Microarray hybridization: Learning about RNA-degradation, G-stacks and binding isotherms

© Hans Binder & Mario Fasold

University Leipzig

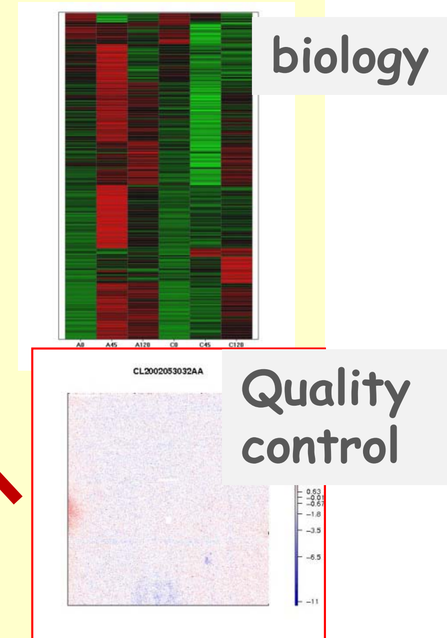
Interdisciplinary Centre for Bioinformatics

Plön May/2011

Microarrays are tools for studying

1. biology
   (expression, copy numbers, genotypes…)

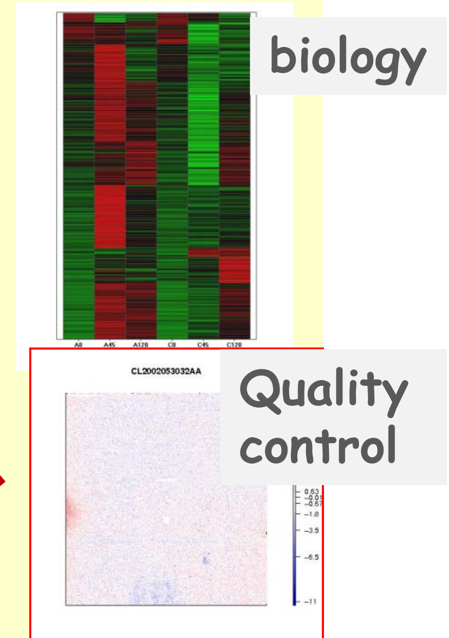
Calibration:
extracting proper expression values



Microarrays are tools for studying

1. biology
   (expression, copy numbers, genotypes…)

Calibration:
extracting proper expression values



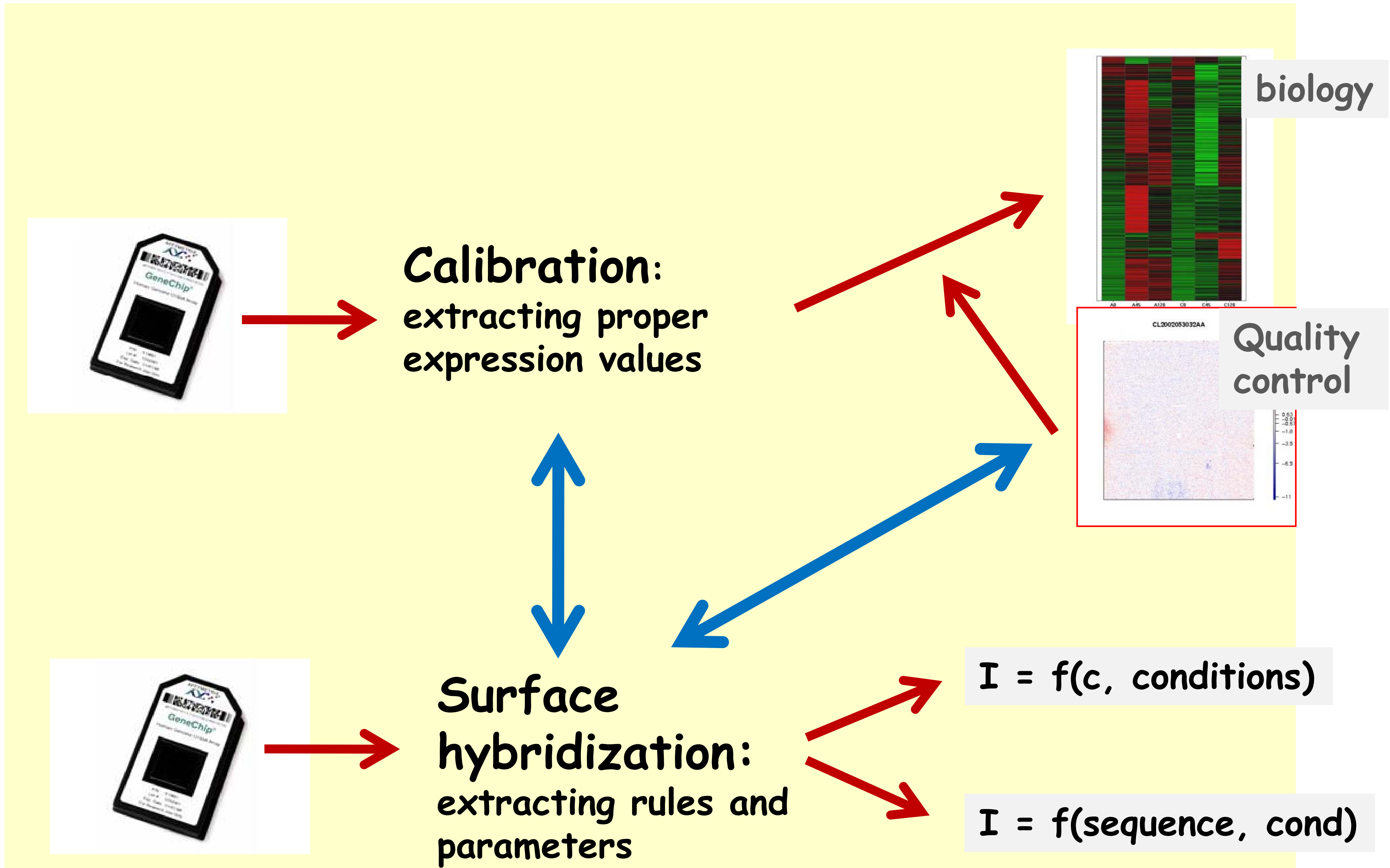
2. surface hybridization
   (isotherms, molecular interactions)

Surface hybridization:
extracting rules and parameters

I = f(c, conditions)

I = f(sequence, cond)

Microarrays are tools for studying



Bad and good things about microarrays

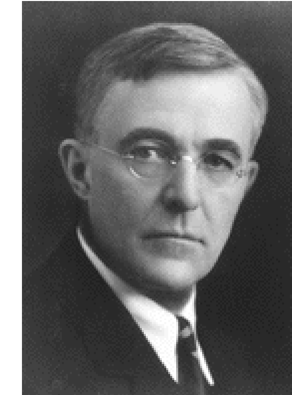
a) bad things about microarrays – effects which distort intensity

- non-specific background
- sequence-specific affinity
- saturation of probe spots
- RNA degradation, 3'/5'-bias
- specific sequence motifs: e.g. GGG

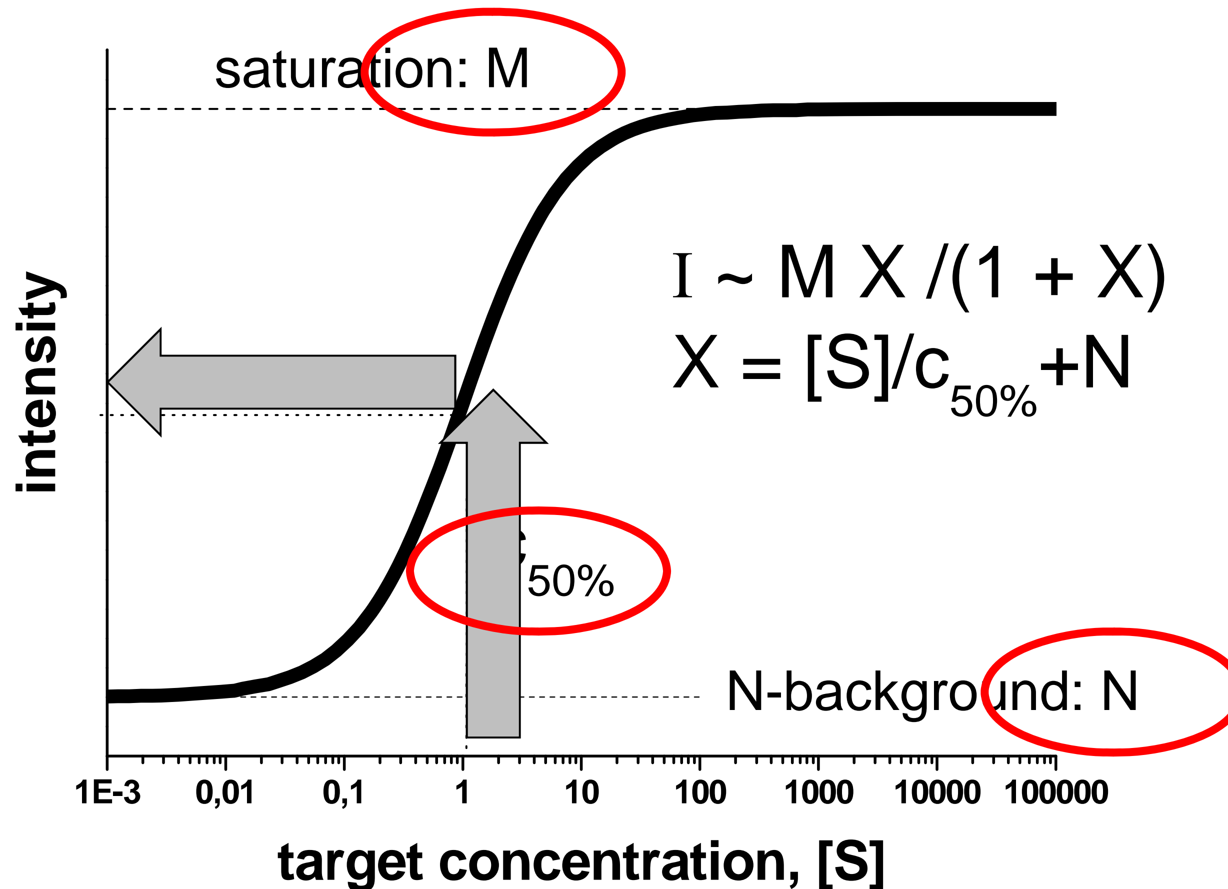
Good things:

