzing single nucleotide polymorphism of pooled samples [2]. In the following we present a theoretical analysis of the errors introduced by such mishybridization utilizing equilibrium thermodynamics. It suggests methods of quantifying the detection of point mutations by DNA chips. In particular, these methods allow to obtain the ratio of concentrations of mutated and wild type DNA in the sample. This ratio is of diagnostic interest and it provides a systematic method for the minimization of false positives. The numerical calculations illustrating this approach are based on recent DNA chip studies of point mutation in the K-ras gene [6, 10].

DNA chips consist of a support surface carrying "spots" [8, 9]. Each spot comprises numerous oligonucleotides of identical and known sequence, "probes", that are terminally anchored to the surface. The spots are placed in a checkered pattern such that each sequence is allocated a unique site. Each probe hybridizes preferentially with a ssDNA containing a complementary sequence

referred to as "target". In a typical experiment the DNA microarray is immersed in a solution containing a mixture of labelled ssDNA fragments of unknown sequence. The presence of a given sequence is signalled by the hybridization on the corresponding spot as monitored by correlating the strength of the label signal with the position of the spot. False positives can occur because each probe can also hybridize with a mismatched sequence. When the DNA microarray is designed to investigate gene expression it is possible to optimize the probe design in order to minimize this effect [11, 12]. However, the implementation of this strategy is more difficult for studies of point mutations. When studying somatic point mutations the situation is further complicated by an excess of wild type, non-mutated, DNA. Two observations substantiate this point. First, solid tumors are heterogenous, containing a mixture both cancerous and normal, stromal, cells. The cancer cells are a minority and the fraction of mutated DNA obtained from homogenized tumor biopsy can be as small as 15% [2]. The fraction is much smaller for noninvasive testing for early stage cancers using body fluids such as urine [5], serum [4] or stool [3, 6]. Since the hybridization is controlled by the mass action law, the excess of wild type sequence will typically contribute to the hybridization on the spots allocated to the point mutations. Accordingly, the ratio of intensities of different spots may not reflect the ratio of concentrations of DNA species in the sample. This mishybridization contribution will increase with the ratio of wild type DNA to mutated DNA thus diminishing the efficacy of the early stage screening. A similar situation is encountered in the analysis of pooled single nucleotide polymorphism samples. This problem can be resolved by determining the contribution of the wild type DNA to the signal of the mutation spots. As we shall discuss, this is possible when three conditions are fulfilled: (i) the hybridization is allowed to reach equilibrium. (ii) The equilibrium hybridization isotherm, relating the hybridization fraction on the spot to the sample composition, is of the Langmuir form i.e., $x/(1-x) = Kc$ where x is the fraction of surface sites that bind a reactant of concentration c and K is the equilibrium constant of the reaction [13]. This regime is expected when the surface density of probes is sufficiently low [14]. (iii) The sample contains an excess of ssDNA as is the case when using asymmetric PCR amplification [15, 16] or following digestion by lambda exonuclease [17]. Under these conditions the fraction of correctly hybridized probes on a spot is obtainable yielding also the concentrations of mutated and wild type DNA in the sample. Two approaches are possible: (a) comparing the signals of two spots with probes that match, respectively, the wild type and mutated sequences. (b) Comparing the signals of the spot corresponding to the mutation of interest at two different temperatures, T_1 and T_2. The first approach requires measurement of the saturation signal of the two spots while for the second this step may be eliminated. Importantly, the experimental set up reported by Fotin et al [18] satisfies the three conditions listed above and allows to implement the two temperature approach.

Our analysis is based on the hybridization isotherms of DNA chips allowing for the role of two types of competitive hybridization [14]. Competitive surface hybridization occurs when two different targets can hybridize with the same probe. Competitive bulk hybridization takes place when the target can hybridize with a complementary sequence in the bulk. The second process is dominant when the samples are produced by symmetric PCR amplification. When asymmetric PCR [15, 16] is used the sample contains an excess of ssDNA which does not experience the effect of competitive bulk hybridization. This situation can also be obtained by digesting the product of symmetric PCR with lambda exonuclease [17]. Under these conditions the excess of ssDNA dominates the hybridization with the probes and competitive surface hybridization becomes the major source of error. In the general case the hybridization isotherm reflects the electrostatic penalty incurred because each hybridization event increases the charge of the probe layer [14, 19]. We will focus on systems where these interactions are screened and the hybridization isotherm assumes the Langmuir form [13] which facilitates quantitative analysis of the data. The use of PCR amplification renders

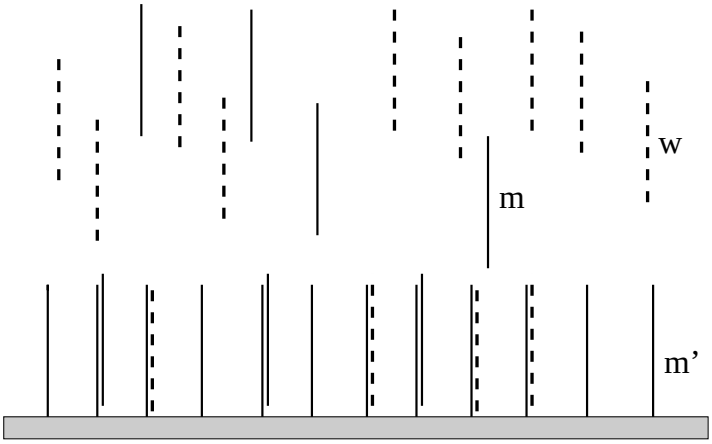

FIG. 1: **Competitive Surface Hybridization.** The m' probes at the mutation spot can hybridize with both the complementary m targets and the mismatched wild type w targets. The fraction of mishybridized, $m'w$, probes is high when the w targets are present in excess. Double stranded targets are not shown.

the absolute concentrations of DNA species meaningless. Our analysis is thus concerned with the ratio c_m/c_w of the excess concentrations of mutated (c_m) and wild form (c_w) of ssDNA.

