Physico-chemical model of surface hybridisation and post-hybridisation washing

Conrad Burden

Mathematical Sciences Institute
Australian National University, Canberra
A comprehensive model of hybridisation plus post-hybridisation washing is set out in

Basic idea: it’s all equilibrium physical chemistry driven by the law of mass-action

**PROBE + TARGET ↔ DUPLEX**

... plus all the other chemical reactions involving nonspecific hybridisation, hybridisation in bulk solution, etc.

Image courtesy of Affymetrix
Table A1. Chemical species present in the model.

<table>
<thead>
<tr>
<th>Species</th>
<th>Unfolded</th>
<th>Folded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific target in the solution</td>
<td>$S$</td>
<td>$S'$</td>
</tr>
<tr>
<td>Non-specific effective target in solution</td>
<td>$N$</td>
<td>$N'$</td>
</tr>
<tr>
<td>Probe at the surface (not bound to the target)</td>
<td>$P$</td>
<td>$P'$</td>
</tr>
<tr>
<td>Duplexes in the solution</td>
<td>$S \cdot S$, $S \cdot N$, $N \cdot N$</td>
<td></td>
</tr>
<tr>
<td>Duplexes at the microarray surface</td>
<td>$P \cdot S$, $P \cdot N$</td>
<td></td>
</tr>
</tbody>
</table>
Assume a set of chemical reactions $r$ each in equilibrium, with equilibrium constant $K_r$:

**Folding**

\[
S \rightleftharpoons S' : \quad [S'] = K_{S\text{fold}}[S].
\]

\[
N \rightleftharpoons N' : \quad [N'] = K_{N\text{fold}}[N].
\]

\[
P \rightleftharpoons P' : \quad [P'] = K_{P\text{fold}}[P].
\]

**Bulk hybridization**

\[
S + N \rightleftharpoons S \cdot N : \quad [S \cdot N] = K_{SN}[S][N].
\]

\[
S + S \rightleftharpoons S \cdot S : \quad [S \cdot S] = K_{SS}[S]^2.
\]

\[
N + N \rightleftharpoons N \cdot N : \quad [N \cdot N] = K_{NN}[N]^2.
\]

**Surface hybridization**

\[
P + S \rightleftharpoons P \cdot S : \quad [P \cdot S] = K_{PS}[P][S].
\]

\[
P + N \rightleftharpoons P \cdot N : \quad [P \cdot N] = K_{PN}[P][N].
\]
Assume a set of chemical reactions $r$ each in equilibrium, with equilibrium constant $K_r$:

**Folding**

\[ S \leftrightharpoons S' : \quad [S'] = K_{S\text{fold}}[S]. \]
\[ N \leftrightharpoons N' : \quad [N'] = K_{N\text{fold}}[N]. \]
\[ P \leftrightharpoons P' : \quad [P'] = K_{P\text{fold}}[P]. \]

**Bulk hybridization**

\[ S + N \leftrightharpoons S \cdot N : \quad [S \cdot N] = K_{SN}[S][N]. \]
\[ S + S \leftrightharpoons S \cdot S : \quad [S \cdot S] = K_{SS}[S]^2. \]
\[ N + N \leftrightharpoons N \cdot N : \quad [N \cdot N] = K_{NN}[N]^2. \]

**Surface hybridization**

\[ P + S \leftrightharpoons P \cdot S : \quad [P \cdot S] = K_{PS}[P][S]. \]
\[ P + N \leftrightharpoons P \cdot N : \quad [P \cdot N] = K_{PN}[P][N]. \]
In addition to the $K_r$, the input parameters are

- **specific target concentration**
  
  $$x_S = [S] + [S'] + [P.S] + [S.N] + [S.S]$$
  
  (= spike-in conc. in spike-in experiments)

- **effective non-specific target concentration**
  
  $$x_N = [N] + [N'] + [P.N] + [S.N] + [N.N]$$

- **effective probe concentration on surface of feature**
  
  $$p = [P] + [P'] + [P.S] + [P.N]$$
The aim (in the first instance) is to determine the coverage fraction $(0 \leq \theta \leq 1)$ on a particular feature of the microarray of fluorescent-dye carrying duplexes

$$\theta = \theta_S + \theta_N = \frac{[P \cdot S]}{p} + \frac{[P \cdot N]}{p}$$

in terms of the input parameters $x_S, x_N, p, K_r (r = PS, PN, SN, ...)$.