- To solve: Suited rules (isotherms, interactions) and sufficient set-parameters that can be estimated for each array
- Physico-chemical stuff: e.g. degradation and GGG-effects

The inverse (Langmuir-) isotherm problem



Irvin Langmuir
(1881-1957)
Nobel prize 1932



$$I \sim M\,X/(1 + X)$$
$$X = [S]/c_{50\%} + N$$

The inverse (Langmuir-) isotherm problem

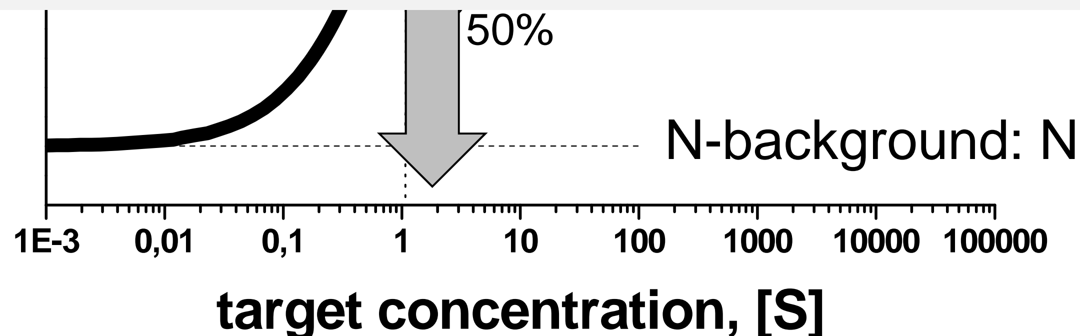
muir
)
1932

saturation: M

Options to estimate the parameters:

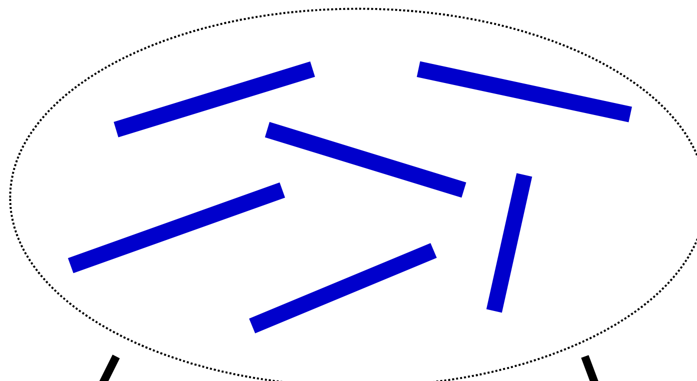
1. Special calibration data (e.g. spiked-in) for each array

2. Use ‚ordinary' probe data as intrisic reference (‚hook' method)
→ Expression values (inverse Langmuir solution)
→ Quality control