Our discussion focuses on the situation following amplification by asymmetric PCR (Figure 1). It concerns two spots: a mutation spot carrying m' probes and a wild type spot carrying w' probes. The sample solution comprises of a wild type ssDNA fragment w and a mutated fragment m as well as their complementary fragments, w' and m'. The number of w and m fragments is however larger. We assume that all free w' and m' fragments hybridize with the free complementary w and m ssDNA fragments. Our analysis focuses on the excess of unhybridized m and w fragments whose concentrations are denoted respectively by c_m and c_w. We consider the small spot limit, when the hybridization at the spot has a negligible effect on the bulk concentrations. In this limit the double stranded DNA ww' and mm' pairs in the bulk do not affect the hybridization equilibrium at the spot. In other words, the hybridization isotherm relates c_m and c_w to the measured hybridization signal of a spot. The targets are actually much larger than the probes. In the K-ras studies we cite the target is comprised of 117 [10] to 157 [6] monomers (nucleotides) while the probes comprise of 13-14 monomers. As we shall discuss, this "size asymmetry" mostly affects the boundary of the Langmuir regime. Its effect on the hybridization isotherms of spots of high probe density is beyond the scope of this article. In section II we present the necessary background on the hybridization isotherm when competitive surface hybridization is dominant. The following two section discuss respectively the two-spot and two-temperature approaches for the determination of c_m/c_w. Experimental considerations, the relation to existing experimental studies as well as open questions are presented in the Discussion.

II. THE ADSORPTION ISOTHERM

The hybridization isotherm relates the equilibrium fraction of hybridized probes on the spot to the composition of the sample [14]. A simple derivation of this isotherm is possible upon equating the rates of hybridization and denaturation at the surface of the spot [13]. In the following we consider the isotherm obtained when competitive surface hybridization is important. In particular, both the wild type sequence, w, and the mutated one, m, can hybridize at the mutated m' spot. Consider a spot carrying a total of N_T m'-probes in a sample containing unhybridized m and w targets of concentrations c_m and c_w moles per liter respectively. The total hybridization reflects both hybridization of the perfectly matched m targets and of the mismatched w ones. The fractions of $m'm$ and $m'w$ hybridized probes are denoted respectively by x and y. The rate of denaturation of mismatched $m'w$ probes is $k_{dm}yN_T$ while the rate of hybridization of m' probes with w targets is $k_{hm}c_w(1-x-y)N_T$ where k_{dm} and k_{hm} are the corresponding rate constants. When electrostatic interactions within the layer are strongly screened, the rate constants do not depend on the probe density and the fraction of hybridized probes. At equilibrium the two rates are equal, $k_{dm}yN_T = k_{hm}c_w(1-x-y)N_T$ leading to

$$\frac{y}{(1-x-y)} = c_w\frac{k_{hm}}{k_{dm}} = c_wK_{m'w} = c_w\exp(-\Delta G^0_{m'w}/RT). \tag{1}$$

where $k_{hm}/k_{dm} = K_{m'w}$ is the equilibrium constant [20]. Similarly, the two rates for the perfectly matched $m'm$ probes are $k_{dp}xN_T$ and $k_{hp}c_m(1-x-y)N_T$. Their equality at equilibrium, $k_{dp}xN_T = k_{hp}c_m(1-x-y)N_T$ leads to

$$\frac{x}{(1-x-y)} = c_m\frac{k_{hp}}{k_{dp}} = c_mK_{m'm} = c_m\exp(-\Delta G^0_{m'm}/RT). \tag{2}$$

Here $K_{m'w}$ and $K_{m'm}$ are respectively the equilibrium constants of hybridization between m' probes and w and m targets, $\Delta G^0_{m'w}$ and $\Delta G^0_{m'm}$ are the corresponding standard Gibbs free energies, T is the absolute temperature and R is the gas constant. This kinetic derivation recovers the results of a rigorous thermodynamic analysis [14] with one caveat: The molar concentrations should be replaced by dimensionless activities $a_i = \gamma c_i$ where γ is the activity coefficient [20]. Since $\gamma \to 1$ as $c_i \to 0$ we will retain expressions used above noting that the c_i are dimensionless having the numerical value of the molar concentrations of the ith species. The two isotherms, (1) and (2), are not helpful because PCR amplification of a given sample does not allow for different labels of the m and w targets. Since both are labeled identically, the measurement yields the total hybridization fraction $\theta = x + y$ rather than x and y separately. Accordingly, the observed isotherm for the m' spot is

$$\Omega_{m'} = \frac{\theta_{m'}}{1-\theta_{m'}} = K_{m'm}c_m + K_{m'w}c_w \tag{3}$$

as obtained by summing (1) and (2). The fraction of mishybridized probes on the m' spot is

$$P_{m'} = \frac{y_{m'}}{\theta_{m'}} = \frac{K_{m'w}c_w}{K_{m'm}c_m + K_{m'w}c_w}. \tag{4}$$

Here, and in the following, the subscript $i = m', w'$ of Ω_i, P_i etc. identifies the spot. $P_{m'} = 1/2$ when $c_w = c_mK_{m'm}/K_{m'w}$ and $P_{m'} > 1/2$ for $c_w > c_mK_{m'm}/K_{m'w}$. In the case of interest, when

$i'j$	$\Delta H^0_{i'j} kcal/mole$	$\Delta S^0_{i'j} cal/mole \cdot \deg$	$\Delta G^0_{i'j}(37^\circ C) kcal/mole$	$K_{i'j(37^\circ C)}$
$m'm$	-99.40	-264.29	-17.43	2.11×10^{12}
$m'w$	-78.20	-214.12	-11.79	2.19×10^{8}
$w'w$	-100.80	-270.55	-16.89	8.75×10^{11}
$w'm$	-93.50	-253.59	-14.85	3.17×10^{10}

TABLE I: The thermodynamic parameters utilized in the numerical calculations correspond to the Alanine 12 and wild type probes utilized by Prix et al [6]: $m' = AGCTG\overline{C}TGGCGTA$, $m = TCGACGACCGCAT$, $w' = CTGGTGGCGTAGG$, $w = GACCACCGCATCC$ as calculated from the nearest neighbor model [27] for $1MNaCl$. Since the targets are longer than the probes two dangling ends are invoked.

$c_w \gg c_m$, $P_{m'}$ of the m' spot is large while $P_{w'}$ of the w' spot

$$P_{w'} = \frac{y_{w'}}{\theta_{w'}} = \frac{K_{w'm} c_m}{K_{w'w} c_w + K_{w'm} c_m} \tag{5}$$

is always small. The values of $P_{m'}$ (Figure 2) for a typical situation (Table 1) confirm that competitive surface hybridization is important only when the competing species is present in large excess. As noted earlier, this is the case in studies of somatic point mutations when the wild type ssDNA is a majority component. The wild type ssDNA then contributes to the hybridization on all the mutation spots. In contrast, the concentrations of the different mutated ssDNA are much smaller. As a result their contribution to the hybridization on the m' spot, and on other mutation spots, is negligible. This observation justifies limiting the analysis to the competition between two species, m and w.

