Diffusion, Mixing, and Associated Dye Effects in DNA-Microarray Hybridizations

Jacob R. Borden, Carlos J. Paredes, and Eleftherios Terry Papoutsakis
Deptartment of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208

ABSTRACT Typical DNA microarrays utilize diffusion of dye-labeled cDNA probes followed by sequence-specific hybridization to immobilized targets. Here we experimentally estimated the distance typical probes travel during static 16-h hybridizations. Probes labeled with Cy3 and Cy5 were individually introduced to opposite sides of a microarray with minimal convective mixing. Oppositely labeled probes diffused across the initial front separating the two solutions, generating a zone with both dyes present. Diffusion-distance estimates for Cy3- and Cy5-labeled cDNAs were 3.8 mm and 2.6 mm, respectively, despite having almost identical molecular masses. In separate 16-h hybridization experiments with oppositely labeled probes premixed, arrays that were continuously mixed had 15–20% higher signal intensities than arrays hybridized statically. However, no change was observed in the Cy3/Cy5 signal intensity ratio between continuously mixed and static hybridizations. This suggests that the observed dye bias in diffusion-distance estimates results from differences in the detection limits of Cy3 and Cy5-labeled cDNA, a potential concern for array data on low-abundance transcripts. Our conservative diffusion-distance estimates indicate that replicate targets $>$7.6 mm apart will not compete for scarce probes. Also, raising the microarray gap height would delay the onset of diffusion-limited hybridization by increasing the amount of available probe.

INTRODUCTION

The impact of DNA microarrays continues to grow, with applications extending beyond transcriptional profiling such as in comparative genomics (1,2), genome replication dynamics (3), and ChIP-on-chip analysis (4–6). Target DNA spots might consist of PCR-amplified full or partial open reading frames (ORFs), or of short oligonucleotides spotted or synthesized directly onto the array surface. Probe species in solution can be cDNA, fragmented cRNA, or fragmented gDNA (genomic DNA) labeled with fluorophores (dyes) through either direct or indirect dye incorporation. Typical array platforms, for instance the transcriptional arrays used in this study, consist of partial ORFs spotted on a microarray slide and hybridized statically with cDNA generated by indirect dye incorporation. Continuous mixing of the hybridization solution is also possible and employed by two major commercial microarray systems (Affymetrix, Santa Clara, CA, and Agilent Technologies, Palo Alto, CA).

The potential for biological insight using microarrays has also stimulated a desire to better understand the underlying processes of reaction and mass transfer kinetics. For instance, the relative importance of diffusion and reaction processes on the extent of hybridization continues to be debated. One school of thought is that the hybridization rate is reaction-rate limited, whereas another argues that it is diffusion limited. If diffusion is fast relative to reaction, then little signal improvement is possible by adding external mixing (because concentration gradients are minimal) and, instead, efforts such as dextran-sulfate based reaction enhancement (7) may be necessary. Also, even closely spaced replicate DNA targets will not compete for labeled probe. However, if slow diffusion in fact limits the overall hybridization rate, then the distance a labeled DNA molecule is capable of diffusing directly impacts both achievable signal intensity and competition for scarce probe between closely spaced replicate DNA targets.

The contention for a reaction limitation stems from initial modeling by Chan et al. (8). An integral assumption in that work, however, was that probe availability is constant during hybridization, a situation acknowledged to be indicative only of the earliest timepoints. Gadgil et al. (9) modeled the development of concentration profiles over time within a static hybridization, without assuming constant probe availability, and examined both rare and abundant probe populations. Their study employed the following parameters: a target spot 100 μm in diameter, a gap height of 140 μm, a sample volume of 20 μl, an average probe length of 2000 nucleotides, and a probe diffusivity of 1×10^{-7} cm^2/s. Based on the concentration profiles generated using the model it was concluded that at 250 μm from a target DNA spot, probe species remain at approximately the initial concentration after 12 h of simulated hybridization.

Using a combination of random walk and continuous reaction/diffusion models, Pappaert et al. (10) show that the determination of which step is limiting, reaction or diffusion, is entirely dependent on the hybridization timescale. Short timescales are characterized by reaction-limited hybridization given readily available probe at the initial bulk concentration. After the initial time period, probe depletion renders the process diffusion-limited.

The goal of this study is to characterize diffusionally relevant properties of dye-labeled probes used in traditional

Submitted June 3, 2005, and accepted for publication August 1, 2005.

Address reprint requests to Eleftherios Terry Papoutsakis, Dept. of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208. Tel.: 847-491-7455; Fax: 847-491-3728; E-mail: e-paps@northwestern.edu.

0006-3495/05/11/3277/08 $2.00

doi: 10.1529/biophysj.105.067934

DNA microarray hybridizations. To this end, we first aim to estimate the distance typical dye-labeled cDNAs diffuse under static hybridization conditions, utilizing a standard microarray platform we developed for transcriptional profiling of the prokaryote *Clostridium acetobutylicum* (11–13). To estimate diffusion distances, microarrays were hybridized with equivalent Cy3- and Cy5-labeled cDNA pools applied to opposite ends of an array to form a stable boundary between the oppositely labeled cDNA solutions. Over time, oppositely labeled probes diffused into each other to form a zone where both dyes were present. The width of this zone was used to determine the distance dye-labeled cDNAs had traveled over the course of the hybridization, and a one-dimensional diffusion model was used to calculate diffusivities from experimentally determined diffusion distances. We also assessed the contribution of initial convective mixing to the estimation of diffusion distances. Our findings that these experimentally determined, labeled cDNA diffusion distances are dye-specific required investigation into the molecular masses of each dye-labeled cDNA population, as well as array hybridizations under continuous mixing conditions.

METHODS

Arrays

Design and validation of the microarrays employed in this study has been described (13). Briefly, PCR products from nonhomologous regions (so as to minimize nonspecific hybridizations) of ~95% of the predicted *C. acetobutylicum* ORFs were spotted in triplicate, yielding an array with 12,672 target spots, each measuring 120 μm in diameter. Replicate targets were placed 17.9 mm apart to prevent competition for labeled probe. The interspot spacing, and therefore the resolution of diffusion-distance estimation detailed below, was 260 μm. Hybridization solutions were applied to the arrays by capillary action between the array and the coverslip, with risers along the long (vertical) axis of the coverslip creating a 57-μm gap height.

Probe generation

RNA from the major transcriptional stages of *C. acetobutylicum* (exponential, transitional, and stationary) was isolated as described (11). cDNA was generated by random hexamer-primed reverse transcription in the presence of amino-allyl (aa) dUTP, followed by indirect coupling of Cy3 and Cy5 dyes to generate labeled probe (13). Finally, individual labeling reactions were pooled according to dye, quantified by measuring absorbance at 260 nm, and dye incorporation estimated at 549 nm and 659 nm for Cy3 and Cy5, respectively.

Experimental overview

In principle, oppositely labeled probe pools, generated as described above, can be applied to opposite sides of a single array (Fig. 1 *a*). When the dye solutions meet, a stable boundary is generated separating Cy3- from Cy5-labeled probes (Fig. 1 *b*). Over time, labeled-probe molecules diffuse into the opposite dye solution, traveling until successful hybridization occurs, and gradually generating a zone where probes of both dyes have hybridized to their target. What results is a gradual transition from one zone where only Cy3-labeled probes have hybridized to a zone where Cy5-labeled probes are the predominant species (Fig. 1 *c*). The height (Δz) of the yellow band in Fig. 1 *c* is indicative of the distance dye-labeled probe is capable of traveling by diffusion.

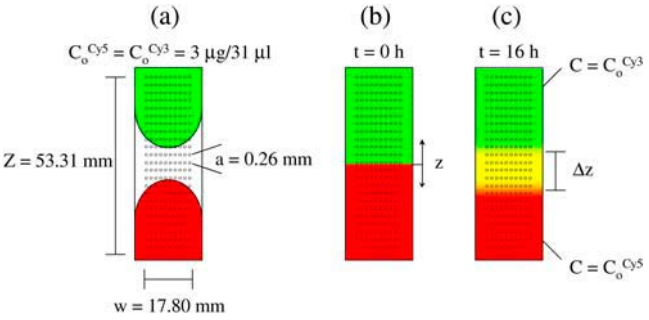

FIGURE 1 Diffusion-slide hybridization process. (*a*) Cy3- (*green*) and Cy5-labeled (*red*) probes are introduced to opposite ends of one array. (*b*) At time 0, the labeled dye solutions meet and form a stable boundary. (*c*) Over the course of a 16-h hybridization, diffusion of labeled probe across the initial boundary generates a zone of height (Δz) where both dyes are present.

Transcript abundance varies across the *C. acetobutylicum* transcriptome; thus, some target DNA spots will have more labeled cDNA with which to hybridize than other target spots. Therefore, in the diffusion arrays discussed above, two neighboring target spots within the diffusion zone hybridizing to different cDNA transcripts may have widely varying signals—not due to diffusion, but due to differences in the abundance of the respective transcripts. Signal variability due to transcript abundance makes it difficult to precisely distinguish the location of the diffusion zone. Therefore, a second set of arrays was hybridized to standardize the observed intensities in the diffusion slides. In these slides, the same dye-labeled cDNAs used in the diffusion slides were completely mixed before application to the array, so that final target signal was a function of transcript abundance alone. Variability in transcript abundance was accounted for by dividing the intensity of each target in a ''diffusion'' slide by the intensity of the corresponding target in a ''premixed'' slide.

Microarray hybridizations for diffusion-distance estimation

Three micrograms of each dye-labeled cDNA sample were individually dissolved in 31 μl of hybridization buffer and introduced to the opposite ends of a diffusion array. Over the course of a 16-h hybridization at 42°C, diffusion and hybridization generate the central zone of Fig. 1 *c*, where both dye-labeled probes were present. To account for transcript abundance as mentioned above, premixed slides were hybridized statically by premixing 3 μg of each dye-labeled cDNA sample into 62 μl of hybridization buffer, which was then introduced onto the microarray.