Situation 1: PM/MM-pairings

targets



identical targets

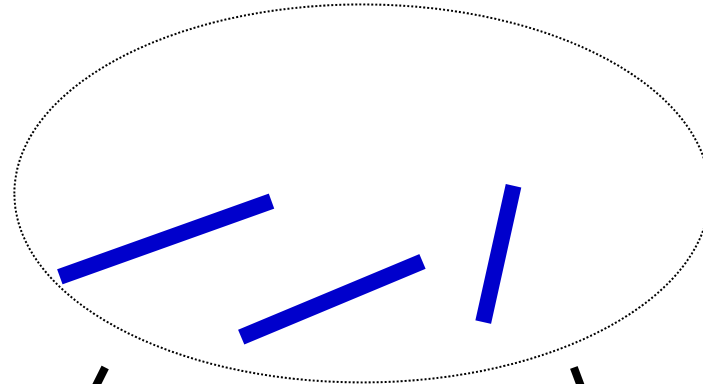
strong affinity : $c_{50\%}^{probe}$

weak affinity : $c_{50\%}^{ref}$

PM	MM	‚old affy' expression
allele A	cross allele B	SNP-arrays

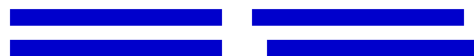
Situation 1: PM/MM-pairings

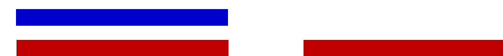
targets



strong affinity : $c_{50\%}^{probe}$

weak affinity : $c_{50\%}^{ref}$





probes

internal reference

microarray

PM MM ‚old affy' expression

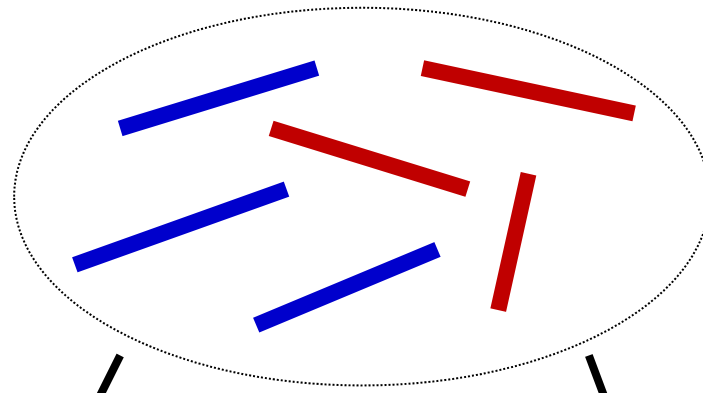
allele A allele B SNP-arrays

Matched mismatched duplexes

Situation 2: PM-only

IZBI University Leipzig

targets



Targets interrogating the same transcripts

strong affinity : $c_{50\%}^{probe}$

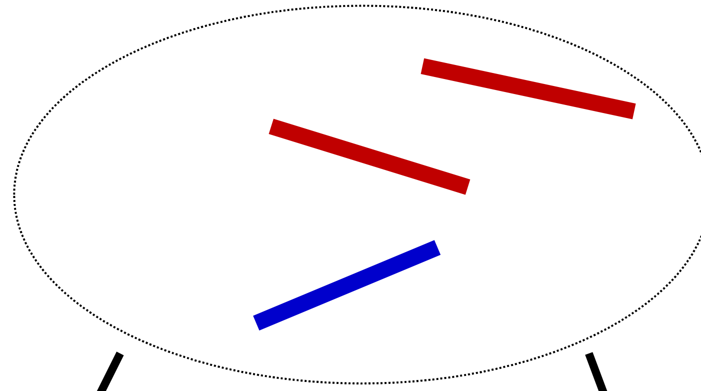
weak affinity : $c_{50\%}^{ref}$

PM(strong) PM(weak) PMonly expression

Strongly-bound weakly-bound matched duplexes

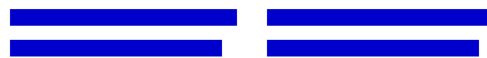
Situation 2: PM/PM-pairings

targets



strong affinity : $c_{50\%}^{probe}$

weak affinity : $c_{50\%}^{ref}$



probes

internal reference

microarray

PM(strong)

PM(weak)

PMonly expression

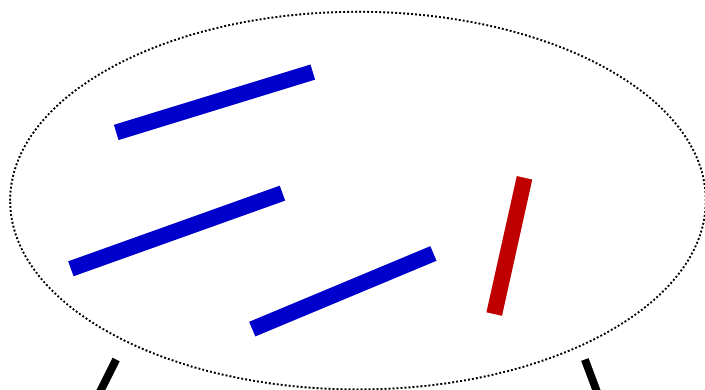
Strongly-bound

weakly-bound

matched duplexes

Situation 3: high/low-conc-pairings

targets



Targets interrogating the same transcripts

High conc.: $[S]^{high}$

Low conc.: $[S]^{low}$

PM(3')
PM(before wash)

