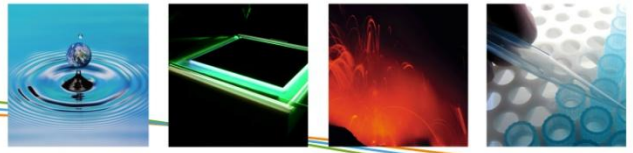
vito
vision on technology



09/05/2011

Detection of point mutations with DNA microarrays: hybridization thermodynamic parameters

Dr. Jef Hooyberghs

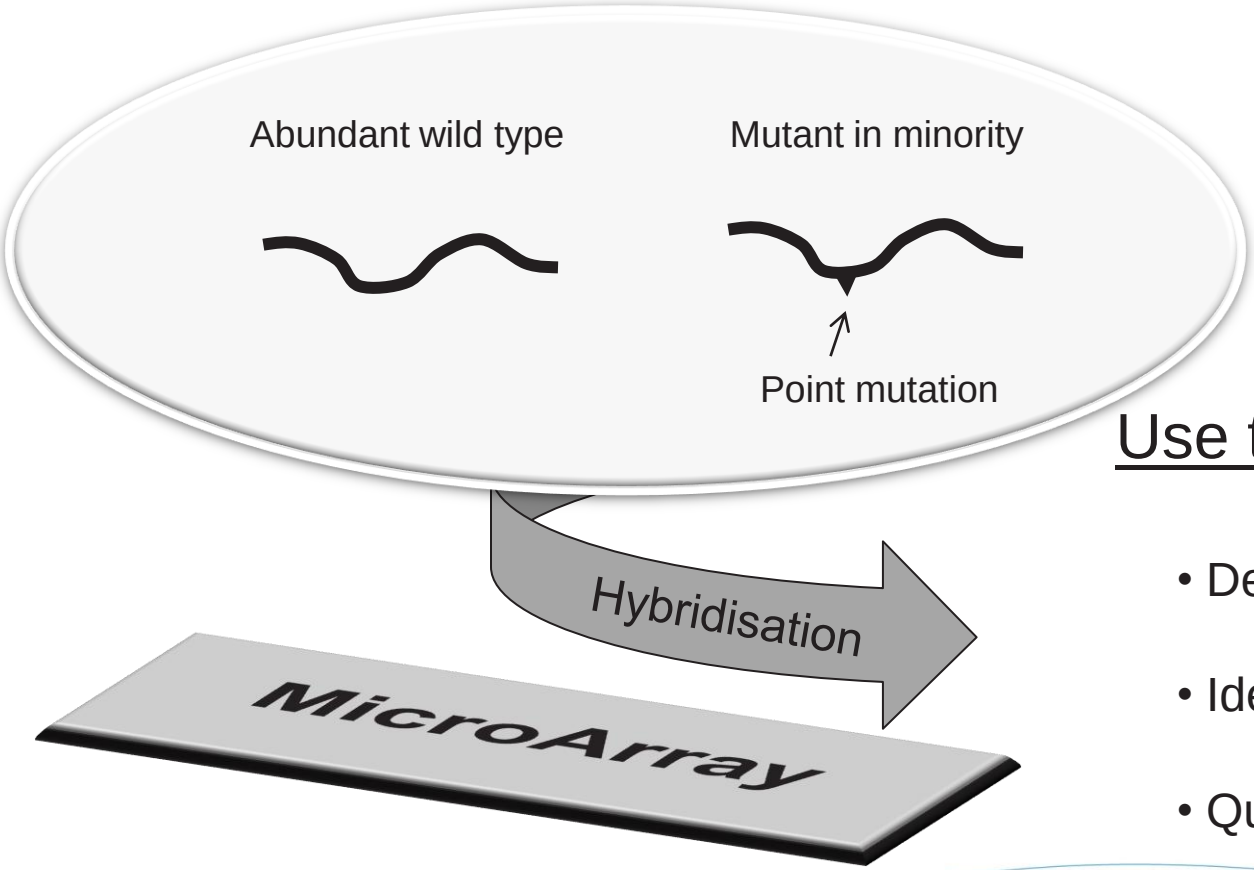
Ploen, 9-12 May 2011

PART I

Detection of point mutations with DNA microarrays

Aim

Solution: mixture of 2 DNA sequences



Use thermodynamics to:

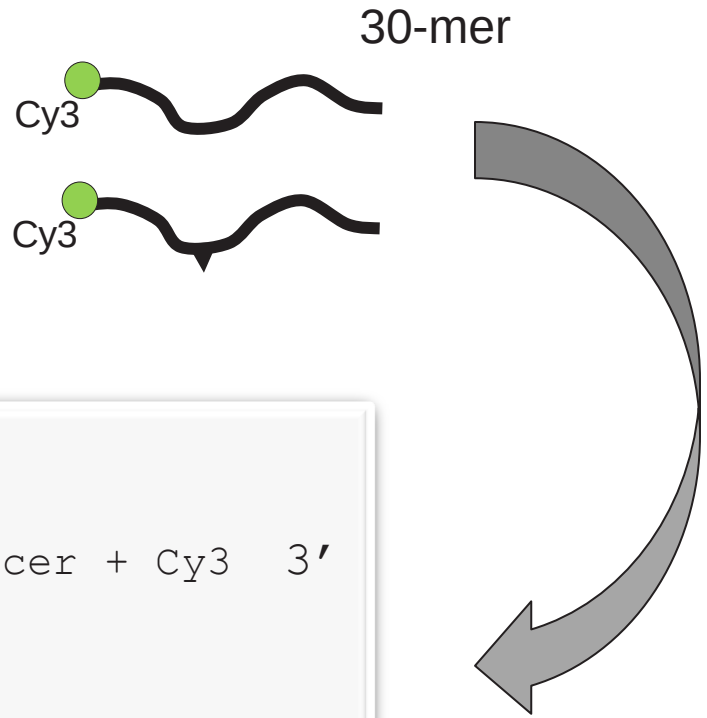
- Detect presence of mutant
- Identify mutation
- Quantify relative concentration

Setup

The solution:

Abundant wild type = synthetic oligo = target ***t***

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



Wild type target:

5′ CAGGGCCTCGTTATCAATGGAGTAGGTTTC + spacer + Cy3 3′

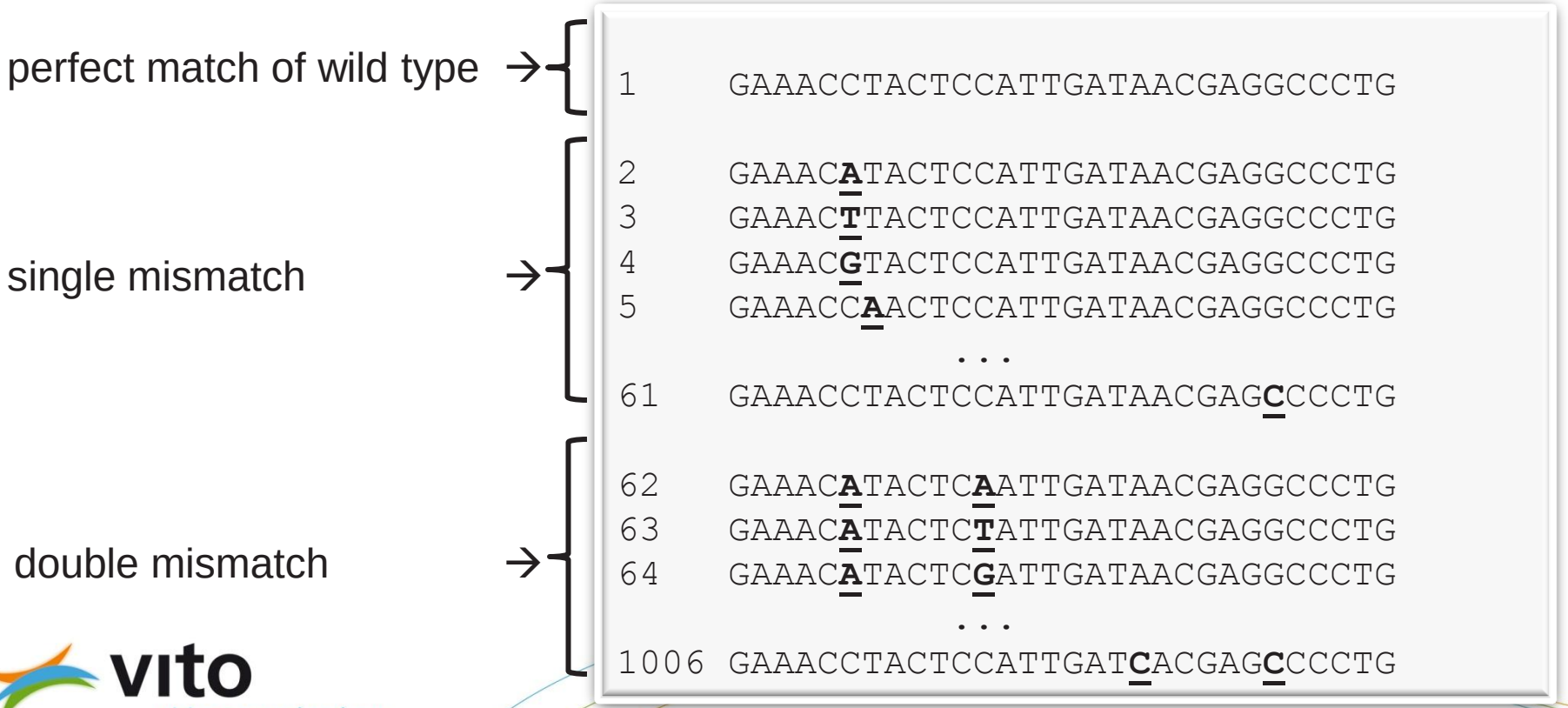
Mutant target:

