

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Thermodynamics of Microarray Hybridization

Raul Măluțan and Pedro Gómez Vilda

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51624>

1. Introduction

Microarrays make the use of hybridization properties of nucleic acids to monitor Deoxyribonucleic acid (DNA) or Ribonucleic acid (RNA) abundance on a genomic scale in different types of cells. The hybridization process takes place between surface-bound DNA sequences - the probes, and the DNA or RNA sequences in solution - the targets. Hybridization is the process of combining complementary, single-stranded nucleic acids into a single molecule. Nucleotides will bind to their complement under normal conditions, so two perfectly complementary strands will bind to each other readily. Conversely, due to the different geometries of the nucleotides, a single inconsistency between the two strands will prevent them from binding.

In oligonucleotide microarrays hundreds of thousands of oligonucleotides are synthesized *in situ* by means of photochemical reaction and mask technology. Probe design in these microarrays is based on complementarity to the selected gene or an expressed sequence tag (EST) reference sequence. An important component in designing an oligonucleotide array is ensuring that each probe binds to its target with high specificity.

The dynamics of the hybridization process underlying genomic expression is complex as thermodynamic factors influencing molecular interaction are still fields of important research [1] and their effects are not taken into account in the estimation of genetic expression by the algorithms currently in use.

2. State of the art

Many techniques have been developed to identify trends in the expression levels inferred from DNA microarray data, and recently the attention was devoted to methods to obtain accurate expression levels from raw data on the underlying principles of the thermodynamics and hybridization kinetics. The development of DNA chips for rapidly

screening and sequencing unknown DNA segments mainly relies on the ability to predict the thermodynamic stability of the complexes formed by the oligonucleotide probes.

The thermodynamics of nucleic acids have been studied from different points of view. Wu *et al.* [2] analyze the temperature-independent and temperature-dependent thermodynamic parameters of DNA/DNA and RNA/DNA oligonucleotide duplexes. The differences between DNA polymer and oligonucleotide nearest-neighbour thermodynamic trends, and the salt dependence of nucleic acid denaturation allowed to SantaLucia [3] to show that there is length dependence to salt effects but not to the nearest-neighbour propagation energies.

An early study on DNA microarray hybridization [4] found that it was strongly dependent on the rate constants for DNA adsorption/desorption in the non-probe covered regions of the surface, the two-dimensional diffusion coefficient, and the size of probes and targets and also suggested that sparse probe coverage may provide results equal to or better than those obtained with a surface totally covered with DNA probes. A theoretical analysis of the kinetics of DNA hybridization demonstrated that diffusion was important in determining the time required to reach equilibrium and was proportional to the equilibrium binding constant and to the concentration of binding sites [5].

Newer studies on hybridization kinetics and thermodynamics reveal that perfect match sequences require less time to reach saturation than mismatches. The experimental results of Dai *et al.* [6] exhibit inverse temporal behaviour, resulting that surface-bound oligos hybridizing primarily with their perfect complement sequence tend to equilibrate more slowly than do those whose binding is dominated by mismatch duplexes. Considering the assumptions, it has been demonstrated [7] that the hybridization time can in fact increase the accuracy of expression ratios, and that this effect may be more dramatic for larger fold changes. Separation between specific and nonspecific binding events can avoid the confusion about what RNA hybridizes the probes. In this case analysis of the perfect match and mismatch intensities in terms of simple single-base related parameters indicates that the intensity of complementary MM introduces a systematic source of variation compared with the intensity of the respective PM probe [8].

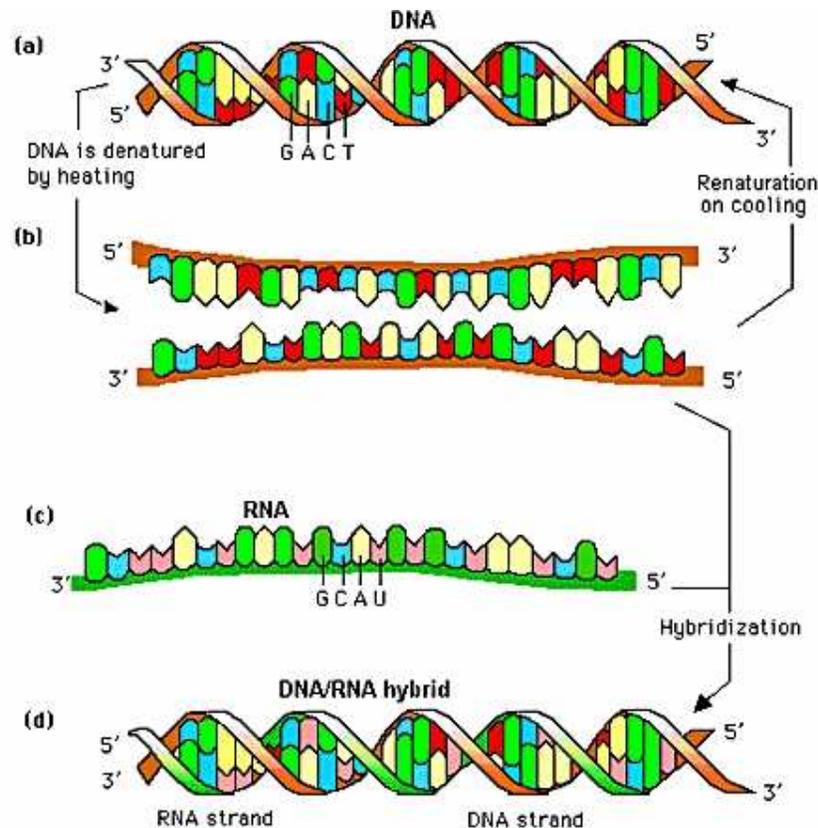
The hybridization of nucleic acids was modelled [9] according with the supposition that the process of hybridization goes through an intermediate state in which an initial short contact region has a single-stranded conformation prior to binding.

The hybridization theory gave the possibility of developing models that can be used to obtain improved measures of expression useful for data analysis. Naef and Magnasco [10] propose a simpler model to describe the probe effect that considers only the sequence composition of the probes. They demonstrate that the interactions between nearest neighbours add much predictive power for specific signal probe effects. The stochastic model proposed by Wu and Irizarry [11] can be used to improve the expression measure or in the normalization and summarization of the data.