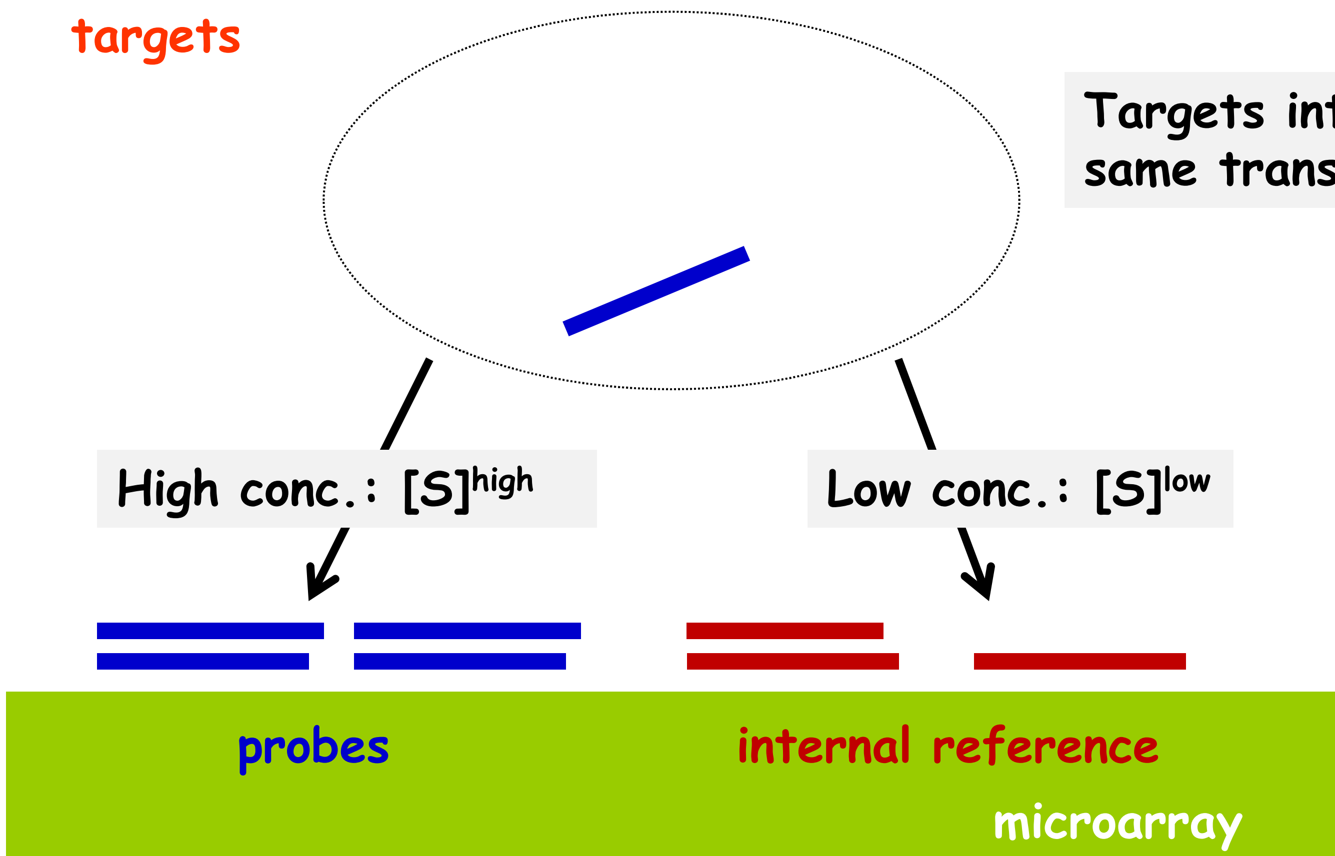
PM(5')
PM(after wash)

PMonly expression

Strongly-bound weakly-bound matched duplexes

Situation 3: high-conc/low-conc-pairings

IZBI University Leipzig



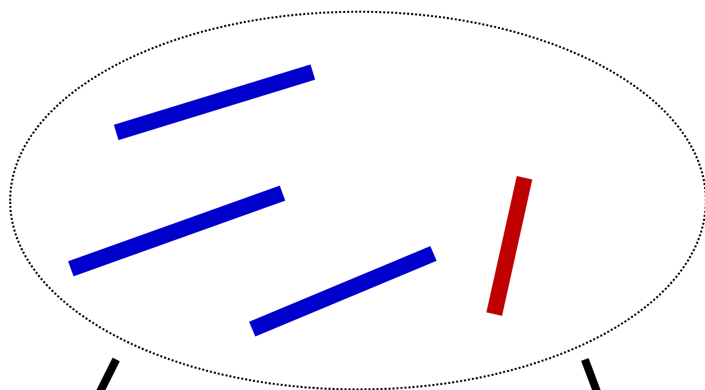
Targets interrogating the same transcripts

PM(3')	PM(5')	P-sets
PM(before wash)	PM(after wash)	