From this the observed fluorescence intensity is simply

$$I(x_S) = a + b\theta$$

physical background  saturation intensity
Solving this system for $\theta(x_S)$ is straightforward algebra:

$$
\theta(x_S) = \frac{X_N + K_S(x_S - p\theta_{\text{sum}})}{1 + X_N + K_S(x_S - p\theta_{\text{sum}})}
$$

where $\theta_{\text{sum}}$ is the solution to

$$
\theta_{\text{sum}} = \sum_{\text{features}_f} \frac{K_S^f(x_S - p\theta_{\text{sum}})}{1 + K_S^f(x_S - p\theta_{\text{sum}})}
$$

with the sum taken over all features contributing to the depletion of specific target species $S$. 
Solving this system for $\theta(x_S)$ is straightforward algebra:

$$\theta(x_S) = \frac{X_N + K_S(x_S - p\theta_{\text{sum}})}{1 + X_N + K_S(x_S - p\theta_{\text{sum}})}$$

Note:

• Three parameters:
  
  $K_S$ (effective equilibrium const.) depends on $x_N$ and $K_r$ ($r = PS, PN, SN, ...$)
  
  $X_N$ (non-specific binding strength) depends on $x_N$ and $K_r$ ($r = PS, PN, SN, ...$)
  
  $p$ (probe density)

• If depletion of target from supernatant solution can be ignored, $p = 0$, and $\theta(x_S)$ is a rectangular hyperbola;

• If $x_N = 0$ then $X_N = 0$ and $\theta(0) = 0$;

• At saturation spec. target concentration $x_S \to \infty$, $\theta(x_S) \to 1$. 
Saturation asymptote \( \theta(\infty) = 1 \) does not agree with spike-in experiments:

For PM/MM pair, theory gives \( \theta_{PM}(\infty) = \theta_{MM}(\infty) = 1 \)

Spike-in experiments *always* give \( \theta_{PM}(\infty) > \theta_{MM}(\infty) \)
Differing asymptotes explained by post-hybridisation washing:

Decaying fluorescence intensity with washing is confirmed by experiment
(Binder, H. et al. BMC Bioinformatics 11:291 (2010))
Differing asymptotes explained by post-hybridisation washing
Differing asymptotes explained by post-hybridisation washing
Saturation asymptote $\theta(\infty) = 1$ does not agree with experiment:

Spike-in experiments *always* give $\theta^{PM}(\infty) > \theta^{MM}(\infty)$.
Putting it all together, the fluorescence intensity on a given feature is related to the specific target concentration by a 4-parameter ‘isotherm’:

\[ I(x_S) = A + B \frac{K(x_S - p\theta_{\text{sum}})}{1 + K(x_S - p\theta_{\text{sum}})} \]

where \( \theta_{\text{sum}} \) is the solution to

\[ \theta_{\text{sum}} = \sum_{\text{features} f} \frac{K^f(x_S - p\theta_{\text{sum}})}{1 + K^f(x_S - p\theta_{\text{sum}})} \]

with the sum taken over all features contributing to the depletion of specific target species \( S \).

The 4 fitting parameters \( A, B, K \) and \( p \) can be fitted well to spike-in data ...
Example of fits to Affy U133a spikes:
The fitting parameters

$$I(x_s) = A + B \frac{K(x_s - p \theta_{\text{sum}})}{1 + K(x_s - p \theta_{\text{sum}})}$$

are complicated functions of physical input parameters

$$x_N, K_r (r = PS, PN, SN, ...), p, a, b, w_S, w_N.$$ 

Is it practical to ‘predict’ these parameters from the available data
(probe sequences, distribution of intensities across microarray, ...)?

Maybe/ Maybe not, but the model does give a physical understanding
which helps with building algorithms (Hook curve, Inverse Langmuir , ...).
Understanding the model bit by bit:

1. With no non-spec. hybridization, target depletion or washing, ‘Langmuir isotherm’ is:

\[ I(x_S) = a + b \frac{K_{PS} x_S}{1 + K_{PS} x_S} \]

\[
\begin{align*}
I(x_S) & = a + b (K_{PS})^{-1} x_S \\
I(x_S) & = a + \frac{1}{2} b (K_{PS})^{-1} x_S \\
I(x_S) & = a
\end{align*}
\]

PROBE + SPEC. TARGET \( \leftrightarrow \) DUPELEX

\( 0 \rightarrow K_{PS}^{-1} \rightarrow x_S \)
Understanding the model bit by bit:

2. Switch on non-specific hybridisation and washing

\[ I(x_S) = A + B \frac{Kx_S}{1 + Kx_S} \]
Understanding the model bit by bit:

2. Switch on non-specific hybridisation and washing

\[ I(x_S) = A + B \frac{Kx_S}{1 + Kx_S} \]
Understanding the model bit by bit:

2. Switch on non-specific hybridisation and washing

\[ I(x_S) = A + B \frac{Kx_S}{1 + Kx_S} \]
Understanding the model bit by bit:

2. Switch on non-specific hybridisation and washing

\[ I(x_s) = A + B \frac{Kx_s}{1 + Kx_s} \]
Understanding the model bit by bit:

3. Switch on target depletion

\[ I(x_S) = A + B \frac{K(x_S - p\theta_{\text{sum}})}{1 + K(x_S - p\theta_{\text{sum}})} \]
Understanding the model bit by bit:

4. Switch on target depletion and consider PM/MM pair

\[ I(x_S) = A + B \frac{K(x_S - p\theta_{\text{sum}})}{1 + K(x_S - p\theta_{\text{sum}})} \]
Example of fits to Affy U133a spikes:
Example of fits to Affy U133a spikes: