Langmuir Adsorption-Desorption Model on Affymetrix Microarrays

Medical College of Georgia
GEORGIA'S HEALTH SCIENCES UNIVERSITY

IWONA PAWLIKOWSKA AND GRZEGORZ REMPALA
Department of Biostatistics, Medical College of Georgia, Augusta GA 30912, USA

ABSTRACT

We use a simple discrete stochastic network model to describe a hybridization reaction on an Affymetrix microarray (GeneChip) obtained by discretization of the standard Langmuir model of adsorption-desorption. The approach allows us to derive some new stochastic laws for filtering microarray signals.

1 INTRODUCTION

The Affymetrix GeneChip design is one of the most common ones for oligonucleotide DNA microarrays. The major limitation of the technology is that rather than the molecular target concentration it only records the empirical measures of expression (i.e., the scanner-measured fluorescence). These fluorescence readings are subject to optical noise, non-specific hybridization, probe-specific effects, and measurement error, and can often lead to imprecise and inaccurate results (see, e.g., [6]). A development of a method of extracting target concentrations from noisy uorescence readings on a GeneChip is therefore of great interest. One approach is to model the chip hybridization process as a network of discrete biochemical reactions following the Langmuir adsorption-desorption model.

Close-up of one probe region on a GeneChip

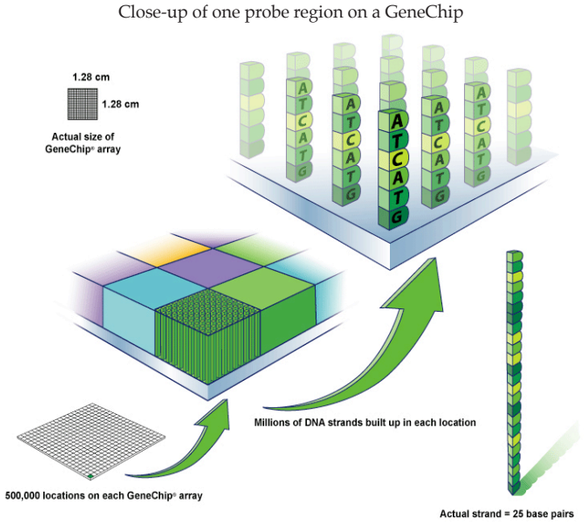


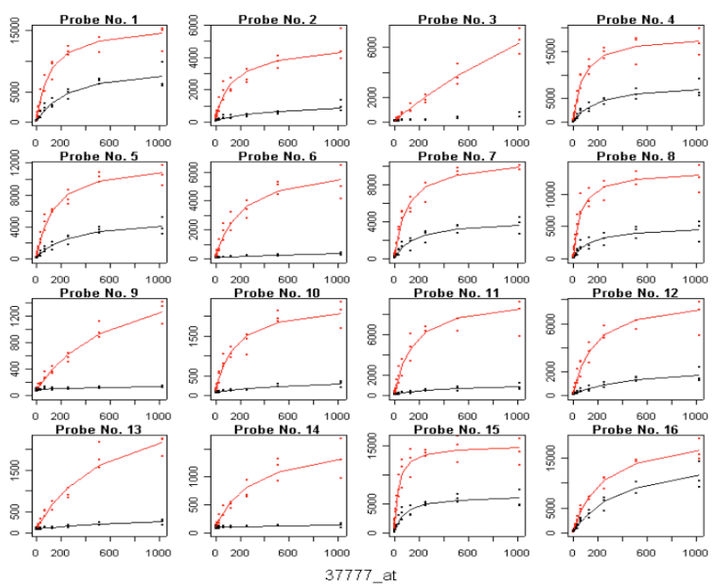
Image courtesy of Affymetrix

One of the most popular adsorption models considered in the context of microarrays (cf. e.g., [3] or [1]) is the so-called Langmuir model which in its simplest deterministic form describes the relationship between concentration and fluorescence levels of probe-target complexes by means of a hyperbolic function. Let $u = u(t) \in (0, 1)$ be a fraction of sites within a probe region occupied by probe-target duplexes at time t after the commencement of hybridization. Adsorption reaction is assumed to occur at a rate $d_1 x(1-u)$, proportional to target concentration x and fraction $(1-u)$ of unoccupied probe sites. Desorption reaction is assumed to occur at a rate $d_2\, u$, proportional to the fraction of occupied probe sites. The fraction of probe sites occupied by probe-target complexes is given by the Langmuir equation

$$\frac{du}{dt} = d_1 x - (d_1 x + d_2)u$$

In order to properly account for the effects of multiple simultaneous hybridizations as well as the cross-hybridization due to competition between similarly sequenced targets for the same probe regions, its seems that the stochastic version of the Langmuir model is needed.

Raw data from .cel files from Affymetrix Latin Square spike-in experiment with fits to Langmuir isotherm, Red = PM, Black = MM



Images courtesy of SCB at MCG (http://scb.mcg.edu).

