Article

Nonspecific Hybridization Scaling of Microarray Expression Estimates: A Physicochemical Approach for Chip-to-Chip Normalization

Hans Binder, Jan Bru#cker, and Conrad J. Burden

J. Phys. Chem. B, **2009**, 113 (9), 2874-2895• DOI: 10.1021/jp808118m • Publication Date (Web): 09 February 2009

Nonspecific Hybridization Scaling of Microarray Expression Estimates: A Physicochemical Approach for Chip-to-Chip Normalization

Hans Binder,*,† Jan Brücker,† and Conrad J. Burden‡

Interdisciplinary Centre for Bioinformatics of Leipzig University, D-4107 Leipzig, Haertelstrasse 16−18, Germany, and Centre for Bioinformation Science, John Curtin School of Medical Research and Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia

Received: September 12, 2008; Revised Manuscript Received: November 25, 2008

The problem of inferring accurate quantitative estimates of transcript abundances from gene expression microarray data is addressed. Particular attention is paid to correcting chip-to-chip variations arising mainly as a result of unwanted nonspecific background hybridization to give transcript abundances measured in a common scale. This study verifies and generalizes a model of the mutual dependence between nonspecific background hybridization and the sensitivity of the specific signal using an approach based on the physical chemistry of surface hybridization. We have analyzed GeneChip oligonucleotide microarray data taken from a set of five benchmark experiments including dilution, Latin Square, and "Golden spike" designs. Our analysis concentrates on the important effect of changes in the unwanted nonspecific background inherent in the technology due to changes in total RNA target concentration and/or composition. We find that incremental changes in nonspecific background entail opposite sign incremental changes in the effective specific binding constant. This effect, which we refer to as the "up−down" effect, results from the subtle interplay of competing interactions between the probes and specific and nonspecific targets at the chip surface and in bulk solution. We propose special rules for proper normalization of expression values considering the specifics of the up−down effect. Particularly for normalization one has to level the expression values of invariant expressed probes. Existing heuristic normalization techniques which do not exclude absent probes, level intensities instead of expression values, and/or use low variance criteria for identifying invariant sets of probes lead to biased results. Strengths and pitfalls of selected normalization methods are discussed. We also find that the extent of the up−down effect is modified if RNA targets are replaced by DNA targets, in that microarray sensitivity and specificity are improved via a decrease in nonspecific background, which effectively amplifies specific binding. The results emphasize the importance of physicochemical approaches for improving heuristic normalization algorithms to proceed toward quantitative microarray data analysis.

1. Introduction

Gene expression profiling using microarrays has become a popular technique in molecular biology with a range of applications that benefit from the large-scale estimation of transcript abundance.[1] This method is based on the hybridization of labeled RNA prepared from samples of interest with gene-specific oligonucleotides on the arrays. It estimates the expression levels of tens of thousands of genes in one measurement, giving a snapshot of transcriptional activity of a cell at the time of RNA extraction.

Quantitative and predictive biology requires the ability to quantify gene expression as absolute transcript abundances, or relative changes in transcript abundances between treatments, if functional interactions between genes are to be understood.[2] Ideally, the raw data provided by microarrays in the form of fluorescence intensity measurements must be converted to an expression measure which is linearly related to transcript abundance, and which is normalized to a common unit of measurement to allow for chip-to-chip comparisons. Existing expression measures, which are predicated primarily on statistical principles, suffer from a number of shortcomings. In general, they are scaled in arbitrary units depending on the particular preprocessing algorithm which corrects raw probe signals for parasitic effects such as nonspecific hybridization and probe-specific affinities (see ref 3 for a minireview). Expression values are related to transcript concentrations, but the mutual dependence is partly nonlinear and the scaling factor context-sensitive. Common procedures for normalizing expression profiles were originally developed to measure differential gene expression between two or more phenotypes. Typically, existing normalization methods assume that relatively few transcripts vary or that any changes that occur are balanced. As a consequence, changes in expression levels are calculated relative to the behavior of most of the transcripts. This does not reflect absolute changes if global shifts in mRNA populations occur.

Moreover, gene expression values are notoriously subject to high variability resulting in poor scaling of the data. Systematic biases in the signals of particular genes can have severe effects on subsequent interpretations. In the context of quantitative modeling therefore, accuracy (i.e., a small systematic bias) takes precedence over other considerations such as precision (due to random scattering) because otherwise it would be impossible to arrive at accurate parameter estimates such as transcript degradation rates. Some of the available preprocessing methods (e.g., RMA,[4] vsn[5]) are optimized for high precision at the expense of accuracy.

* Corresponding author. E-mail: binder@izbi.uni-leipzig.de. Fax: + +49-341-9716679.
† Interdisciplinary Centre for Bioinformatics of Leipzig University.
‡ Australian National University.

10.1021/jp808118m CCC: $40.75
Published on Web 02/09/2009

To overcome these shortcomings, a thorough understanding of the physicochemical processes involved in microarray technology is required.[8−12] We have recently proposed a new analysis, called the hook method,[8] which enables a linear estimate of transcript abundances across a single chip. The method is based on the competitive two-species Langmuir adsorption isotherm and uses natural-metrics to estimate expression values of a given array. The hook method accounts for the nonspecific background due to cross hybridization of parasitic transcript sequences, for saturation of the probe spots with bound transcripts and for sequence-specific variations of the binding constant, and applies appropriate corrections. Here we apply this method to estimate the degree of specific and of nonspecific hybridization independently for each investigated chip. These data are required to study their mutual dependence to develop appropriate corrections not considered so far.

The purpose of this paper is to tackle the problem of chip-to-chip normalization, aiming at correcting expression values for chip-specific biases to enable direct comparison between chips. The impact of this issue is high because downstream expression analyses depend sensitively on proper scaling of the data. Naively one might assume that fixed experimental conditions, such as performing the measurements on the same chip type with the same preparation protocols, entail invariant binding constants and thus identical scaling factors of the transcript concentration. However, the physicochemical theory describing the subtle interplay of molecular interactions at the chip surface and in bulk solution predicts that the affinity of the transcripts for binding to the probes depends on the quantity and particular composition of RNA targets in the sample.[6,7] Thus, the scaling factor is a sample-specific property, which is expected to vary from preparation to preparation. A recent analysis of hybridization data taken from a special Latin Square spike-in experiment has confirmed the theory,[8] clearly indicating marked changes of the specific binding constant after addition of a complex RNA mixture to the samples. Spike-in experiments are model studies with known transcript concentrations and are very helpful for studying basic principles of microarray hybridization. The transformation of these results into analysis algorithms for typical chip applications with unknown transcript concentrations, however, requires further efforts.

The remainder of the paper is laid out as follows. Section 2 sets out the physicochemical theory and necessary background material on the hook-curve method. A description is given of five benchmark studies, which mimic different experimental situations ranging from simple dilution of the RNA to the addition of a complex RNA cocktail which completely changes background hybridization. In section 3, data from each of the five studies is analyzed in the light of the theory set out in section 2. The discussion in section 4 formulates the quantitative patterns observed in the data analysis as a general principle which we term the "up−down effect": incremental changes in nonspecific binding entail incremental opposite sign changes in specific binding as a result of the interplay of various chemical reactions. A critique of existing normalization methods is given, and a set of rules is set out for the proper normalization of expression values. As a special case we explore the effect of substituting RNA by DNA targets and discuss how this modification affects the scaling of expression measures.

2. Theory and Data Analysis

Surface and Bulk Hybridization on Microarrays. Consider one of the thousands to millions of probe spots on a typical microarray, which we assume to be independent of one another. For the description of its hybridization one has to take into account essentially three reaction partners, the probe oligonucleotides (P) attached on the chip surface, and the specific (S) and nonspecific (N) polynucleotide transcript fragments in the supernatant solution which either completely or partially match the probe sequence, respectively. We consider the following chemical reactions: unimolecular folding (P-fold, S-fold, and N-fold) and pairwise-bimolecular dimerization (P−S, P−N, N−N, S−S, S−N).[7−10] We neglect P−P dimerization. Under equilibrium conditions, the occupancy of the probe oligonucleotides with specific and nonspecific transcripts is governed by the effective binding constants of specific and nonspecific hybridization

$$K^{S} \approx \frac{K^{P-S}(1 + K^{P-fold})^{-1}}{(1 + K^{S-fold} + K^{S-N}[N] + \sqrt{K^{S-S}[S]})}$$

and

$$K^{N} \approx \frac{K^{P-N}(1 + K^{P-fold})^{-1}}{(1 + K^{N-fold} + K^{S-N}[S] + K^{N-N}[N])} \quad (1)$$

respectively.[7] The effective "overall" constants are functions of the individual equilibrium constants of the reactions introduced above as indicated by the respective superscript of the K's on the right-hand side of the equations. The total concentrations of specific and nonspecific transcripts are denoted by [S] and [N], respectively, where all target sequences not complementary to the specific probe sequence are assumed to be represented by a single effective species N. Equation 1 simplifies to

$$K^{h} \approx \frac{K_{A}^{h}}{1 + K_{B}^{h}[N]} \quad \text{with h} = S, N$$

$$K_{A}^{h} = \frac{K^{P-h}}{(1 + K^{P-fold})(1 + K^{h-fold})},$$

$$K_{B}^{h} = \frac{K^{h-N}}{(1 + K^{h-fold})} \quad (2)$$

in the reasonable limit of negligible bulk hybridization due to specific transcripts ($K^{S-S}[S]$, $K^{S-N}[S] \ll 1$). The effective equilibrium constants of specific (h = S) and nonspecific (h = N) hybridization, K_{A}^{h} and K_{B}^{h}, are directly related to the dimerization constants of the respective reaction on the chip surface (subscript A) and in the bulk supernatant solution (subscript B). The folding reactions are inversely related to the effective constant of the respective reaction. Note that the folding and bulk-dimerization reactions reduce the concentration of free specific transcripts and this way decrease the overall specific hybridization constant compared with the respective surface and bimolecular constants, i.e., $K^{S} \le K_{A}^{S} \le K^{P-S}$. Bulk dimerization thus predicts the decrease of K^{N} and K^{S} with increasing concentration of nonspecific transcripts.

The product of binding constant and total concentration of the respective reacting species defines the so-called binding strength. The total binding strengths X^{S} and X^{N} can be written as hyperbolic functions of the surface binding strength for nonspecific hybridization X_{A}^{N}

$$X^{h} \equiv K^{h}[h] \approx \frac{X_A^h}{1 + X_A^N/X_\infty^h} \quad \text{with } X_A^h \equiv K_A^h[h] \quad \text{and} \quad X_\infty^h \equiv \frac{K_A^N}{K_B^h} = \frac{K^{P-N}}{K^{h-N}(1 + K^{P\text{-fold}})} \frac{(1 + K^{h\text{-fold}})}{(1 + K^{N\text{-fold}})} \quad (3)$$

The scaling factor X^h_∞ is defined as the ratio of the nonspecific surface constant to the respective bulk constant. For h = N it specifies the limiting value of the binding strength at infinite background concentration (and binding strength), $X^N|_{[N]\to\infty} = X^N_\infty$. The specific binding strength vanishes under these conditions, i.e., $X^S|_{[N]\to\infty} = 0$.

The behavior of the binding strength for specific binding upon changing [N] depends on the relation between the concentrations of specific and nonspecific targets, [S] and [N]. Let us introduce the S/N ratio as the ratio of the specific and nonspecific binding strengths

$$R \equiv \frac{X^S}{X^N} = \frac{K^S}{K^N}\frac{[S]}{[N]} \approx R_A \frac{1 + X_A^N/X_\infty^N}{1 + X_A^N/X_\infty^S} \quad \text{where } R_A \equiv \frac{X_A^S}{X_A^N} \quad (4)$$

Equation 4 shows that the S/N ratio of the surface reactions, R_A, is scaled nonlinearly by the binding strength of nonspecific surface hybridization, X^N_A.

Special Classes of Probes. Microarray probes are designed to detect specific transcripts via dimerization and optical labeling. One can classify the probes of a chip according to different criteria, for example, in relation to the abundance of the specific transcripts [S] which they are intended to estimate in a series of at least two hybridizations:

(i) "Invariant-expressed" (or "control" probes) probes referring to an invariant concentration of specific transcripts, i.e., [S] = const > 0. Equation 4 applies with X^S_A = const. to these probes. The increment of the specific binding strength between the samples is $\delta \log X^S = \delta \log K^S$.

(ii) "Differentially-expressed" probes referring to a variable concentration of specific transcripts, i.e. [S] = var > 0. The change of the binding strength is $\delta \log X^S = \delta \log K^S + \delta \log[S]$.

(iii) "Dilution" probes are a special case of "invariant-expressed" probes referring to a constant concentration ratio of specific and nonspecific transcripts, i.e., [S]/[N] = const. and R_A= const in eq 4. The latter condition provides $\delta R_A = (\delta \ln X^S_A - \delta \ln X^N_A)R_A = 0$ and thus $\delta \log X^S_A = \delta \log X^N_A$.

(iv) "Empty" (invariant, not expressed) probes with absent specific transcripts [S] = 0 in all preparations.

The respective conditions are realized in special benchmark experiments of the dilution or spike-in design. In the dilution experiment a complex RNA mixture is hybridized to a series of chips in different concentrations, thus leaving the concentration ratio [S]/[N] constant. In the spike-in experiments, transcripts referring to selected probes are added in definite concentrations into the hybridization solution. We refer to these probes as "spiked" probes or "spikes". If no other transcripts are added then the remaining "nonspiked" probes are empty probes. One can, however, add a complex RNA mixture to the hybridization solution containing the spikes to mimic the nonspecific background. In this case, the nonspiked probes are either invariant-expressed or empty probes with respect to the abundance of the respective specific transcripts.

In general, all the experimental conditions comprise the variation of the amount of nonspecific transcripts, [N] = var, between the preparations. The respective increment of the nonspecific binding strength is $\delta \log X^N = \delta \log K^N + \delta \log[N]$. Spike-in experiments of the Latin-square design ensure [N] ≈ const and thus $\delta \log[N] \approx 0$ to minimize the interference between changing nonspecific and specific hybridization (see below).

Our classification of probes so far relies on the "true" abundance of specific transcripts which is known in the benchmark experiments. An alternative classification is based on the decision whether a probe detects specific transcripts or not, i.e., whether the specific probe signal is beyond the detection limit of the method or not. The respective probes are called present or absent by accepted convention originally introduced by Affymetrix.[11] In the present context we will neglect false negative (e.g., expressed but absent due to small transcript concentrations) and false positive (e.g., not expressed but present due to strong cross hybridization) measurements and consider the paired terms "expressed"/"present" and "empty"/"absent" as synonyms.

Correction for Bulk Hybridization. The signal intensities of the probe spots measured in the microarray experiment are governed by the binding strengths for both specific and nonspecific binding.[3,9] Appropriate analysis allows separation of the two contributions (see below). The obtained estimate of the specific binding strength represents a relative measure of transcript abundance scaled by the effective binding constant for specific hybridization (see eq 3). In the previous section, we showed that the probe-specific scaling factor K^S, however, is not a constant but it depends on the binding strength of nonspecific hybridization owing to bulk reactions. Such bulk effects can vary from experiment to experiment and from probe to probe because the level of nonspecific background usually depends in a sequence-dependent fashion on the particular preparation.

On the other hand, the respective surface binding constant K^S_A is independent of bulk effects. The surface binding strength X^S_A thus represents a better choice for comparing expression values between different samples and hybridization conditions because it scales with the specific transcript concentration independently of the background hybridization. In the next step, we therefore rearrange eq 3 to express the surface binding strength as a function of the effective nonspecific binding strength, X^N, which in turn can be estimated from the chip data (see below):

$$X_A^h = X^h C^h \quad \text{with h} = \text{N, S}; \quad C^N \equiv \frac{1}{1 - X^N/X_\infty^N} \quad \text{and} \quad C^S \equiv \left(1 + \frac{X^N/X_\infty^S}{1 - X^N/X_\infty^N}\right) = C^N(1 - X^N(1/X_\infty^N - 1/X_\infty^S)) \quad (5)$$

Here C^S and C^N are the correction factors transforming the (hybridization dependent) overall binding strengths into the respective (hybridization independent) binding strength of surface hybridization, which refers to the overall binding strength in the absence of nonspecific background hybridization.

Let us assume for the sake of convenience $X_\infty = X^N_\infty \approx X^S_\infty$ and thus $C = C^S = C^N$ (see eq 5), or equivalently $K^{S-N} \approx K^{N-N}$ and $K^{S\text{-fold}} \approx K^{N\text{-fold}}$, i.e., similar bulk dimerization and folding constants for specific and nonspecific transcripts (see eq 3).

To estimate C we make use of the special classes of probes defined in the previous section:

(i) Invariant-present probes with constant abundance of the respective transcripts in a series of hybridizations, i.e., $X_A^S \approx CX^S \approx$ const (see (i) above). For the simplest experimental design of two groups referring, e.g., to a reference (#r) and treatment (#t) preparations we get from eq 5 and the condition $X_A^S(\#r) = X_A^S(\#t)$

$$X_\infty^N = X^N(\#r)\frac{(10^{\delta \log X^N} - 10^{\delta \log X^S})}{(1 - 10^{\delta \log X^S})} \quad (6)$$

with

$$\delta \log X^N \equiv \log\left(\frac{X^N(\#t)}{X^N(\#r)}\right) > 0 \quad \text{and}$$

$$\delta \log X^S \equiv \log\left(\frac{X^S(\#t)}{X^S(\#r)}\right) < 0$$

The opposite sign of δ log X^N and δ log X^S is chosen in accordance with eq 3, which predicts that the increase of [N] (and of X^N) is paralleled by the decrease of K^S (and X^S) and vice versa. We will refer to these opposite changes of nonspecific and specific hybridizations as the "up−down effect". Note that for the invariant-present probes we have δ log $X^S = \delta$ log $K^S \equiv \log(K^S(\#t)/K^S(\#r))$; i.e., the change of the specific binding strength is completely given by the change the respective binding constant (see above).