The ideal intensity profile

To discuss how the diffusion/premixed (*d*/*m*) slide intensity ratio in each channel changes along the long (vertical) axis of the array slide, consider three targets—one at the top of the diffusion array in Fig. 1 *c*, one in the middle, and one at the bottom. At the top of the diffusion slide, far from the central diffusion zone, the target is exposed to only Cy3-labeled probe and at a twofold higher concentration than in the premixed slide. It should therefore have a *d*/*m* intensity ratio of 2 in the Cy3 channel:

$$\left(\frac{d}{m}\right) = \frac{3\ \mu g/31\ \mu l}{3\ \mu g/62\ \mu l} = \frac{62}{31} = 2. \quad (1)$$

At the center of the yellow band of Fig. 1 *c*, where diffusional mixing has taken place, a target spot in the diffusion slide and premixed slide have

approximately equivalent Cy3 (and Cy5) probe concentrations, giving a Cy3 (and Cy5) *d*/*m* ratio of 1. Finally, at the bottom of the slide, no Cy3 is present, so the *d*/*m* ratio falls to 0 for Cy3 but is now 2 for Cy5. Thus, along the vertical axis of the slide in Fig. 1 *c*, the ideal *d*/*m* intensity ratio in either channel has limits of 2 and 0 with a transition region in between that defines the diffusion zone.

Estimating the diffusion-zone length

The low end of the intensity spectrum in both channels is accompanied by increased experimental noise. To minimize this noise, log-log data transformation of the *d*/*m*-ratio values along the vertical position within each slide was carried out, which also transformed the *d*/*m*-ratio profile within the diffusion zone to one that can be approximated as linear. Next, we use this straight-line approximation to calculate the distance between *d*/*m*-ratio values of 2 and 0. First, however, it is recognized that because log(0) is undefined, we instead define the low-ratio boundary to be when the diffusion slide intensity is 1:100 of the intensity of the premixed slide, below which the ratio values show greater scatter because of increased signal noise. The diffusion distance in either channel is then taken as one-half of the straight-line distance between the two *d*/*m*-ratio extrema. Within the technical limits of detection, the liquid-phase concentration of dye-labeled cDNA is assumed to be directly proportional to the amount of hybridized cDNA probe.

Microarray hybridizations for estimating the impact of continuous mixing

Microarrays were also hybridized with continuous mixing throughout the hybridization. These microarrays (not to be confused with ''premixed'' microarrays described previously) were hybridized with 3 μg each of Cy3- and Cy5-labeled cDNA dissolved in 490 μl hybridization buffer. During the 16-h hybridization, arrays were continuously mixed using the Agilent microarray system (Agilent Technologies) of introducing an air bubble and rotating the array at 4 rpm throughout the hybridization (Agilent microarray processing protocol, Version 4.1, http://www.chem.agilent.com/scripts/literaturePDF.asp?iWHID=34961). The signal intensities of target spots in these ''continuously mixed'' slides were compared with the target-spot intensities in slides hybridized statically with 3 μg of each dye-labeled probe in 490 μl of hybridization buffer.

RESULTS AND DISCUSSION

Estimation of the initial convective mixing

The contacting of dye-labeled probe solutions is accompanied by some convective mixing. Convective mixing may contribute to the generation of a zone whereby probes of both dye species are present, leading to possible overestimation of the diffusion distance of labeled cDNA. To estimate the extent of initial convective mixing, a digital camera was used to capture images at the time the two dye solutions came into contact. As Cy dyes under natural light have red (Cy3) and blue (Cy5) hues, initial convective mixing was estimated using the blue/green/red content of pixels within the digital pictures taken just as the advancing dye solutions came into contact. The results of such experiments (Fig. S1, Supplementary Material) showed that convective mixing generates a zone of ~300 μm (10 pixels) having both blue and red hues. Although this shows a small amount of convective mixing, the impact is on the order of the resolution limit within the array because of intrafeature spacing (260 μm), and thus no corrections were made to subsequent diffusion-distance estimates, with the understanding that they may be overestimated by up to 300 μm.

Molecular weight distribution of cDNA

To relate diffusion-distance and diffusivity estimates to the molecular mass of labeled cDNAs, the size range of cDNAs used in the diffusion experiments was also determined. *C. acetobutylicum* transcripts for generation of cDNA vary in length, from as small as 250 basepairs to operons that exceed 5 kb (14,15). Additionally, reverse transcription to generate cDNAs is conducted with a molar excess of random primers, thus resulting in multiple short cDNAs generated from a single transcript. To estimate the length distribution of cDNAs applied to the microarray, RNA samples were reverse transcribed, as described in Materials and Methods, without the incorporation of dye molecules. This unlabeled cDNA population was then analyzed by electrophoresis on a 0.7% agarose gel, showing an average fragment size of 500–600 basepairs, and maximum and minimum fragment sizes of ~3000 and 100 basepairs, respectively (Fig. S2, Supplementary Material).