2 THE LANGMUIR BIRTH DEATH MODEL

Let assume no probe interactions. We consider a simple one dimensional birth-death process described by one chemical species $Dplx$ (the amount of probe-target duplex). Hence, we consider two coupled chemical reaction

$$\begin{aligned} \emptyset &\xrightarrow{b(\cdot)} Dplx \\ Dplx &\xrightarrow{d(\cdot)} \emptyset. \end{aligned}$$

Langmuir Birth-Death Process is any BD process with the set of states $\{0, \ldots, N\}$ and the birth and death rates of the form

$$\begin{aligned} b(k) &= c_1(N-k) + \mathcal{C}(k, N) \\ d(k) &= c_2\, k + \mathcal{C}(k, N) \end{aligned}$$

for $k = 0, \ldots, N$, where $c_1, c_2 > 0$ are some constants and the function $\mathcal{C}(\cdot, N)$ is intended to model the noise of the non-target adsorption and desorption annd is assumed to satisfy the boundary conditions $\mathcal{C}(0, N) = \mathcal{C}(N, N) = 0$.
In GineChip array $\mathcal{C}(\cdot, N)$ accounts for the competition for the same RNA targets between different probe regions with similar nucleotide sequences. Herein we consider only $\mathcal{C}(k, N)$ given by the functions $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ defined below with the corresponding models henceforth referred to as M_1–M_3

$$\begin{aligned} &\mathcal{C}_1(k, N) = c_3 Nk && \text{for } 0 \le k < N \text{ and } \mathcal{C}_1(N, N) = 0 && (M_1) \\ &\mathcal{C}_2(k, N) = c_3 N(N-k) && \text{for } 0 < k \le N \text{ and } \mathcal{C}_2(0, N) = 0 && (M_2) \\ &\mathcal{C}_3(k, N) = c_3 k(N-k) && \text{for } 0 \le k \le N && (M_3) \end{aligned}$$

Two assumptions on hybridization under M_1

- the level of the target-specific signal in the probe region has lower magnitude than the level of non-specific signal (i.e., signal noise)
- the non-specific signal noise is proportional to the total system (i.e., probe region) size as well and the current system state and the target concentration

3 LIMIT THEOREM

For $z, \gamma > 0$ denote

$$\Gamma(z, \gamma) = \int_0^\gamma s^{z-1} \exp(-s)\, ds$$

For any $\alpha, \beta > 0$ let $IG(\alpha, \beta, 1)$ denote an Incomplete Gamma Distribution with the density function

$$f(x) = \Gamma(\alpha, \beta)^{-1}\, \beta^\alpha\, x^{\alpha-1}\, \exp(-x\beta)$$

for $x \in (0, 1)$ and zero otherwise.
We say that the random variable Z has the Langmuir-Incomplete Gamma Distribution with parameters α, β ($LIG(\alpha, \beta)$) satisfying $\beta > \alpha > 0$ if the following equality in distribution holds

$$Z \stackrel{D}{=} (1 - \pi_{\alpha,\beta})W + \pi_{\alpha,\beta}\, \delta_1$$

where the random variable W is distributed according to $IG(\alpha, \beta, 1)$, δ_x is the Dirac delta function at x and

$$\pi_{\alpha,\beta} = \frac{\beta^\alpha}{\beta^\alpha + \Gamma(\alpha, \beta)\,(\beta - \alpha)\exp(\beta)}$$

Limit Theorem for LBD Process Let $X_N^{(i)}$ be the stationary distributions of LBD Process under M_i for $i = 1, 2, 3$, and let $a = c_1/c_3$ and $b = (c_1 + c_2)/c_3$, as well as $Y_N^{(i)} = X_N^{(i)}/N$. Then, as $N \to \infty$ we have weak convergence

$$Y_N^{(i)} \stackrel{D}{\to} \mathcal{Z}_i \quad i = 1, 2, 3$$

where the limiting random variables $\mathcal{Z}_i$ are as follows
(i) $\mathcal{Z}_1$ is distributed as $LIG(a, b)$
(ii) $\mathcal{Z}_2$ is such that $1 - \mathcal{Z}_2$ is distributed as $LIG(b - a, b)$
(iii) $\mathcal{Z}_3$ is $Beta(a, b - a)$

4 CONCLUSIONS

- Our results imply that the (incomplete) gamma and beta type distributions could be used as approximations to the observed uorescence readings of the oligo-probes on a GeneChip microarray.
- Both M_1 and M_2 models are amenable to the gamma-type approximation of their stationary distributions (with some proper adjustment for the boundary probability)
- M_3 with 'boundary symmetric' noise term yields a beta stationary distribution with no boundary effects.

References

[1] C. Burden, Y. Pittelkow, and S. Wilson, Statistical analysis of adsorption models for oligonucleotide microarrays, Statistical Applications in Genetics and Molecular Biology, **3**(35), (2004)

[2] B. Dennis, and G. P. Patil, The gamma distribution and weighted multimodal gamma distributions as models of population abundance, Mathematical Biosciences, **68**:187-212, (1984)

[3] D. Hekstra, A.R. Taussig, M. Magnasco, and F. Naef, Absolute mrna concentrations from sequence-specific calibration of oligonucleotide arrays, Nucleic Acids Research **31**:1962-1968, (2003)

[4] M. Newton, A. Noueiry, D. Sarkar, and P. Ahlquist, Detecting differential gene expression with a semi-parametric hierarchical mixture method, Biostatistics, **5**:155-176, (2004)

[5] M. A. Newton, C. M. Kendziorski, C. S. Richmond, F. R. Blattner, and K. W. Tsui, On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data, Journal of Computational Biology, **8**:37-52, (2001)

[6] Z. Wu, and R. Irizarry, Stochastic models inspired by hybridization theory for short oligonucleotide arrays, Journal of Computational Biology, **12**(6):882-893, (2005)

[7] G. Rempala, and I. Pawlikowska, Limit Theorems for Hybridization Reactions on Oligonucleotide Microarrays, Journal of Multivariate Analysis, **99**(9):2082-2095, (2008)