Highly- weakly- populated duplexes

General: high/low-binding strength

targets



Targets interrogating the same transcripts

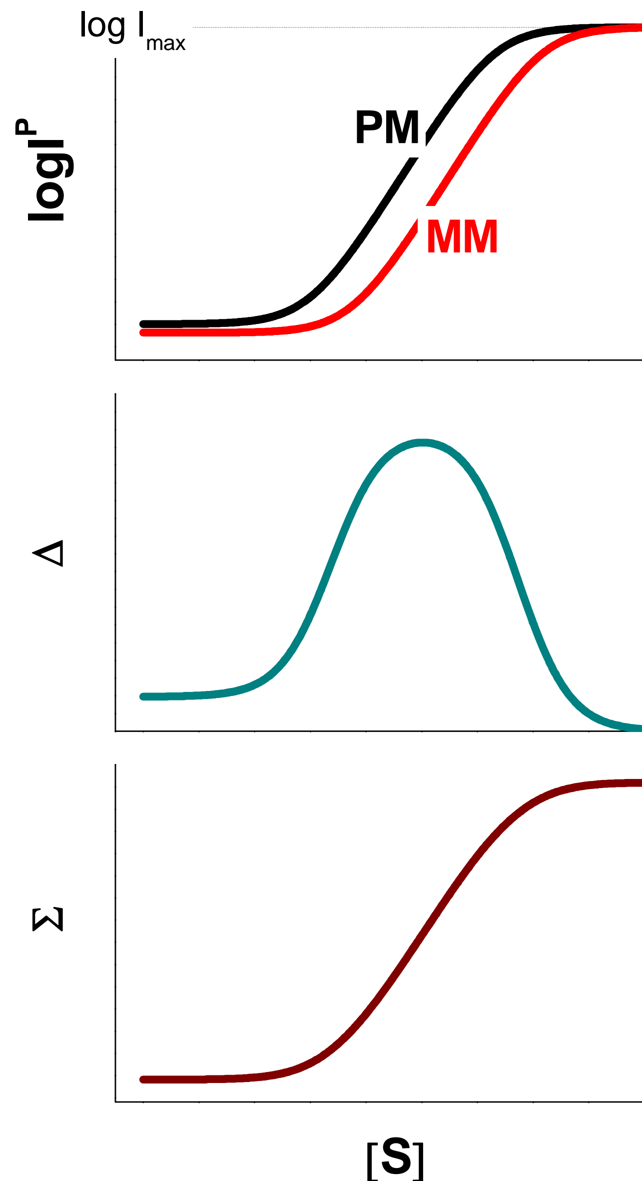
High : X

Low: X^{ref}

$X = [S]/c_{50\%}$

Highly- weakly- populated duplexes

Example: PM/MM-pairings



Let's use the MM as internal reference, by calculating