Diffusion distances from global analysis

The global diffusion distance in each channel can be estimated by plotting the intensity ratio of all array features in a single graph as $\log_{10}$(*d*/*m* ratio) versus $\log_{10}$(vertical target position within array) as shown in Fig. 2 for the Cy3 channel of one diffusion/premixed slide comparison. It can be seen that, within experimental error, the data agree with the expected extrema of the *d*/*m* intensity ratio (0 and 2), the *d*/*m*-

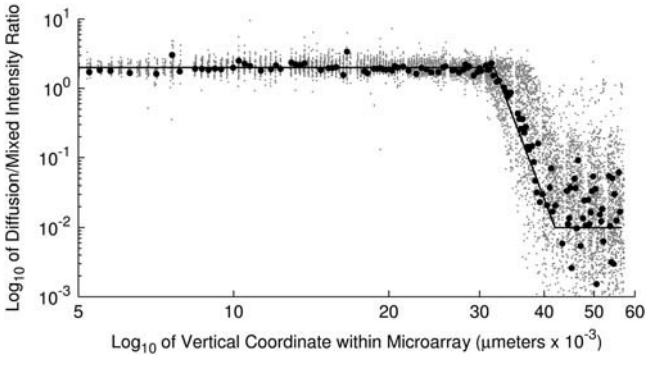

FIGURE 2 Global analysis for estimation of diffusion distance in the Cy3 channel of a representative diffusion and mixed slide comparison. Data are shown after log-log transformation as discussed in the text. Data from a single column within the array (*large black dots*) are shown for comparison of one column profile to the accumulated profile of all targets. The theoretical profile is also shown, with unitless extrema of 2 and 10^{-2}, and a linear transition through the diffusion zone (see text for further explanation).

ratio profile appears linear in the diffusion zone after log-log data transformation, and the signal/noise ratio quickly increases as *d*/*m* intensity ratios approach 1:100.

In total, three diffusion slides were hybridized to estimate the diffusion-zone length, and two mixed slides hybridized to standardize against transcript variability. Each diffusion slide was then compared to each mixed slide for calculation of intensity ratios and, thereby, diffusion distances, yielding a total of six data sets per channel (Table 1). From the global analysis of the six comparisons, the diffusion distances for Cy3 and Cy5 were estimated as 3.9 and 2.6 mm (with standard deviations of 0.4 and 0.5 mm), respectively. These estimates of the Cy3 and Cy5 diffusion distances were different with 99% confidence. This result was unexpected given the similarity in molecular mass and chemical structure of the Cy3 and Cy5 dye molecules (16). Initial skepticism centered on the use of a global analysis approach, which possibly concealed experimental biases such as wall effects close to the array edge. Therefore, a more detailed analysis was undertaken to estimate diffusion distances in individual array columns.

Moving average distance calculations and determination of wall effects

To account for possible spatial biases within a diffusion slide, the following adjustments were made to the global estimation procedure. First, the *d*/*m* intensity ratio profiles of individual columns were examined one by one rather than globally looking at the profile of all columns of a diffusion/premixed array comparison. Second, the comparison of one diffusion slide to multiple mixed slides was combined onto a single profile. Third, the profile for each column in a diffusion slide was analyzed alongside its neighboring columns to the left and right, yielding a ''moving average'' for the diffusion zone in any single column. Thus, the diffusion distance calculated for a given diffusion slide column (k) is in fact the average of six profiles, the profiles for columns $k-1$, k, and $k+1$ standardized using the first mixed slide, and the profiles for these same columns standardized using the second mixed slide. Finally, log-log data transformation was carried out and the diffusion zone data were fit to a straight line by trimmed least squares robust regression (17). The outcome of this procedure is depicted in Fig. 3 for both channels of a representative diffusion-slide column.

Diffusion-distance estimates for each column of a representative diffusion slide are shown in Fig. 4. Estimates near the edges of the array are clearly different from estimates made near the array center, prompting iterative exclusion of the columns closest to the microarray edges. That is, for each channel, distance estimates in each column were made, and the average distance for all the columns of an array was calculated without edge exclusion (Table 2). Then, the number of columns to be excluded (n_c) was determined by dividing the average diffusion distance per channel by the spacing between columns (260 μm). After excluding n_c columns from both edges of an array, a new average distance (per channel) was calculated. Finally, the process was repeated until the number of excluded columns remained constant. Table 2 shows that the mean distance estimate approached the median value with iterative exclusion of the edge columns. The diffusion-distance estimates for Cy3- and Cy5-labeled cDNA were 3.8 mm and 2.6 mm, respectively, different with >99% confidence, and in line with a global analysis (see Table 1), despite eliminating the bias due to wall effects. These diffusion-distance estimates are an order of magnitude larger than those reported by Gadgil et al. (9), determined such that probe concentration was ''....almost equal to the initial concentration at a distance of ~250 mm from the center of the spot....'' Here, however, we used a 1:100 *d*/*m*-ratio cutoff to provide a conservative estimate of the distance beyond which there is no significant competition for labeled probe between

TABLE 1 Diffusion-distance estimates using the global analysis approach

Diffusion slide	Mix slide	Diffusion distance in the Cy-3 channel (mm)	Diffusion distance in the Cy-5 channel (mm)
A	1	4.5	2.8
A	2	4.0	3.0
B	1	4.3	3.3
B	2	3.7	2.7
C	1	3.8	2.1
C	2	3.4	2.1
Group averages			
Diffusion slide A ($n = 2$)		4.2	2.9
Diffusion slide B ($n = 2$)		4.0	3.0
Diffusion slide C ($n = 2$)		3.6	2.1
Mixed slide 1 ($n = 3$)		4.2	2.7
Mixed slide 2 ($n = 3$)		3.7	2.6
Overall average		3.9	2.6
Overall standard deviation		0.4	0.5

See text for details of the global analysis approach.

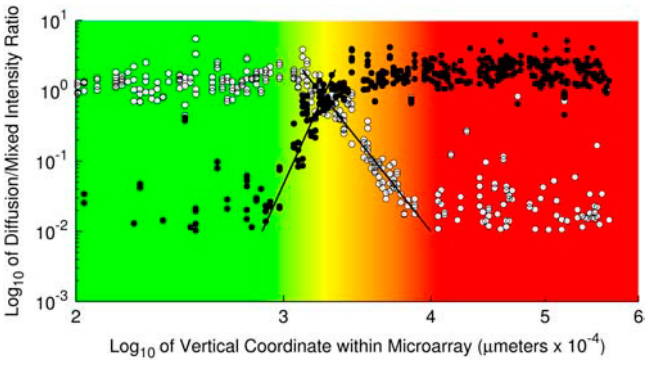

FIGURE 3 Diffusion/mixed intensity ratio profiles for both channels of a single column using the moving average approach (see text for details). The best-fit lines for the Cy3 channel (*open circles*) and Cy5 channel (*closed circles*) are also shown. Background shading depicts the profile of color transition as it appears on a scanned diffusion slide.

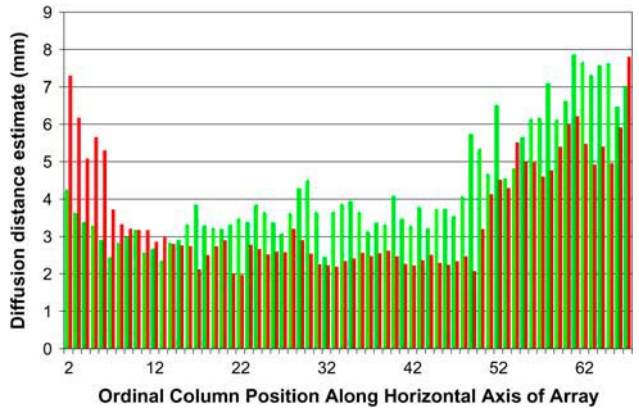

FIGURE 4 Profile of distance estimates along the horizontal (short) axis of a representative diffusion/mixed slide comparison. The calculated distance for each column is plotted relative to the ordinal position of that column within the array. Diffusion distance values for both Cy3 (*green bars*) and Cy5 (*red bars*) are shown.

replicate target DNA spots. A suitable model for our experimental system was then necessary for comparing our experimental values to theoretical predictions.

Comparison with diffusion model

To model the evolution of diffusion-driven concentration profiles in our system, the solution for transient one-dimensional diffusion was used along with the accompanying initial and boundary conditions, with the z coordinate being zero at the initial point of contact between oppositely labeled probes in the diffusion microarrays discussed above:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial z^2} \tag{2}$$

$$\begin{aligned} &\text{at} \quad t = 0 \quad C = 0 \quad \text{for all} \quad z > 0 \\ &\text{at} \quad z = \infty \quad C = 0 \quad \text{for all} \quad t \geq 0 \\ &\text{at} \quad z = -\infty \quad C = C_o \quad \text{for all} \quad t \geq 0. \end{aligned}$$

Since the diffusion zone (Δz) is only a small fraction of the microarray-slide length, for mathematical convenience we took the z domain as infinite in length. The solution methodology for this type of boundary-value problem is well established (18), and the following solution can be readily obtained for this specific problem:

$$\frac{C}{C_o} = \frac{1}{2}\left[1 - erf\left(\frac{z}{2\sqrt{Dt}}\right)\right]. \tag{3}$$

The concentration profiles that develop in the z-direction across the initial dye boundary can be calculated for different hybridization times using Eq. 3, and the diffusivities for Cy3- and Cy5-labeled cDNAs can also be estimated. As discussed in Materials and Methods, for diffusion-distance estimation, the cutoff for diffusion/mixed intensity ratios was 0.01; however, the end-point for diffusion distance using Eq. 3 is where C/C_o approaches 0.005 after 16 h of hybridization due to the mixed slide being only one-half the C_o concentration in the diffusion slide. Contours for simulated diffusion in both channels are depicted for two hybridization times (Fig. 5). The initial diffusion boundary serves as the point where z equals zero, shown as the vertical line in Fig. 5. Diffusivities of 1.9×10^{-7} and 0.9×10^{-7} cm^2/s in Eq. 3 yield diffusion distances of 3.8 and 2.6 mm for Cy3 and Cy5, respectively.

What could possibly result in the dye bias of diffusion-distance estimation?

The experimentally observed dye bias in diffusion distances and thus diffusivities could reflect either differences in the molecular mass of dye-labeled cDNAs, or instead dye dif-

TABLE 2 Diffusion-distance estimates using the moving average approach

	Cy-3 channel			Cy-5 channel		
Diffusion slide	Average diffusion distance (mm)	Median diffusion distance (mm)	Standard deviation (mm)	Average diffusion distance (mm)	Median diffusion distance (mm)	Standard deviation (mm)
A	4.3	4.3	1.1	2.7	2.6	0.9
B	4.2	3.6	1.5	3.5	2.8	1.5
C	3.8	3.4	1.1	2.9	2.7	1.0
A*	4.3	4.2	0.9	2.4	2.2	0.7
B*	3.7	3.6	0.6	2.6	2.5	0.6
C*	3.3	3.2	0.7	2.8	2.7	0.6
Group averages						
Uncorrected	4.1	3.7	1.3	3.0	2.7	1.2
Border corrected	3.8	3.6	0.9	2.6	2.5	0.7
Diffusivity (cm^2/s)		1.9×10^{-7}			0.9×10^{-7}	

See text for details of the moving average approach. Diffusivity for each channel is shown, calculated using Eq. 3 and the average diffusion distance after correcting for wall effects.
*Border corrected (see text for details).

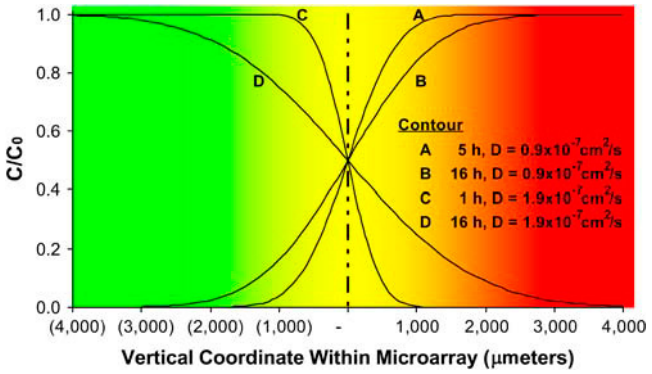

FIGURE 5 Model of the diffusion process. Contours are calculated in each channel using Eq. 3, with the diffusivities and hybridization times shown for each contour. Background shading depicts the profile of color transition after 16-h hybridization, as in Fig. 3.

ferences in physical properties affecting the experimental observations. Differences in the molecular masses of Cy3- and Cy5-labeled cDNAs could be due either to molecular mass differences between Cy3 and Cy5 dyes (766 and 792, respectively) or possibly to the different dye-incorporation rates for Cy3- and Cy5-labeled cDNA probes (18.6 and 24.0 fluor molecules/1000 nucleotides, respectively). According to the Stokes-Einstein correlation for large spherical molecules, the diffusivity is inversely proportional to the hydrodynamic radius (19), or inversely proportional to the cube root of molecular mass. Therefore, to account for the ~2-fold difference in diffusivities found when Eq. 3 was applied to the experimental data, Cy5-labeled cDNA would need to have an ~8-fold larger molecular mass than Cy3-labeled cDNA.

To investigate this possibility, a commercially available RNA ladder (Millenium Marker, Ambion, Austin, TX) was reverse-transcribed in the presence of aa dUTP and a primer

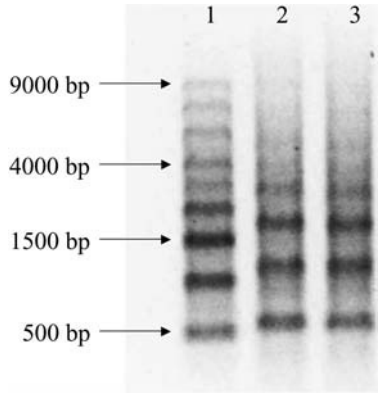

FIGURE 6 A 0.7% agarose gel of aa incorporated reverse transcription products. (*Lane 1*) aa-cDNA standard generated by reverse transcription of the Millenium Marker RNA standard. (*Lanes 2* and *3*) Cy3- and Cy5-labeled cDNA generated by cross-linking of the respective dye with the aa-cDNA standard shown in Lane 1.

that anneals with all the fragment lengths of the ladder (0.5–9.0 kb). This aa-cDNA ladder was then labeled with Cy3 and Cy5 dyes and analyzed by gel electrophoresis on a 0.7% agarose gel (Fig. 6). Although a shift in electrophoretic mobility between the unlabeled and dye-labeled cDNAs was observed for all fragment lengths (Fig. 6, compare lane *1* with lanes *2* and *3*), no difference in mobility was detected between the Cy3- and Cy5-labeled cDNAs for any fragment length (Fig. 6, compare lane *2* with lane *3*). This shows that the molecular masses of Cy3- and Cy5-labeled cDNAs are nearly identical, and thus molecular-mass differences could not account for the observed dye bias in diffusivities.