After insertion of eq 6 into eq 5 we get the correction factors for the preparations #r and #t to adjust the specific binding strength

$$C(\#r) \equiv \frac{X_\infty^N}{X_\infty^N - X^N(\#r)} = \frac{10^{\delta \log X^N} - 10^{\delta \log K^S}}{10^{\delta \log X^N} - 1} \geq 1 \quad \text{and}$$

$$C(\#t) \equiv \frac{X_\infty^N}{X_\infty^N - X^N(\#t)} = 10^{-\delta \log K^S} C(\#r) \geq 1 \quad (7)$$

Bulk hybridization reduces the effective binding strength of specific binding ($X_A^S \geq X^S$). Both correction factors therefore adopt values greater than unity. They are simply related to the asymptotic and the actual values of the nonspecific binding strength (see eq 7). Figure 1 shows that $C(\#r)$ and $C(\#t)$ change nonlinearly with the log-difference of the signals, δ log X^S. The inverse relation between the correction factors and δ log X^N can be rationalized by the trend that a larger change of X^N at δ log K^S = const implies the smaller effect of bulk hybridization and thus the smaller value of the correction factor. The correction becomes linear in the limit δ log $X^N \to \infty$: In this situation the signals are scaled by the ratio of the specific binding constants $C(\#t) = K^S(\#t)/K^S(\#r) =$ const and $C(\#r) = 1$. The correction trivially vanishes in the limit δ log $K^S \to 0$ ($C(\#r) = C(\#t) = 1$).

(ii) "Spiked" probes are suited for background correction if one knows the nominal concentration change, δ log[S] = log[S](#t) − log[S](#r) and the increment of the binding strength. Then eq 7 applies with δ log K^S replaced with δ log $X^S - \delta$ log[S].

(iii) For "dilution" probes, eq 4 provides $R_A \approx R =$ const in the limit of the approximation $X_\infty^N \approx X_\infty^S$. The S/N ratio of the total and of the surface binding strengths is consequently invariant for these probes which are therefore not suited for estimating the scaling coefficient C.

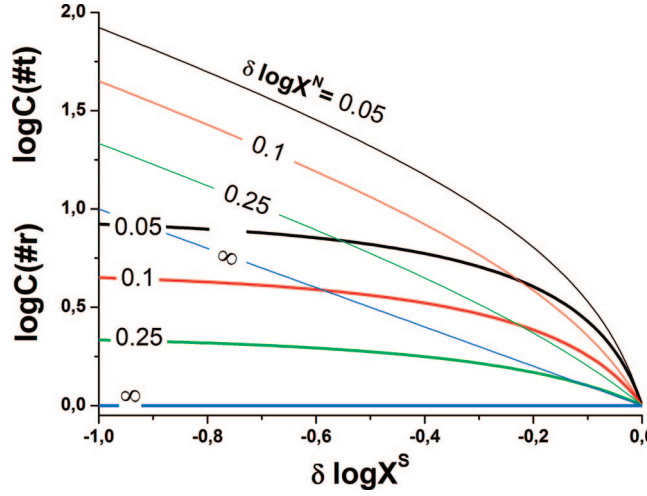

Figure 1. Correction factors $C(\#r)$ (thick lines) and $C(\#t)$ (thin lines) as a function of the log-difference of the specific signal at different log increments of the nonspecific background level. The curves are calculated using eq 7 which assumes the nonlinear scaling of the signal due to bulk hybridization (see eq 3).

(iv) The intensity signal of invariant-absent probes is solely determined by nonspecific hybridization. These probes therefore potentially provide information about the change of background hybridization between the samples, δ log X^N, used in eq 7 (see below).

Hook Analysis. The correction factor requires the determination of the specific and nonspecific binding strengths for the selected probes in each of the considered preparations (see eq 7). For this task we make use of the so-called hook method (see refs 3, 12, and 13 for a detailed description; a brief summary is given as Supporting Information) which applies to microarrays of the GeneChip-type containing pairs of perfect match (PM) and mismatch (MM) probes to estimate the abundance of each transcript. Importantly, this method estimates the required characteristics without using any explicit concentration dependence of selected transcripts (as, for example, realized in spiked-in experiments). Instead, the method uses the mismatched probes as an internal reference for each PM probe to estimate its occupancy by bound transcripts. The hook method transforms the PM and MM probe intensities (I^{PM} and I^{MM}, respectively) to two new coordinates, (Δ, Σ), defined by

$$\Delta = \log I^{PM} - \log I^{MM}, \quad \Sigma = \frac{1}{2}\langle \log I^{PM} + \log I^{MM} \rangle_{set} \quad (8)$$

Subsequent smoothing and correction for sequence-specific effects provides the so-called hook curve in Δ vs Σ coordinates. The mean nonspecific binding strength of the particular hybridization is simply given by the width of the hook curve, $\beta \approx -\log X^{PM,N}$ (one value per chip, see also panel b of Figure 2 for illustration). In the final step of the hook analysis the sequence-corrected probe-level intensity data are corrected for the nonspecific background, for sequence-specific effects, and for saturation and then summarized for each probe set to get one transcript-related estimate of the specific binding strength, log X^S, as the second input values required in eq 7 (one value per transcript). In addition, the method provides a "detection call" per transcript; i.e., it estimates whether the intensity of the transcript-specific intensity is beyond the detection limit or

not. The total fraction of not-detectable transcripts per chips is given as "percent-absent probes" (%N).

Comparison of the Hook Coordinates. The proposed rescaling approach of correcting the specific binding strength for bulk effects relies on the comparison of at least two experimental conditions of the type treatment vs reference, #*t* and #*r*, each of which provides one hook curve. Moreover, each probe set is characterized by its hook-co ordinates (Σ, Δ) which can be compared between both samples to evaluate the increment of the binding constants by visual inspection. Differentiation of the hook equation (see the Supporting Information) provides the increments of the Δ and Σ coordinates as a function of the parameter increments δR (increment of the S/N ratio, see eq 4) and $\delta\beta$ (increment of the width of the hook curve as a measure of the nonspecific binding strength, see eq 6) in the linear hybridization range ($B^{PM} \approx B^{MM} \approx 1$)

$$\delta\Delta \equiv \Delta(\#t) - \Delta(\#r) \approx d_\Delta \delta R$$
$$\delta\Sigma \equiv \Sigma(\#t) - \Sigma(\#r) \approx -\delta\beta + d_\Sigma \delta R \qquad (9)$$

with

$$d_\Delta = \frac{1 - 10^{-\alpha}}{\varphi(R^2)} \propto \frac{1}{\varphi(R^2)},$$

$$d_\Sigma = \frac{R \cdot 10^{-\alpha} + \frac{1}{2}(1 + 10^{-\alpha})}{\varphi(R^2)} \approx$$

$$\frac{R \cdot 10^{-\alpha}}{\varphi(R^2)} + \frac{1}{2} d_\Delta \propto \frac{R \cdot 10^{-\alpha} + \frac{1}{2}}{\varphi(R^2)}$$

$$\varphi(R^2) = (\ln 10)^{-1}(R \cdot 10^{-\alpha} + 1)(R + 1)$$

The parameter α specifies the vertical height of the hook in the absence of saturation. Equation 9 implies, for example, $\delta\Delta = 0$ for $\delta R = 0$. In other words, dilution probes are expected to possess invariant Δ-coordinates upon invariant binding constants (see eq 4). The horizontal shift between the compared hooks is given by the change of the nonspecific background level, $\delta\Sigma \approx \delta\beta = -\delta \log X^N$.

With $\delta R \ln 10 = (\delta \log X^S - \delta \log X^N)R = (\delta \log X^S + \delta\beta)R$ and $\delta \log X^S = 0$, one gets from eq 9

$$\delta\Delta(R) \approx (\ln 10)^{-1} d_\Delta \delta\beta R$$
$$\delta\Sigma(R) \approx -\delta\beta(1 - (\ln 10)^{-1} d_\Sigma R) \qquad (10)$$

the variation of the hook coordinates owing to the change of the nonspecific background under the assumption of invariant specific binding strength.

Banana-Plot Analysis. The hook plot of the PM and MM intensities represents a special variant of a difference vs sum presentation of paired data. Alternatively, one can plot the paired intensity data taken from two experimental conditions of the treatment vs reference design in a similar fashion using the so-called *M* vs *A* plot with the coordinates

$$M \equiv \langle \log I^{PM}(\#t) - \log I^{PM}(\#r) \rangle_{set}$$
$$A \equiv \frac{1}{2} \langle (\log I^{PM}(\#t) + \log I^{PM}(\#r)) \rangle_{set} \qquad (11)$$

The probe intensities are hyperbolic functions of the total binding strength *X* according to the two-species Langmuir model (see, e.g., refs 9, 10, 12, and 14):

$$I^{PM}(\#i) = I_{max} X(\#i)/(1 + X(\#i)) \quad \text{with } i = r, t$$
$$X(\#t) = X^N(\#t)(R + 1) \qquad (12)$$

and

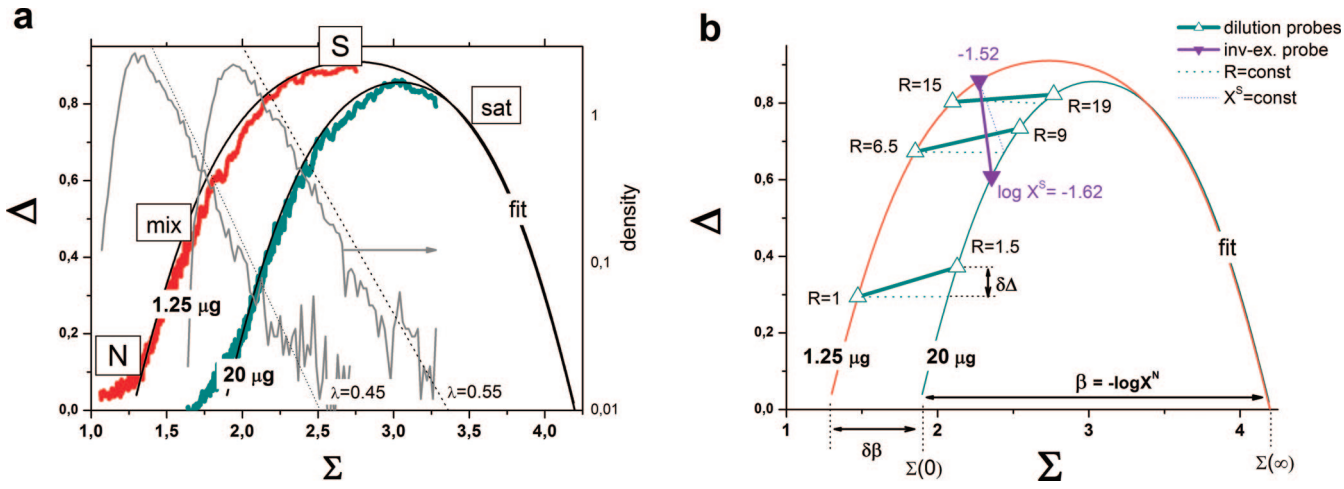

Figure 2. Hook analysis of the dilution experiment: the hybridization regimes (N, mix, S, and sat)) are indicated in part a. Only the hooks of the smallest and largest dilution steps are shown (see ref 13 for the full set of curves). The increasing part of the hook curve shifts to the left by the increment $\delta\beta$ upon dilution of the sample from 20 to 1.25 μg of added RNA. The width of the curve, β, is related to the nonspecific binding strength and thus to the degree of dilution. The curved lines are fits of the hook equation (see Supporting Information and ref 12). The right axis in panel a refers to the frequency distributions which are calculated for the highest and lowest dilution steps. The dotted lines are exponential decays with the decay rate λ given in the figure. Panel b of the figure shows the respective theoretical curves with horizontal lines referring to invariant S/N ratios R = const (thin dotted lines). The open triangles are the positions of selected dilution probe sets in both dilution steps which clearly indicate the slight increase of R with increasing amount of RNA. The solid triangles refer to a control probe set which is spiked in equal concentrations to the samples.

$$X(\#r) = X^{N}(\#t)(R/10^{\delta \log X^{S}} + 1/10^{\delta \log X^{N}})$$

$$R \equiv \frac{X^{S}(\#t)}{X^{N}(\#t)}$$

Insertion of eq 12 into eq 11 provides an analogous equation with the hook equation (see Supporting Information)

$$M(R) = M^{\text{start}} + \log\left\{\frac{(R+1)}{(R\cdot 10^{-\alpha'}+1)}\right\} - \log\left\{\frac{B^{\#t}(R)}{B^{\#r}(R)}\right\}$$

and

$$A(R) = A^{\text{start}} + \frac{1}{2}\log\{(R+1)(R\cdot 10^{-\alpha'}+1)\} - \frac{1}{2}\log\{B^{\#t}(R)B^{\#r}(R)\} \quad (13)$$

with the saturation terms $(B^{\#t}(R) = 1 + 10^{-(\beta'-(1)/(2)M^{\text{start}})}(R+1)$ and $B^{\#r}(R) = 1 + 10^{-(\beta'+(1)/(2)M^{\text{start}})}(R\cdot 10^{-\alpha'}+1)$, and the parameters

$$\alpha' = \delta \log X^{S} - \delta \log X^{N}, \quad \beta' = \frac{1}{2}\delta \log X^{N} - \log X^{N}(\#r)$$

and

$$M^{\text{start}} \approx M(0) = \delta \log X^{N}, \quad A^{\text{start}} \approx A(0) = A(\infty) - \beta' \quad (14)$$

Most importantly, the "vertical" coordinate of the starting point $M(0)$ thus provides the increment of the nonspecific binding strength whereas the vertical dimension α' of the graph in addition depends on the increment of the specific binding strength.

Note that $M(0)$ can be estimated using empty probes, the intensity of which is exclusively determined by nonspecific hybridization. For illustration, let us also examine the course of the $M-A$ plot for invariant-expressed probes with nonnegligible increment of the nonspecific binding strength between the two conditions owing to the up−down effect, i.e., $\delta \log X^{N} > 0$ and $\delta \log X^{S} < 0$ (see for illustration Figures 12 and 13 in the Discussion section). The respective $M-A$ plot starts at positive $M(0)$, turns with increasing A to smaller, possibly even negative M values (note that α' defines the maximum possible vertical difference with respect to $M(0)$) and finally levels off toward the ordinate under the assumption of equal asymptotic intensity values $\Sigma(\infty)$ of both measurements. This behavior resembles the banana-like shape of $M-A$ plots known from numerous chip data analyses.

Benchmark Data. In this study we analyzed the following benchmark data sets:

(a) Dilution-data set taken from the Genelogic dilution experiment:[15,16]

In this experiment cRNA extracted from human liver tissue was hybridized on HG-U95 GeneChips in various dilutions using 1.25, 2.5, 5.0, 7.5, 10.0, or 20.0 μg. Selected transcripts of bacterial RNA for hybridization control were added to all samples in equal concentrations. Each condition was realized in five replicates.

(b) LS-BG and LS+BG data sets taken from the Affymetrix U95 Latin square spike-in experiments with and without nonspecific background:[17,8]

This spiked experiment was designed to study the intensity response of selected probe sets to changes of the respective transcript concentrations. In total 14 transcripts are spiked at 14 concentrations (0, 0.25, 0.5,..., 1024 pM) using a cyclic Latin-square design to ensure a constant total concentration of the spikes (2.05 nM) in each hybridization. The experiment was performed in two modifications, namely, without and with the addition of background. In the latter case a complex cRNA mixture extracted from human pancreas was added in equal amounts to each of the samples to mimic background hybridization. Each hybridization condition was realized in triplicate.

(c) GS-data set taken from the Golden spike experiment:[18]

This large-scale spike experiment mimics a treatment vs reference group scenario to identify genes which are differentially expressed. PCR products from Drosophila Gene Collection were hybridized in different amounts onto Drosophila DrosGenome1 GeneChip arrays in triplicate for each group. Particularly, a total of 3860 individual cRNAs were divided into "invariant present" (2551) and "differentially expressed" (1309) sets. The latter sets were spiked in with differing concentrations between the "treatment" and "reference" samples whereas the former sets are present at identical concentration in each sample. The amount of cRNA material used for hybridization was 3.23 μg (referring to invariant present probes) and 5.38 μg (differentially expressed probes) in both the reference and treatment samples, and additionally 7.93 μg (differentially expressed probes) in the treatment sample. In total, the reference and treatment samples contain 8.97 and 16.54 μg cRNA, respectively. Out of the 14 116 probe sets on the Drosophila DrosGenome1 array, 10 131 probe sets are called "empty" (nonspikes) because they are not assigned to any of the added cRNA spikes. In addition, 20 μg of unlabeled poly-C RNA was added to the reference sample to adapt the amount of RNA to that of the treatment sample.

(d) RNA/DNA data set taken from the DNA/RNA spike experiment:[19]

This experiment uses a baseline RNA specimen extracted from a human T-cell leukemia-derived Jurkat cell line with or without an added set of spikes comprising hemoglobin transcripts HbA1, HbA2, and HbB. The experiment was performed in two modifications (each in triplicate), using either the standard in vitro transcription protocol to produce cRNA targets or using an isothermal protocol to produce cDNA targets (see ref 19 and references cited therein). The cRNA and cDNA sample solutions of nominally identical composition were hybridized onto human genome HG-U133 2.0 arrays with 22 277 probe sets, nine of which were intended to detect the spikes.

3. Results

Benchmark Experiments with Variable Nonspecific Background. In this study, we analyze microarray data taken from five benchmark experiments to investigate the effect of changing nonspecific "background" level on the specific binding constant which scales the expression degree. Table 1 summarizes and compares selected array and sample characteristics of these experiments which are described below. A schematic overview of the essential components contributing to the respective probe signal is given in the Supporting Information.