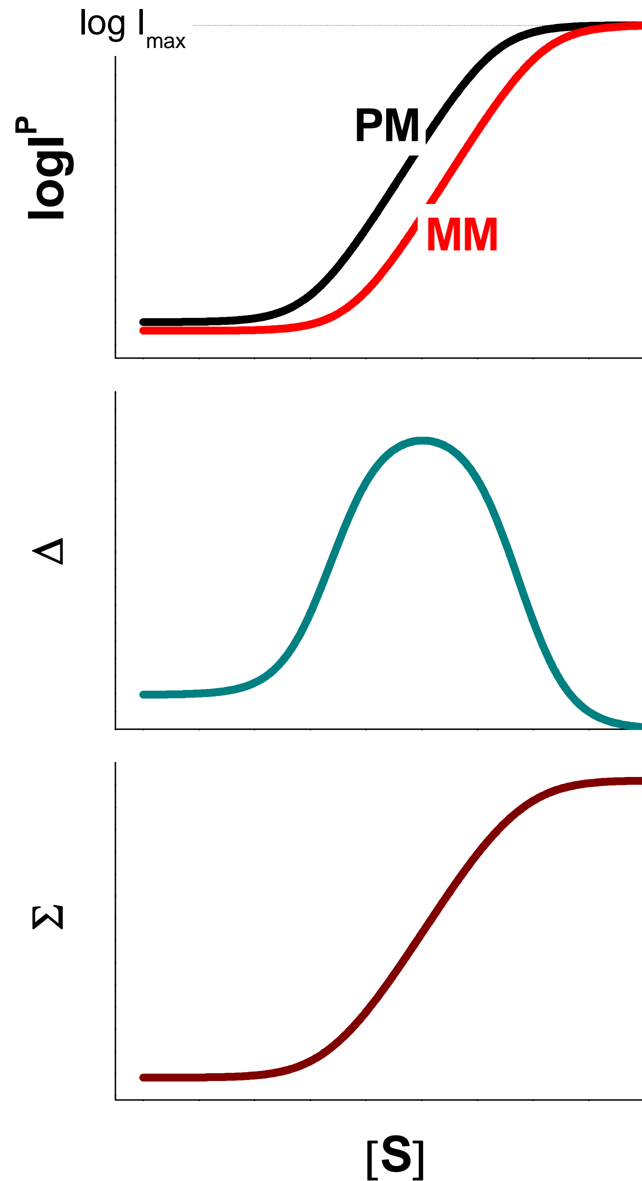
- the difference

$$\Delta \equiv \log PM - \log MM$$

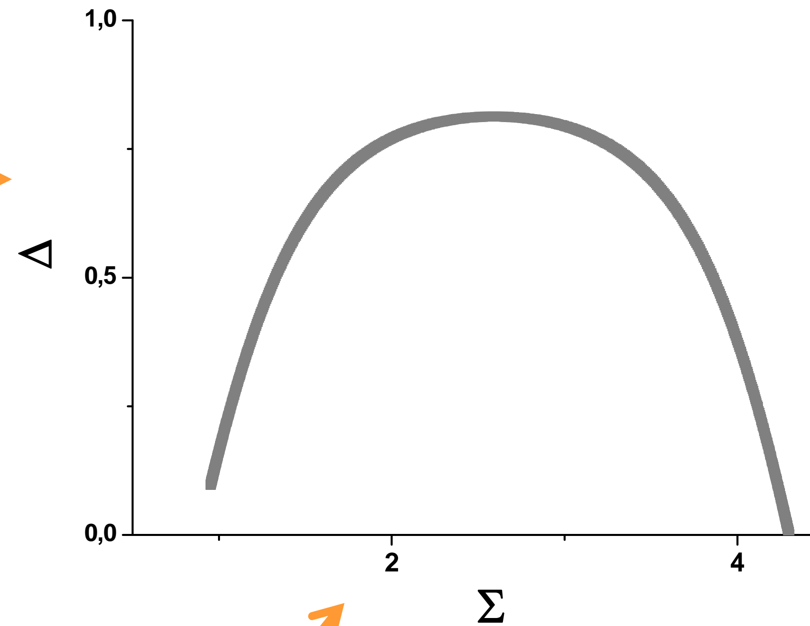
- the mean

$$\Sigma \equiv \tfrac{1}{2}\left(\log PM + \log MM\right)$$

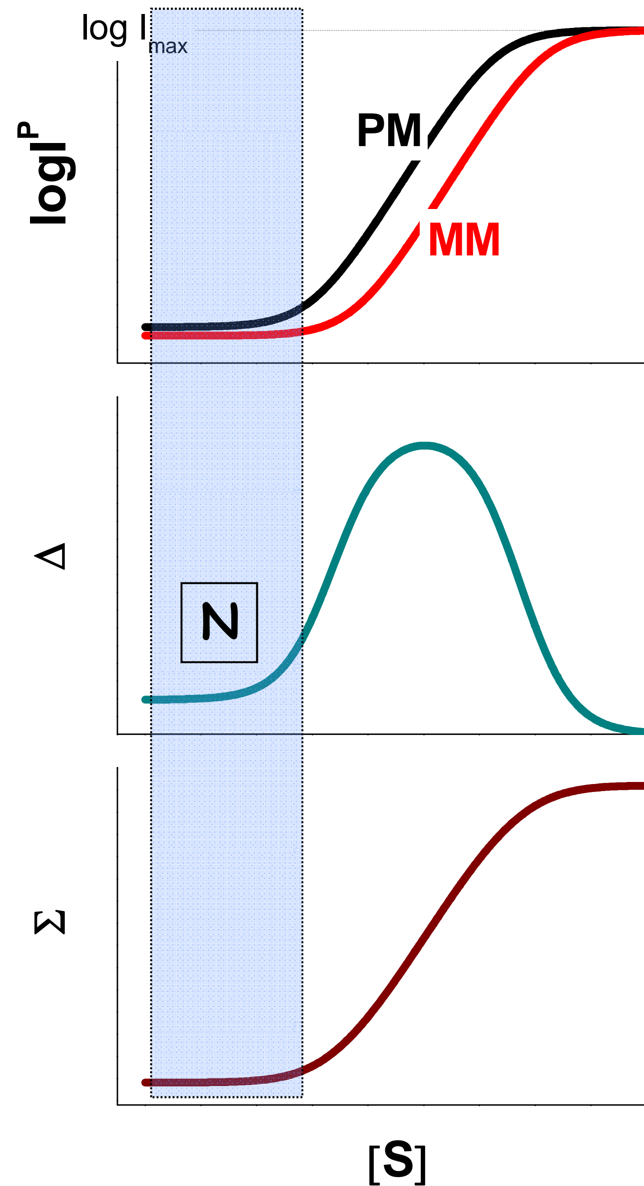
The Δ-vs-Σ plot