3. DNA hybridization

DNA is a nucleic acid that contains the genetic instructions monitoring the biological development of all cellular forms of life, and many viruses. DNA is a long polymer of nucleotides and encodes the sequence of the amino-acid residues in proteins using the genetic code, a triplet code of nucleotides. DNA is organized as two complementary strands, head-to-tail, with the hydrogen bonds between them. Each strand of DNA is a chain of chemical “building blocks”, called nucleotides, of which there are four types: adenine (A), cytosine (C), guanine (G) and thymine (T). Between the two strands, each base can only bond with one single predetermined other base: A with T, T with A, C with G, and G with C, being the only possible combination.

Hybridization refers to the annealing of two nucleic acid strands following the base pairing rule. As shown in Figure 1, at high temperatures approximately 90°C to 100°C the complementary strands of DNA separate, denature, yielding single-stranded molecules. Two single strands under appropriate conditions of time and temperature e.g. 65°C, will re-nature to form the double stranded molecule. Nucleic acid hybrids can be formed between two strands of DNA, two strands of RNA or one strand of DNA and one of RNA. Nucleic acids hybridization is useful in detecting DNA or RNA sequences that are complementary to any isolated nucleic acid.



Nucleic Acid Hybridization

Figure 1. DNA-RNA hybridization. Hybridization is the process of combining complementary, single-stranded nucleic acids into a single molecule. (from [12])

Finding the location of a gene or gene product by adding specific radioactive or chemically tagged probes for the gene and detecting the location of the radioactivity or chemical on the chromosome or in the cell after hybridization is called *in-situ* hybridization.

In the same way, in microarray technology, hybridization is used in comparing mRNA abundance in two samples, or in one sample and a control. RNA from the sample and control are extracted and labeled with two different fluorescent labels, *e.g.* a red dye for the RNA from the sample population and green dye for that from the control population. Both extracts are washed over the microarray and gene sequences from the extracts hybridized to their complementary single-strand DNA molecule previously attached to the microarray. Then, to measure the abundance of the hybridized RNA, the array is excited by a laser.

In the oligonucleotide microarrays the hybridization process occurs in the same way, the only difference here is that the sequences to be laid over the chip are sequences of 25 nucleotides length, perfect complementary to same length sequence of the gene, PM – perfect match, and sequences of 25 nucleotides length, designed to correspond to PM, but having the middle base - the 13th one, changed by its complementary base, MM – mismatch, as in Figure 2. The MM probes give some estimates of the random hybridization and cross hybridization signals. One principle to be followed in the design of oligonucleotide arrays is ensuring that the probes bind to their target with high accuracy. When the two strands are completely complementary they will bind by a specific hybridization, as it can be seen in Figure 3. On the contrary if there are mismatches between the nucleotides of the strands and they bind, a process called non-specific hybridization or cross-hybridization occurs.

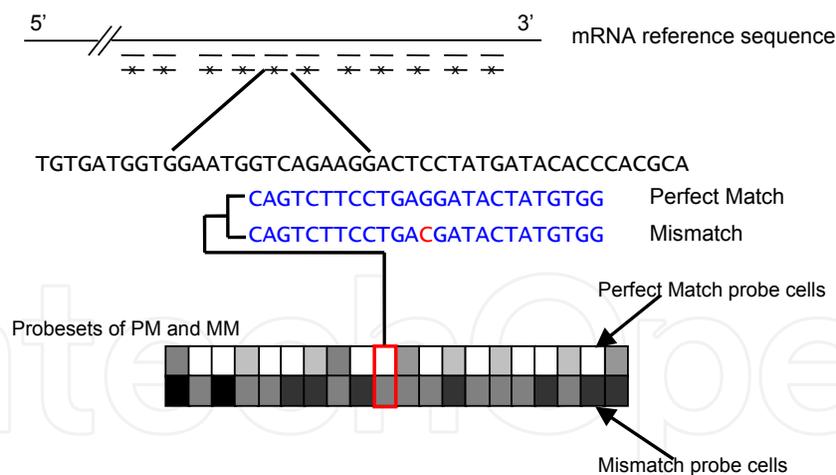


Figure 2. Perfect Match – Mismatch probeset strategy. Sequence of 25-mer length complementary to the selected part of mRNA sequence form a Perfect Match probe, while the Mismatch probe is artificially created by changing middle base with its complementary. In an oligonucleotide array a gene is represented by 11 to 20 probes. (modified from [13])

The hybridization process has been studied from point of view of interaction between base pairs, the interaction with unintended targets and also from its kinetics processes. Because in practice the DNA chips are immersed in the target solution for a relatively short time, the arrival to equilibrium is not guaranteed. Yet full analysis of the reaction kinetics requires

knowledge of the equilibrium state. An understanding of the equilibrium state is also necessary to identify the relative importance of kinetic controls of the performance of the DNA microarrays. The effect of the cross-hybridization on probe intensity is predictable in the oligonucleotide microarrays, and models for avoiding this have been developed [14], [15], [16] some aspects of it going to be described in the following section.

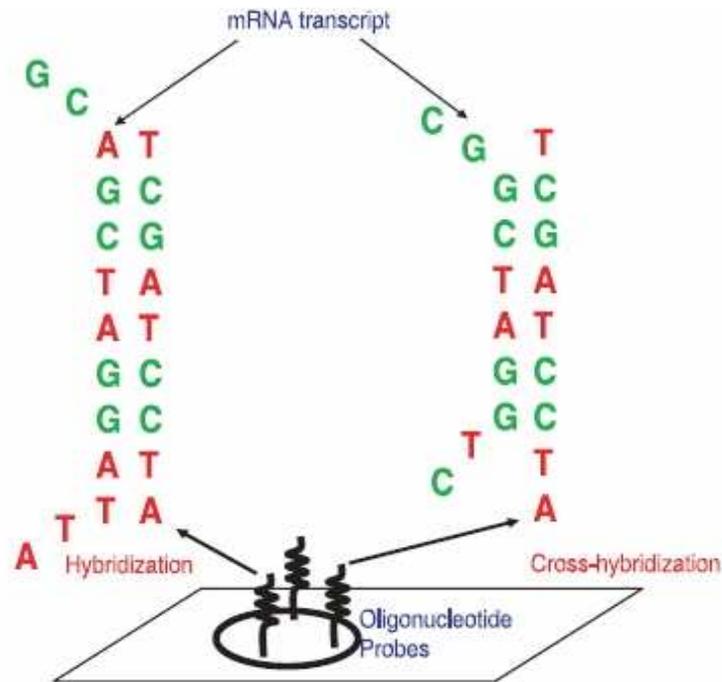


Figure 3. Cross-hybridization on a nucleotide probe. In specific hybridization the sequences are completely complementary, while in non-specific or cross hybridization the sequences contain mismatches. (from [17])

4. Technical factors affecting gene expression

4.1. Thermodynamics parameters

Black and Hartley [18] define enthalpy as the sum of the internal energy of a thermodynamic system plus the energy associated with work done by the system on the atmosphere, which is the product of the pressure times the volume, as in equation (1)

$$H = U + pV \quad (1)$$

Because enthalpy is a property, its value can be determined for a simple compressible substance once two independent, intensive thermodynamic properties of the substance are known, and the change in enthalpy is independent of the path followed between two equilibrium states

In [18] the entropy, S , was defined using the following equation:

$$dS = \frac{\delta Q}{T} \quad (2)$$

where δQ is an amount of heat introduced to the system and T is a constant absolute temperature. Since this definition involves only differences in entropy, the entropy itself is only defined up to an arbitrary additive constant.