5′ CAGGGCC**G**CGTTATCAATGGAGTAGGTTTC + spacer + Cy3 3′

Setup

The microarray: Agilent 15K custom array

1006 different probes (15 replicates):



Setup

The microarray: Agilent 15K custom array

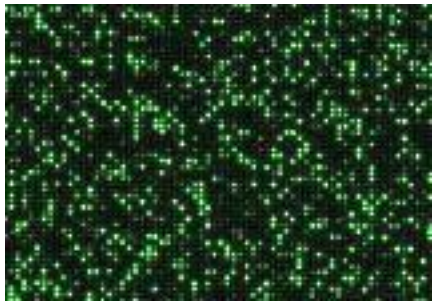
1006 different probes (15 replicates):

Affinity to wild type target

perfect match of wild type →

single mismatch →

double mismatch →

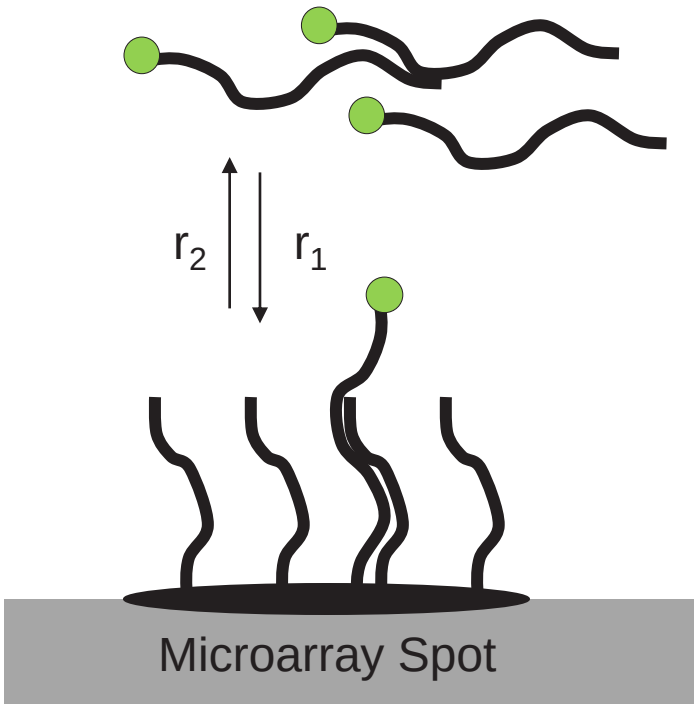


Covers the entire intensity-range of the scan device

Also affinity range for mutant target

Thermodynamics, pure target

Pure target ***t***, concentration ***c***



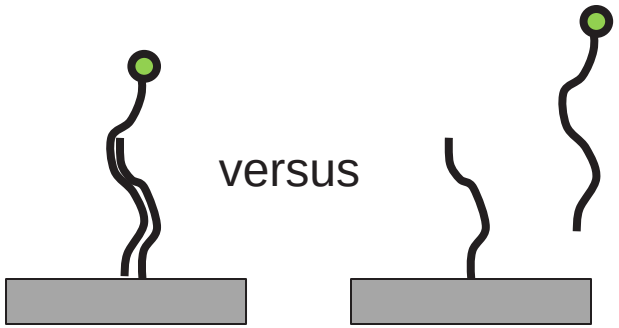
Probe with or without mismatche(s)

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

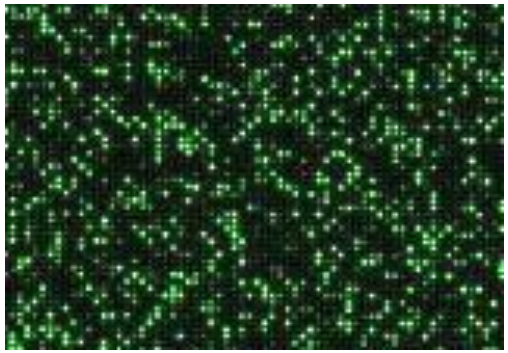
→ Langmuir Isotherm

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

$\Delta G =$ Probe-target affinity



Thermodynamics, pure target



Intensity of each spot (probe):

$$I = F(\Delta G)$$

ΔG Depends on probe sequence

For short sequences ΔG can be calculated from:

Nearest neighbour model

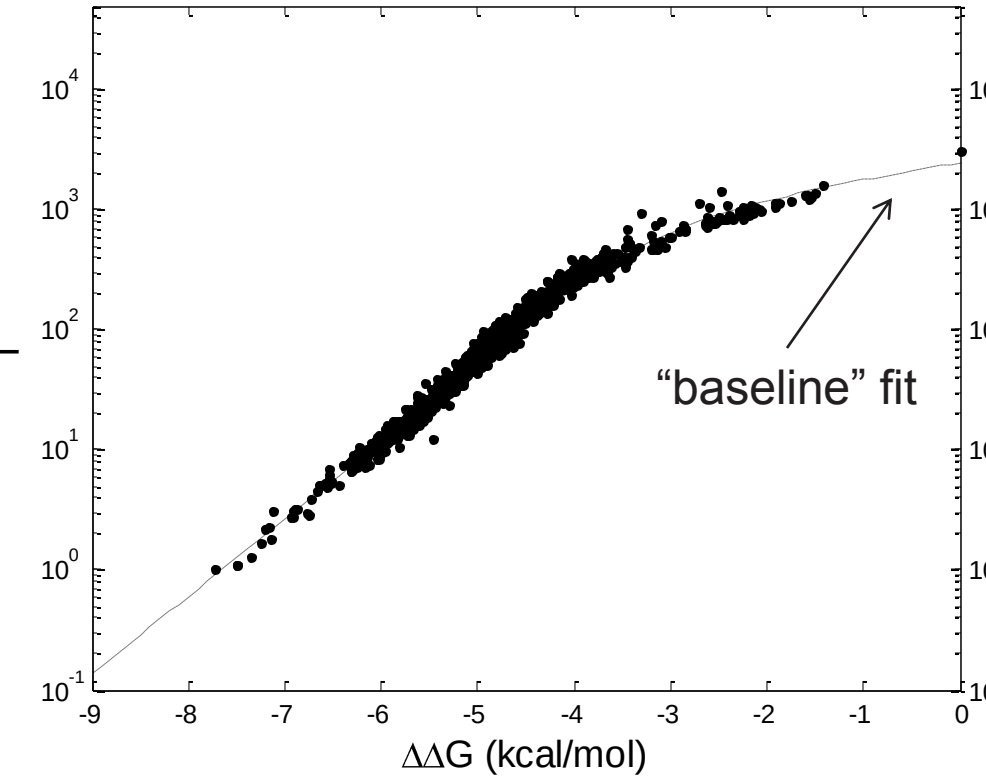
Thermodynamic parameters fitted from microarray experiments

Hooyberghs J., Van Hummelen P. and Carlon E.
*Nucleic Acids Research, **37**, 7 e53 (2009)*