This led us to independently reassess whether the true diffusion distances and diffusivities were indeed different for the two populations of dye-labeled cDNAs. To this effect, standard microarray hybridizations (i.e., with the oppositely dye-labeled cDNAs premixed) were conducted to compare the Cy3/Cy5 signal intensity ratios in hybridizations where the hybridization solution is continuously agitated to ratios in microarrays hybridized statically. Assuming that the difference in diffusion distances of Cy3- and Cy5-labeled cDNAs is correct, Cy3 labeled probes are able to migrate farther during hybridization than Cy5-labeled probes. As a result, under static hybridization conditions any given target spot should be able to hybridize with more Cy3- than Cy5-labeled probe, resulting in a diffusion-biased Cy3/Cy5 intensity ratio. By contrast, continuous agitation of the hybridization solution would eliminate any potential diffusion bias, thus resulting in a Cy3/Cy5 intensity ratio different from the Cy3/Cy5 ratio in a statically-hybridized microarray. Comparison of Cy3/Cy5 signal intensity ratios in agitated arrays to ratios in static arrays showed no change across the two hybridizations (Fig. S3, Supplementary Material). Although the signal intensity in both channels increased as a result of agitation (thus indicating that diffusion indeed limits the amount of dye-labeled cDNAs available for target hybridization), the intensity in both channels increased to roughly the same extent: 15–20% improvement in signal intensity above static hybridizations in both channels. We therefore conclude that Cy3- and Cy5-labeled cDNAs have the same diffusional properties, and that therefore the dye bias in the estimated diffusion distances is an experimental artifact resulting from dye properties (other than molecular masses) affecting the experimental estimation of diffusion distances. We argue below that the artifact most likely derives from the use of a single intensity cutoff for both channels, and the associated issue of different detection limits for the two dyes.

An issue of detection sensitivity?

One aspect of the experimental design that may contribute to the measured diffusion-distance bias is possibly the use of a single lower intensity cutoff (0.01) with which to estimate diffusion distances. That is, for both the Cy3 and Cy5 channels, the end-point for diffusion distance was taken as the

point where intensity in the diffusion slide decreased to 1/100 of the intensity in the mixed slide. This cutoff was envisioned to be the point where a target in the diffusion slide has bound some minimal but discernable amount of labeled cDNAs. It was assumed that a single intensity threshold for both channels would be a fair representation of the same number of hybridized molecules in both channels. Two dye properties that impact the use of a single low-end cutoff are the quantum yield and the limit of detection of Cy3 and Cy5 dyes, with each contributing to the amount of labeled probe that must be hybridized to achieve a signal sufficiently greater than background. However, the quantum yield of Cy5 is in fact greater than that of Cy3 (16). This then leaves differences in the limit of detection as the possible source of the apparent dye bias in the diffusion-distance estimation using the same lower intensity threshold. It has indeed been reported that the limit of detection for Cy5-labeled cDNAs is twice that for Cy3-labeled cDNAs (6 and 3 fluor molecules/μm^2, respectively) (20). This also implies that the low signal intensities in the Cy5 channel (for the same number of Cy3- or Cy5-labeled cDNA molecules) will be accompanied by greater noise than low signals in the Cy3 channel.

CONCLUSIONS

The diffusion distances calculated using the approach described here apply to a cDNA population having a distribution of molecular weights. Although the average cDNA fragment size was shown to be 500–600 basepairs, the distance estimates likely represent the diffusion distance of the smaller strands of the cDNA population with larger diffusivities due to their lower molecular mass. However, the diffusion-distance estimate that results is a conservative approximation (i.e., slight overestimate) of the distance labeled cDNAs travel during standard microarray experiments. Furthermore, our method slightly overestimates the diffusion distance by up to 300 μm due to the initial convective mixing.

The apparent difference in estimated diffusivities between Cy3- and Cy5-labeled cDNA should bias the amount of dye-labeled cDNA hybridizing to a given target spot during static hybridization, in favor of the apparently farther-traveling Cy3-labeled species. When diffusion limitations of the amount of labeled probe were overcome by continuous mixing, however, the Cy3 and Cy5 channels showed the same relative increase in signal intensity, indicating no difference in the diffusional properties of Cy3- and Cy5-labeled cDNAs. Also, the molecular masses of Cy3- and Cy5-labeled cDNAs were found to be the same. Therefore the finding of a difference in the diffusional distances of Cy3- and Cy5-labeled cDNAs is likely due to the use of a single low-end intensity ratio cutoff for both channels, which, in perspective, is inappropriate in view of the differences in detection limits between the two dyes. This has important implications for the analysis of array data for rare transcripts at or near the threshold detection limit. Because Cy5 may be more susceptible to random signal fluctuation at low signal intensity, weakly expressed transcripts visualized in the Cy5 channel may be erroneously considered overexpressed. Even with a dye-swap design to account for the inherent dye bias, random noise fluctuations cannot be assured to be consistent across different hybridizations.

It seems practical and sensible, then, to consider the Cy3 value for diffusion distances as the conservative estimate for static hybridizations. Additionally, consider two replicate target DNA spots that each attract labeled probe from within an area having a 3.8-mm radius. To eliminate competition for labeled probe, a distance of two radii must be established between these two replicate spots. It is therefore our recommendation to place replicate spots $>$7.6 mm apart, ensuring that no competition for labeled probe will occur in either channel.