When competitive hybridization is negligible, equation (3) reduces to $\Omega_{m'} = K_{m'm} c_m$ and the corresponding isotherm for the wild type is $\Omega_{w'} = K_{w'w} c_w$. In such cases c_m/c_w can be determined from

$$\frac{\Omega_{m'}}{\Omega_{w'}} = \frac{K_{m'm}}{K_{w'w}} \frac{c_m}{c_w} \tag{6}$$

This situation can be realized in gene expression studies with proper probe design and in the absence of overwhelming excess of one target species. Note that θ of the mutation spot can be tuned to be low, $\theta \ll 1$ by adjusting the hybridization temperature. In such regimes one may invoke a "weak spot" approximation

$$\Omega_i \simeq \theta_i \tag{7}$$

As we shall discuss, when applicable this simplifies the analysis of the two temperature experiments. Finally, in most cases the signal of the spot i at equilibrium, I_i is proportional to $\theta_i N_T$ and

$$\theta_i = \frac{I_i}{I_i^{\max}} \tag{8}$$

where $I_i^{\max}$ is the saturated signal of the i spot as obtained upon equilibration with a concentrated solution of the perfectly matched target. It is important to note that $I_i^{\max}$ can depend on temperature.

The melting temperatures of double stranded DNA are used as design criterion for probes [6, 12]. It is thus useful to note that competitive surface hybridization affects the effective melting temperature, T_M. T_M is defined by the condition $\theta_i = 1/2$ or $\Omega_i = 1$. In the general case T_M

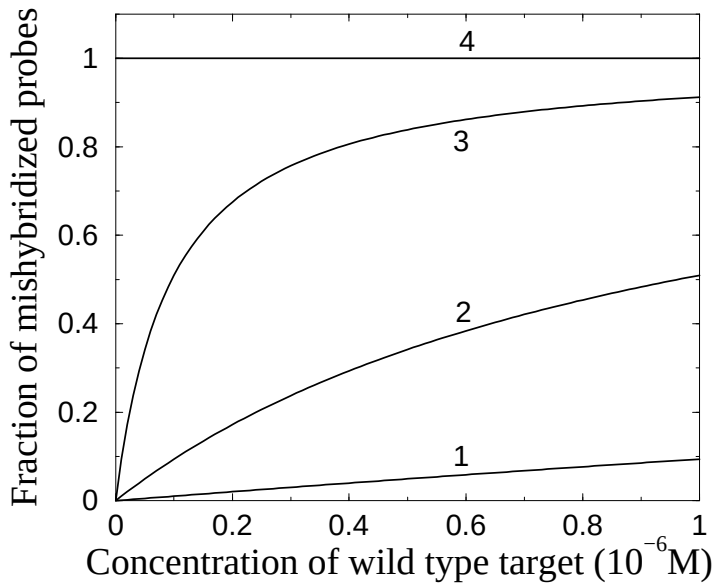

FIG. 2: **The fraction of mishybridized probes on the mutation spot as function of the concentration of the wild type target.** $P_{m'}$ vs. c_w curves calculated at $T = 47^\circ C$ for (1) $c_m = 10^{-9}M$, (2) $c_m = 10^{-10}M$, (3) $c_m = 10^{-11}M$ and (4) $c_m = 0$.

depends on both c_m and c_w i.e., $T_M = T_M(c_m, c_w)$. Note that there are two targets that can hybridize with the same probe. This situation, involving three ssDNA species and two different dsDNA, is different from the one invoked in the definition of the melting temperature of a dsDNA. In this last case two complementary ssDNA hybridize to form one type of dsDNA and the total number of species is three rather than five. Furthermore, we are considering the small spot limit where the hybridization with the probes has negligible effect on the bulk composition. In the absence of competitive surface hybridization T_M reduces to its familiar forms: for $c_w = 0$, this leads $T_{m'm}(c_m) = \Delta H^0_{m'm}/(\Delta S^0_{m'm} + R \ln c_m)$ while for $c_m = 0$ it leads to $T_{m'w}(c_w) = \Delta H^0_{m'w}/(\Delta S^0_{m'w} + R \ln c_w)$. It is not possible to obtain an explicit analytical expression for $T_M(c_m, c_w)$. However, in the regime of interest, $c_m/c_w \ll 1$ the melting temperature is well approximated by the first two terms of the Taylor expansion in c_m around $c_m = 0$ i.e., $T_M(c_m, c_w) \simeq T_{m'w} + c_m \frac{\partial T_M}{\partial c_m}|_{c_m=0}$ thus specifying the melting temperature at the mutation spot

$$T_M(c_m, c_w) = T_{m'w}(c_w) + c_m \frac{K_{m'm}(T_{m'w})}{|\Delta H^0_{m'w}|} R T^2_{m'w} \tag{9}$$

Importantly, in this regime T_M increases steeply with c_m (Figure 3) and its initial value, for $c_m = 0$ is $T_{m'w}$ rather than $T_{m'm}$. In marked distinction, for $c_m/c_w \gtrsim 1$ the effect of the competitive surface hybridization is negligible and $T_M \approx T_{m'm}(c_m)$ with a weak logarithmic dependence on c_m.

When the probe density is relatively high and the salt concentration is sufficiently low, it is necessary to allow for the electrostatic penalty incurred upon hybridization because of the extra charge deposited at the probe layer. In this regime the hybridization isotherms assume a different