The following models to be described use the state function parameters, enthalpy and entropy. State functions define the properties of a thermodynamic state. In a change between two thermodynamic states, the change in value of the state function is given by the symbol Δ .

The standard enthalpy change, ΔH° , is the difference in the standard enthalpies of formation between the products and the reactants. This state function is associated with changes in bonding between reactants and products. Changes in enthalpy during reactions are measured by calorimetry experiments.

The standard entropy change, ΔS° , is the difference in standard entropies between reactants and products. Entropy is a measure of the degree of order in a chemical system due to bond rotations, other molecular motions, and aggregation. The more random a system (disorder), the greater the entropy is. The larger a structure, the more degrees of freedom it has, and the greater its entropy.

4.2. Interaction between pairs

The nucleic acid duplex stability can be endangered by the interaction between the nucleotide bases. Thermodynamics for double helix formation of DNA/DNA, RNA/RNA or DNA/RNA can be estimated with nearest neighbour parameters. Enthalpy change, ΔH° , entropy change, ΔS° , free energy change, ΔG° , and melting temperature, T_m , were obtained on the basis of the nearest-neighbour model.

The nearest-neighbour model for nucleic acids, known as the NN model, assumes that the stability of a given base pair depends on the identity and orientation of neighbouring base pairs [3]. Previous studies in NN model parameters were brought forth in [15] and [19].

In the NN model, sequence dependent stability is considered in terms of nearest-neighbour doublets. In duplex DNA there are 10 such unique internal nearest-neighbour doublets. Listed in the 5'-3' direction, these are AT/AT TA/TA AA/TT AC/GT CA/TG TC/GA CT/AG CG/CG GC/GC and GG/CC. Dimmer duplexes are represented with a slash separating strands in antiparallel orientation *e.g.* AC/TG means 5'-AC-3' Watson-Crick base-paired with 3'-TG-5'.

The total difference in the free energy of the folded and unfolded states of a DNA duplex can be approximated at 37°, with a nearest-neighbour model:

$$\Delta G^\circ(\text{total}) = \sum_i n_i \Delta G^\circ(i) + \Delta G^\circ(\text{init w/term G} \cdot \text{C}) + \Delta G^\circ(\text{init w/term A} \cdot \text{T}) + \Delta G^\circ(\text{sym}) \quad (3)$$

where $\Delta G^\circ(i)$ are the standard free-energy changes for 10 possible Watson-Crick nearest neighbours, *e.g.* $\Delta G^\circ(1) = \Delta G_{37}^\circ(\text{AA/TT})$, $\Delta G^\circ(2) = \Delta G_{37}^\circ(\text{TA/AT})$, n_i is the number of occurrences of each nearest neighbour, i , and $\Delta G^\circ(\text{sym})$ equals +0.43 kcal/mol if the duplex

is self complementary and zero if it is not self-complementary. The total difference in the free energy at 37°, ΔG_{37}^0 , can be computed from ΔH^0 and ΔS^0 parameters using the equation:

$$\Delta G_{37}^0 = \Delta H^0 - T\Delta S^0 \quad (4)$$

For a specific temperature one can compute the total free energy using the values from Table 1. As described in [19] the melting temperature T_m is defined as the temperature at which half of the strands are in double helical and half are in the random-coil state. A random-coil state is a polymer conformation where the monomer subunits are oriented randomly while still being bonded to adjacent units.

For self-complementary oligonucleotides, the T_m for individual melting curves was calculated from the fitted parameters using the following equation:

$$T_m = \Delta H^0 / (\Delta S^0 + R \ln C_T) \quad (5)$$

where R is the general gas constant, *i.e.* 1.987 cal/K mol, the C_T is the total strand concentration, and T_m is given in K. For non-self-complementary molecules, C_T in equation (5) was replaced by $C_T/4$.

Sequence	ΔH^0 kcal/mol	ΔS^0 kcal/mol
AA/TT	-7.9	-22.2
AT/TA	-7.2	-20.4
TA/AT	-7.2	-21.3
CA/GT	-8.5	-22.7
GT/CA	-8.4	-22.4
CT/GA	-7.8	-21.0
GA/CT	-8.2	-22.2
CG/GC	-10.6	-27.2
GC/CG	-9.8	-24.4
GG/CC	-8.0	-19.9
Init. w/term G•C	0.1	-2.8
Init. w/term A•T	2.3	4.1
Symmetry correction	0	-1.4

Table 1. Unified oligonucleotide ΔH^0 and ΔS^0 nearest neighbour parameters in 1M NaCl. The table shows the values of the total enthalpy and entropy for the dimer duplexes as used in [3].

The nearest-neighbour parameters of Delcourt et al. (1991) [20], SantaLucia et al. (1996) [19], Sugimoto et al. (1996) [15] and Allawi et al. (1997) [21] were evaluated from the analysis of optical melting curves of a variety of short synthetic DNA duplexes in 1 M Na⁺.

The observed trend in nearest-neighbor stabilities at 37°C is $GC/CG = CG/GC > GG/CC > CA/GT = GT/CA = GA/CT = CT/GA > AA/TT > AT/TA > TA/AT$, as in Table 2. This trend suggests that both sequence and base composition are important determinants of DNA duplex stability. It has long been recognized that DNA stability depends of the percent G-C content.