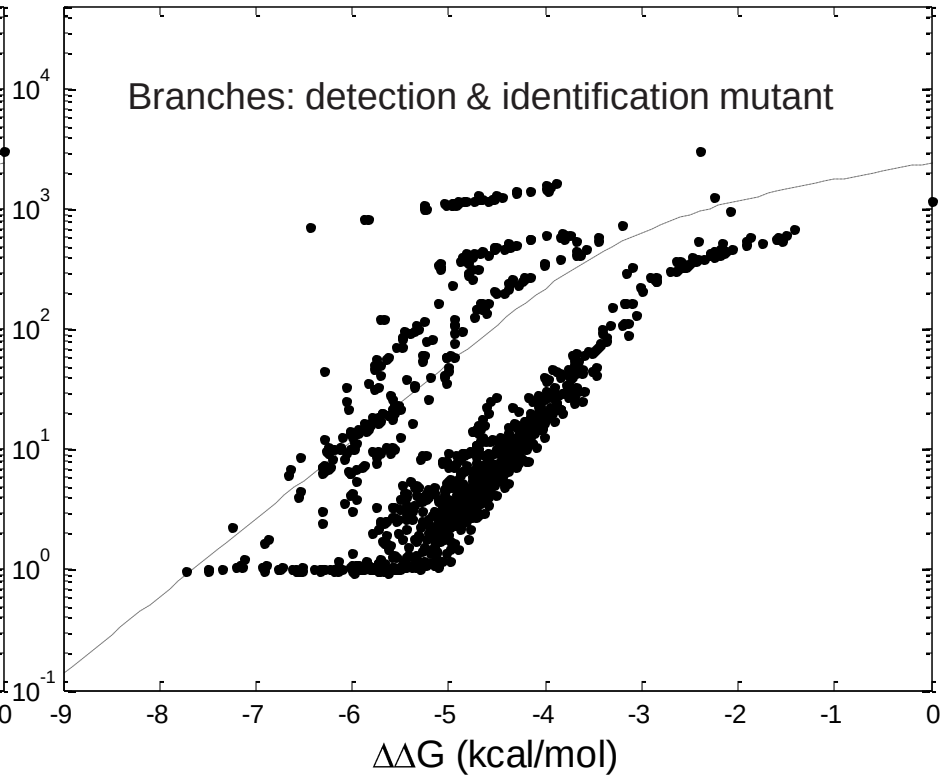
Experimental result, pure target

Target concentration c = 50 pM; T = 65° C; hybridisation time = 17h

Wild type target ***t***



Mutant target ***t'***

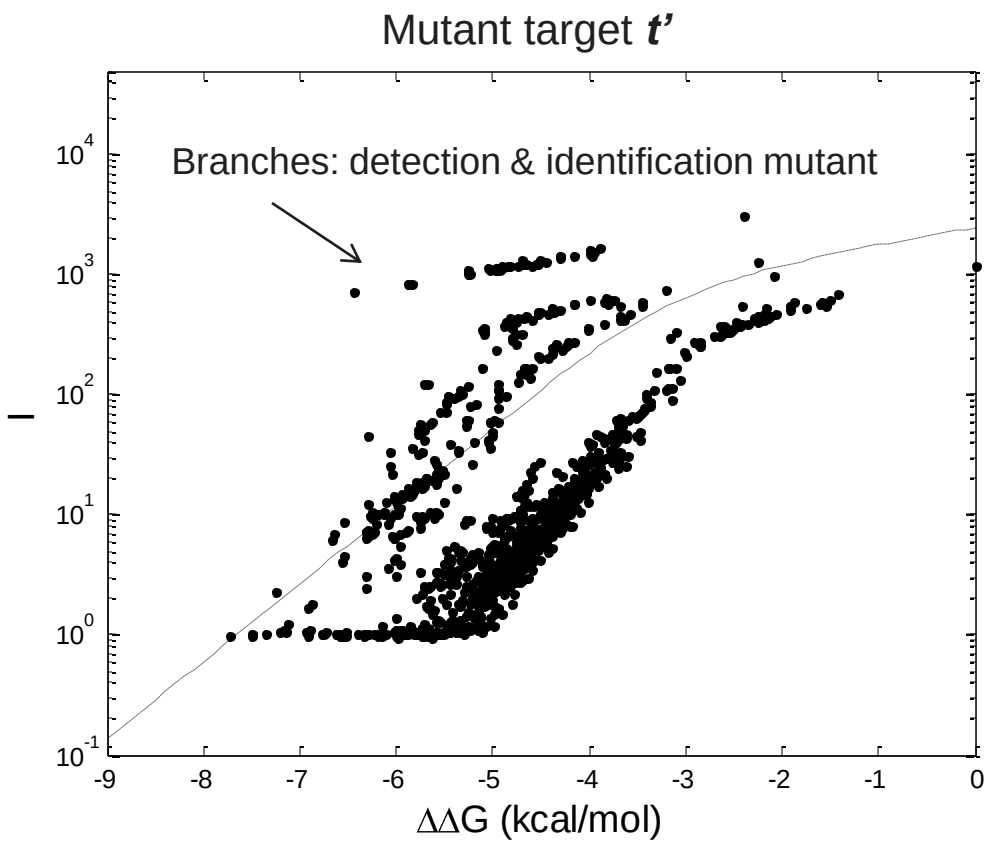


Experimental result, pure target

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

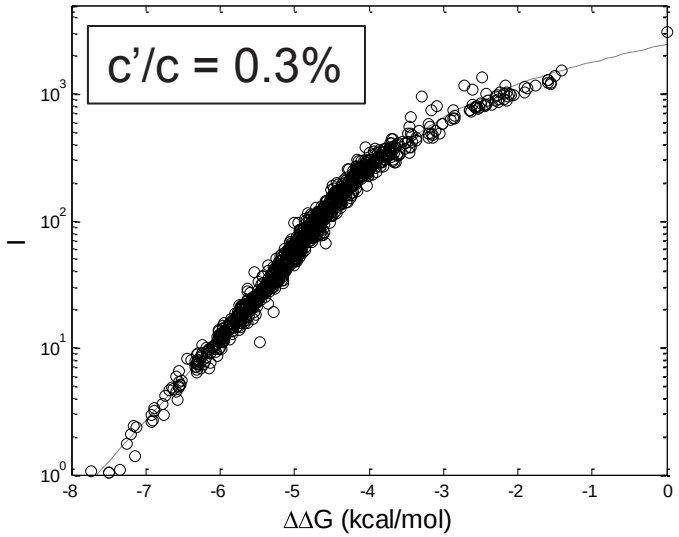
$$I = c.F(\Delta G) + c' F(\Delta G')$$

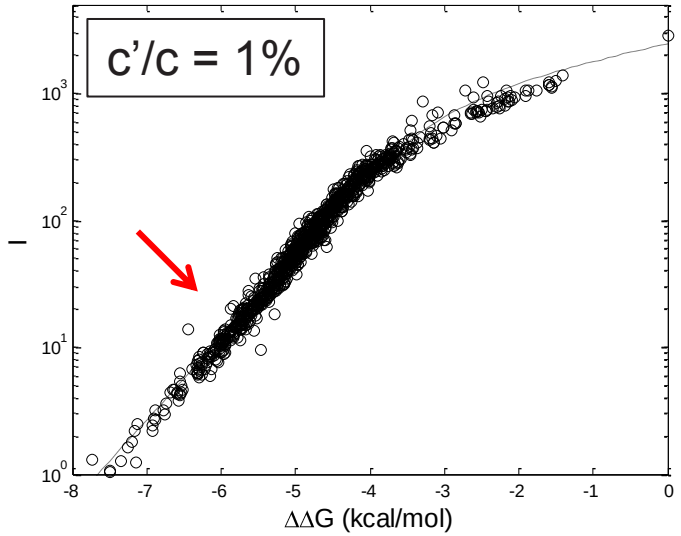
F = baseline fit

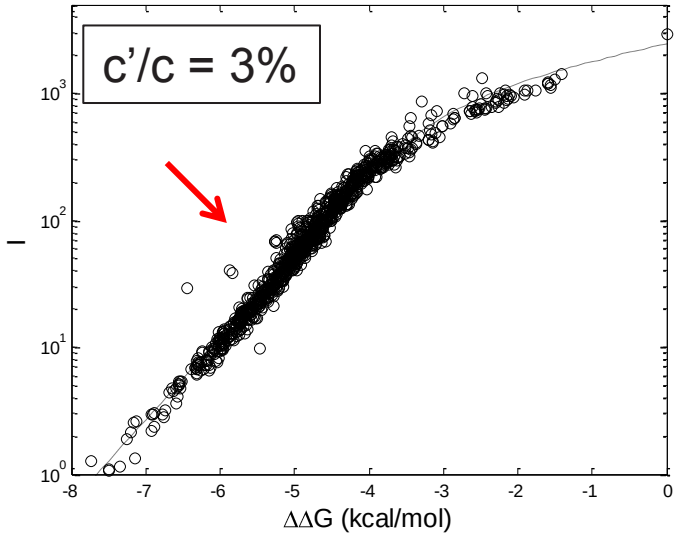
<u>In theory:</u>
when c' << c the mutant contribution can be significant for probes with $\Delta G' > \Delta G$