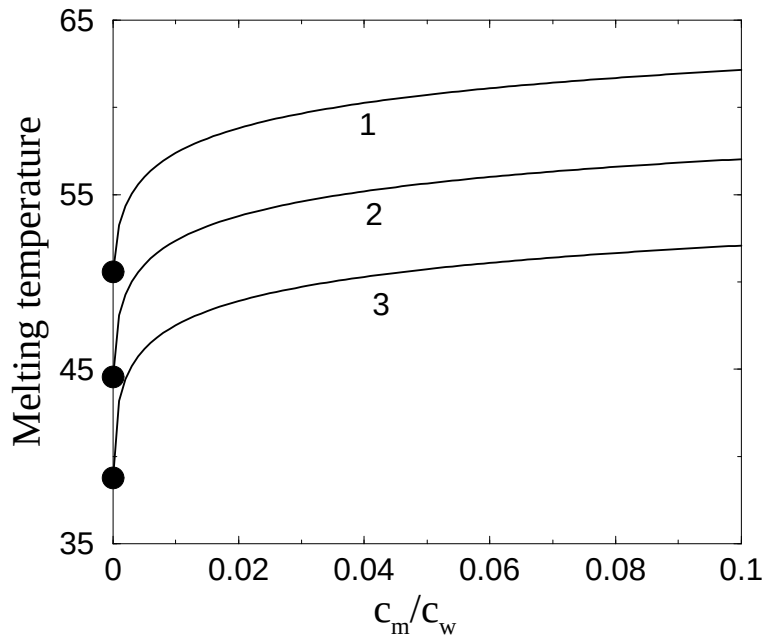

FIG. 3: **The dependence of the melting temperature of the mutation spot on the concentration of the mutation target.** $T_M(c_w, c_m)$ in $^\circ C$ is plotted vs. c_m/c_w for various $c_m + c_w = c$: (1) $c = 10^{-6}M$ (2) $c = 10^{-7}M$ and (3) $c = 10^{-8}M$. Dots correspond to the melting temperatures $T_{m'w}(c_w)$.

form exemplified by

$$\Omega_{m'} = \frac{\theta_{m'}}{1-\theta_{m'}} = (K_{m'm}c_m + K_{m'w}c_w)\exp\left[-\Gamma_{m'}(1+\theta_{m'})\right]. \tag{10}$$

For the case of probes and targets of equal and short length $\Gamma_{m'} = constant \times \sigma_{m'}$ where $\sigma_{m'}$ is the charge density of the unhybridized m' spot and the constant is set by the ionic strength and the length of the probe [14]. In the following we consider the opposite limit, when the electrostatics interactions are screened and the chains do not interact. The hybridization is then described by a simple Langmuir isotherm that is, the case of $\Gamma_i \simeq 0$. Importantly, in studies of point mutations the targets are typically longer than the probes. Usually, the hybridization site is situated roughly in the middle of the target. Accordingly, a hybridized probe incorporates two unhybridized target sections of similar length. The span of such segment is $R_0 \approx (nal)^{1/2}$ where n is the number of bases in the section, a is the base size and l is the persistence length. Typical values are $a \approx 0.6nm$ and $l \approx 0.75nm$ [21]. The Langmuir regime occurs when the area per probe, Σ, exceeds R_0^2 thus ensuring that the hybridized probes do not interact. For the target-probe pairs considered this condition is satisfied when $\Sigma \gtrsim 65nm^2$ or less than 1.5×10^4 grafted probes per μm^2. Operation in this range allows us to benefit from the absence of the non-linear behavior introduced by the $\exp\left[-\Gamma_{m'}(1+\theta_{m'})\right]$ term.

The hybridization isotherm is only applicable at equilibrium. In turn, this implies two conditions: stationary signal and path independence i.e., independence of the preparation method and sample history. The reported times required to attain stationary signal vary between minutes [18] to 14

hours [22, 23]. Importantly, the equilibration time may depend on the probe density, hybridization conditions, the equilibrium θ_i etc. Path independence requires that the measured signal at equilibrium will not change with the thermal history (heating and cooling cycles). As stated earlier, these conditions are satisfied by the system of Fotin et al [18]. Under conditions of thermodynamic equilibrium the quantitative methods of analysis we describe below are applicable. However, these methods require knowledge of θ_i and Ω_i. In turn, to obtain θ_i and Ω_i it is necessary to ascertain the saturation value of the signal of the i spot, $I_i^{\max}$ or, equivalently, the N_T of the i spot.

Numerical implementation of the analysis we describe requires knowledge of the equilibrium constants $K_{m'm}$ and $K_{m'w}$ as well as $K_{w'm}$ and $K_{w'w}$. These can be best obtained experimentally from the hybridization isotherm of a spot in contact with single component samples. However, this approach is time consuming when a number of equilibrium constants is required. Fortunately, in the Langmuir regime it is reasonable to approximate the equilibrium constants by their bulk values. The nearest neighbor model, with the unified set of parameters compiled by Santa Lucia and collaborators [24, 25, 26], allows to calculate ΔH^0 and ΔS^0 for the hybridization of perfectly matched oligonucleotide pairs as well as for pairs containing a single mismatch. In the numerical calculations we use probe target pairs incorporating the 12th and 13th K-ras codons. The ΔH^0 and ΔS^0 values are obtained by use of HYTHERTM software implementing the nearest neighbor approach [27]. The identity of the probes and targets considered in the numerical calculations as well as the corresponding standard enthalpies, entropies are listed in Table 1. The table also lists the free energies of formation and the equilibrium constants of hybridization at 37^0C. The hybridization temperatures considered are chosen in view of the conditions in the cited experiments. Fotin et al [18] studied the hybridization at the range of $T = -20^0C$ to $T = 60^0C$. In the K-ras studies the hybridization temperatures were 37^0C [6] and 50^0C [10]. The c_m values in the numerical examples vary around $3 \cdot 10^{-10}M$, while the concentration of the control target c_w is 10^2 to 10^3 larger [6].

III. THE TWO SPOT APPROACH

When two spots carrying respectively wild type, w', and mutation, m', probes are placed in contact with a solution of w and m targets, the corresponding isotherms are

$$\Omega_{w'} = c_w K_{w'w} + c_m K_{w'm} \tag{11}$$

$$\Omega_{m'} = c_w K_{m'w} + c_m K_{m'm} \tag{12}$$

where $\Omega_i = \theta_i/(1-\theta_i)$ of spot $i = w', m'$ and θ_i is the corresponding fraction of hybridized probes, irrespective of their identity (mismatched or perfectly matched). These two equations immediately determine c_w and c_m

$$c_w = \frac{\Omega_{w'} K_{m'm} - \Omega_{m'} K_{w'm}}{K_{w'w} K_{m'm} - K_{w'm} K_{m'w}} \tag{13}$$

$$c_m = \frac{\Omega_{m'} K_{w'w} - \Omega_{w'} K_{m'w}}{K_{w'w} K_{m'm} - K_{w'm} K_{m'w}}. \tag{14}$$

Accordingly

$$\frac{c_m}{c_w} = \frac{\Omega_{m'} K_{w'w} - \Omega_{w'} K_{m'w}}{\Omega_{w'} K_{m'm} - \Omega_{m'} K_{w'm}} = \frac{K_{w'w} - \alpha K_{m'w}}{\alpha K_{m'm} - K_{w'm}} \tag{15}$$

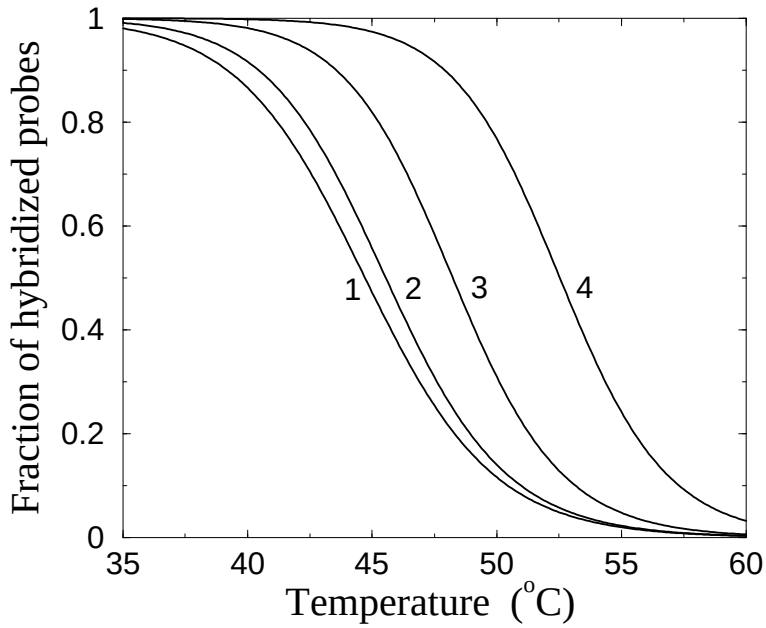

FIG. 4: **Melting curves of the mutation spot.** $\theta_{m'}$ vs. temperature curves are plotted for $c_m + c_w = c = 10^{-7}M$ and (1) $c_m/c_w = 0$, (2) $c_m/c_w = 10^{-4}$, (3) $c_m/c_w = 10^{-3}$ and (4) $c_m/c_w = 10^{-2}$. For a given $c_w \simeq c$ the melting temperature and the saturation temperature increase with c_m/c_w.

where, in the regime considered, $\alpha = \Omega_{w'}/\Omega_{m'} \gg 1$. The range of α values varies between $\alpha_{\max} = K_{w'w}/K_{m'w} \gg 1$, corresponding to $c_m = 0$ and $\alpha_{\min} = K_{w'm}/K_{m'm} \ll 1$ when $c_w = 0$. In the realistic limit of $\alpha K_{m'w} \ll K_{w'w}$ this expression reduces to $c_m/c_w \approx K_{w'w}/\alpha K_{m'm}$ that is

$$\frac{c_m}{c_w} = \frac{K_{w'w}}{K_{m'm}} \frac{I_{m'}}{I_{w'}} \frac{I_{w'}^{\max} - I_{w'}}{I_{m'}^{\max} - I_{m'}} \tag{16}$$

where $I_{m'}$ and $I_{w'}$ are the measured intensities of the m' and w' spots while $I_{m'}^{\max}$ and $I_{w'}^{\max}$ are the corresponding saturation values. Thus, the implementation of this approach requires knowledge of N_T, or equivalently $I_{\max}$, for the two spots. In the general case it is necessary to know four equilibrium constants while in the simplest case, of $\alpha K_{m'w} \ll K_{w'w}$, knowledge of $K_{m'm}$ and $K_{w'w}$ is sufficient.

The two spots approach is feasible when the values of $\theta_{m'}$ and $\theta_{w'}$ for different sets of c_m and c_w are distinguishable. In practical terms this imposes two requirements. First, it is necessary to avoid the saturation regimes of the melting curve (Figure 4) and the hybridization isotherm (Figure 5). Second, the $\theta_{i'}$ values corresponding to the sample composition must be large enough in comparison with the experimental errors. One may optimize the performance by tuning the hybridization temperature T: Increasing T lowers the hybridization degree thus preventing saturation at the price of weaker $\theta_{i'}$ and a higher noise to signal ratio.

The operational conditions are thus determined by two inputs: The typical c_w of the amplicon and the desired range of c_m/c_w. Once these two parameters are specified it is possible to calculate the relevant melting curves and hybridization isotherms in order to choose the hybridization

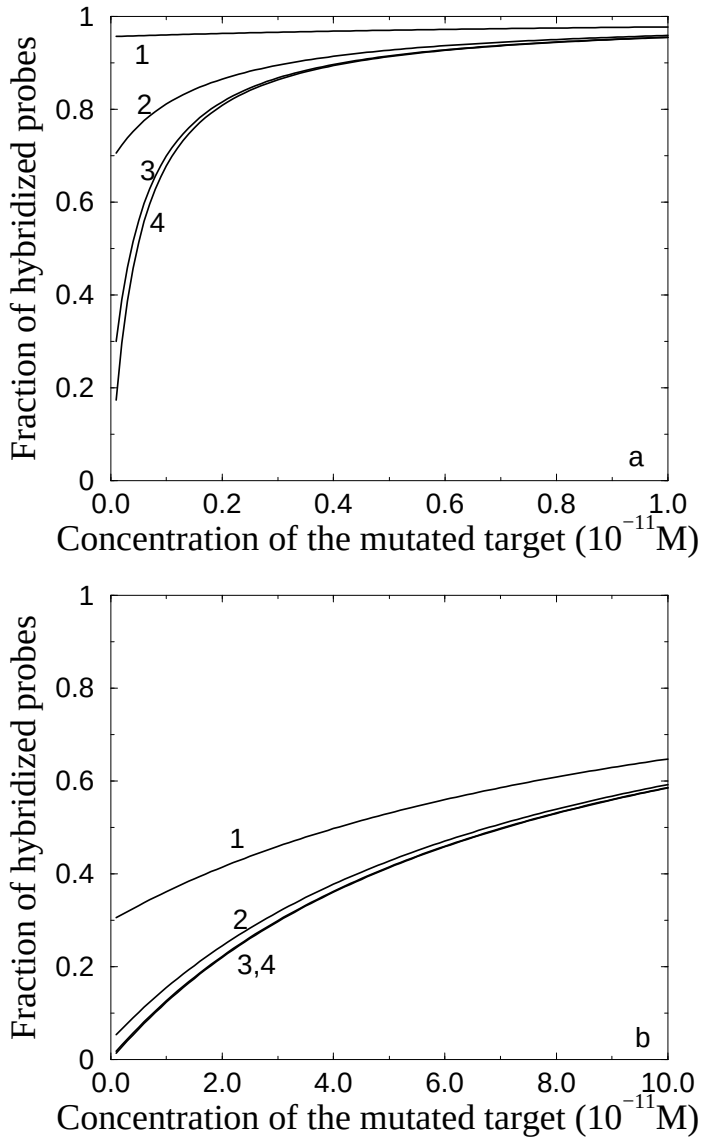

FIG. 5: **The hybridization isotherm of the mutation spot.** $\theta_{m'}$ vs. c_m curves are plotted at (a) $T = 37°C$ (b)$T = 47°C$ for (1) $c_w = 10^{-7}M$ (2) $c_w = 10^{-8}M$ (3) $c_w = 10^{-9}M$ and (4) $c_w = 0$. The saturation regime can be avoided by increasing the hybridization temperature.

temperature. As noted before (Eq. 15), c_m/c_w is extracted from $\alpha = \Omega_{w'}/\Omega_{m'}$. The temperature dependence of α is relatively weak (Figure 6). The accessible range of α depends however on the detection limit of $\theta_{m'}$ which sets a lower bound of $\theta_{m'}$, $\theta_{m'}(min)$. Since $\Omega_{w'} \leq 1$ the maximal range of α is roughly $1/\theta_{m'}(min)$.

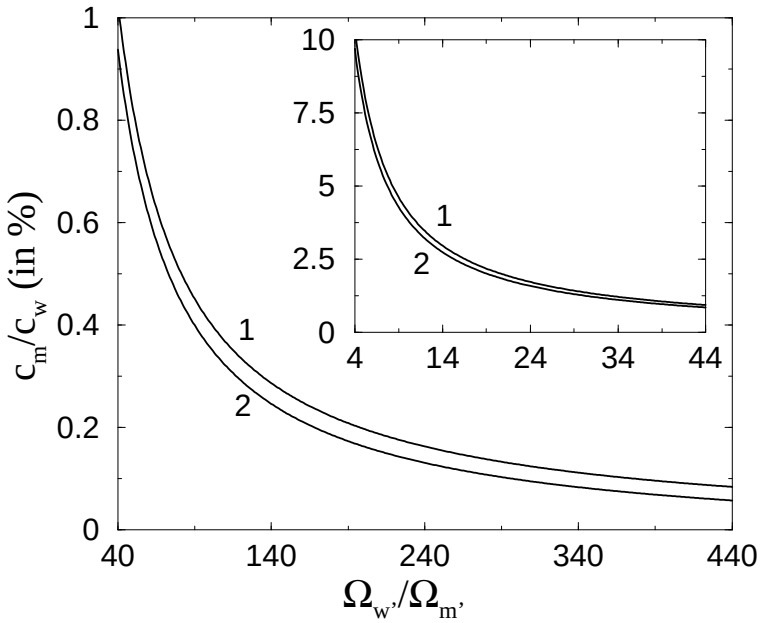

FIG. 6: **The ratios of concentrations obtainable from a "2-spot" experiment.** c_m/c_w vs. $\Omega_{w'}/\Omega_{m'}$ curves are plotted at (1) $T = 37°C$ and (2) $T = 47°C$ in the range $0.05\% \leq c_m/c_w \leq 1\%$. Inset: Range $1\% \leq c_m/c_w \leq 10\%$.

IV. THE TWO TEMPERATURE APPROACH

An alternative approach involves equilibration of the DNA chip with the biological sample at two different temperatures, T_1 and T_2. Focusing on the m' spot, we have

$$\Omega_{m'}(T_1) = c_w K_{m'w}(T_1) + c_m K_{m'm}(T_1), \tag{17}$$

$$\Omega_{m'}(T_2) = c_w K_{m'w}(T_2) + c_m K_{m'm}(T_2). \tag{18}$$

Solving equations (17) and (18) leads to

$$c_w = \frac{\Omega_{m'}(T_2)K_{m'm}(T_1) - \Omega_{m'}(T_1)K_{m'm}(T_2)}{K_{m'w}(T_2)K_{m'm}(T_1) - K_{m'w}(T_1)K_{m'm}(T_2)}, \tag{19}$$

$$c_m = \frac{\Omega_{m'}(T_1)K_{m'w}(T_2) - \Omega_{m'}(T_2)K_{m'w}(T_1)}{K_{m'w}(T_2)K_{m'm}(T_1) - K_{m'w}(T_1)K_{m'm}(T_2)} \tag{20}$$

and

$$\frac{c_m}{c_w} = \frac{\Omega_{m'}(T_1)K_{m'w}(T_2) - \Omega_{m'}(T_2)K_{m'w}(T_1)}{\Omega_{m'}(T_2)K_{m'm}(T_1) - \Omega_{m'}(T_1)K_{m'm}(T_2)}. \tag{21}$$

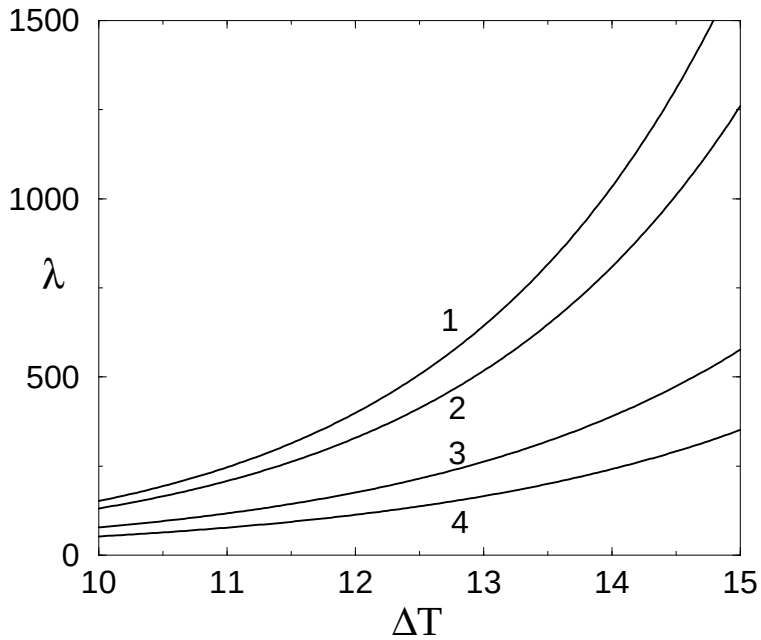

FIG. 7: **Effect of the concentration ratio on the temperature dependence of** $\Omega_{m'}(T_1)/\Omega_{m'}(T_2)$**.** $\lambda = \Omega_{m'}(T_1)/\Omega_{m'}(T_2)$ is plotted vs. ΔT for the interval $10 \leq \Delta T \leq 15$ for $T_1 = 37°C$. The concentration ratios are (1) $c_m/c_w = 10^{-2}$ (2) $c_m/c_w = 10^{-3}$ (3) $c_m/c_w = 10^{-4}$ and (4) $c_m/c_w = 0$.