The overall hybridization process is indeed diffusion-limited, as indicated by a 15–20% increase in signal intensities in both channels with continuous agitation. Depending on the gap height, probe species are quickly depleted perpendicular to the array surface, leaving only lateral diffusion as the source of probe for a large time fraction of a typical 16-h hybridization. This was also supported by the models by Pappaert et al. showing that the time for the shift from reaction to diffusion-limited hybridization is a function of the gap height (10). For static hybridizations, increasing the coverslip gap height will increase probe availability and signal sensitivity, as greater advantage is taken of diffusion perpendicular to the slide surface and, in effect, delaying the shift to a diffusion limited hybridization. Although this may necessitate dilution of limited cDNA, which decreases the driving force for diffusion, an overall net increase in the volume available for a given target is anticipated, resulting in an increase in the amount of hybridized probe. Indeed, experiments in our laboratory have shown that to be the case (13). During the process of validating and optimizing the arrays utilized in this study, several sets of hybridizations were performed where a constant mass of probe DNA was hybridized under coverslips of different heights (25 μm and 57 μm), thereby generating different hybridization volumes (35 μl and 75 μl). A scoring system was then used to compare the background subtracted signal intensities in the small and large gap height hybridizations. The larger, 57-μm gap height consistently showed statistically improved signal in both channels (despite more dilute probe concentration) (13). Therefore, it appears that the larger gap height provides for more probe diffusion perpendicular to the array surface, which more than compensates for the decrease in probe concentration used to achieve the larger gap height.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

The authors thank Keith Alsaker for his assistance, and also the Keck Biophysics Facility.

This work was supported by a National Science Foundation grant (BES-0331402) and a National Institutes of Health/National Institute of General Medical Sciences Biotechnology Training grant (T32-GM08449-11) fellowship to J.R.B.

REFERENCES

1. Baptista, C. S., R. Z. Vencio, S. Abdala, M. P. Valadares, C. Martins, C. A. de Braganca Pereira, and B. Zingales. 2004. DNA microarrays for comparative genomics and analysis of gene expression in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 138:183–194.
2. Cassat, J. E., P. M. Dunman, F. McAleese, E. Murphy, S. J. Projan, and M. S. Smeltzer. 2005. Comparative genomics of *Staphylococcus aureus* musculoskeletal isolates. *J. Bacteriol.* 187:576–592.
3. Raghuraman, M. K., E. A. Winzeler, D. Collingwood, S. Hunt, L. Wodicka, A. Conway, D. J. Lockhart, R. W. Davis, B. J. Brewer, and W. L. Fangman. 2001. Replication dynamics of the yeast genome. *Science*. 294:115–121.
4. Iyer, V. R., C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*. 409:533–538.
5. Weinmann, A. S., and P. J. Farnham. 2002. Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods*. 26:37–47.
6. Odom, D. T., N. Zizlsperger, D. B. Gordon, G. W. Bell, N. J. Rinaldi, H. L. Murray, T. L. Volkert, J. Schreiber, P. A. Rolfe, D. K. Gifford, E. Fraenkel, G. I. Bell, and R. A. Young. 2004. Control of pancreas and liver gene expression by HNF transcription factors. *Science*. 303:1378–1381.
7. Ku, W. C., W. K. Lau, Y. T. Tseng, C. M. Tzeng, and S. K. Chiu. 2004. Dextran sulfate provides a quantitative and quick microarray hybridization reaction. *Biochem. Biophys. Res. Commun.* 315:30–37.
8. Chan, V., D. J. Graves, and S. E. McKenzie. 1995. The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophys. J.* 69:2243–2255.
9. Gadgil, C., A. Yeckel, J. J. Derby, and W. S. Hu. 2004. A diffusion-reaction model for DNA microarray assays. *J. Biotechnol.* 114:31–45.
10. Pappaert, K., P. Van Hummelen, J. Vanderhoeven, G. V. Baron, and G. Desmet. 2003. Diffusion-reaction modelling of DNA hybridization kinetics on biochips. *Chem. Eng. Sci.* 58:4921–4930.
11. Tomas, C. A., K. V. Alsaker, H. P. J. Bonarius, W. T. Hendriksen, H. Yang, J. A. Beamish, C. J. Parades, and E. T. Papoutsakis. 2003. DNA-array based transcriptional analysis of asporogenous, non-solventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *J. Bacteriol.* 185:4539–4547.
12. Alsaker, K. V., T. R. Spitzer, and E. T. Papoutsakis. 2004. Transcriptional analysis of spo0A overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *J. Bacteriol.* 186:1959–1971.
13. Alsaker, K. V, and E. T. Papoutsakis. 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. *J. Bacteriol.* 187:7103–7118.
14. Nölling, J., G. Breton, M. V. Omelchenko, K. S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, and D. R. Smith. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183:4823–4838.
15. Paredes, C. J., I. Rigoutsos, and E. T. Papoutsakis. 2004. Transcriptional organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Res.* 32:1973–1981.
16. Mujumdar, R. B., L. A. Ernst, S. R. Mujumdar, C. J. Lewis, and A. S. Waggoner. 1993. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug. Chem.* 4:105–111.
17. Rousseeuw, P. J. 1984. Least median of squares regression. *J. Am. Stat. Assoc.* 70:871–881.
18. Middleman, S. 1998. An Introduction to Mass and Heat Transfer: Principles of Analysis and Design. John Wiley & Sons, New York.
19. Bird, R. B., W. E. Stewart, and E. N. Lightfoot. 1960. Transport Phenomena. John Wiley & Sons, New York.
20. Schena, M. 2000. Microarray Biochip Technology. Eaton Publishing, Natick, MA.