In the experiments of the *spiked-in with complex background*-type, a limited number of transcripts (the "spikes") matching selected probes (the "spike probes") were mixed with a complex RNA cocktail of unspecified composition and subsequently hybridized onto microarrays using a Latin-square design to ensure a constant total concentration of the spikes in all preparations. Each spike constitutes the specific transcript of concentration $[S]_{sp}$ which binds to the respective spike probe with binding strength $X^{S}_{sp} = K_{sp}{}^{S}[S]_{sp}$. Each spiked transcript also constitutes a nonspecific transcript for all other probes not matching its sequence. Our model pools all not-matching spikes

TABLE 1: Chip, Sample, and Hybridization Characteristics of the Benchmark Experiments

experiment	dilution	Latin-square		Golden spike		RNA/DNA spike			
		−BG	+BG			RNA		DNA	
design[a]	dilution	Latin-square		treat vs ref		treat vs ref		treat vs ref	
this study	−	#*r*	#*t*	#*r*	#*t*	#*r*	#*t*	#*r*	#*t*
Array Characteristics									
chip		HGU95A		DrosGenome 1		HG133A2.0			
# psets		12626		14010		22277			
Sample Characteristics									
complex BG	yes	no	yes	no		yes			
no. spikes	<10	14	14	3860		9		9	
no. diff-expr.	∼7000	∼3000		1309		9		9	
no. inv-expr.	<10	14		2551		∼17000		∼17000	
no. empty[b]	∼5000	∼12500	∼9500	10131		∼5000		∼5000	
RNA/μg[c]	1.25 − 20	<1 μg (spikes)[d]		8.97 + 20[e]	16.54	not specified			
Hybridization Characteristics									
PM/MM gain (log *s*)	1.0 ± 0.05	0.89 ± 0.04	0.89 ± 0.04	1.09 ± 0.02	1.05 ± 0.03	1.12 ± 0.03	1.05 ± 0.02	0.95 ± 0.03	0.93 ± 0.03
% N (absent probes)	41 ± 5	>98	75 ± 10	72 ± 2	74 ± 2	24 ± 2	33 ± 2	22 ± 1	23 ± 2
log X^{N}_{∞}	−2.2	−2.4 ± 0.1		−1.6 ± 0.05		−2.7 ± 0.05		−2.8 ± 0.05	
log X^{N}(#*r*)	<−2.2	−3.0 ± 0.1		−2.4 ± 0.05		−3.0 ± 0.05		−3.0 ± 0.05	
δ log X^{N}		+0.5 ± 0.1		+0.25 ± 0.05		+0.2 ± 0.05		0.03 ± 0.05	
δ log K^{S}		−0.7 ± 0.2		−0.07 ± 0.05		−0.3 ± 0.05		−0.06 ± 0.05	
C(#*t*)/*C*(#*r*)		3−8		1.1		2		1.1	

[a] Dilution experiment (Dil), Latin-square (LS), treatment vs reference (#*t* vs #*r*). [b] Invariant spikes are added in equal concentrations to the #*r* and #*t* samples whereas the concentration of the variable spikes is changed. [c] The hybridization volume is 220 μL; the amount of RNA thus refers to concentrations in the range of 5 pg/μL (1 μg) to 90 pg/μL (20 μg). [d] The total concentration of spikes is 2.05 nM. [e] 20 μg of unlabeled poly-C RNA was added to the #*r* sample.

into one effective species which binds to a selected probe in a sequence-specific fashion with the mean binding strength $X^{\mathrm{N}}_{\mathrm{sp}} = K_{\mathrm{N}}[\mathrm{N}]_{\mathrm{sp}}$. Also the transcripts of the complex background split into specific and nonspecific ones depending on whether a probe matches their sequence or not. The respective binding strengths are $X^{\mathrm{S}}_{\mathrm{bg}} = K^{\mathrm{S}}_{\mathrm{bg}}[\mathrm{S}]_{\mathrm{bg}}$ and $X^{\mathrm{N}}_{\mathrm{bg}} = K_{\mathrm{N}}[\mathrm{N}]_{\mathrm{bg}}$, respectively. The nonspecific contributions add together giving rise to the overall nonspecific, probe-specific binding strength $X^{\mathrm{N}} = \langle X^{\mathrm{N}}_{\mathrm{sp}} + X^{\mathrm{N}}_{\mathrm{bg}} \rangle_{\mathrm{N}}$ where the angular brackets denote averaging over all relevant nonspecific fragments. The total binding strength combines the specific and nonspecific contribution, i.e., $X_{\mathrm{sp}} = X^{\mathrm{S}}_{\mathrm{sp}} + X^{\mathrm{N}}$ for the spike probes and $X_{\mathrm{bg}} = X^{\mathrm{S}}_{\mathrm{bg}} + X^{\mathrm{N}}$ for the remaining nonspike probes. The intensity of the probes is given to a good approximation by the Langmuir adsorption model which predicts the hyperbolic function[9,10,14]

$$I_{\mathrm{p}} = I_{\max}\frac{X_{\mathrm{p}}}{1 + X_{\mathrm{p}}} \quad \text{with p} = \mathrm{sp}, \mathrm{bg} \tag{15}$$

The maximum intensity at saturation, $I_{\max}$, is approximated as a common constant for all probes of the chip. In general, $I_{\max}$ is a probe-specific value because of post-hybridization washing of the microarrays which removes less strongly bound MM probes more efficiently than the PM. Preliminary results on the effect of washing, however, show that this effect is relatively weak and eq 15 well approximates the intensity response of the probes in the relevant range of transcript concentrations.

The studied microarrays of the GeneChip-type use perfect match (PM) and mismatch (MM) probes. Each MM possesses a reduced specific binding strength compared with the respective PM owing to the mismatched middle base. The joint processing of both intensities allows us to judge the presence of the respective specific transcript in the studied sample using algorithms such as MAS5[11] or hook which use the MM as an internal reference for the PM (see above). Abundant transcripts with specific binding strengths exceeding a certain detection limit which is comparable in magnitude with the respective background contribution, i.e., $X^{\mathrm{S}} > X^{\mathrm{N}}$, are usually called present (and absent otherwise).

The *spiked without background* Latin square experiment exactly reruns the "spiked with background" experiment, however, without addition of the complex RNA mixture. As a consequence, the respective mean nonspecific binding strength considerably decreases by the factor 0.5 ($10^{\log(0.5)}$) and nearly all nonspiked probes are detected absent: The absent rate is 41% in experiments with the complex background, and greater than 98% in those without (see Table 1 and the respective scheme in the Supporting Information). Cross hybridization of the spikes to most of the nonspike probes can be virtually neglected in this special case.

In the benchmark experiments the concentration and/or concentration changes of selected transcripts in the different preparations are explicitly known and can be used to study methodical issues. For the two types of spiked experiments, comparing the specific binding strength of the spiked probes referring to transcripts of the same concentration in both preparations enables us to study the effect of the nonspecific background level on the specific binding constant. These spiked probes of constant specific transcript concentration in both preparations are designated "invariant-expressed" probes (see the section 2).

The *Golden spike experiment* provides another, alternative access to study the relation between nonspecific and specific hybridization: It is designed in analogy with the Latin-square spike-in experiment without complex background. The number of spikes is, however, markedly increased from about a dozen

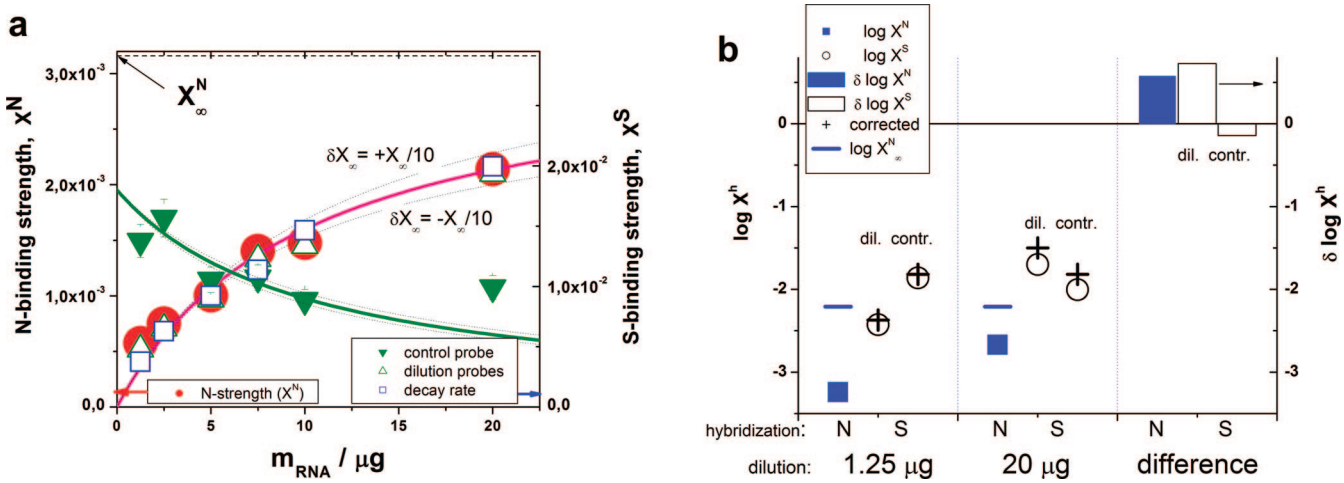

Figure 3. Nonspecific and specific binding strengths as a function of the amount of added RNA in the dilution experiment (part a): X^N was estimated from the width of the hook curves (solid circles). Three options are used to estimate X^S as the mean value over a selection of 294 dilution probe sets with S/N ratios $7 < R < 13$ in the first dilution step (open triangles, selections from other R intervals provide virtually identical results); using the decay constants of the frequency distributions, $X^S = \lambda X^N$ (squares); and using the control probe (AFFX-BioC-3_at, solid triangles). The curves are theoretical functions using eqs 16 and 17 (see text). Part b compares the binding strengths of the nonspecific background and of the dilution and control probe sets at two dilution steps. Note that X^N and X^S of the dilution probes increase with increasing concentration of RNA whereas X^S of the invariant expressed control probes decreases (see bars). Correction for bulk hybridization (eq 7) increases the specific (crosses) and background (blue horizontal line) binding strengths.

to nearly 4000 (see Table 1) and, moreover, the concentration of more than a thousand of these are, on average, doubled in the "treatment" preparation. This considerable increase of the amount of hybridized RNA is expected to change the degree of nonspecific cross hybridization which, in turn, potentially affects the specific binding strength. The invariant-expressed spikes in both preparations allow direct analysis of this effect.

The *DNA/RNA spike-in experiment* uses a complex background mixture and a few spikes which are added in relatively high amounts to the "treatment" samples, potentially modifying the nonspecific background level compared with the reference samples without the spikes (see below). Invariant-expressed probes can be recruited from the nonspiked probes because of the constant amount of added background-RNA. This experiment attracts special interest because it has been realized in two virtually identical modifications using either cRNA or cDNA targets. These different chemical entities give rise to different interactions at the chip surface (DNA/RNA vs DNA/DNA) and in the bulk solution (RNA/RNA vs DNA/DNA) with different binding constants which us enable to study the resulting consequences for specific and nonspecific hybridizations.

In the *dilution experiment* an RNA mixture of constant composition is diluted in six steps. Selected spike controls (e.g., BioB controls) are added to all samples in equal amounts, so providing a set of invariant-expressed probes. This type of experiment allows us to study how the specific and nonspecific signals depend on the amount of added RNA, and to compare with the respective theoretical predictions.

Dilution Experiment. In the dilution experiment the total RNA concentration is changed in a series of hybridizations. Figure 2 shows the smoothed plot of the sensitivity-corrected PM and MM probe intensities of four dilution steps in Δ vs Σ coordinates (see eq 8). We called this presentation hook curve because of the characteristic shape of the resulting graphs. It can be interpreted in terms of different hybridization regimes: the curve "starts" at small abscissa values with the N range of almost exclusively nonspecifically hybridized probes followed the mix range of steep positive slope. A breakpoint in the course of the curve separates the N and mix range. The amount of probes in the former range defines the fraction of absent probes because their signal is only weakly affected by specific hybridization. The increasing part of the hook curve reflects the progressively increasing fraction of specific hybridization contributing to the intensity of the respective probes. The hook curve reaches its maximum in the S range where the probes become predominantly hybridized with specific transcripts. Subsequently, the curve starts to decrease owing to the onset of saturation (sat. range).

The essential characteristics of this curve can be well fitted using the competitive two-species Langmuir isotherm which predicts the "parabola-like" graphs shown in Figure 2 (see Supporting Information and ref 12): The hook curve spans a certain width between its start and end points, $\beta = \Sigma(\infty) - \Sigma(0)$, which estimates the mean binding strength of nonspecific hybridization of the chosen chip (see Figure 2 and ref 12).

Figure 2 shows that dilution essentially broadens the hook curves. This trend trivially reflects the decreasing extent of nonspecific binding upon dilution (see eq 9). It results exclusively from the shift of the increasing part and of the start point toward smaller abscissa values at virtually invariant position of the end point, which reflects a constant saturation intensity of the probes in the series of measurements.

In part a of Figure 3 we plot X^N and X^S as functions of the diluted mass of RNA in solution in μg, m_{RNA}. This diluted mass is directly proportional to both the nonspecific transcript concentration [N] and the specific transcript concentration [S] for noncontrol probes. For X^N, the theory predicts the hyperbolic function (see eq 3 with h = N)

$$X^N = \frac{a m_{RNA}}{1 + a m_{RNA}/X^N_\infty} \quad \text{with } a \propto K^N_A \tag{16}$$

where a denotes the proportionality constant linking m_{RNA} with $X^N_A = K^N_A[N]$. This function describes well the experimental data with the asymptotic nonspecific binding strength $\log(X^N_\infty) \approx -2.5$, or $X^N_\infty \approx 3 \times 10^{-3}$ (see the full circles in Figure 3). The negative deviation of X^N from the linear increase indicates the

decrease of K^N with increasing amount of RNA owing to bulk hybridization which effectively decreases the available concentration of free transcripts.

Equation 3 (with h = S) predicts the specific binding strength for the special case of dilution ([S] $\propto m_{RNA}$) and invariant-expressed control ([S] $\approx$ const) probes

$$X^S(\text{dilution}) = \frac{R_A a m_{RNA}}{1 + a m_{RNA}/X^S_\infty}$$

$$X^S(\text{control}) = \frac{X^S_A(\text{control})}{1 + a m_{RNA}/X^S_\infty}$$

$$\text{with } X^S_A(\text{control}) = R_A(\text{control}) a m_{RNA} \approx \text{const} \qquad (17)$$

where R_A is defined in eq 4. Figure 3 compares experimental X^S data for selected dilution (open triangles) and control (solid triangles) probe sets with the theoretical curves. A group of 294 dilution probe sets was selected with S/N ratios $7 < R < 15$ and a mean $< R > = 9.3$ in the first dilution step ($m_{RNA} = 20$ μg). Their specific binding strength is estimated using the hook method (see Analysis section) and log-averaged. The obtained mean values follow essentially a hyperbolic function of the same curvature as that for the respective X^N data which justifies the relation $\log(X^S_\infty) \approx \log(X^N_\infty)$ in a first-order approximation (compare eqs 16 and 17). This result is further confirmed by the invariant-expressed control probes, the specific binding strength of which decreases with increasing amount of RNA owing to the depletion of free specific transcripts (solid triangles in Figure 3).

Figure 2 (part a) also shows the frequency density of data points with respect to the Σ coordinate. The right flank of the distributions can be approximated by an exponential decay law as indicated by the dotted lines. Its slope defines the decay constant λ referring to the decrease of the density by 1 order of magnitude.[12] It can be interpreted as the mean S/N ratio of each hybridization. It provides an estimate of the mean specific binding strength of all dilution probes $\langle X^S \rangle_{\text{dilution probes}} \approx \lambda X^N$. These data agree well with the predicted function (see open squares in part a of Figure 3).

Detailed inspection shows that the λ values decrease slightly by about 20% from $\lambda = 0.55$ to 0.45 in the considered dilution range (see part a of Figure 2). This change of the mean S/N ratio indicates subtle differences between the asymptotic binding strengths X^N_∞ and X^S_∞ and/or the variation of the S/N ratio of surface hybridization, R_A = var. Note that $X^N_\infty = X^S_\infty$, R_A = const, and [S]/[N] = const predict R = const and λ = const upon varying [RNA] and m_{RNA} (see eqs 4, 16, and 17).

To demonstrate further the change of the S/N ratio upon dilution we compare the position of the mean hook coordinates averaged over sets of dilution probes referring to S/N ratios $1 < R < 3$ (1199 probe sets, weak expressed probes), $7 < R < 13$ (294 probe sets, medium expressed probes) and $15 < R < 25$ (102 probe sets, highly expressed probes) respectively (see part b of Figure 2). For R = const one expects invariant Δ-coordinates (i.e., $\delta\Delta = 0$ for $\delta R = 0$, see eq 9), i.e. the respective data points are predicted to align horizontally between both hook curves. The results, however, indicate the decrease of R upon dilution by 20−30% in agreement with the respective change of λ.

Note that our approximation (eq 2) explicitly neglects bulk dimerization of the specific transcripts. This reaction is expected to affect the nonspecific binding constant K^N ($\sim 1/[S]$) more strongly than K^S ($(\sim 1/([S])^{0.5})$ upon changing [S] (see eq 1) which possibly explains the observed change of the S/N-ratio. The dotted curves in Figure 2 (part b) refer to the variation of X^N_∞ by $\pm$10%. They illustrate that the consequence is relatively small and therefore application of the proposed approximation seems justified.

