Detection of point mutations with DNA microarrays: hybridization thermodynamic parameters

Dr. Jef Hooyberghs

Ploen, 9-12 May 2011
PART I

Detection of point mutations with DNA microarrays
Use thermodynamics to:

- Detect presence of mutant
- Identify mutation
- Quantify relative concentration
Setup

The solution:

Abundant wild type = synthetic oligo = target $t$

Mutant in minority = synthetic oligo = target $t'$

Wild type target:

5’ CAGGGCCCTCGTTATCAATGGAGTAGGTTC + spacer + Cy3 3’

Mutant target:

5’ CAGGGCCCGCGTTATCAATGGAGTAGGTTC + spacer + Cy3 3’
Setup

The microarray: Agilent 15K custom array

1006 different probes (15 replicates):

- Perfect match of wild type
- Single mismatch
- Double mismatch
Setup

The microarray: Agilent 15K custom array

1006 different probes (15 replicates):

- Perfect match of wild type
- Single mismatch
- Double mismatch

Affinity to wild type target

Covers the entire intensity-range of the scan device

Also affinity range for mutant target
Thermodynamics, pure target

Pure target $t$, concentration $c$

- Two-state process
- In thermodynamic equilibrium
- Far from spot saturation

$\rightarrow$ Langmuir Isotherm

$I = A.c.e^{\frac{\Delta G}{RT}}$

$\Delta G =$ Probe-target affinity

Microarray Spot

Probe with or without mismatch(es)
Thermodynamics, pure target

For short sequences $\Delta G$ can be calculated from:

Nearest neighbour model

Thermodynamic parameters fitted from microarray experiments

Hooyberghs J., Van Hummelen P. and Carlon E.
Nucleic Acids Research, 37, 7 e53 (2009)
Experimental result, pure target

Target concentration \( c = 50 \text{ pM} \); \( T = 65^\circ \text{C} \); hybridisation time = 17h

Wild type target \( t \)

Mutant target \( t' \)

Branches: detection & identification mutant

"baseline" fit

\[ \Delta \Delta G \text{ (kcal/mol)} \]

\[ -9 -8 -7 -6 -5 -4 -3 -2 -1 0 \]

\[ -1 \]

\[ 0 \]

\[ 10^0 \]

\[ 10^1 \]

\[ 10^2 \]

\[ 10^3 \]

\[ 10^4 \]
Experimental result, pure target

Deviating branch =

Cluster of probes with a mutation
fixed position
fixed type
versus the wild type target

Most deviating branch
→ identification of mutant
Thermodynamics, two targets

Wild type (abundant) = target $t \rightarrow \Delta G$
Mutant (in minority) = target $t' \rightarrow \Delta G'$

Competitive hybridisation

$$I = c \cdot F(\Delta G) + c' \cdot F(\Delta G')$$

$F =$ baseline fit

In theory:
when $c' \ll c$ the mutant contribution can be significant for probes with $\Delta G' > \Delta G$
Experimental result, two targets

Total concentration $c'+c = cte = 50\text{pM}$
Conclusions

DNA microarray & Thermodynamics of hybridisation

Point mutation, up to a minority of a few %

• Detection
• Identification
• Estimation of concentration

The detection limit

• will depend on mutation type
• will improve if system is in equilibrium
PART II

Hybridization thermodynamic parameters

Fitting with optimal design
Setup

Agilent 44K custom array

Target, \( L = 25 \): \( \text{CTGGTCTTAGATGCAGCGACTGTTT} \)

\( \Rightarrow \) equilibrium is reached after 17 h of hybridization at 65°C

<table>
<thead>
<tr>
<th>Probeset: (15 spot replicates/probeseq.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( \underline{\text{AAACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>2 ( \underline{\text{TAACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>3 ( \underline{\text{CAACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>4 ( \underline{\text{GAACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>5 ( \underline{\text{ATACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>( \cdots )</td>
</tr>
<tr>
<td>76 ( \underline{\text{AAACAGTCGCTGCATCTAAGACCAGC}} )</td>
</tr>
<tr>
<td>77 ( \underline{\text{TTACAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>78 ( \underline{\text{TATCAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>79 ( \underline{\text{TAAGAGTCGCTGCATCTAAGACCAG}} )</td>
</tr>
<tr>
<td>( \cdots )</td>
</tr>
<tr>
<td>2776 ( \underline{\text{AAACAGTCGCTGCATCTAAGACCGC}} )</td>
</tr>
</tbody>
</table>
Optimal design

Procedure to fit the free energy parameters $\Delta G_\alpha \alpha = 1 \ldots 58$ $(10 + 48)$

Intensity of spot i: 1 ... 2776:

$$ I_i = Ace^{\Delta G/RT} $$

Relative to PM: $A \epsilon$

$$ y_i = \ln(I_i) - \ln(I_{PM}) = -\frac{\Delta G - \Delta G_{PM}}{RT} = -\frac{\Delta \Delta G}{RT} $$

NN model parameters:

$$ y_i = \sum_{\alpha=1}^{58} X_{i\alpha} \omega_\alpha $$

$$ \omega_\alpha = \frac{\Delta G_\alpha}{RT} $$  

The 58 NN dinucleotide parameters we want to fit

$$ X_{i\alpha} $$  

The design matrix, determined by sequence of target & probes
Optimal design

In matrix notation: \( \bar{y} = X\hat{\omega} \)

Overdetermined set of equations, linear in \( \omega_{\alpha} \)

Least square estimate: the estimator \( \hat{\omega} \) which minimizes \( S = (\bar{y} - X\hat{\omega})^2 \)

\[
\hat{\omega} = (X^TX)^{-1}X^T\bar{y} \\
\text{var}(\hat{\omega}_\alpha) = \sigma \left[(X^TX)^{-1}\right]_{\alpha\alpha}
\]

\[\rightarrow\text{Variance matrix: related to design (sequences)}\]

\[\rightarrow\text{Variance of residuals: related to model and measurement}\]
Optimal design

Different popular criteria:

D-optimal design: Determinant

minimize $|(X^T X)^{-1}|$

A-optimal design: Average variance

minimize $\text{Trace} (X^T X)^{-1}$

E-optimal design: Eigenvalue

minimize maximal eigenvalue of $(X^T X)^{-1}$
Optimal design

Procedure:

Note: Probeset is fixed when target sequence is chosen

• (part of) human genome was scanned for possible targets
• The candidates were ranked according to the 3 criteria
• $\rightarrow$ CTGGTCTTAGATGCAGCGACTGTTT
Results

Analysis is still ongoing

We have recent results concerning

the free energy penalty due to a MM

when more than one MM is present
Results

Spot Intensity

\[
\begin{align*}
I_{PM} & \quad \text{CTGGTCTTAGATGCAGCGACTGTTT} \\
& \quad \text{GACCAGAATCTACGTCGCTGACAAA} \\
I_{MM} & \quad \text{CTGGTCTTAGATGCAGCGACTGTTT} \\
& \quad \text{GACCAGAATCTCGTCGCTGACAA} \\
I_{PM+} & \quad \text{CTGGTCTTAGATGCAGCGACTGTTT} \\
& \quad \text{GACCAGAATCTACGTCGCTGCACA} \\
I_{MM+} & \quad \text{CTGGTCTTAGATGCAGCGACTGTTT} \\
& \quad \text{GACCAGAATCTCGTCGCTGCACA} \\
\end{align*}
\]

\[
\Delta \Delta G \left( \begin{array}{c} \text{ATG} \\ \text{TTC} \end{array} \right) = RT \frac{\ln(I_{PM})}{\ln(I_{MM})}
\]

\[
\Delta \Delta G \left( \begin{array}{c} \text{ATG} \\ \text{TTC} \end{array} \right) = RT \frac{\ln(I_{PM+})}{\ln(I_{MM+})}
\]

Provided MM1 and MM2 do not interact
Results

Note:

Data is filtered on

\( I > \text{noise} \)
\( I < \text{saturation} \)
Preliminary conclusion

The free energy penalty due to isolated MM’s appears to be non-additive.

The penalty due to a MM is larger when a second MM is present, except when the second MM is at the boundary.
Acknowledgements:

Prof. Dr. Enrico Carlon
Dr. Jean-Charles Walter

Drs. Wahyu Wijaya Hadiwikarta