To simplify this equation it is convenient to express $\frac{\Delta H_{m'w}}{RT_2}$ as $\frac{\Delta H_{m'w}}{RT_1}(1 - \frac{\Delta T}{T_1+\Delta T})$ where $T_2 = T_1 + \Delta T$. Upon defining $\lambda = \Omega_{m'}(T_2)/\Omega_{m'}(T_1)$ we obtain c_m/c_w in the form

$$\frac{c_m}{c_w} = \frac{K_{m'm}(T_1)}{K_{m'w}(T_1)} \frac{\exp\left(\frac{\Delta H_{m'w}}{RT_1}\frac{\Delta T}{T_1+\Delta T}\right) - \lambda}{\lambda - \exp\left(\frac{\Delta H_{m'm}}{RT_1}\frac{\Delta T}{T_1+\Delta T}\right)}. \tag{22}$$

Equation (22) can be simplified further when $\Delta T/T_1 \ll 1$ thus allowing to approximate $\frac{\Delta T}{T_1+\Delta T} \approx \frac{\Delta T}{T_1}$. In the weak spot limit $\lambda = \frac{I_{m'}(T_2)}{I_{m'}(T_1)}\frac{I^{\max}_{m'}(T_1)}{I^{\max}_{m'}(T_2)}$ and the saturation signals cancel when $I^{\max}_{m'}(T)$ is independent of T. When $I^{\max}_{m'}(T)$ does vary with T it is possible to eliminate this contribution by using an appropriate calibration method [18]. Hence, the measurement of the saturation values can be eliminated in the weak spot regime. This is of interest when the measurement technique allows to study the $\theta_i \ll 1$ range.

As we discussed, the optimization of the two spot approach is achieved by tuning a single hybridization temperature. In the two temperature approach it is attained by choosing T_1 and $\Delta T = T_2 - T_1$. The general requirements are the same: Avoiding the saturation regimes on one hand and the high noise to signal ratios on the other. Two additional observations merit comments. First, increasing ΔT magnifies the range of λ corresponding to the c_m/c_w range of interest (Figure 7). Second, it also allows to avoid the divergence regions in the c_m/c_w vs. λ curves (Figure 8) where ultra precise values of λ are required.

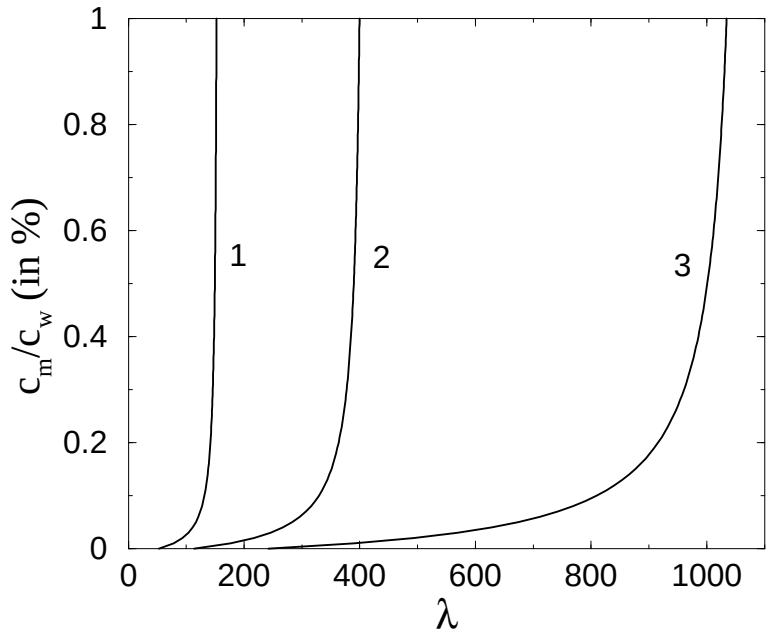

FIG. 8: **Ratio of concentrations obtainable from a two temperature experiment.** c_m/c_w vs. $\Omega_{m'}(T_1)/\Omega_{m'}(T_2)$ for $T_1 = 37°C$ and (1) $T_2 = 47°C$, (2) $T_2 = 49°C$ and (3) $T_2 = 51°C$.

V. DISCUSSION

Studies of somatic point mutations inevitably concern samples with a large excess of wild type DNA. This excess is propagated by the PCR amplification unless peptide nucleic acid (PNA) clamps [28] are used. The utilization of DNA chips to characterize the composition of such samples is hampered by competitive surface hybridization due to the pairing of the wild type target with the mutation probe. This mishybridization contributes significantly to the intensity of the mutation spot signal and thus gives rise to false positives. Our theoretical analysis suggests a systematic approach to the minimization of such errors. At best one can extract c_m/c_w thus obtaining additional clinical data. At least, this method provides a rational approach for devising criteria for identification of false positives.

DNA chips designed to detect point mutation in solid tumors were developed by Lopez-Crapez et al [10]. The sample preparation in this case involves asymmetric PCR or symmetric PCR followed by digestion with lambda exonuclease. DNA chips of higher sensitivity for the detection of K-ras mutation in stool were reported recently by Prix et al [6]. The amplification method utilizes PNA-mediated PCR clamping [28]. The PNA binds to the wild type codons of interest thus inhibiting their amplification. The PNA binding to the mutated DNA is weaker and does not inhibit its amplification. The overall result is a selective amplification of the mutated DNA. The methods we propose can be used to improve the performance of the approach of Lopez-Crapez et al. In this case the c_m/c_w values may provide supplementary information of diagnostic value. In marked contrast, our methods are of little use within the approach of Prix et al, which relies on preferential

amplification of the mutated DNA thus avoiding problems due to excess of wild type DNA.

Our analysis suggested two methods to obtain c_m/c_w. The advantage of the two-spot approach is that it does not involve a change in T. This approach requires however knowledge of the saturation value of the two spots. Obtaining the saturation values can be time consuming. However, for chips designed for multiple use this step can be performed only once provided the saturation value does not change with the hybridization regeneration cycles. The two-temperature approach is attractive when using a label free detection scheme such as surface plasmon resonance. In this case the two measurements can be carried out with no washing steps. Importantly, Fotin et al [18] demonstrated the feasibility of studying the temperature dependence of hybridization isotherms even for targets labelled with fluorescent tags.

Thermodynamic equilibrium is a necessary condition for the implementation of the quantification methods proposed in this article. The equilibration times of DNA chips and their variation with the hybridization conditions are not yet fully understood. The reported equilibration times for hybridization on DNA chips vary widely. In comparing the results it is important to note differences in hybridization conditions, length of probes, type of chip and detection techniques. Two studies are of special interest. Peterson et al reported equilibration times of up to 14 hours [22, 23]. In marked contrast, the results of Fotin et al [18] suggest equilibration within minutes. In the present context, the work of Fotin et al is of special interest since it reports equilibrium hybridization isotherms at different temperatures. Importantly, the isotherms indeed satisfy the conditions of equilibrium: Stationary state and lack of hysteresis. Furthermore, the system exhibits a Langmuir type hybridization isotherm and the thermodynamic parameters extracted from the temperature dependence of the hybridization isotherms are in good agreement with the reported bulk values. Note however that the hybridization constants describing other types of DNA chips do not always exhibit such agreement[29, 30].

The proposed methods require DNA chips obeying Langmuir type hybridization isotherms. In turn, this requires spots with relatively low density of probes. Clearly, there are advantages to chips carrying spots with high density of probe. One is that high density enable smaller spots thus allowing for greater number of different spots. Importantly, in DNA microarrays the high probe density gives rise to an electrostatic modification of the isotherms. The resulting nonlinearities are undesirable for obtaining quantitative results.

Acknowledgement

The authors benefitted from insightful discussions with T. Livache. The work of EBZ was funded by the CNRS.

[1] Lodish H, Berk A, Zipursky L, Matsudaira P, Baltimore D, Darnell J. Molecular Cell Biology. WH Freeman 4th Ed 2000.

[2] Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F. Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res 2002; 30:3295-3311.

[3] Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. Science 1992; 256:102-105.

[4] Hibi K, Robinson CR, Booker S, Wu L, Hamilton SR, Sidransky D, Jen J. Molecular detection of genetic alternations in the serum of colorectal cancer patients. Cancer Res 1998; 58:1405-1407.

[5] Sánchez-Carbayo M. Use of high-throughput DNA microarrays to identify biomarkers for bladder cancer. Clin Chem 2003; 49:23-31.
[6] Prix L, Uciechowski P, Böckmann B, Giesing M, Schuetz AJ. Diagnostic biochip array for fast and sensitive detection of K-ras mutations in stool. Clin Chem 2002; 48:428-435.
[7] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and prediction by gene expression monitoring. Science 1999; 286:531-537.
[8] Graves DJ. Powerful tools for genetic analysis come of age. Trends Biotechnol 1999; 17:127-134.
[9] Wang J. From DNA biosensors to gene chips. Nucleic Acids Res 2000; 28:3011-3016.
[10] Lopez-Crapez E, Livache T, Marchand J, Grenier J. K-ras mutation detection by hybridization to a polypyrrole DNA chip. Clin Chem 2001; 47:186-194.
[11] Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high density oligonucleotide arrays. Nature Biotechnol 1996; 14:1675-1680.
[12] Li F, Stormo GD. Selection of optimal DNA oligos for gene expression arrays. Bioinformatics 2001; 17:1067-1076.
[13] Adamson AW. Physical Chemistry of Surfaces. J Wiley, New York, 4th Ed, 1982.
[14] Halperin A, Buhot A, Zhulina EB. Sensitivity, specificity and the hybridization isotherms of DNA chips. Biophys J 2004; 86:718-730.
[15] Gyllensten UB, Erlich HA. Generation of single stranded DNA by the polymerase chain reaction and its application to the direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci USA 1988; 85:7652-7656.
[16] Gao HF, Tao SC, Wang D, Zhang C, Ma XM, Cheng J, et al. Comparison of different methods for preparing single stranded DNA for oligonucleotide microarray. Analytical Lett 2003; 36:2849-2863.
[17] Higuchi RG, Ochman H. Production of single stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res 1989; 17:5865.
[18] Fotin AV, Drobeyshev AL, Proudnikov DY, Perov AN, Mirzabekov AD. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchip. Nucleic Acids Res 1998; 26:1515-1521.
[19] Vainrub A, Pettitt MB. Coulomb blockage of hybridization in two-dimensional DNA arrays. Phys Rev E 2002; 66:041905.
[20] Moore WJ. Physical Chemistry. Longman, London 1972.
[21] Smith SB, Cui Y, Bustamante C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science 1996; 271:795-799.
[22] Peterson AW, Heaton, RJ, Georgiadis RM. The effect of surface probe density on DNA hybridization. Nucleic Acids Res 2001; 29:5163-8.
[23] Peterson AW, Wolf LK, Georgiadis RM. Hybridization of mismatched or partially matched DNA at surfaces. J Am Chem Soc 2002; 124:14601-7.
[24] SantaLucia J Jr. A unified view of polymer, dumbbell and oligonucleotide DNA nearest neighbor thermodynamics. Proc Natl Acad Sci USA 1998; 95:1460-5.
[25] Peyret N, Senerviratne PA, Allawi HT, SantaLucia J Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G and T·T mismatches. Biochemistry 1999; 38:3468-3477.
[26] Bommarito S, Peyret N, SantaLucia J Jr. Thermodynamic parameters for DNA sequences with dangling ends. Nucleic Acids Res 2000; 28:1929-1934.
[27] HYTHERTM version 1.0, Nicolas Peyret and John SantaLucia Jr. Wayne State University available at http://ozone2.chem.wayne.edu/Hyther/hytherm1main.html (Accessed April 2004).
[28] Thiede C, Bayerdörffer E, Blasczyk R, Wittig B, Neubauer A. Simple and sensitive detection of mutations in ras proto-oncogenes using PNA-mediated clamping. Nucleic Acids Res 2001; 24:983-4.
[29] Held GA, Grinstein G, Tu Y. Modeling of DNA microarray data by using physical properties of hybridization. Proc Natl Acad Sci USA 2003; 100:7575-7580.
[30] Zhang L, Miles MF, Aldape, KD. A model of molecular interactions on short oligonucleotide microarrays. Nature Biotechnol 2003; 21:818-821.