Sequence	ΔG 37 (kcal/mol)			
	<i>Delcourt et al.</i>	<i>SantaLucia et al.</i>	<i>Sugimoto et al.</i>	<i>Allawi et al.</i>
AA/TT	-0.67	-1.02	-1.20	-1.00
AT/TA	0.62	-0.73	-0.90	-0.88
TA/AT	-0.70	-0.60	-0.90	-0.58
CA/GT	-1.19	-1.38	-1.70	-1.45
GT/CA	-1.28	-1.43	-1.50	-1.44
CT/GA	-1.17	-1.16	-1.50	-1.28
GA/CT	-1.12	-1.46	-1.50	-1.30
CG/GC	-1.87	-2.09	-2.80	-2.17
GC/CG	-1.85	-2.28	-2.30	-2.24
GG/CC	-1.55	-1.77	-2.10	-1.84
Average	-1.20	-1.39	-1.64	-1.42
Init. w/term G•C	NA	0.91	1.70	0.98
Init. w/term A•T	NA	1.11	1.70	1.03

Table 2. Comparison of computed NN free energy parameters at 37°C

On the other hand, the nearest neighbour ΔH° parameters from Table 1, do not follow this trend. This suggests that stacking, hydrogen bonding, and other contributions to the ΔH° present a complicated sequence dependence.

4.3. Interaction with unintended targets

As seen in previous sections the major issue in microarray oligonucleotide technology is the selection of probe sequences with high sensitivity and specificity. It has been shown [22] that the use of MM probes for assessment of non-specific binding is unreliable. Since the duplex formation in solution has been studied using the nearest neighbour model [3], [15] the microarray design in terms of probe selection has been achieved by using a model based on the previously mentioned nearest neighbour model [16]. The model of Zhang *et al.* presents some modification to the nearest neighbour model, firstly to assign different weight factors at each nucleotide position on a probe with the scope of reflecting that the binding parts of a probe may contribute differently to the stability of bindings, and secondly to take into account two different modes of binding the probes: gene specific binding, *i.e.* formation of DNA-RNA duplexes with exact complementary sequences, and non-specific binding, *i.e.* formation of duplexes with many mismatches between the probe and the attached RNA molecule. They called their model, the positional-dependent-nearest-neighbour model.

According with their method, the observed signal I_{ij} for probe i in the probe set for gene j is modelled as:

$$I_{ij} = \frac{N_j}{1 + e^{E_{ij}}} + \frac{N^*}{1 + e^{E_{ij}^*}} + B \quad (6)$$

where B is the background intensity, N_j is the number of expressed mRNA molecules contributing to gene specific binding, N^* represents the number of RNA molecule contributing to nonspecific binding, E and E^* are the binding energies for gene specific and respectively nonspecific binding. These energies are calculated as the weighted sum of stacking energies:

$$E_{ij} = \sum \omega_k \varepsilon(b_k, b_{k+1}) \quad (7)$$

$$E_{ij}^* = \sum \omega_k^* \varepsilon^*(b_k, b_{k+1}) \quad (8)$$

where ω_k and ω_k^* are the weight factors that depend on the position along the probe from the 5' to 3' end, and $\varepsilon(b_k, b_{k+1})$ is the same as the stacking energy used in nearest neighbour model [15].

The positional-dependent-nearest-neighbour model appears to indicate that the two ends of a probe contribute less to binding stability according to their weight factors, see Figure 4. a). It also can be observed that there is a dip in the gene specific binding weight factors of MM probes around the mismatch position, probably due the mismatch which destabilizes the duplex structure. In Figure 4. b) it can be noted that stacking energies in the positional-dependent-nearest-neighbour model can give an explanation for the presence of negative probe pair signals.

This model, together with the nearest neighbour model solves the problem of binding on microarrays, but still there are factors that affect the gene expression measuring. One of them affects the process of competing adsorption and desorption of target RNA to from probe-target duplexes at the chip surface.

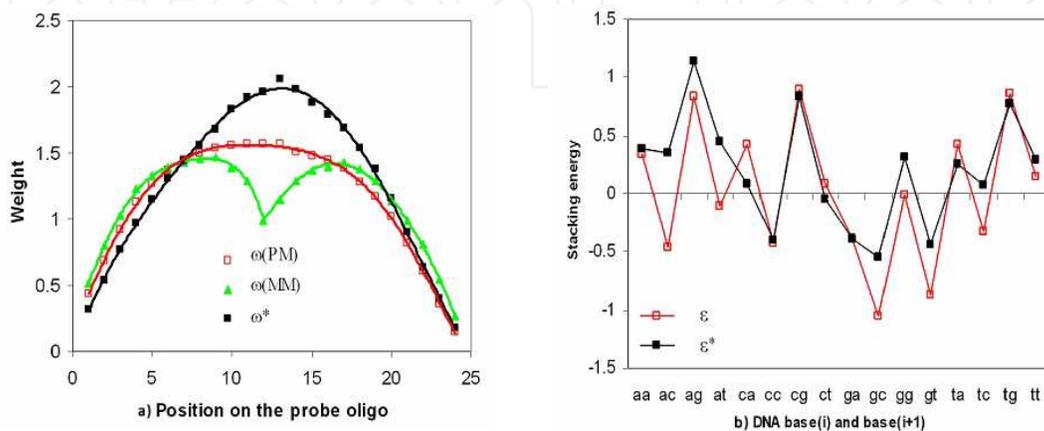


Figure 4. a) weight factors; b) nearest-neighbour stacking energy. (from [16])

4.4. Kinetic processes in hybridization thermodynamics

4.4.1. Derivation of the Langmuir isotherm

For molecules in contact with a solid surface at a fixed temperature, the Langmuir Isotherm, developed by Irving Langmuir in 1916, describes the partitioning between the gas phase and adsorbed species as a function of applied pressure.

The adsorption process between gas phase molecules, A, vacant surface sites, S, and occupied surface sites, SA, can be represented by the following chemical equation, assuming that there are a fixed number of surface sites present on the surface, as in Figure 5.



When considering adsorption isotherms it is conventional to adopt a definition of surface coverage (θ) which defines the maximum (saturation) surface coverage of a particular adsorbate on a given surface always to be unity, *i.e.* $\theta_{\max} = 1$.

4.4.2. Thermodynamic derivation

An equilibrium constant k can be written in terms of the concentrations of “reactants” and “products”:

$$k = \frac{[SA]}{[S][A]} \quad (10)$$

where:

[SA] is proportional to the surface coverage of adsorbed molecules, or proportional to θ ;

[S] is proportional to the number of vacant sites, $(1 - \theta)$;

[A] is proportional to the pressure of gas, P .

Thus it is possible to define another equilibrium constant, b :

$$b = \frac{\theta}{(1 - \theta)P} \quad (11)$$

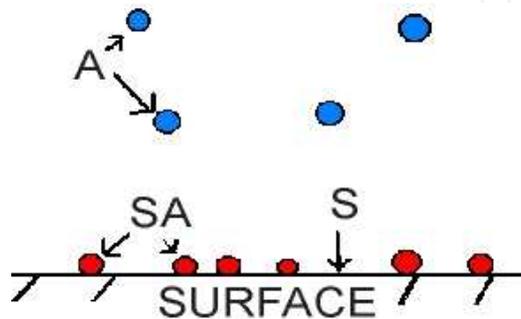


Figure 5. Absorption process.

Rearranging the equations (10) and (11) one can obtain the expression for surface coverage:

$$\theta = \frac{bP}{1 + bP} \quad (12)$$

4.4.3. Kinetic derivation

The equilibrium that may exist between gas adsorbed on a surface and molecules in the gas phase is a dynamic state, *i.e.* the equilibrium represents a state in which the rate of adsorption of molecules onto the surface is exactly counterbalanced by the rate of desorption of molecules back into the gas phase. It should therefore be possible to derive an isotherm for the adsorption process simply by considering and equating the rates for these two processes.

The rate of adsorption will be proportional to the pressure of the gas and the number of vacant sites for adsorption. If the total number of sites on the surface is N , then the rate of change of the surface coverage due to adsorption is:

$$\frac{d\theta}{dt} = k_a p N (1 - \theta) \quad (13)$$

The rate of change of the coverage due to the adsorbate leaving the surface (desorption) is proportional to the number of adsorbed species:

$$\frac{d\theta}{dt} = -k_d N \theta \quad (14)$$

In these equations, k_a and k_d are the rate constants for adsorption and desorption respectively, and p is the pressure of the adsorbate gas. At equilibrium, the coverage is independent of time and thus the adsorption and desorption rates are equal. The solution to this condition gives us a relation for θ , equation (12), where $b = k_a/k_d$. Here b is only a constant if the enthalpy of adsorption is independent of coverage.

4.4.4. Dynamic absorption model

Burden *et al.* [14] develop a dynamic adsorption model based on Langmuir isotherm. If x is the concentration of mRNA target and $\theta(t)$ is the fraction of sites occupied by probe-target duplex, then in the forward absorption, target mRNA attaches to probe at a rate $k_f x (1 - \theta(t))$ proportional to the concentration of specific target mRNA and the fraction $(1 - \theta(t))$ of unoccupied probes; and in the backward desorption reaction, target mRNA detaches from probes at a rate $k_b \theta(t)$ proportional to the fractions of occupied probes. The fraction of probe sites occupied by probe-target duplexes is then given by the differential equation:

$$\frac{d\theta(t)}{dt} = k_f x (1 - \theta(t)) - k_b \theta(t) \quad (15)$$

For the initial condition $\theta(0) = 0$, equation (15) has the following solution:

$$\theta(t) = \frac{x}{x+K} \left[1 - e^{-(x+K)k_f t} \right] \quad (16)$$

where $K = k_b/k_f$.

Using equation (16) Burden *et al.* estimate the measured fluorescence intensity I , with I_0 as the background intensity at zero concentration, to be:

$$I(x, t) = I_0 + \frac{bx}{x+K} \left[1 - e^{-(x+K)k_f t} \right] \quad (17)$$

At equilibrium, the intensity $I(x)$ at target concentration x follows *Langmuir Isotherm* (12):

$$I(x) = I_0 + \frac{bx}{x+K} \quad (18)$$

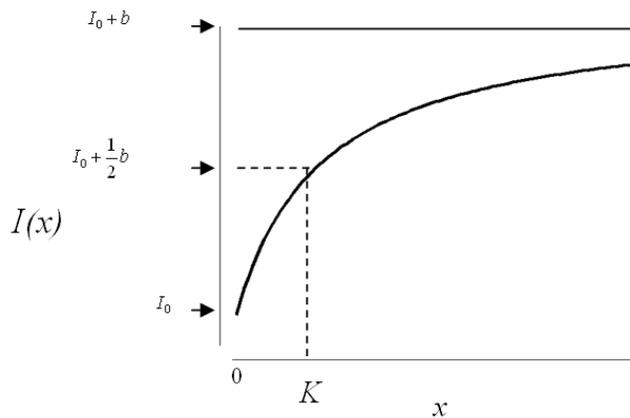


Figure 6. Hyperbolic response function for the intensity $I(x)$ according to the Langmuir isotherm.

5. Hybridization dynamics compensation

5.1. Modelling hybridization by thermodynamics

It is well known that hybridization processes may be seen under the point of view of general thermodynamic conditions [23], meaning that the hybridization probability of a given test segment will be defined by its thermodynamic conditions, *i.e.* by its hybridization temperature. Regarding this, one can state that hybridization process will respond to the dynamic equation:



where P represents the number of oligonucleotides available for hybridization, T the concentration of free RNA target, C the number of bound complexes, k_f and k_b are the respective forward and backwards rate constants for the reaction. This equation has as a natural solution the following expression in the time domain:

$$C(t) = \frac{T}{T+K} [1 - \exp(-t/\tau)] \quad (20)$$

where K defined as in equation (16) is an equilibrium dissociation constant, and $\tau = \frac{1}{k_f(T+K)}$ denoting a characteristic time over which the system reaches equilibrium.

Recent studies [24], [25] confirm the hypothesis that the hybridization process for the each of the probe pairs follows a time model according to the one from Figure 7. This model of evolution predicts that the probability of hybridization will be almost zero if not enough time interval is provided for the experiment to take place, and that in the limit, if enough time is allowed saturation will take place.

A practical solution to the different hybridization dynamics can be solved by using multiple regressions to convey PM-MM probe pairs to equivalent thermodynamic conditions by processing diachronic hybridization experiments [26].

The last procedure will be explained in more detail in the following paragraphs.

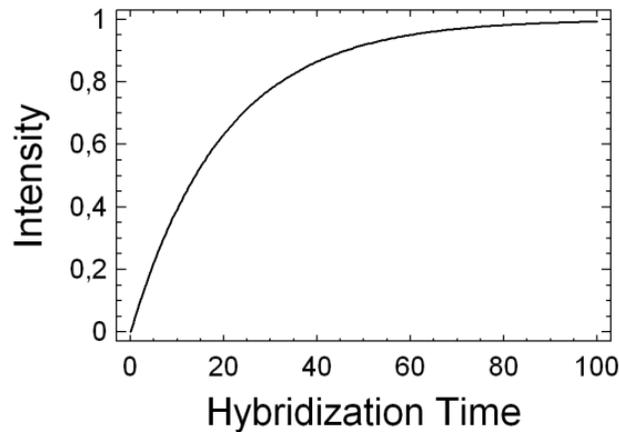


Figure 7. Theoretical model for perfect match hybridization. Intensity of perfect match versus hybridization time. (adapted from [24])

5.2. Exponential regression model

From equation (20) one can assume that a model to solve the multiple regression problem implicit in this study will have the following form:

$$y = a(1 - e^{-bx}) \quad (21)$$

where a and b are parameters to be estimated adaptively using *least square fitting* and the *gradient method*.

Vertical least square fitting proceeds by finding the sum of the squares of the vertical deviations R^2 of parameters a and b :

$$R^2 = \sum_i \left[y_i - a(1 - e^{-bx_i}) \right]^2 \quad (22)$$

where:

$$\varepsilon_i = y_i - a(1 - e^{-bx_i}) \quad (23)$$

is the estimation error incurred for each component.

With this notation equation (22) will become:

$$R^2 = \sum_i \varepsilon_i^2 \quad (24)$$

The condition of R^2 to be at a minimum is that

$$\frac{\partial(R^2)}{\partial a} = 0 \quad (25)$$

$$\frac{\partial(R^2)}{\partial b} = 0 \quad (26)$$

From equations (24), (25) and (26) one will obtain:

$$\frac{\partial(R^2)}{\partial a} = \sum_i \varepsilon_i \frac{\partial \varepsilon_i}{\partial a} = -\sum_i \varepsilon_i (1 - e^{-bx_i}) = 0 \quad (27)$$

$$\frac{\partial(R^2)}{\partial b} = \sum_i \varepsilon_i \frac{\partial \varepsilon_i}{\partial b} = -\sum_i \varepsilon_i a x_i e^{-bx_i} = 0 \quad (28)$$

A solution for equations (27) and (28) can be found using the gradient method. In this case the parameters are going to be computed adaptively:

$$a_{k+1} = a_k - \beta_a \frac{\partial(R^2)}{\partial a} = a_k + \beta_a \sum_i \varepsilon_{i,k} (1 - e^{-b_k x_i}) \quad (29)$$

$$b_{k+1} = b_k - \beta_b \frac{\partial(R^2)}{\partial b} = b_k + \beta_b \sum_i \varepsilon_{i,k} a_k e^{-b_k x_i} \quad (30)$$

where $\varepsilon_{i,k}$ is defined as in equation (23) and β is a parameter used as an adjust step.

5.3. Application for experimental data

The experimental part has been complemented with artificially simulated test probes used for algorithmic validation. A diachronic database was also being produced to estimate hybridization time constants for different gene segments.

Considering these assumptions data records have been created from experimental data fitted by the above mentioned models. The time dynamics of hybridization for both probe sets and their profiles were evaluated at certain time intervals.

Firstly, the diachronic data distribution for an evolution from 0 to 30 minutes is shown in Figure 8 in both cases, for the PM probe set and the MM probe set, and in the following figures, *i.e.* Figure 9 and Figure 10, show this time evolution for 60 and 120 minutes is also shown following the model in equation (20).

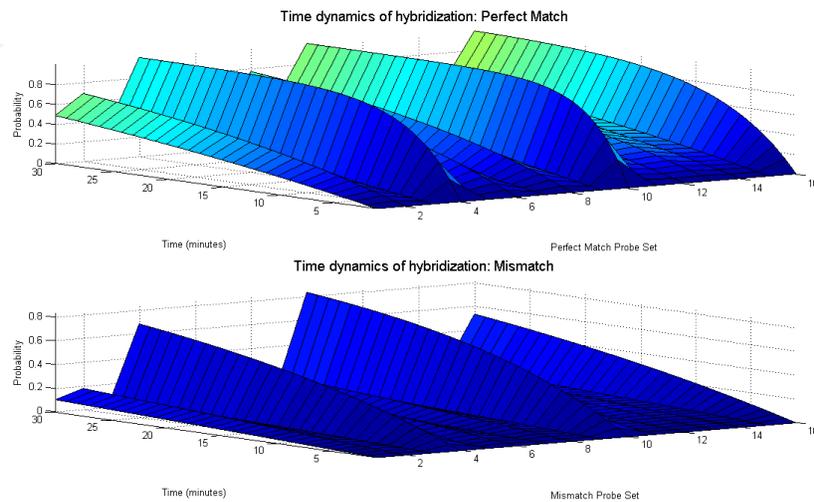


Figure 8. Time dynamics of hybridization corresponding to perfect and mismatch probes, for a maximum of 30 minutes.

The next step on data analysis was to look at the probe profiles, at certain times. Figure 11 shows the regression parameters obtained for time constants. The profiles of the perfect and mismatch were extracted for two different time values underlining the fact that if enough time is allowed to some probes, the mismatches will also hybridize completely.

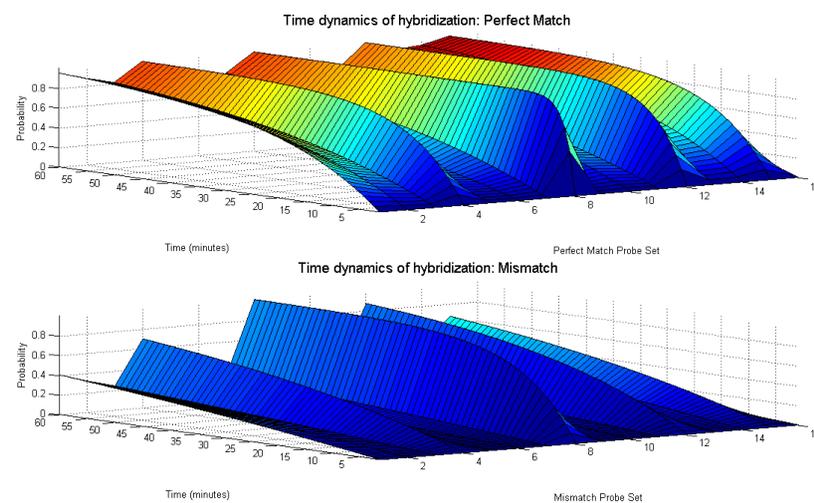


Figure 9. Time dynamics of hybridization corresponding to perfect and mismatch probes, for a maximum of 60 minutes.

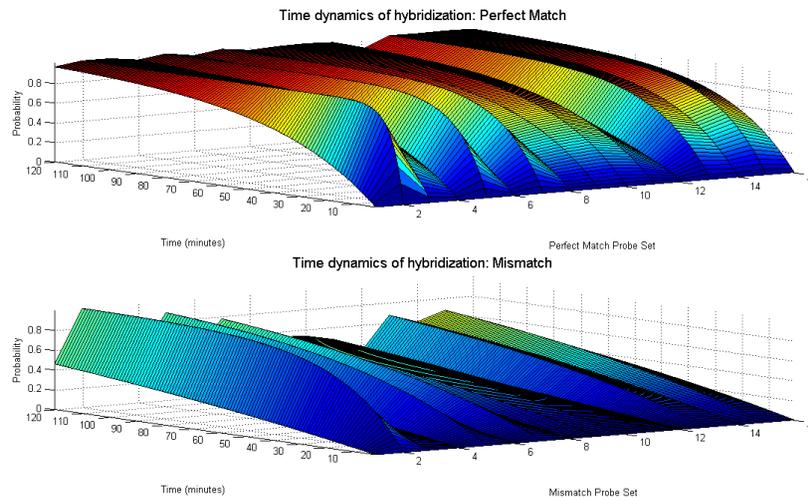


Figure 10. Time dynamics of hybridization corresponding to perfect and mismatch probes, for a maximum of 120 minutes.

Considering this and applying the regression algorithm, we observed that this algorithm searches for the matching values of expression levels of probes sets and for estimated values of perfect and mismatch probes. One of the steps of this iterative algorithm can be seen in Figure 12.

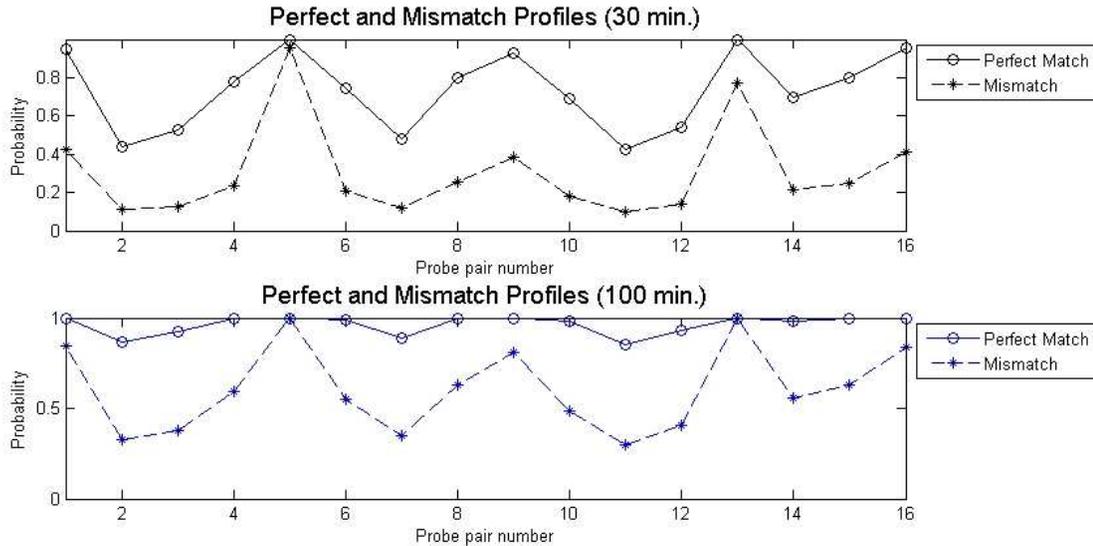


Figure 11. Profiles corresponding to perfect and mismatch probes for time constants, at 30 and 100 minutes.

Once the iterative process was complete, certain probes have reached their target. In the expression level estimation most of the perfect match probes obtained the expected values, while some of the mismatch probes did not reach their target, Figure 13. Similar results were obtained in the case of matching hybridization for time constants.

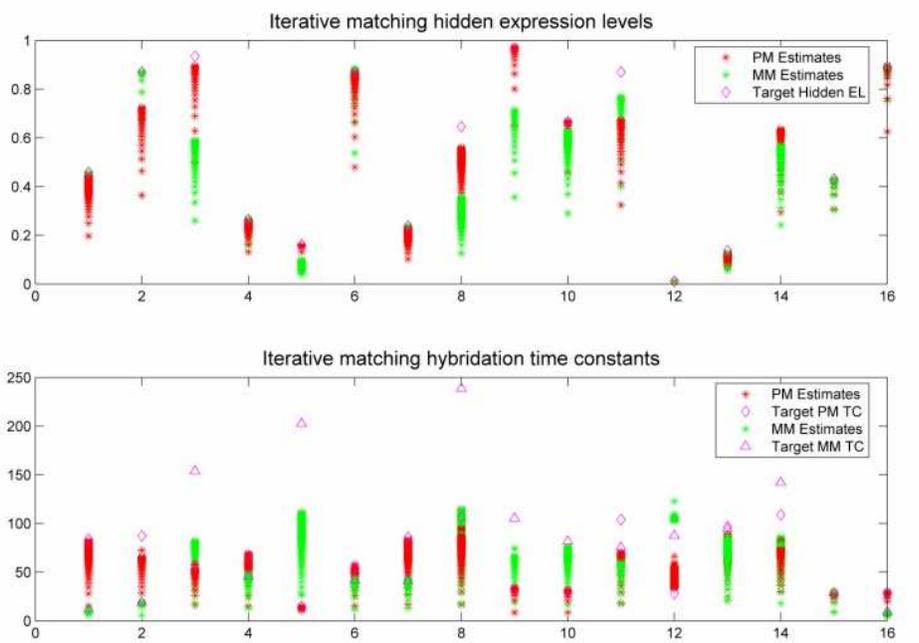


Figure 12. Top template shows the iterative matching for hidden expression levels. Bottom template shows the iterative matching for perfect and mismatch hybridization.

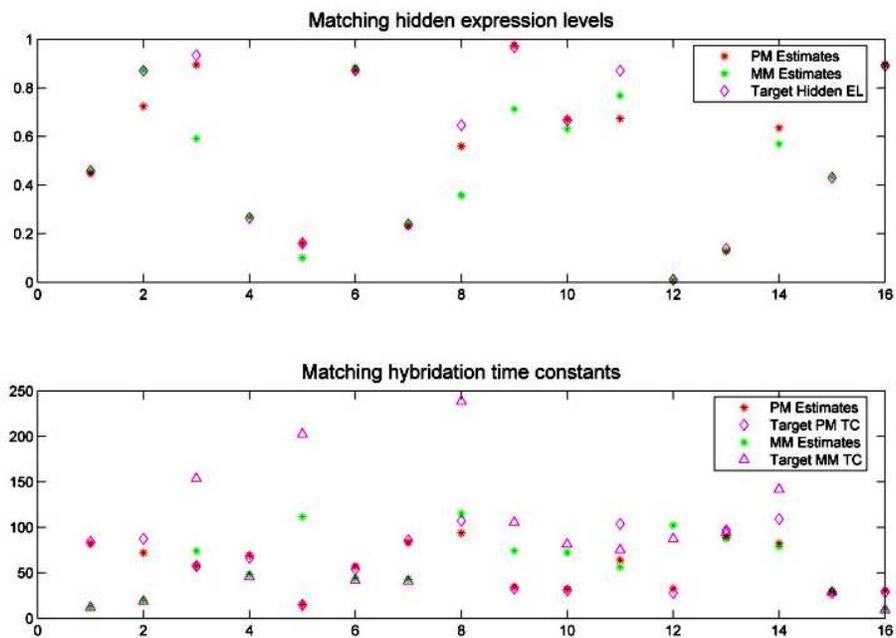


Figure 13. Results for the iterative process of matching.