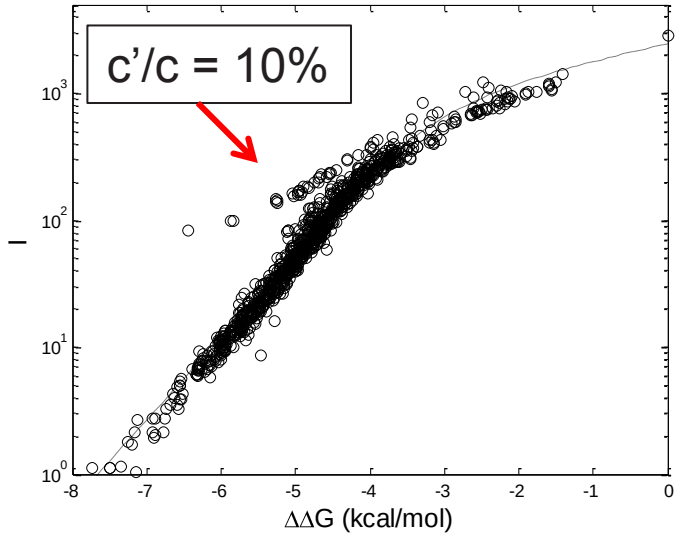
Experimental result, two targets

Total concentration c'+c = cte = 50pM









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: CTGGTCTTAGATGCAGCGACTGTTT

equilibrium is reached after 17 h of hybridization at 65°C

Probeset: (15 spot replicates/probeseq.)

perfect match of wild type →

single mismatch →

double mismatch →

```
1     AAACAGTCGCTGCATCTAAGACCAG

2     TAACAGTCGCTGCATCTAAGACCAG
3     CAACAGTCGCTGCATCTAAGACCAG
4     GAACAGTCGCTGCATCTAAGACCAG
5     ATACAGTCGCTGCATCTAAGACCAG
                 ...
76    AAACAGTCGCTGCATCTAAGACCAC

77    TTACAGTCGCTGCATCTAAGACCAG
78    TATCAGTCGCTGCATCTAAGACCAG
79    TAAGAGTCGCTGCATCTAAGACCAG
                 ...
2776  AAACAGTCGCTGCATCTAAGACCGC
```

vito vision on technology

Optimal design

Procedure to fit the free energy parameters ΔG_α α = 1 … 58 (10 + 48)

Intensity of spot i: 1 … 2776: $I_i = Ace^{\Delta G / RT}$

Relative to PM: A c $y_i = \ln(I_i) - \ln(I_{PM}) = -\frac{\Delta G - \Delta G_{PM}}{RT} = -\frac{\Delta\Delta G}{RT}$

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, determinded by sequence of target & probes

Optimal design

In matrix notation: $\vec{y} = X\vec{\omega}$

Overdetermined set of equations, linear in ω_α

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

$$\hat{\vec{\omega}} = (X^T X)^{-1} X^T \vec{y}$$

$$\mathrm{var}(\hat{\omega}_\alpha) = \sigma\,[(X^T X)^{-1}]_{\alpha\alpha}$$

Variance matrix: **related to design (sequences)**

Variance of residuals: related to model and measurement

Optimal design

Different popular criteria:

D-optimal design: Determinant $\text{minimize} \left|(X^T X)^{-1}\right|$

A-optimal design: Average variance $\text{minimize} \ \text{Trace}\,(X^T X)^{-1}$

E-optimal design: Eigenvalue $\text{minimize} \ \text{maximal eigenvalue of} \ (X^T X)^{-1}$

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
- → 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

I_{PM}

CTGGTCTTAGATGCAGCGACTGTTT
GACCAGAATCTACGTCGCTGACAAA

I_{MM}

CTGGTCTTAGATGCAGCGACTGTTT
GACCAGAATCT**T**CGTCGCTGACAAA

$$\Delta\Delta G\begin{pmatrix} \text{ATG} \\ \text{T}\underline{\mathbf{T}}\text{C} \end{pmatrix} = RT\,\frac{Ln(I_{PM})}{Ln(I_{MM})}$$

I_{PM+}

CTGGTCTTAGATGCAGCGACTGTTT
GACCAGAATCTACGTCGCT**C**ACAAA

I_{MM+}

CTGGTCTTAGATGCAGCGACTGTTT
GACCAGAATCT**T**CGTCGCT**C**ACAAA

$$\Delta\Delta G\begin{pmatrix} \text{ATG} \\ \text{T}\underline{\mathbf{T}}\text{C} \end{pmatrix} = RT\,\frac{Ln(I_{PM+})}{Ln(I_{MM+})}$$

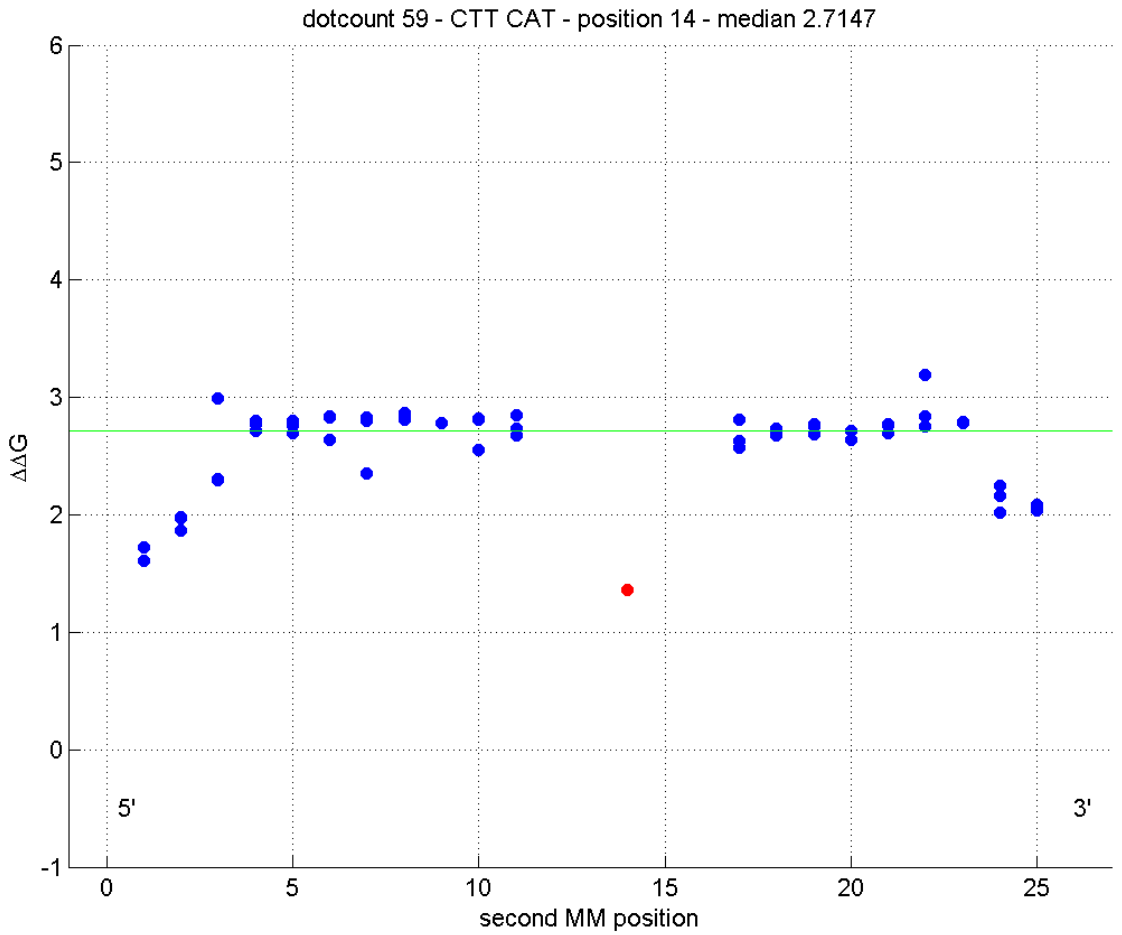
Provided
MM1 and MM2
do not interact

Results

Note:

Data is filtered on

I > noise
I < 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:

KATHOLIEKE UNIVERSITEIT LEUVEN

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

vito vision on technology

Drs. Wahyu Wijaya Hadiwikarta