Part b of Figure 3 summarizes the essential results of our analysis of the dilution experiment in terms of the comparison of two experimental conditions in a treatment vs reference design (20 μg vs 1.25 μg). The N-binding strength and the S-binding strength of the dilution probes increase with increasing RNA amount whereas the S-binding strength of the invariant-expressed control probe decreases. These opposite changes reflect the up−down effect, i.e., the decrease of K^S upon increasing nonspecific background due to bulk hybridization which is predicted theoretically (see eq 2). The increments of the binding strengths between both preparations provide the limiting value of the N-binding strength, $X^N|_{[N]\to\infty} = X^N_\infty$, which enables calculation of corrected values of the specific binding strength referring to the absence of bulk hybridization, or alternatively, to the respective surface binding strength, $X^S|_{[N]\to 0} = X^S_A$. The values of both, X^N_∞ and X^S_A, exceed the respective actual values of the total binding strengths, X^N and X^S, respectively.

Latin-Square Experiment. Figure 4 (part a) shows the hook curves of microarray data taken from the Latin-square spike-in experiment with and without complex background. The hook of each of these experiments was calculated in two versions using either the nonspiked probes (thick curves, taken from one chip of this series) or the spikes only (open circles, taken from all chips of the respective series).

The addition of complex background increases the amount of total RNA in the sample. As for the dilution series, this change shifts the increasing part of the hook curve to the right toward larger abscissa values at virtually invariant decaying part (compare with Figure 2). The observed narrowing simply reflects the increase of the nonspecific binding strength upon addition of background (see eq 10).

Note also another difference between both hook curves: for the system with complex background the respective curve which was calculated from the nonspiked probes covers virtually the same Σ range as the hook calculated using the spikes only. For the system without background the hook of the nonspikes is, however, considerably narrowed and essentially restricted to the N range of the hook curve obtained from the spikes. The hook algorithm calls the probes in the N range as absent. The respective fraction in the latter system is consequently markedly larger than in the samples with complex background (98% vs 75%, see Table 1). Hence, the complex background reduces the percentage of absent probes as expected because a certain fraction of the nonspiked probes become specifically hybridized by RNA background fragments (see also the respective schemes in the Supporting Information). This trend is also reflected by the larger decay constant of the density distribution which more than doubles from $\lambda = 0.07$ to $\lambda = 0.18$.

To illustrate the effect on the specific binding strength of changing nonspecific hybridization, we replot the theoretical hook curves referring to the two considered systems in Figure 4 (part b) together with the coordinates of selected spikes (see symbols). The removal of the complex background obviously shifts the relative position of the probe sets referring to the same spiked-in concentration away from start point of the hook. This shift can be partly explained by the reduction of X^N which increases the respective S/N ratio R. The respective shift for invariant X^S and K^S was estimated using eq 10) and indicated by the dotted lines between the two hooks in Figure 4 (part b). The observed shifts by far exceed the expected shifts for K^S =

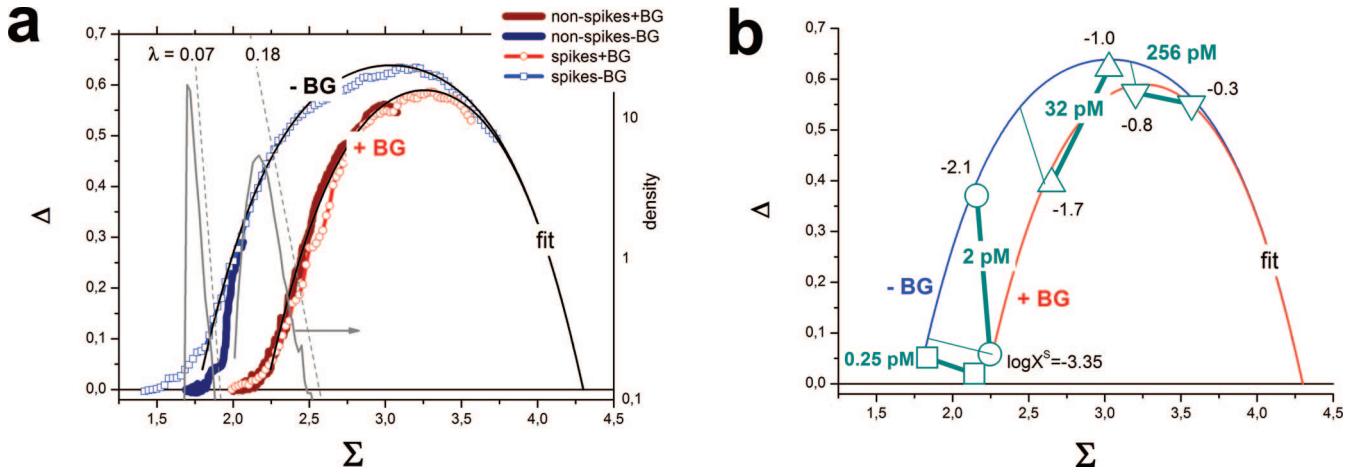

Figure 4. Latin-square spike-in experiment with (+BG) and without (−BG) background: the hooks are separately calculated from the spikes (14 probe sets at 14 concentrations in triplicate) and nonspikes (12 627 probe sets) (part a). Part b shows selected spikes of common concentration in the experiments without and with background by symbols of the same type. The respective specific binding strengths (in log scale) are indicated in the figure. Note that the specific binding strength of the spikes increases upon removing the background from the sample. Iso-X^S lines (δ log X^S = const; thin lines, see eq 10) are shown for comparison.

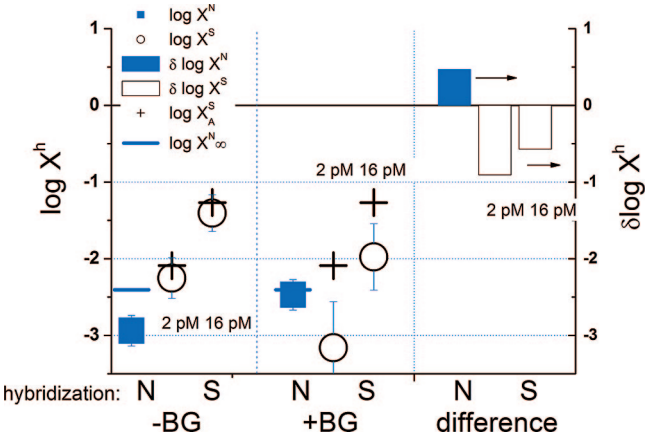

Figure 5. Binding strength of nonspecific (N) and specific (S) hybridization of the LS experiment with (+BG) and without (−BG) complex background. The specific data refer to spike-in concentrations [S] = 2 and 16 pM. The log-differences (−BG) − (+BG) of the binding strengths are shown as bars in the right part of the figure. Note that the N- and S-binding strengths change into opposite directions. The corrected data are shown by crosses (log X^S_A) and horizontal bars (log X^N_∞).

const. Hence, this simple graphical analysis clearly indicates the increase of the specific binding constant in the absence of the specific background.

In Figure 5 we directly compare the nonspecific and the specific binding strengths for two spiked-in concentrations (2 and 16 pM) in the absence and presence of the complex RNA background. This comparison meets the condition for invariant-expressed probes and thus we have δ log $X^S = \delta$ log K^S. The "up−down" effect again becomes obvious by the changes of X^N and K^S in opposite directions.

Note that for the preparation without background the values of the specific binding constants are close to the estimated surface binding constants (compare the open circles with the crosses) because the nonspecific binding strength is markedly smaller than its limiting value X^N_∞ (compare solid squares with the horizontal bar). In the preparation with complex background this relation reverses, i.e., X^N is close to X^N_∞ whereas X^S is distinctly smaller than X^S_A. Hence, both situations are close to the limiting cases of very large and very small nonspecific hybridization, respectively.

The respective decrease of the specific binding strength gives rise to the relative correction factor $C(\#t)/C(\#r) = 10^{(-\delta \log K_S)} = 3-8$ (see eq 7 and Table 1). In other words, bulk dimerization can cause underestimation of the expression degree by up to nearly 1 order of magnitude without appropriate correction.

Golden Spike Experiment. The increase of the amount of spiked-RNA in the "treatment" sample causes the narrowing of the corresponding hook curve compared with that of the "reference" sample (see Figure 6). In this experiment, no complex background was added explicitly. Instead, the increase of the concentration of selected transcripts nearly doubles the amount of spiked-RNA in the treatment sample. Part b of Figure 6 shows "sub-hooks" and probe-density distributions for each sample which have been separately calculated for the nonspikes (empty probes) and spiked probes, respectively. The empty probes accumulate essentially in the N range of the hooks whereas the spikes preferentially cover the range of larger Σ values with virtually constant probe density because of the special experimental design. A considerable amount of the added spikes consequently hybridizes nonspecifically to the empty probes.

Note that the reference sample contains, in addition to the spikes, a certain amount of unlabeled poly-C RNA which was added to level out the total RNA concentrations in both samples (see Table 1). Potentially, it constitutes a special reservoir of nonspecific transcripts, which upon binding "optically" dilutes the respective intensity contribution owing to the absence of labels. As a consequence, the Σ(0) coordinate of the reference sample is expected to shift toward smaller values. This "optical" effect alone is not related to changes of the binding constants. On the other hand, the poly-C reservoir possesses relatively specific binding properties: owing to their homogeneous sequences, the respective RNA fragments are poorly approximated by a unique background contribution to the nonspecific binding strength of all probes, as assumed by our two-species binding model. Instead, they will hybridize, if at all, to a small subset of probes with partly complementary sequences, leaving the remaining majority of probes unaffected. In this case the poly-C background remains essentially indiscernible in the reference sample, having only a small effect on the binding properties, which are governed by the smaller amount of spikes compared with the treatment sample. Both effects, the optical dilution of

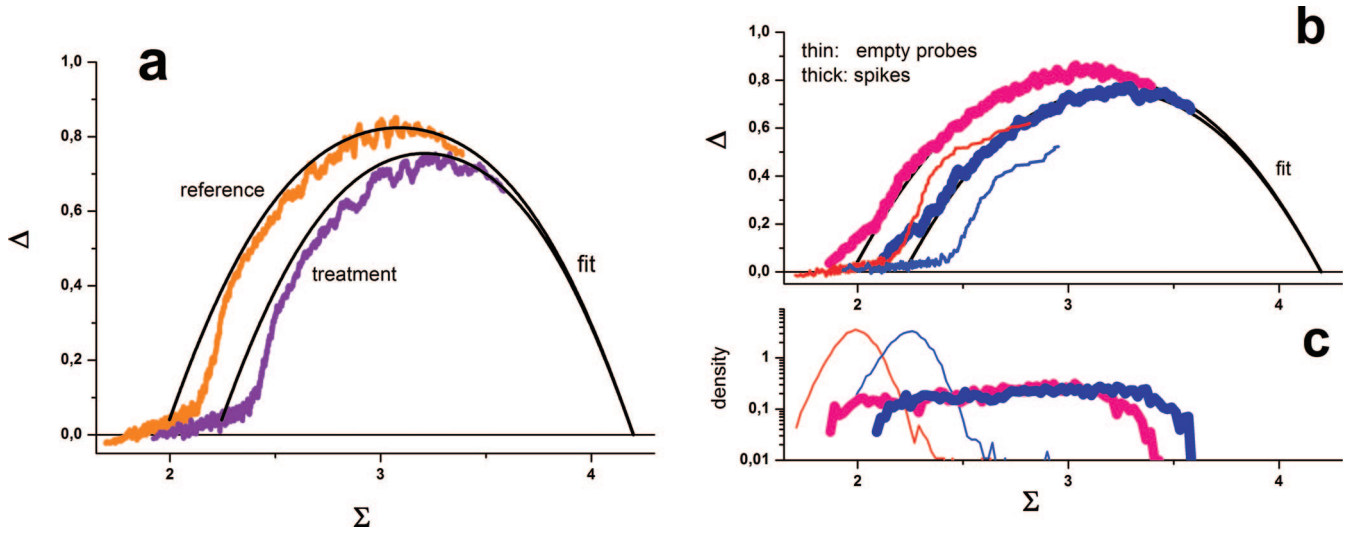

Figure 6. Large-scale spike-in experiment (golden spike): The hooks refer to the reference and treatment conditions. In the treatment sample the amount of added spike-RNA is nearly duplicated. The hooks are calculated for all probe sets (14 020, part a) and separately for spiked (3839) and empty probe sets (part b). The treatment of the samples with additional RNA narrows the hooks. The respective density distributions of the empty and spiked probes indicate that the empty probes preferentially accumulate in the left horizontal part of the hook assigned to absent probes (part c). Contrarily, the spikes preferentially cover the region of the hook to the right from the breakpoint which is assigned to the present probes.

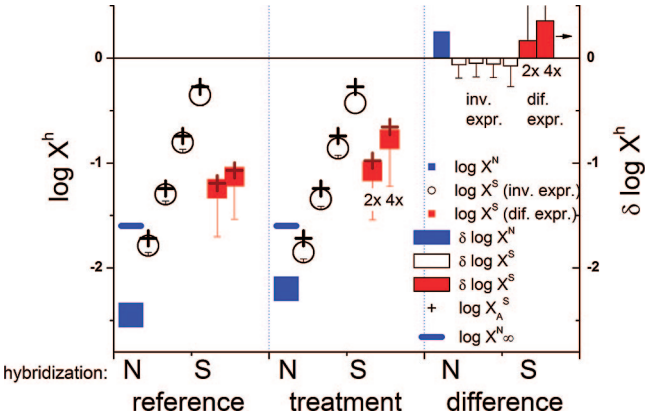

Figure 7. Binding strength of nonspecific (N) and specific (S) hybridization of the reference and treatment samples of the GS experiment. The specific data refer to four collections of invariant expressed control probes of increasing S/N ratio (log $R(\#r) = 0.5 \pm 0.1$; 1.0 ± 0.1; 1.5 ± 0.1; 2.0 ± 0.1) and to two collections of differentially expressed spikes (fold changes: 2× and 4×). The increments of the binding strengths between the treatments and reference samples are shown as bars in the right part of the figure. Note that the N- and S-binding (of invariant expressed probes) strengths change in opposite directions. The corrected data are shown by crosses (specific binding strength) and horizontal bars (nonspecific binding strength).

the background and the inefficient hybridization of the poly-C transcripts, predict the smaller $\Sigma(0)$ coordinate of the reference and virtually invariant $\Sigma(\infty)$ values. Thus, the positions of the start and end points of the hook curve do not enable us to distinguish between these two effects. Optical dilution of the background, on the other hand, will not affect the position of the maximum of the hook because it is determined nearly exclusively by specific binding of labeled transcripts. The shift of the maximum position of the hook plots between the samples observed in Figure 6 lets us conclude that poly-C RNA contributes little to the nonspecific background level. Consequently the increase of the nonspecific background in the treatment sample is induced predominantly by the additional spikes.

Figure 7 shows the binding strengths of nonspecific and specific hybridization in the reference and treatment preparations and the respective increments. The X^S values were calculated for invariant and differently expressed spikes. The former ones were divided into four bins of increasing expression (see legend of Figure 7). The log-increments of X^N and of X^S for the invariant-expressed probes clearly show the up−down effect where the increase of nonspecific hybridization is accompanied by the common reduction of their specific binding constant which is virtually independent of the level of specific binding. For the differently expressed probes the apparent fold change is smaller than the nominal one because of this effect, i.e., $\delta \log X^S(\mathrm{fc}) = \delta \log K^S + \log(\mathrm{fc})$ with $\mathrm{fc} = [S](\#t)/[S](\#r)$. For example, for the 2-fold increased spikes one gets $\delta \log X^S(\mathrm{fc}=2) = 0.2 \pm 0.1$ instead of the nominal value $\log(2) = 0.3$. The difference between both values is roughly given by the decrease of the binding constant obtained from the invariant expressed probes, $\delta \log K^S \cong -0.07$. Note also that the level of nonspecific background due to the spikes is far from saturation (compare blue squares with the horizontal bars in Figure 7).

RNA/DNA Spike Experiment. The last experiment considered in this paper uses a treatment vs control design where both preparations contain a complex background. Altogether nine spikes are added into the treatment samples in concentrations not specified in the original publication.[19] The experiment has been performed in two versions using either cRNA (R) or cDNA (D) as targets for the DNA probes. The comparison of the hook plots reveals essentially three differences between the R and D hybridizations (see Figure 8, compare part a with part b): first, the hooks of the R hybridizations are shifted toward larger Σ values, indicating a generally larger intensity level for the R hybridization; second, the decay constant of the probe-density distribution is larger for the D samples, indicating a more advantageous S/N ratio; and third, the hooks of the #*t* and #*r* samples are shifted relative to each other in the R experiments but not in the D experiments, which seems to be a puzzling difference between both systems.

Figure 9 summarizes the binding strengths of all four preparations (part a) and the respective increments in the treatment vs reference comparisons separately for each chemical entity (R and D); and in the R vs D comparisons separately for the reference and treatment preparations (#*r* and #*t*; part b). The

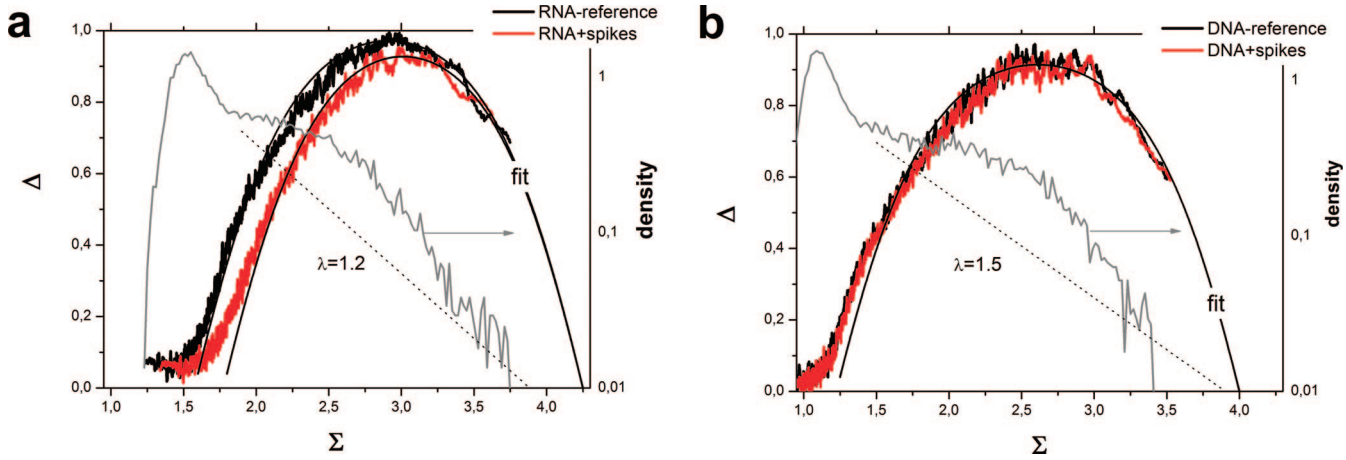

Figure 8. RNA/DNA spike-in experiment: the hooks refer to the hybridization before and after adding the spikes for RNA (part a) and DNA (part b) hybridizations. The mean decay of the density distributions of the reference samples is indicated by the dotted lines (λ is the decay constant).

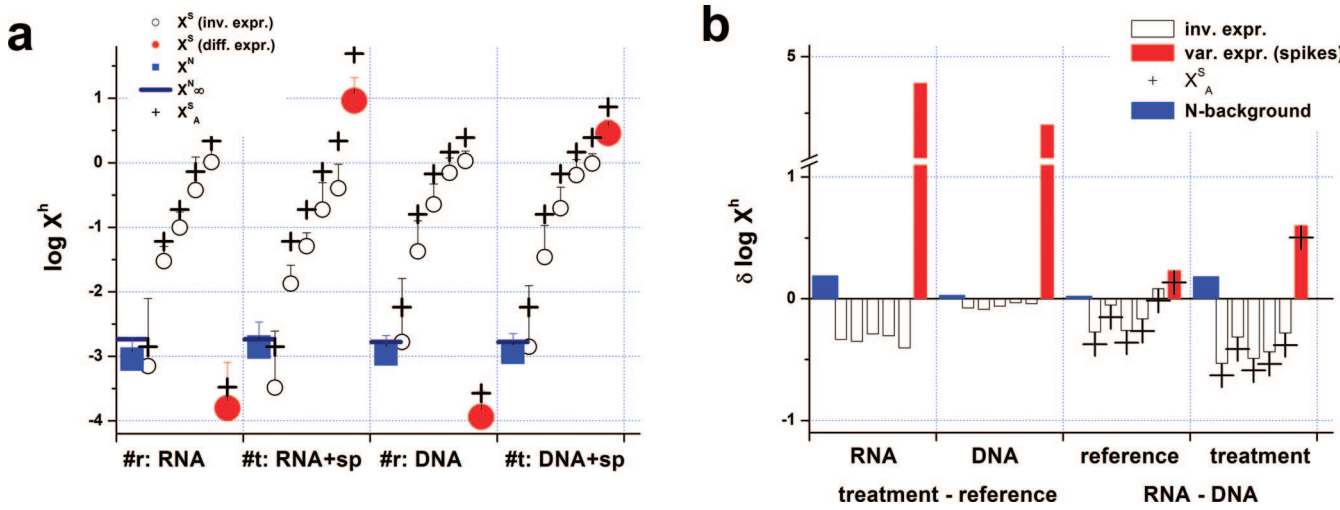

Figure 9. Binding strengths of the background (X^N) and of invariant-expressed control and of differentially expressed spiked probe sets (X^S) before (reference) and after (treatment) adding the spikes in RNA and DNA hybridizations (part a). Part b shows the respective differences of the binding strengths treatment minus reference for RNA and DNA hybridizations and RNA minus DNA for the reference and treatment, respectively. The spikes (differentially expressed) are mean values over all 9 spikes and the five groups of invariant-expressed data are averages over 18 invariant-expressed probe sets each. These groups were chosen from different intervals of X^S to provide representative values ranging from small to large expression values.

N-binding strength of the background was estimated from the width of the respective hook curves. The S-binding strength was separately determined for five groups of invariant-expressed probe sets collected from the nonspikes to cover the range from small to high expression values; and the spikes collected into the group of differentially expressed probes.

As in the previous examples, the addition of targets (spikes) increases the N-binding strength of the background and decreases the S-binding strength of the invariant-expressed control probes. For D targets this trend virtually disappears. Hence, the up−down effect is clearly manifested in the R experiments but hardly detectable for D as already suggested by the comparison of the respective hook curves. The actual nonspecific binding strengths of the latter systems are close to their estimated limiting values, X^N_∞ (compare horizontal bars and squares in the part a of Figure 9). This implies that the given D samples are relatively insensitive to variations of the nonspecific binding strength.

Note that the spikes are virtually absent in the #*r* samples because their specific binding strength is markedly smaller than the respective X^N values. The binding strengths of the few spikes adopt extraordinary large values of $X^S > 3$: in the #*t* samples these transcripts are obviously added in very large amounts. Typical values of the binding strengths rarely exceed unity. The impact of substituting RNA by DNA targets will be discussed below.

4. Discussion

The Up−Down Effect. We studied the effect of changing nonspecific background hybridization on the specific signal. Microarray data taken from different benchmark experiments were analyzed using a simple theoretical approach which expresses the effective reaction constants of specific and nonspecific hybridization as functions of the equilibrium constants of relevant molecular processes such as the folding of target (specific and nonspecific) and probes and dimerizations between probe and targets at the chip surface and between the targets in bulk solution (see eq 1). Part a of Figure 10 shows the expected dependencies of the specific and nonspecific binding strengths upon increasing background

level for invariant expressed probes ([S] = const; eq 3). Accordingly, theory predicts opposite signs for the changes of X^S and X^N. The respective functions are characterized by the limiting "start" value of X^S at vanishing background hybridization, X_A^S (the specific surface binding strength) and the asymptotic value of X^N at infinite background concentration, X_∞^N. The latter value also defines the abscissa value referring to 50% of the asymptotic value.

For further discussion we introduce the reduced binding strengths

$$\xi^N \equiv \frac{X^N}{X_\infty^N}, \quad \xi^S \equiv \frac{X^S}{X_A^S}, \quad \text{and} \quad \xi_A \equiv \frac{X_A^N}{X_\infty^N} = \frac{X_A^N}{X_\infty^S} \tag{18}$$

which rescale the coordinate axes according to

$$\xi^S = \frac{1}{1+\xi_A} \quad \text{and} \quad \xi^N = \frac{\xi_A}{1+\xi_A} \tag{19}$$

and use the log−log representation (see part b of Figure 10). In eq 18 we have used the approximation $X_\infty^N \approx X_\infty^S$ introduced in the discussion following eq 5. The respective log-increments for invariant-expressed probes, e.g., between treatment and reference conditions, $\delta \log \xi \equiv \log \xi(\#t) - \log \xi(\#r)$, then simply become (see part c of Figure 10)

$$\delta \log \xi^S = -\xi^N \delta \log \xi_A \quad \text{and} \quad \delta \log \xi^N = \xi^S \delta \log \xi_A \tag{20}$$

Samples prepared using identical hybridization protocols, target chemistry, and chip types meet the condition of invariant chemical interactions ($\delta \log K^{P-h} = \delta \log K^{h-N} = \delta \log K^{h-fold} = 0$). For the invariant-expressed probes one gets in this case (with $\delta \log[S] = 0$; see eqs 2 and 3)

$$\delta \log \xi^S \approx \delta \log K^S = \delta \log X^S,$$
$$\delta \log \xi^N \approx \delta \log X^N, \quad \text{and} \quad \delta \log \xi_A \approx \delta \log X_A^N \tag{21}$$

i.e., the increments of the reduced binding strengths directly provide the increments of the respective binding constants and strengths.

The relation between the absolute values of the changes of the specific and nonspecific binding strength varies with increasing nonspecific background level as follows (see parts b and c of Figure 10):

(i) at small background levels, $\xi_A \ll 1$, large changes of the nonspecific background induce relatively small changes of the specific binding constant, i.e., $|\delta \log \xi^N| > |\delta \log \xi^S|$;

(ii) the intermediate range is characterized by comparable magnitudes of both processes, $|\delta \log \xi^N| \approx |\delta \log \xi^S|$; and

(iii) upon saturating background level one expects a relative large effect of ξ^S but only small changes of the background, i.e., $|\delta \log \xi^N| < |\delta \log \xi^S|$.

In all the four benchmark experiments studied, we have confirmed that the up−down effect agrees with theory. Two of them, the dilution and the Golden spike data, correspond to situation (i) whereas the Latin-square and the RNA/DNA spiked data refer to situations (ii) and (iii), respectively. The hook plots enable the simple and straightforward identification of the respective increments of the binding strength by visual inspection and subsequent detailed analysis. These values then provide estimates of the specific surface binding strength, X_A^S, and of the limiting nonspecific binding strength, X_∞^N (see horizontal bars and crosses in Figures 3, 5, 7, and 9 and also Table 1).

The magnitude of the latter values was $\log X_\infty^N = -2.8$ to -2.4 for the experiments that use a complex background of "biological" origin extracted from cell lines. It increases by roughly 1 order of magnitude for the Golden spike experiment to $\log X_\infty^N = -1.6$ for unknown reasons. Note, however, that the background level in this experiment is measured by cross hybridization of a relatively large but nevertheless limited number of more than 3000 spikes of predefined and relatively high concentrations to the empty probes. The marked increase

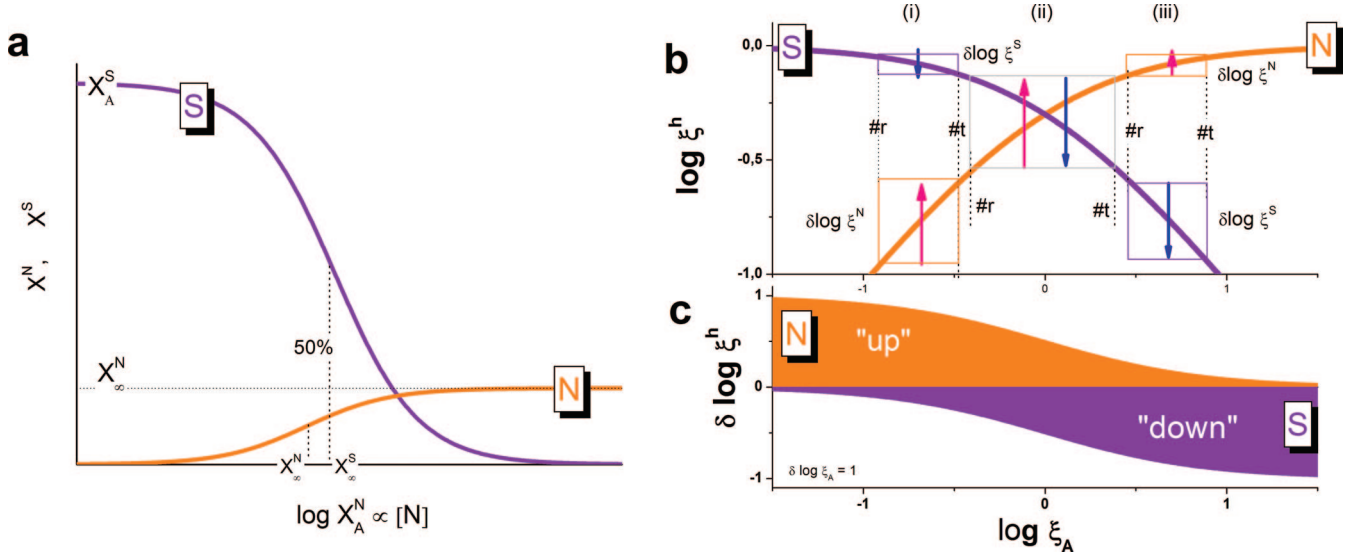

Figure 10. Specific and nonspecific binding strength as a function of the nonspecific background given in units of the surface binding strength of nonspecific hybridization (part a). The curves are calculated using eq 3. Part b shows the same curves in reduced coordinates (see eqs 18 and 19). The arrows illustrate the log-increments of the binding strengths between treatment (#*t*) and reference (#*r*) conditions upon increasing nonspecific hybridization. Note the opposite signs of the changes of the specific and nonspecific values. This up−down effect is illustrated in part c which shows the slope of curves shown in part b (eq 20). The three ranges (i)−(iii) are explained in the text.

of X^N_∞ possibly reflects the smaller propensity of theses transcripts for bulk dimerization in this experiment (see eq 3). This difference shows that the limiting value X^N_∞ in general depends on the composition and heterogeneity of the RNA cocktail used for hybridization.

Vice versa, X^N_∞ scales the up−down effect: Rearrangement of eq 6 and setting $X^N = X^N(\#r)$ in eq 21 gives

$$\delta \log X^S = \log\left(\frac{X^N_\infty - X^N(\#t)}{X^N_\infty - X^N(\#r)}\right) = \log\left(\frac{1 - \xi^N \cdot 10^{\delta \log X^N}}{1 - \xi^N}\right) \quad (22)$$

which explicitly links the increments of the specific and nonspecific binding strengths.

Proper Scaling of Expression Measures. Microarray experiments endeavor to estimate selected transcript abundances from the complete set of intensity responses of the probes across the microarray surface. There are a multitude of different so-called preprocessing methods which aim at correcting raw probe intensities for systematic biases due to parasitic effects such as nonspecific hybridization, sequence-specific binding affinities, and/or saturation (see, e.g., refs 4, 20, and 21−24, and ref 3 for a minireview). These methods transform the probe intensity into an "expression degree" (or "expression index") by inverting the selected intensity function. The obtained measure is at best (e.g., if properly corrected for parasitic effects such as nonspecific background and/or saturation) directly related to the specific transcript concentration (and specific binding strength) as follows

$$E = FK^S[S] \quad (23)$$

where F is a proportionality constant owing to "technical" factors depending, for example, on the scanner settings, array design and the protocol for RNA extraction, preparation, hybridization, and labeling. The differential "expression" is given as the increment of the expression degree (in logarithmic scale) between two conditions which thus additively depends on all considered factors, e.g., $\delta \log E \equiv \log E(\#t) - \log E(\#r) = \delta \log(FK^S) + \delta \log[S]$.

Note that the biologically relevant "expression change" is the increment of specific transcript concentration, $\delta \log[S]$. Hence, differential expression, $\delta \log E$, is potentially biased by $\delta \log(FK_S)$ and therefore error prone.

Any transformation of the expression values which aims at ensuring $\delta \log(FK^S) = 0$ is called normalization. This way normalization makes the expression values of different chip experiments "comparable" in the common scale of transcript abundance. Proper normalization thus removes potential biases due to changes of the scaling of the transcript concentration between the experiments.

It should be mentioned at this point that our hook method estimates the technical factor F, which is simply given by the maximum intensity upon saturation of the probes, i.e., $\log F = \log I_{max} = \Sigma(\infty)$ (see the hook equation in the Supporting Information and the linear part of eq 15, i.e., the $X_p \ll 1$ regime relevant to most intensity measurements within a given experiment). As a consequence, the hook-expression values are scaled in natural units of the specific binding strength, $\delta \log X^S = \delta \log K^S + \delta \log[S]$, which corresponds to $\delta \log F = 0$.

Our analysis clearly shows that nonspecific hybridization affects the specific binding constant and in this way scales the expression degree even at virtually invariant technical conditions (i.e., $\delta \log K^S \neq 0$). The level of nonspecific hybridization depends on the overall expression level of the studied cells and on the RNA yield of extraction and subsequent amplification. Both factors can systematically vary from experiment to experiment. The up−down down effect therefore should be seen as an inherent property of the microarray techniques. It implies the following algorithmic framework and basic rules for its proper correction:

(a) The expression degree (eq 23), i.e., quantities designed to scale linearly with the specific target concentration) and not raw intensities should be used in adequate correction/normalization transformations (see also next section).

(b) Selected probes with special properties such as invariant-expressed probes (and not all probes of the chip or absent probes) provide the value(s) of the expression degree or specific binding strength for proper correction.

(c) Correction applies to the expression degree (and not to the intensity) of all remaining probes. Absent probes with tiny expression values consequently remain virtually uncorrected.

On the basis of (a)−(c) we propose a simple multiplicative correction in a treatment vs reference experimental design. The correction factor rescales the apparent expression degrees into their limiting value referring to the absence of nonspecific background (eq 7). The extension for more than two conditions can be simply realized, e.g., by decomposing the conditions into multiple treatment vs reference pairings. Below we modify our approach using M vs A plots to compare and correct the expression data of paired conditions.

Note that most popular normalization algorithms violate rules (a)−(c). In the next sections we illustrate and discuss the consequences for the obtained expression estimates and substantiate the proposed rules.

Normalization of Microarray Data. In addition to background correction and summarization, normalization is one of the three basic steps of microarray preprocessing which transforms raw intensity data into expression measures. There exist a multitude of normalization methods which, with respect to the processed ensemble of probe data, can be roughly classified into complete-set and subset normalizations. The former ones use the whole intensity information of all probes of the chip whereas the latter ones select a subset of probes suited for normalization. The probe signals of the respective sets are then summarized e.g. by log-averaging in the simplest case, and then leveled between the arrays of the considered series, e.g., by proportional scaling.

Complete-set normalization methods such as Global (based on a trimmed mean summarization[25,20]), LOESS (leveling the local mean[26,27]), Quantile,[28,29] and Zipfs law[30] (both leveling the signal-frequency distributions) normalizations assume that the expression changes are "balanced", i.e., up- and down regulations are virtually equal in magnitude. Violation of this assumption gives rise to biased expression estimates: for example, an unbalanced up regulation becomes counterbalanced by the apparent down regulation of in reality invariant genes. Complete set normalizations validly apply to situations where only a small fraction of genes is differently expressed between the compared arrays.

Subset normalizations aim at avoiding this "center of gravity" bias of the complete-set methods by using either external controls which are spiked onto the arrays with known concentrations[31,32] or by selecting probes of invariant expression as reference for leveling the signals of all probes.[33,23,5] Typically, normalizations using external controls are not very precise

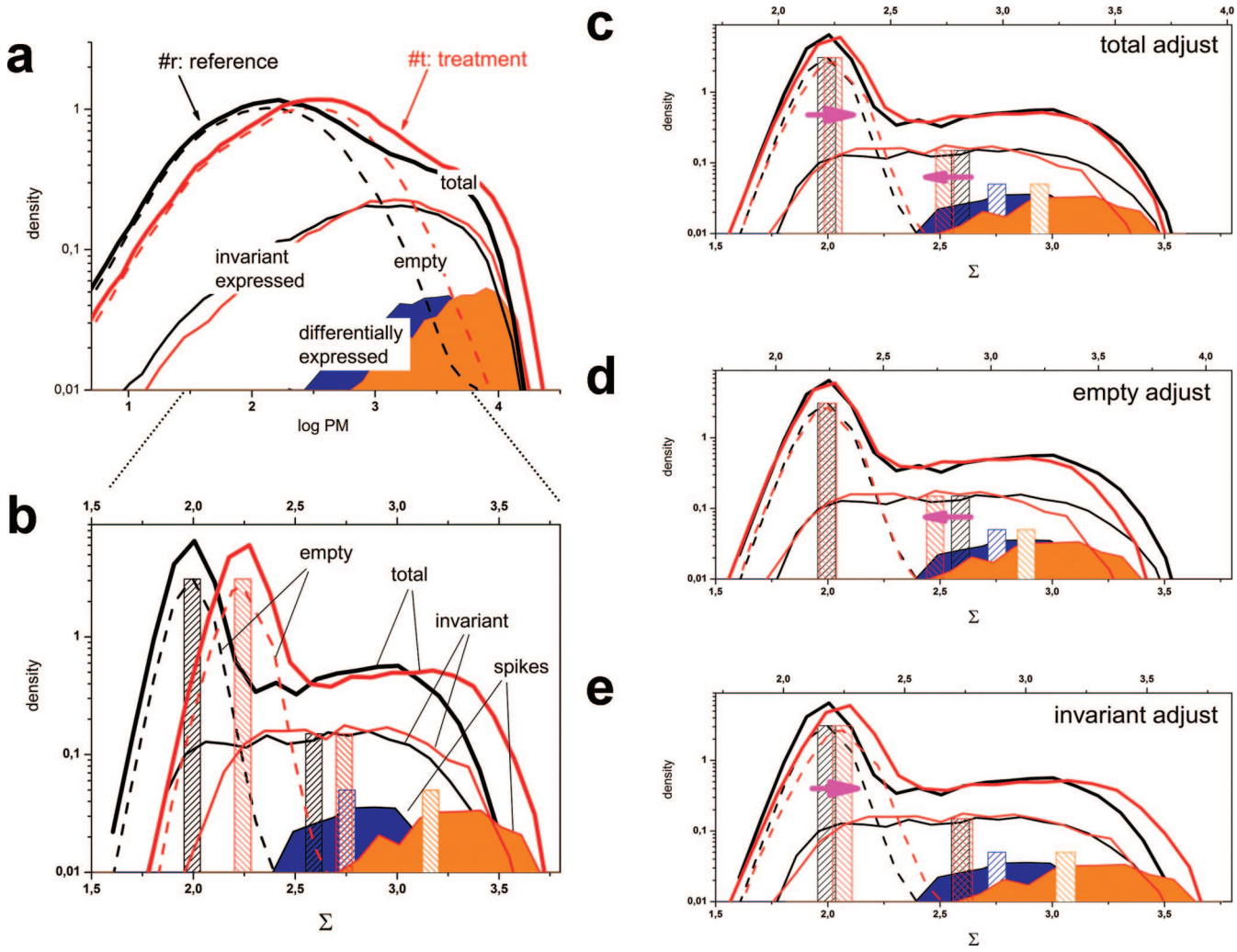

Figure 11. Density distributions of the reference (black lines) and treatment (red lines) probes of the Golden spike experiment. The distributions are calculated for all probes (total) and for subsets of the empty, invariant and differentially expressed probes (see part a for assignments). Part a shows the intensity distributions and part b the distributions along the set-averaged Σ-coordinate. Note that set-averaging considerably narrows the widths of the distributions. The vertical bars indicate the center of gravity of the subdistributions. Three different normalizations are shown in parts c−e: The distributions of the total (part c), the empty (d), or invariant (e) probes of the treatment sample are shifted and scaled in such a way that they align with the respective distribution of the reference sample. Note that none of these transformations aligns all subdistributions. Shifts of the centers of gravity between the samples are indicated by the arrows.

because of the relatively small number of spikes. Invariant set methods pose not only the problem how to identify invariant genes from the probe signals but they are conceptually incorrect: Apparently invariant-expressed probes are in fact differently expressed probes where the change of the true expression is compensated by the change of the specific binding constant due to the up−down effect (see eq 23 with $E = \text{const}$ and $\delta \log(FK^S) = -\delta \log[S] \neq 0$; see also next section).

Most of the normalization algorithms can be applied in a standalone fashion independently of the other preprocessing steps, for example, either before or after background correction. In most applications, normalization directly applies to intensity data to adjust them for subsequent downstream analyses. The rationale behind this approach is the assumption that intensity contributions due to specific and nonspecific hybridization scale in a similar fashion and, consequently, they can be corrected in parallel.

Reference Probe Selection—Pitfalls of Complete and Subset Normalizations. Normalization techniques employed by existing expression indices are heuristic approaches which fail to account explicitly for the underlying hybridization chemistry, and, in particular, for the up−down effect. The resulting consequences are illustrated here using as an example signal distributions of the Golden spike data (Figure 11). In addition to the total distribution of all probes, we also depict the distributions of selected subsets of probes, namely that of the empty (i.e., invariant-absent), invariant-expressed and differentially expressed spiked (referring to a 3.5−4 fold increase upon treatment) probes, together with the center of gravity of the subset-distributions (see the vertical bars).

The distributions of the raw intensities ($\log I^{PM}$) are relatively broad and largely overlap (see part a of Figure 11). To filter out subtle differences, we substitute the probe-level data by the set-averaged Σ-coordinate used also for the hook plots (see eq 8). This transformation significantly narrows the distributions because it partly corrects the data for sequence-specific variations of the probe intensities. Part b of Figure 11 shows that all distributions of the treatment are shifted to the right compared with the reference owing to the stronger nonspecific background (compare also with Figure 6).

Part c (total adjust) mimics complete set normalization methods by shifting and scaling the total distribution of the treatment in horizontal direction until it aligns roughly with that of the reference sample, and then transforming the subset

distributions in the same way. This total-adjust method does not successfully align the subdistributions with each other. In particular, the right flank and the mean of the distribution of invariant expressed genes of the treatment become "overcompensated"; i.e., they are shifted to the left with respect to the respective reference distribution. This difference simply reflects the decrease of the specific binding constant in the treatment sample. As a consequence, invariant-expressed probes appear to be down regulated and the apparent fold change of the spikes underestimates the true value as indicated by the arrow (see also below). Contrarily, the empty (i.e., invariant absent) probes are apparently up-regulated. Hence, both kinds of invariant probes register as being differentially expressed, though in opposite directions.

The panels d (empty adjust) and e (invariant adjust) of Figure 11 mimic subset algorithms which use either the empty (invariant absent) or invariant-expressed probes for normalization. Again we shift and scale the respective subset distribution of the treatment sample along the Σ-axis to match that of the reference and then transform the remaining distributions of the treatment sample in the same way. The results show that the empty adjust transformation also overcompensates, in fact even more than the total-adjust alignment: the mean position of the invariant-expressed probes is clearly down regulated because the up−down effect shifts the distributions of empty and expressed probes in opposite directions. Contrarily, the alignment of the invariant distributions upon empty-adjust transformation undercompensates the signals of the empty probes, the expression of which appears to be up-regulated after treatment.

While the finding that neither complete-set nor subset normalization can accurately adjust all probe intensities may seem disappointing, it is not surprising given the underlying up−down effect which drives specific and nonspecific hybridization in opposite directions. Consequently, any parallel scaling of absent and present probes must fail. Instead, both kinds of probes must be processed differently, as proposed in the previous section by rules (b) and (c). Accordingly, invariant expressed probes are used for estimating a correction factor which is then applied to all present probes. Absent probes are simply filtered out from further analysis.

Note that the apparent expression of both empty (i.e., invariant absent) and invariant ("truly") expressed probes does not change between the preparations. Available invariant subset-normalization methods[5,23,33] do not in general remove absent probes from the controls, and the consequent mixing of absent and present invariant probes renders these methods prone to errors.

Normalization of Intensity or Expression Data? Probe intensities are superpositions of contributions due to specific and nonspecific hybridization, each of which are differently affected by the up−down effect. Differentiation of eq 15 provides the respective intensity increment

$$\delta \log I \equiv \log I(\#t) - \log I(\#r) = \frac{x^{S}\delta \log X^{S} + (1 - x^{S})\delta \log X^{N}}{1 + X} \quad (24)$$

where $x^{S} = X^{S}/X$ denotes the fraction of specific hybridization and $X = (X^{S} + X^{N})$ is the total binding strength. Equation 24 shows that each intensity change between preparations represents a weighted superposition of the opposite up and the down shifts. The intensity change of weakly expressed low-intensity probes with $x^{S} \ll 1$ is consequently dominated by the shift of the nonspecific binding strength $\delta \log X^{N}$, whereas strongly expressed probes with $x^{S} \cong 1$ are mainly affected by the opposite sign shift of $\delta \log X^{S}$.

This effect becomes evident in Figure 11 (see, e.g., part c, total adjust) if one compares the widths of the distributions of the invariant expressed probes between the #*t* and #*r* samples: The relative narrowing of the #*t* distribution reflects the progressive compensation of $\delta \log X^{N} > 0$ by $\delta \log X^{S} < 0$ with increasing abscissa value. As a consequence, the right high-intensity flank of the distribution shifts to a less degree than the left, low-intensity flank.

Intensity-based normalizations therefore tend to underestimate the extent of correction because the up and the down contributions partly compensate each other. Proper normalization consequently requires foregoing correction of the intensity for the nonspecific background and direct scaling of the expression measures as suggested by rule (a) (see above). In the next paragraph, we illustrate the importance of this rule together with a simple correction approach based on the *M* vs *A* plot of the data.

M vs A Correction. *M* vs *A* plots accent specific differences in treatment vs reference comparisons in a signal-dependent fashion. Panel a of Figure 12 shows the probe-set-averaged raw intensity data of the Golden spike experiment in *M* vs *A* coordinates (see eq 11). The invariant-expressed and empty probes form highly nonlinear data clouds which reveal the so-called "banana effect" which is known especially from *M* vs *A* plots of two-color microarray platforms. It results from differences of the sensitivity and the offset between the two dye-channels.[34,35] The up−down effect can be interpreted in analogy if one considers the increment of the nonspecific binding strength between the samples as offset-bias and the increment of the specific binding constant as sensitivity-bias. Accordingly, the empty probes are offset in up-direction at small A values ($\delta \log X^{N} > 0$). The points referring to the invariant expressed probes shift downward with increasing abscissa values owing to the progressively increasing contribution of specific hybridization (with $\delta \log X^{S} < 0$) to the respective probe-intensities (see eq 24).

A second analogy links the *M* vs *A* plot with the hook plot, which is essentially a smoothed *M* vs *A* plot of the affinity-corrected PM and MM intensities (see eqs 8 and 11). Applying eq 13, one can model the treatment vs reference intensity data shown in Figure 12 to obtain the curves in part a. The vertical position of the starting point of the theoretical curves is given by the log-increment of the nonspecific binding strength, whereas their "height" in the intermediate range is related to the log-increment of the specific binding strength (see eq 14). The obtained curves render both the behavior of the invariant expressed probes and the differentially expressed probes accurately. The curves for the latter probes are calculated with $\delta \log[S] > 0$ according to selected fold changes. Note that all curves converge at the start and end points owing to the common nonspecific binding and saturation properties, respectively. In the intermediate range, they diverge along nonparallel trajectories, thus complicating any direct interpretation of the intensity *M* vs *A* plot. Note that the affine transformation used for normalizing two-color microarray intensity data implicitly accounts for diverging *M* vs *A* curves, assuming the superposition of two additive components (due to the offset and the sensitivity).[35] This method, however, neglects saturation and is therefore unable to reproduce the convergence of the curves to a common right-hand end point.

We use an alternative approach and first calculate the expression degree of the probe sets in units of the S/N ratio *R*

obtained from the hook analysis (eq 4). This quantity is then plotted in M vs A coordinates, as shown in part b of Figure 12. This transformation essentially straightens the curved banana-like shape of the intensity data to virtually horizontal data clouds of the expression degree. Their M-coordinate simply provides the log-increment of the S/N ratio, $M = \delta \log R = (\delta \log K^{S} + \delta \log[S] - \delta \log X^{N})$. The difference between differentially ($|\delta \log[S]| > 0$) and invariant-expressed ($\delta \log[S] = 0$) data then provides the increment of the transcript concentration between the #t and #r samples, $\delta M \equiv M_{\text{diffexpr}} - M_{\text{invexpr}} = \delta \log[S]$. In other words, the M coordinates of the invariant expressed probes constitute the reference level for estimating differential expression as indicated by the horizontal lines in part b of Figure 12. The particular δM levels systematically underestimate the larger nominal fold changes FC > 3 for reasons which are unclear. Similar discrepancies have been reported in previous analyses of the Golden spike data.[18,36]

We emphasize again that this simple relation does not hold for intensity data. Local baseline-correction methods like LOESS are inadequate to correct intensity based M vs A data for the up−down effect, because the nonlinear differences between differentially expressed and invariant probes remain uncorrected. LOESS normalization becomes, however, an appropriate correction method for (small) baseline biases *after* transforming intensity into of the expression values.

This result gives reasons for our normalization rule (a) (see above). Most popular preprocessing methods such as dChip,[23] RMA,[4] gcRMA,[37] vsn,[5] and PLIER[24] are multichip algorithms using the intensity information from whole chip series to extract expression values. For these methods rule (a) raises inherent problems because they require intensity normalization prior to the estimation of the expression values in the summarization step. On the other hand, the single-chip methods MAS5[25] and hook[12] provide expression values without the necessity of normalizing intensity data from different chips. These methods are therefore preferential candidates for the adequate correction of the up−down effect. However, in MAS5, in order to make microarrays comparable, the trimmed average of all signals on a microarray is set to some predefined value which intrinsically overrules our normalization rule (a).

Rule (c) requires the removal of not-expressed probes from normalization. The hook algorithm calculates expression values for all probe sets independent of their present call. The expression values of absent probes shown in Figure 12 (part b) reveal that these data are indeed ineligible for proper correction of the up down effect because they systematically deviate from the M level given by the invariant expressed probes. Absent flagged probe sets are per definition beyond the resolution limit of the method and their quantitative analysis in terms of expression values consequently must fail.

For estimating the reference M level one has therefore to identify invariant expressed (and not invariant absent) probes among the microarray data. This task is beyond the topic of this article. Note that rule (b) applies to "truly" invariant-expressed probes. Their "apparent" expression value prior to normalization changes, however, due to up−down effect whereas differently expressed probes of a fold change that compensates the increment of the specific binding constant become apparently invariant. For example, part b of Figure 12 indicates that invariant-expressed probes are apparently down regulated ($M < 0$) whereas the spikes with a 2-fold change are apparently invariant ($M = 0$). Consequently, algorithms which select probe sets of small variance of their expression values in a chip series as invariant sets for subsequent normalization purposes[33,38] are not suited for proper correction of the up−down effect.

Impact of Substituting RNA by DNA Targets. Figure 13 compares the intensity- and expression-based M vs A plots of the DNA/RNA-spiked experiments which use either RNA or DNA targets for hybridization of the DNA probes. The RNA sample clearly shows the stronger curved "banana" of the raw data, revealing the larger up−down effect in agreement with the results obtained above (see Figures 8 and 9). After hook analysis, the data clouds of the expression values (S/N ratio R) run virtually parallel with the abscissa. As argued in the previous section, the vertical distance between the data levels given by the invariant and by the differentially expressed probes provides increment of the expression value. With $\delta \log[S] \cong 4.1$, it is in the same order of magnitude as the measuring range of the method. Note the different average M levels of the invariant-expressed probes of both preparations, again indicating the different effect of the same amount of spikes on the up−down effect.

Direct comparison of the binding strengths of both chemical entities shows characteristic differences (see part b of Figure 9): RNA targets give rise to a larger nonspecific background level accompanied by smaller specific binding constants compared with DNA targets. Hence, the substitution of DNA (D) by RNA (R) resembles the up−down effect for changing the concentration of nonspecific transcripts discussed above. In contrast to this situation, the condition of invariant chemical interactions (eq 21) does not hold for D-by-R substitution. Differentiation of eq 3 provides after some algebra with $X^{N}_{\infty} = X^{S}_{\infty}$ (see discussion following eq 5) the increment of the binding strengths,

$$\delta \log X^{h} = \delta \log K^{h}_{A} + \delta \log[h] - \xi^{N}(\delta \log K^{N}_{B} + \delta \log[N]) \quad (25)$$

where ξ^{N} is defined in eq 18. Decoding the effective binding strengths in terms of the reaction constants (eq 2), one gets for the increment of the binding strength owing to the D-by-R substitution at invariant specific and nonspecific transcript concentrations, [S] = const and [N] = const, respectively,

$$\delta \log X^{h}|_{[h]=\text{const}} = \delta \log K^{h} \equiv \log K^{h}(R) - \log K^{h}(D) = \delta \log K^{P-h} - (\xi^{N}\delta \log K^{N-N} + (1 - \xi^{N})\delta \log(1 + K^{N-\text{fold}}) + \delta \log(1 + K^{P-\text{fold}})) \quad (26)$$

According to eq 26, the increment of the binding strengths changes linearly with the nonspecific background level, ξ^{N}. For the limiting cases of vanishing ($\xi^{N} = 0$) and saturated ($\xi^{N} = 1$) background and assuming invariant folding propensities of the DNA probes ($\delta \log (1 + K^{P-\text{fold}}) = 0$), eq 26 provides

$$\delta \log K^{h} = \delta \log K^{P-h} - \begin{cases} f^{N-\text{fold}}\delta \log K^{N-\text{fold}} & \text{for } \xi^{N} = 0 \\ \delta \log K^{N-N} & \text{for } \xi^{N} = 1 \end{cases} \quad \text{with} \quad f^{N-\text{fold}} \equiv \frac{K^{N-\text{fold}}}{(1 + K^{N-\text{fold}})} \quad (27)$$

Hence, the value of the bimolecular association constant of probe/target duplex formation, K^{P-h}, becomes effectively modified due to the folding and/or bulk dimerization of the respective

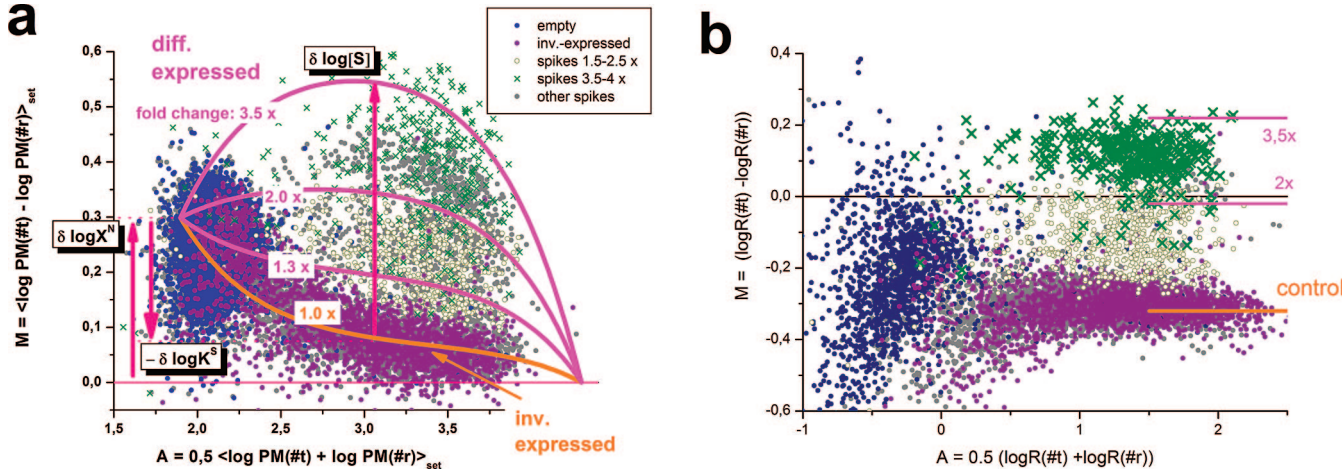

Figure 12. Treatment vs reference comparisons of the Golden spike experiment: *M* vs *A* plots of probe set-averaged log-intensity values of the PM probes (part a) and of the S/N ratio (log *R*, part b). Empty, invariant expressed and spiked (differentially expressed) probes are color coded (see figure). The curves in the first plot are calculated using the eq 13 with the parameters (eq 14); see text). The curves refer to the invariant expressed probes with δ log[S] = 0 and to the spikes with different fold changes. The data cloud formed by the former probes resembles a banana-like shape which is caused by the up−down changes of X^N and X^S (see arrows) after treatment. The S/N ratio of the probe sets were obtained by the fit of the hook equation to the Δ vs Σ plots shown in Figure 6. The hook analysis bends the curved bananas of the expressed probes into virtually horizontal data clouds (see part b).

transcripts in solution. The effect of folding is maximal at vanishing nonspecific background whereas bulk dimers affect target binding mostly at saturated background level. Note that the folding effect can be neglected for $K^{\text{N-fold}} \ll 1$.

Equations 26 and 27 show that the changes of the effective bindings constants depend on the increments of the different reaction rates caused by the substitution of the target. Recall that the R and D targets give rise to different interactions on the chip surface and in the bulk solution, namely D/R vs D/D base pairings in the probe/target duplexes and R/R vs D/D interactions for bulk dimerization and folding of the targets, respectively.

A previous study shows that the thermodynamic stability of specific 27-meric duplexes increases with D/D < D/R < R/R where the mixed case D/R is closer to D/D than to R/R.[44] For an alternative evaluation we calculated mean Watson−Crick dimerization free energies over all probes of a GeneChip using the nearest-neighbor model (Table 2). The resulting general pattern of interaction strength is D/D ≅ D/R < R/R, i.e., D/D and D/R interactions are of similar magnitude whereas R/R interactions are significantly stronger. The binding free energies are strongly correlated between the different chemical entities with correlation coefficients $\rho > 0.85$, indicating similar sequence-specific sensitivities of the probes in all cases (see footnote in Table 2).

The negative logarithm of the equilibrium constants is directly related to the respective free energy, i.e., $\delta\Delta G \equiv \Delta G(R) - \Delta G(D) \propto -\delta \log K$. The obtained ΔG values therefore suggest

$$\begin{aligned}
\delta \log K^{\text{P-h}} &\equiv \log K^{\text{P-}h}(R/D) - \log K^{\text{P-h}}(D/D) \approx 0 \\
\delta \log K^{\text{N-N}} &\equiv \log K^{\text{N-N}}(R/R) - \log K^{\text{N-N}}(D/D) > 0 \\
\delta \log(1 + K^{\text{N-fold}}) &\equiv \log(1 + K^{\text{N-fold}}(R)) - \\
&\qquad \log(1 + K^{\text{N-fold}}(D)) \geq 0 \qquad (28)
\end{aligned}$$

These relations predict in combination with eqs 26 and 27 the parallel decrease of the specific and nonspecific binding constants, i.e., $\delta \log K^{\text{h}} < 0$. After substitution of the D by R targets, one does indeed observe a decrease of the specific binding strength of the invariant-expressed probes (see part b of Figure 9). Hence, the stronger R/R interactions obviously amplify bulk dimerization and target folding, which effectively weakens hybridization of the probes with specific targets.

This argument also applies to nonspecific target binding which is apparently in contradiction with the observed slight increase of the background level, i.e., $\delta \log K^{\text{N}} > 0$ (see Figure 9, part b). Note, however, three arguments which might explain this effect as follows.

First, our estimation of the binding strengths summarized in Table 2 is based on perfect-matched 25-meric dimers. This situation does not apply to nonspecific duplexes with mismatched pairings which potentially contribute to duplex stability in a different fashion. For example, G/U mismatches form relatively stable wobble base pairings in DNA/RNA hybrid duplexes (and RNA secondary structures as well) which are expected to enhance the mean binding constant of mismatched nonspecific R/D duplexes compared with D/D complexes.

Second, the two-species binding model approximates nonspecific binding of a large variety of fragments by one effective N species which represents an affinity-weighted mean value averaged over all nonspecific fragments.[45] Possibly, the wider distribution of R/D interactions (compared with D/D) and the larger weighting of stronger interactions are expected to enhance the mean binding constant of the nonspecific R fragments.

Third, the ΔG estimates of duplex stability in Table 2 are calculated using bulk solution parameters which are not necessarily appropriate for surface binding, for example, because of position-dependent effects. The so-called specific PM/MM gain estimated from the height of the hook curve represents a relative measure of specific probe/target stability which compares the central Watson−Crick pairing in the PM probes with the self-complementary mismatches formed in the specific duplexes of the MM probes. This PM/MM gain is significantly larger in the presence of RNA targets (1.05−1.12 vs 0.93−0.95, see Table 1) which implies the slightly stronger surface binding.

Taking together relatively stable mismatches, the stronger interactions in the wider tail of the density distribution and/or surface effects suggest $\delta \log K^{\text{P-N}} > 0$ which compensates or

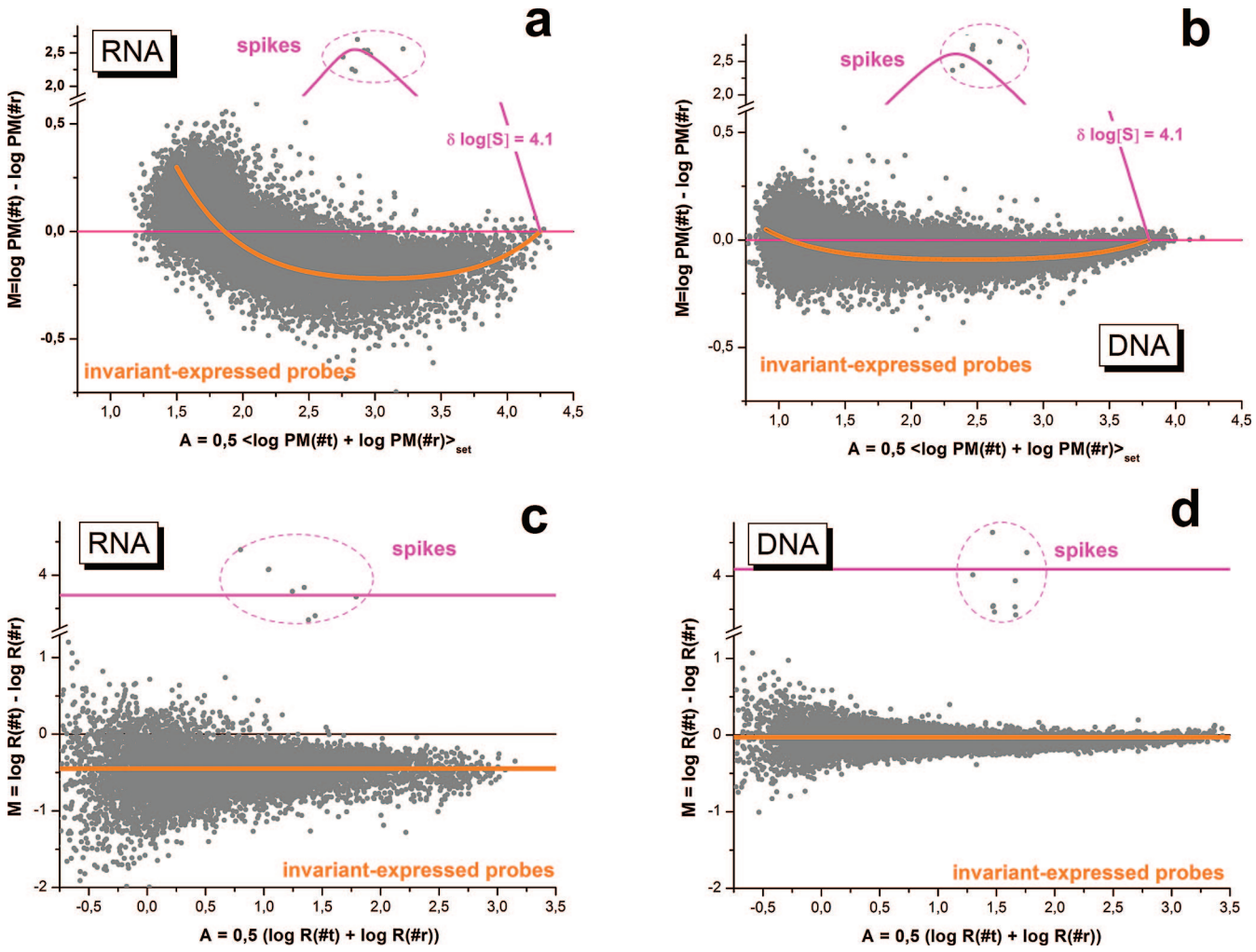

Figure 13. Treatment vs reference comparisons of the RNA/DNA spike experiment: $M-A$ plots of probe-set-averaged log-intensity values of the PM probes (panels a and b) and of the S/N ratio log R (panels c and d). The theoretical curves in parts a and b refer to the invariant and differentially expressed probes (i.e., spikes). The spikes are enriched by about 4 orders of magnitude after treatment. The banana-like shape of the DNA hybridization is much less pronounced than that of the RNA hybridization owing to the weaker up−down effect (see also Figure 9). After hook analysis the curves straighten into the horizontal lines as shown in parts c and d.

TABLE 2: Mean Free Binding Energies and Sdandard Deviation (SD) of Watson−Crick Paired DNA/DNA, DNA/RNA, and RNA/RNA 25-meric Duplexes in Bulk Solution[a]

interaction	mean ΔG/ kcal/mol	SD(ΔG)/ kcal/mol	NN parameters taken from ref
RNA/RNA	−64.7	7.9	39
DNA/RNA	−42.5	7.0	40
DNA/DNA(1)	−42.6	4.4	41
DNA/DNA(2)	−47.6	4.5	42

[a] The duplex free binding energies in 1 M NaCl bulk solution at 45 °C were calculated for each of the 403 614 25-mer probe sequences on an Affymetrix HGU95a GeneChip using nearest neighbor stacking models[43] and averaged. SD denotes the respective standard deviation. In these models, the total free energy at temperature T is given by $\Delta G = \Delta H - T\Delta S$, where ΔH and ΔS are each calculated as a sum of an initiation term plus experimentally determined stacking parameters corresponding to neighbor pairs of nucleotides taken from the references listed in the last column of the table. Binding energies are highly correlated across different probe sequences, with correlation coefficients $\rho(\Delta G^{R/R},\Delta G^{D/R}) = 0.92$, $\rho(\Delta G^{R/R},\Delta G^{D/D}) = 0.95$, and $\rho(\Delta G^{D/R},\Delta G^{D/D}) = 0.89$.

even overcompensates the effect bulk hybridization and effectively increases K^N. Note also the corresponding slight difference of the limiting values $X^N_\infty(R) \geq X^N_\infty(D)$ (see Table 1) which might be explained in an analogous fashion by the larger surface effect (see eq 3).

In summary, the effect of substituting RNA by DNA targets is essentially 2-fold. First, it slightly decreases the nonspecific background level owing to weaker binding of the nonspecific D fragments to the D probes. Second, it effectively boosts specific binding because free accessible D fragments deplete to a smaller degree than R fragments because of their weaker folding and bulk dimerization. Both effects are expected to improve the performance of the microarrays in terms of sensitivity and specificity. Note that the sensitivity of the method is directly related to K^S and its specificity to K^S/X^N.[6,7] Note also that the S/N ratio of the D hybridization spans a larger measuring range as reflected by the larger decay constant λ of the density distribution (see above). This characteristic parameter aggregates both K^S and X^N into one characteristic signal-to-noise ratio.

Adequacy of Benchmark Experiments. In this study we reanalyzed the raw-data of five benchmark experiments which, in general, aim at generating and verifying theoretical aspects of microarray hybridization and/or at judging the performance of microarray analysis methods. Different test scenarios can be used to model either a particular experimental application (e.g., a treatment vs control comparison) or the calibration function

linking the input with the output quantities of the measurement, transcript concentration and intensity, respectively.

Spike-in experiments of the Latin-square-type were successfully used in many methodical studies comparing preprocessing methods (see, e.g., refs 4 and 46 and references cited therein) and developing hybridization models (see, e.g., refs 8− 10, 14, 45, and 47). The small number and amount of variable transcripts affecting less than 1% of the available probe sets and the cyclic permutations of the spikes among the chips ensure a constant total RNA concentration in all preparations which results in a rather small interchip variability of the nonspecific background level. More importantly, these chip data virtually lack systematic shifts of the mean background hybridization. As a result the concentration dependence of the spikes is not perturbed by the up−down effect. On the other hand, this type of experiment is not optimally designed for judging normalization algorithms owing to the virtual invariance of chip effects.

Contrarily, a relatively high number of transcripts referring to ∼25% of all probe sets are hybridized in benchmark experiments of the Golden spike type where the concentration of about one-half of these spikes is varied in a treatment vs control design.[18] The parallel change of the total RNA concentration of the spikes gives rise to a significant up−down effect which complicates "naive" analyses to reproduce the nominal fold changes. For example, in their original report Choe et al. suggest multiple normalization steps on raw intensities and expression values before and after summarization, respectively, combined with signal-dependent filtering of the data.[18] This heuristic approach partly accounts for our normalization rules (a)−(c). The presented *M* vs *A* plots (see, e.g., Figures 4 and 6 in ref 18) clearly indicate that this correction method remains incomplete owing to the reasons discussed above.

Problems with analysis of the Golden spike experiment related to nonuniform distributions of null *p* values[48] and "unusual" feature characteristics (such as unrealistic high spike concentrations and high percentage of spikes)[49] raise doubts about the adequacy of the design of this benchmark experiment. It has, however, been demonstrated that problems with statistical null hypotheses are related to inadequate normalization of intensity data and can be avoided using improved methods.[50,51] Such intuitive approaches use filtering of the raw signals,[38] renormalization of expression values,[50,51] and probe-level mixture models.[36] In agreement with our normalization rule (b), normalization based on invariant expressed probes provides best results.[51,50]

Our study reveals the up−down effect as the rationale behind these heuristic approaches. On the basis of the proposed hybridization model we are able to explain and to quantify the observed effects and to propose simple rules for proper normalization of the data. Moreover, we show that the underlying up−down effect is not specific to the Golden spike experiment. Instead, it represents an inherent property of microarray hybridizations which has been unambiguously identified in another four benchmark studies. These experiments of unbalanced expression changes thus provide valuable data for studying the hybridization physics beyond the concentration dependence of the intensity addressed by the balanced expression changes of the Latin-square design. Despite some controversial discussion about possible flaws of a particular benchmark study, there is consensus about the demand in benchmark data addressing different aspects of microarray technology.[18,49,48]

The relevance of the up−down effect for biological data will be addressed in our forthcoming research. Unbalanced expression changes in large-scale knockout studies and/or chip-to-chip variations of the RNA amount in samples of different biological origin (e.g., different tissue types or developmental stages in time course studies) are potential candidates for the up−down effect with consequences for expression analysis. In a recent publication, preliminary results were presented which clearly demonstrate the presence of the up−down effect in samples of different RNA quality.[13]

Probe-Specific Up−Down Effect. This paper addresses the up−down effect at the chip level, i.e., the variation of the mean specific binding constant, averaged over all probes on the chip, upon variation of the mean nonspecific background level. However, at the probe level, both the specific binding constant and the nonspecific background level are highly dependent on particular probe and target sequences.[9] Accordingly, eq 2 applies to each probe in a sequence specific manner, where, for example, a large propensity for folding and/or bulk dimerization is accompanied by a corresponding reduction in the effective binding constant for the particular probe. This probe-level up−down effect has previously been demonstrated at the individual probe level and explained by physical models.[52,53,8]

One can formally decompose the probe-specific binding constants into the mean chip contribution ($K \equiv K_c$) and a probe-specific incremental term

$$\log K_p = \log K + \delta_p \log K \tag{29}$$

The former value is given by eq 2 whereas the latter term can be approximated in analogy with eq 26 by

$$\delta_p \log K^h|_{[h]=\text{const}} = \delta_p \log K^{P-h} - (\xi^N \delta_p \log K^{N-N} + (1 - \xi^N) f^{N-\text{fold}} \delta_p \log K^{N-\text{fold}} + f^{P-\text{fold}} \delta_p \log K^{P-\text{fold}}) \tag{30}$$

where the increment now refers to the sequence effect. Accordingly, the incremental term changes linearly with the background level between the limiting values

$$\delta_p \log K^h = \delta_p \log K^{P-h} - f^{P-\text{fold}} \delta_p \log K^{P-\text{fold}} - \begin{cases} f^{N-\text{fold}} \delta_p \log K^{N-\text{fold}} & \text{for } \xi^N = 0 \\ \delta_p \log K^{N-N} & \text{for } \xi^N = 1 \end{cases} \tag{31}$$

The hook methods uses the position-dependent sensitivity model for correction of sequence-specific effects of the probe intensities. It approximates the increments of the binding constants by the sum

$$\delta_p \log K^h = \sum_{k=1}^{25} \sigma_k^h(B_{pk}) \tag{32}$$

with the constraints $\langle \delta_p \log K^h \rangle_{\text{all p}} = 0$ and $\sum_{B=A,T,G,C} \sigma_k^h(B) = 0$. In eq 32, B_{pk} denotes the nucleotide letter at sequence position k of probe p. For the sake of simplicity we consider here a single-base model for the sensitivities σ_k^h, whereas in general the hook method uses a nearest-neighbor or even next-nearest-neighbor sensitivity model as standard. Comparison of eq 32 with eq 30 shows that the sensitivity terms can be decomposed in an analogous fashion into a weighted sum of reaction-related base and position-dependent terms

$$\sigma_k^{h}(B) = \sigma_k^{P-h}(B) - (\xi^{N}\sigma_k^{N-N}(B) + (1 - \xi^{N})f^{N-fold}\sigma_k^{N-fold}(B) + f^{P-fold}\sigma_k^{P-fold}(B)) \quad (33)$$

The sensitivity terms on the right side of this equation and the fractions of the folding constants are virtually constant for a given ensemble of probes and targets. On the one hand, one expects a certain degree of similarity between the reaction-related profiles owing to the underlying common base-pair interactions. On the other hand, reaction-specific differences, e.g., between folding and duplexing, are expected to cause subtle differences between the profiles. Their different scaling upon changing background contribution ξ^{N} is also expected to vary the total profiles $\sigma_k^{h}(B)$. This prediction has been confirmed theoretically:[54] the profiles referring to different nucleotides scale mutually relative to one another depending on whether or not one considers bulk hybridization.

Figure 14 compares the total sensitivity profiles taken from selected chips of the considered benchmark experiments at different background levels. In general, the profiles are also sensitive to other factors such as the optical background correction, and saturation.[45,54] To minimize these effects, we selected the profiles of nonspecifically hybridized probes (h = N) and used the standard Affymetrix zone algorithm for optical background correction in all cases.[25]

The data show that the observed effect is small (dilution experiment) or moderate (Golden spike and RNA spike). The negligible effect in the DNA-spike experiment is not surprising given that the background level remains essentially invariant after treatment (see above). Although there is no unique trend in the other examples, the observed changes give rise to the conclusion that nonspecific background hybridization modifies probe/target binding in a probe specific fashion. The correction for this effect consequently requires a sample-specific approach which takes into account subtle differences due to changing bulk hybridization.

5. Summary and Conclusions

Microarray expression measures are scaled by unwanted nonspecific hybridization through an effect we refer to as the up−down effect. In this effect, changes in the nonspecific background component of total RNA target induce an opposite sign change in the effective specific binding constant. The effect can equivalently be seen as a variation in each probe's sensitivity to specific transcript concentration, defined as the increment in a probe's intensity response per unit increment in specific target concentration. Because the total RNA background changes from chip to chip, owing to technical and/or biological reasons, a

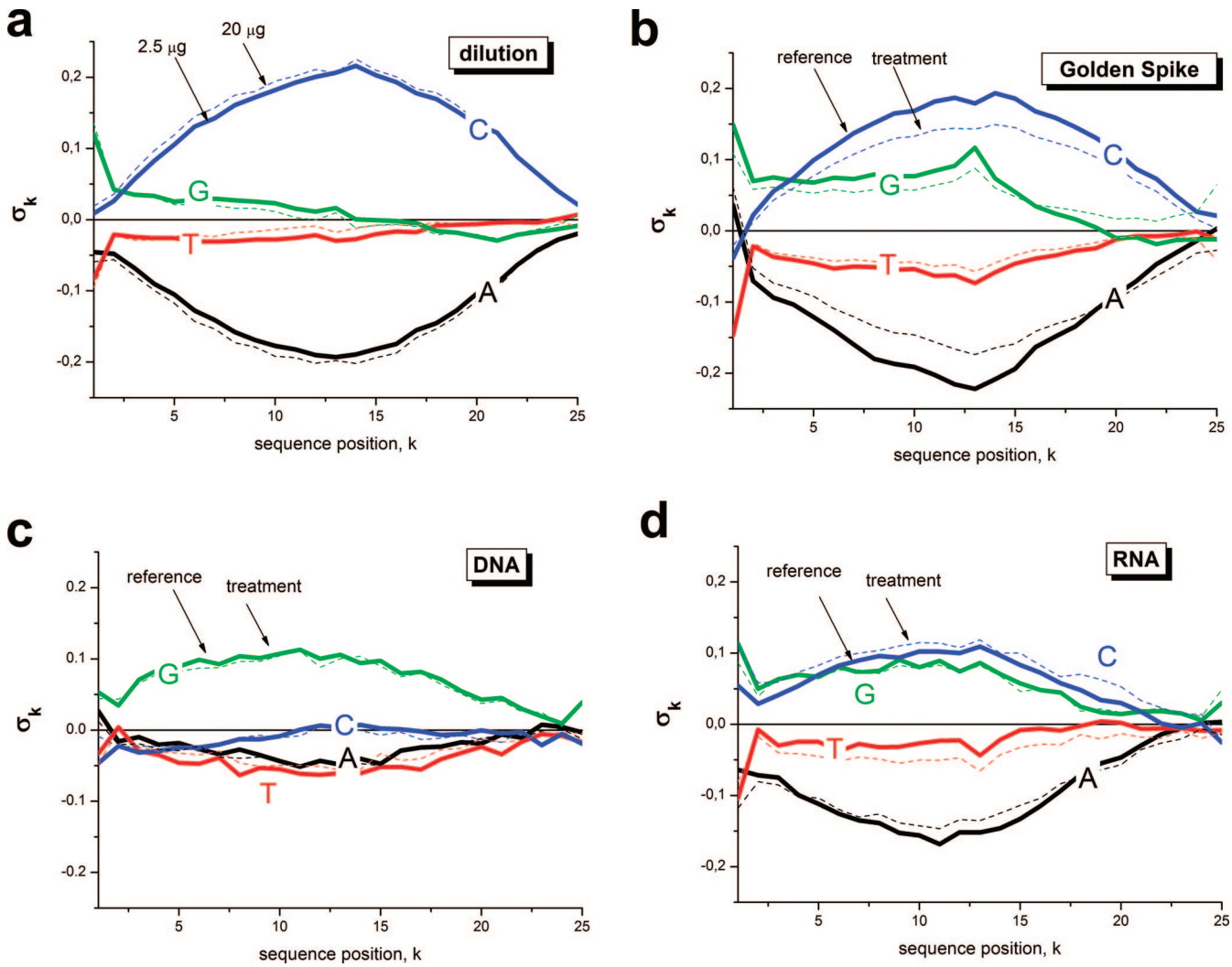

Figure 14. Position-dependent single-base sensitivity profiles of samples of different nonspecific background levels taken from different benchmark experiments. Compare the solid with the dotted curves.

normalization which corrects for the up−down effect is needed. In this paper we have proposed a set of special rules for this purpose.

In short, our normalization relies on leveling expression values obtained from the proven hook method. Only probes known to be invariantly expressed are used in the leveling process. By comparison, existing heuristic normalization techniques are biased by a number of shortcomings including failure to exclude absent (i.e., nonexpressed) probes, leveling raw fluorescence intensities rather than more appropriate expression values, and using low-variance criteria for identifying invariant sets.

The rationale behind the up−down effect is recognition of the subtle interplay of competing interactions between the probes and specific and nonspecific targets at the chip surface and in bulk solution, which can be understood in the framework of equilibrium thermodynamics. The results emphasize the importance of physicochemical approaches to improving microarray data analysis. Our next tasks include the conversion of the normalization rules into practical algorithms and their verification and application in the context of quantitative microarray analyses, for example, of wide-scale knockout studies or tissue comparisons.

Acknowledgment. The work was supported by the Deutsche Forschungsgemeinschaft under grant no. BIZ 6/4. Part of the work was kindly supported by ANU (Centre for Bioinformation Sciences) during the working stay of HB in Canberra in 2007. H.B. especially thanks Sue Wilson for discussing aspects of the work.

Supporting Information Available: The equation used to fit the hook plots and a schematic overview of the benchmark experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Lipshutz, R. J.; Fodor, S. P. A.; Gingeras, T. R.; Lockhart, D. J. *Nat. Genet.* **1999**, *21*, 20.
(2) Stark, J.; Callard, R.; Hubank, M. *Trends Biotechnol.* **2003**, *21*, 290.
(3) Binder, H.; Preibisch, S.; Berger, H. *Methods Mol. Med.* **2008**, in press (preprint: www.izbi.de/izbi/working_papers.php).
(4) Irizarry, R. A.; Bolstad, B. M.; Collin, F.; Cope, L. M.; Hobbs, B.; Speed, T. P. *Nucleic Acids. Res.* **2003**, *31*, e15.
(5) Huber, W.; von Heydebreck, A.; Sueltmann, H.; Poustka, A.; Vingron, M. *Bioinformatics* **2002**, *18*, 96.
(6) Halperin, A.; Buhot, A.; Zhulina, E. B *Biophys. J.* **2004**, *86*, 718.
(7) Binder, H. *J. Phys. Condens. Matter* **2006**, *18*, S491.
(8) Burden, C. J. *Phys. Biol.* **2008**, *5*, 016004.
(9) Binder, H.; Kirsten, T.; Loeffler, M.; Stadler, P. *J. Phys. Chem. B* **2004**, *108*, 18003.
(10) Burden, C. J.; Pittelkow, Y. E.; Wilson, S. R. *Stat. Appl. Gen. Mol. Biol.* **2004**, *3*, 35.
(11) Affymetrix. Affymetrix Microarray Suite 5.0. In *User Guide*; Affymetrix, Inc.: Santa Clara, CA, 2001.
(12) Binder, H.; Preibisch, S. *Algorithms Mol. Biol.* **2008**, *3*, 12.
(13) Binder, H.; Krohn, K.; Preibisch, S. *Algorithms Mol. Biol.* **2008**, *3*, 11.
(14) Held, G. A.; Grinstein, G.; Tu, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7575.
(15) GeneLogic dilution data: http://www.GeneLogic.com/.
(16) GeneLogic report 2003.
(17) Affymetrix spiked-in data set: http://www.affymetrix.com/support/technical/sample_data/datasets.affx.
(18) Choe, S.; Boutros, M.; Michelson, A.; Church, G.; Halfon, M. *Genome Biol.* **2005**, *6*, R16.
(19) Eklund, A. C.; Turner, L. R.; Chen, P.; Jensen, R. V.; deFeo, G.; Kopf-Sill, A. R.; Szallasi, Z. *Nat. Biotechnol.* **2006**, *24*, 1071.
(20) Affymetrix Affymetrix Microarray Suite 5.0, Affymetrix, Inc., 2001.
(21) Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U.; Speed, T. P. *Biostatistics* **2003**, *4*, 249.
(22) Wu, Z.; Irizarry, R. A.; Gentleman, R.; Murillo, F. M.; Spencer, F. *John Hopkins Univ., Dept. Biostat. Working Pap.* **2003**, 1.
(23) Li, C.; Wong, W. H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 31.
(24) Affymetrix. Technical Note 2005.
(25) Affymetrix. Technical Note 2002, p 28.
(26) Dudoit, S.; Yang, Y. H.; Callow, M. J.; Speed, T. P. *Stat. Sin.* **2002**, *12*, 111.
(27) Smyth, G.; Speed, T. *Methods* **2003**, *31*, 265.
(28) Bolstad, B. Low Level Analysis of High-Density Oligonucleotide Array Data: Background, Normalization and Summarization, Ph.D. Thesis, University of California, Berkeley, 2004.
(29) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, *19*, 9.
(30) Lu, T.; Costello, C.; Croucher, P.; Hasler, R.; Deuschl, G.; Schreiber, S. *BMC Bioinformatics* **2005**, *6*, 37.
(31) Andrew, A. Hill; Brown, E. L.; Whitley, M. Z.; Tucker-Kellogg, G.; Hunter, C. P.; Slonim, D. K. *Genome Biol.* **2001**, *2*, 0055.1.
(32) van de Peppel, J.; Kemmeren, P.; van Bakel, H.; Radonjic, M.; van Leenen, D.; Holstege, F. C. P. *EMBO Rep.* **2003**, *4*, 387.
(33) Barenco, M.; Stark, J.; Brewer, D.; Tomescu, D.; Callard, R.; Hubank, M. *BMC Bioinformatics* **2006**, *7*, 251.
(34) Haldermans, P.; Shkedy, Z.; Van Sanden, S.; Burzykowski, T.; Aerts, M. *Stat. App. Gen. Mol. Biol.* **2007**, 6.
(35) Bengtsson, H.; Hossjer, O. *BMC Bioinformatics* **2006**, *7*, 100.
(36) Lemieux, S. *BMC Bioinformatics* **2006**, *7*, 391.
(37) Wu, Z.; Irizarry, R. A.; Gentleman, R.; Murillo, F. M.; Spencer, F. *J. Am. Stat. Appl.* **2004**, *99*, 909.
(38) Calza, S.; Raffelsberger, W.; Ploner, A.; Sahel, J.; Leveillard, T.; Pawitan, Y. *Nucleic Acids Res.* **2007**, *35*, e102.
(39) Xia, T.; SantaLucia, J., J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719.
(40) Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. *Biochemistry* **1995**, *34*, 11211.
(41) SantaLucia, J.; Hicks, D. *Annu. Rev. Biomol. Struct.* **2004**, *33*, 415.
(42) Sugimoto, N.; Nakano, S.; Yoneyama, M.; Honda, K. *Nucleic Acids Res.* **1996**, *24*, 4501.
(43) SantaLucia, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460.
(44) Barone, F.; Cellai, L.; Matzeu, F.; Pedone, F. *Biophys. Chem.* **2000**, *86*, 37.
(45) Binder, H.; Preibisch, S.; Kirsten, T. *Langmuir* **2005**, *21*, 9287.
(46) Irizarry, R. A.; Wu, Z.; Jaffee, H. A. *Bioinformatics* **2006**, *22*, 789.
(47) Binder, H.; Preibisch, S. *J. Phys. Condens. Matter* **2006**, *18*, S537.
(48) Dabney, A.; Storey, J. *Genome Biol.* **2006**, *7*, 401.
(49) Irizarry, R.; Cope, L.; Wu, Z. *Genome Biol.* **2006**, *7*, 404.
(50) Gaile, D.; Miecznikowski, J. *BMC Genomics* **2007**, *8*, 105.
(51) Schuster, E.; Blanc, E.; Partridge, L.; Thornton, J. *Genome Biol.* **2007**, *8*, R126.
(52) Heim, T.; Tranchevent, L.-C.; Carlon, E.; Barkema, E. T. *J. Phys. Chem. B* **2006**, *110*, 22786.
(53) Carlon, E.; Heim, T. *Physica A* **2006**, *362*, 433.
(54) Heim, T.; Wolterink, J. K.; Carlon, E.; Barkema, G. T. *J. Phys. Condens. Matter* **2006**, *18*, S525–S536.

JP808118M