6. Conclusions

The thermodynamics of oligonucleotide hybridization processes where PM-MM results do not show the expected behaviour, thus affecting to the reliability of expression estimation, was studied in this chapter and the following conclusions were emphasized:

- Modelling the hybridization process through thermodynamical principles reproduces exponential-like behaviour for each P-T segment pair.
- The hybridization process should be confined to the time interval where linear growth is granted, this is, at the beginning of the exponential curve shown in Figure 6.
- Adaptive fitting may be used to predict and regress expression levels on a specific test probe to common thermodynamic conditions. Time constants may be inferred from the regression parameters adaptively.
- The main features of the PM-MM probe sets may be reproduced from probabilistic modelling.
- It may be expected that more precise and robust estimations could be produced using this technique with diachronically expressed hybridization experiments.

Author details

Raul Măluțan*

Communications Department, Technical University of Cluj Napoca, Cluj Napoca, Romania

Pedro Gómez Vilda

DATSI, Universidad Politécnica de Madrid, Madrid, Spain

Acknowledgement

This work was supported by the project "Development and support of multidisciplinary postdoctoral programmes in major technical areas of national strategy of Research - Development - Innovation" 4D-POSTDOC, contract no. POSDRU/89/1.5/S/52603, project co-funded by the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013.

7. References

- [1] Malutan R, Gómez Vilda P, Berindan Neagoe I, Borda M (2011) Hybridization Dynamics Compensation in Microarray Experiments. *Advances in Intelligent and Soft Computing*. 93: 255-261
- [2] Wu P, Nakano S, Sugimoto N (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *European Journal of Biochemistry*. 269:2821 – 2830
- [3] SantaLucia Jr. J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *PNAS on Biochemistry*. 95:1460 – 1465
- [4] Chan V, Graves D.J., McKenzie S.E. (1995) The Biophysics of DNA Hybridization with Immobilized Oligonucleotides Probes. *Biophysical Journal*. 69:2243 – 2255

* Corresponding Author

- [5] Livshits M A, Mirzabekov A D (1996) Theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligonucleotides. *Biophysical Journal*. 71:2795 – 2801
- [6] Dai H, Meyer M, Stepaniants S, Ziman M, Stoughton R (2002) Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Research*. 30(16): e86.1 – e86.8
- [7] Dorris D R, *et al.* (2003) Oligodeoxyribonucleotide probe accessibility on a three-dimensional DNA microarray surface and the effect of hybridization time on the accuracy of expression ratios. *BMC Biotechnology*. 3:6
- [8] Binder H, Preibisch S (2005) Specific and Nonspecific Hybridization of Oligonucleotide Probes on Microarrays. *Biophysical Journal*. 89:337 – 352
- [9] Wang J Y, Drlica K (2003) Modelling hybridization kinetics. *Mathematical Bioscience*. 183:37 – 47
- [10] Naef F, Magnasco M O (2003) Solving the riddle of the bright mismatches: Labeling and effective binding in oligonucleotide arrays. *Physical Review E*, 68:011906-1 – 011906-4
- [11] Wu Z, Irizarry R A (2004) Stochastic models inspired by hybridization theory for short oligonucleotide arrays. *Proc. of the 8th Annual International Conference on Research in Computational Molecular Biology*. 98 – 106
- [12] www.accessexcellence.org/RC/VL/GG/index.html
- [13] Lipshutz R L, Fodor S P A, Gingeras T R, Lockhart D J (1999) High density synthetic oligonucleotide arrays. *Nature Genetics Supplement*. 21:20 – 24
- [14] Burden C, Pittelkow Y E, Wilson S R (2004) Statistical Analysis of Adsorption Models for Oligonucleotide Microarrays. *Statistical Applications in Genetics and Molecular Biology*. 3:35
- [15] Sugimoto N *et al* (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Research*. 24:4501 – 4505
- [16] Zhang L, Miles M F, Aldape K D (2003) A model of molecular interactions on short oligonucleotide microarrays, *Nature Biotechnology*. 21(7):818 – 821
- [17] Huang J C, Morris Q D, Hughes T R, Frey B J (2005) GenXHC: a probabilistic generative model for cross-hybridization compensation in high-density genome-wide microarray data. *Bioinformatics*. 21:i222 – i231
- [18] Black W Z, Hartley J G (1991) *Thermodynamics*. Second Edition. SI Version. Harper Collins Publisher
- [19] SantaLucia Jr J, Allawi H T, Seneviratne P A (1996) Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry*. 35(11): 3555 – 3562
- [20] Delcourt S G, Blake R D (1991) Stacking Energies in DNA. *The Journal of Biological Chemistry*. 266: 15160 – 15169
- [21] Allawi H T, SantaLucia Jr J. Thermodynamics and NMR of Internal G•T Mismatches in DNA. *Biochemistry*. 36:10581 – 10594
- [22] Li C, Wong W H (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *PNAS USA*. 98(1):31 – 36
- [23] El Samad H, Khammash M, Petzold L, Gillespie D (2005) Stochastic Modelling of Gene Regulatory Networks. *Int. Journal of Robust and Nonlinear Control*. 15(15):691 – 711

- [24] Dai H, Meyer M, Stepaniants S, Ziman M, Stoughton R (2002) Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Research*. 30(16):e86.1 – e86.8
- [25] Zhang Y, Hammer D A, Graves D J (2005) Competitive Hybridization Kinetics Reveals Unexpected Behavior Patterns. *Biophysical Journal*. 89:2950 – 2959
- [26] Diaz F, Malutan R, Gomez P, Martinez R, Stetter B, Fe Paz M, Garcia E, Pelaez J (2006) Estimating Oligonucleotide Microarray Expression by Hybridization Process Modelling. *Proc. of IEEE/NLM Life Science Systems and Applications Workshop*. 1 – 2

IntechOpen