… and by plotting Δ-vs-Σ into a sort of M-A-plot

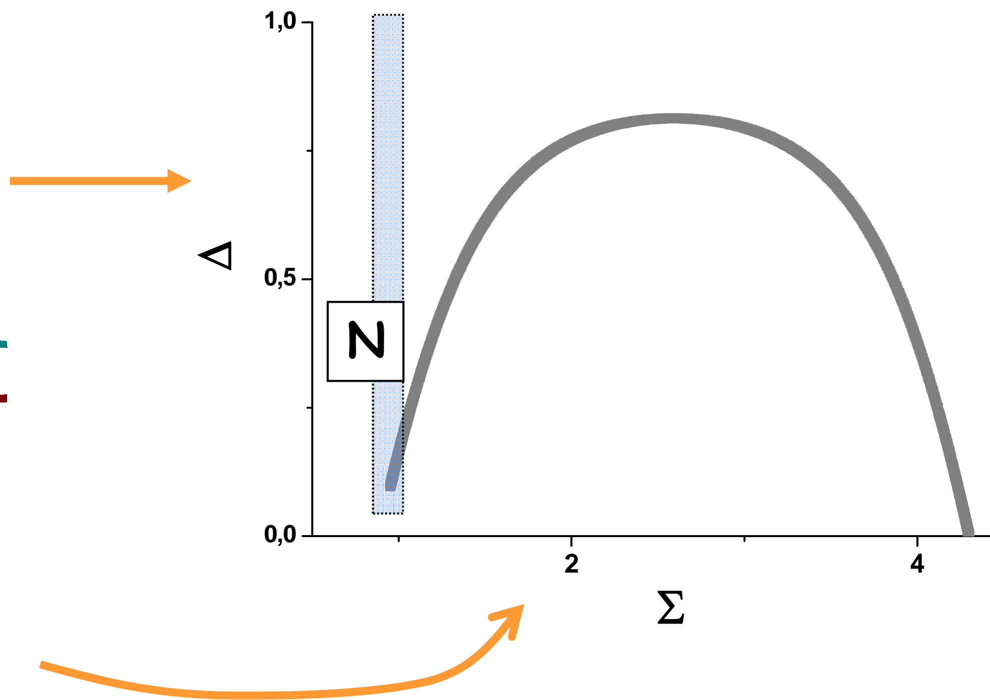


The hybridization regimes



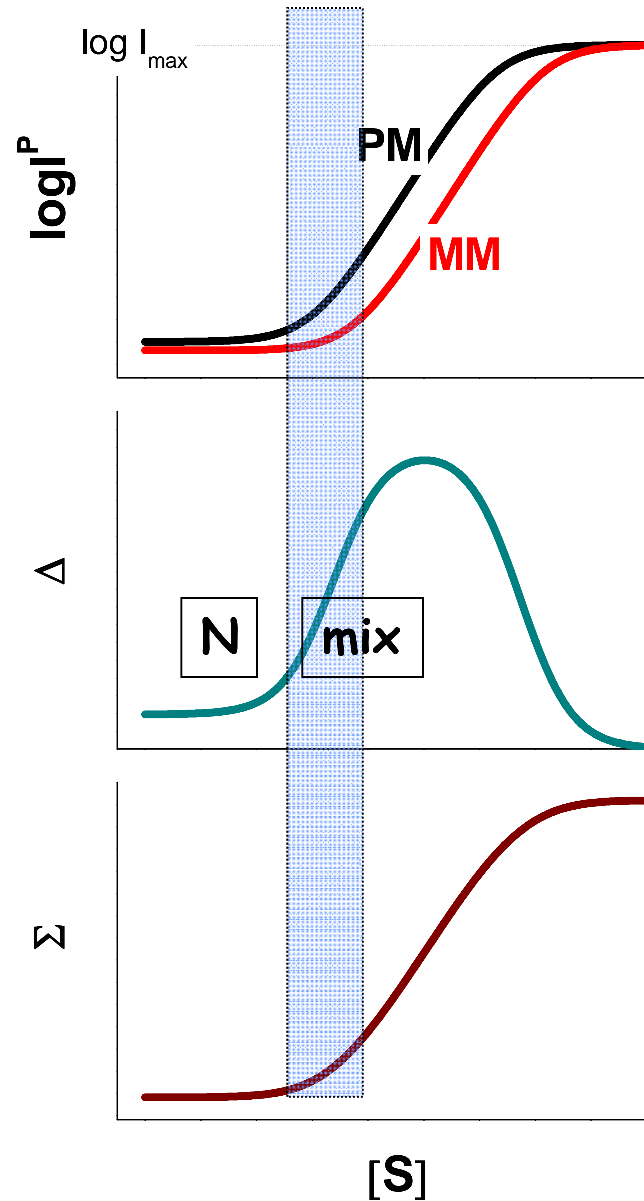
... can be clearly identified:

N... non-specific binding



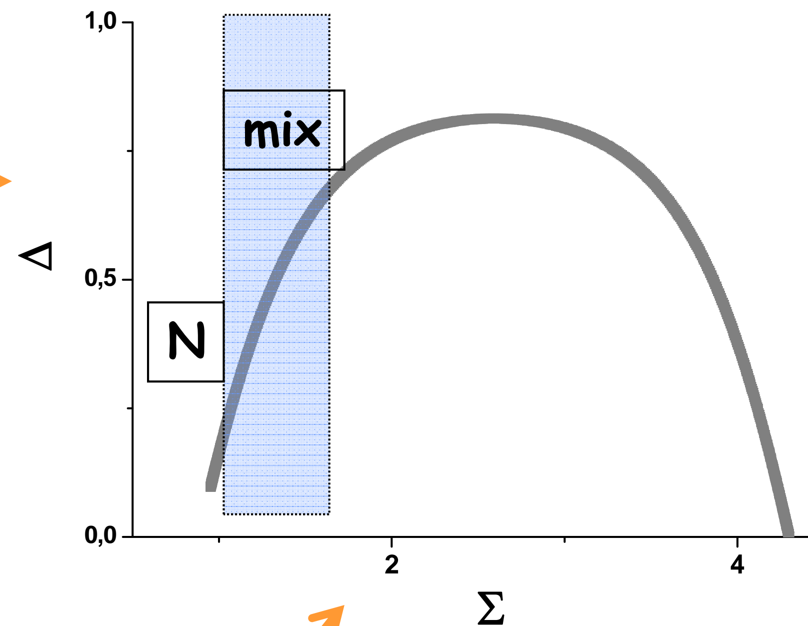
The hybridization regimes

IZBI University Leipzig

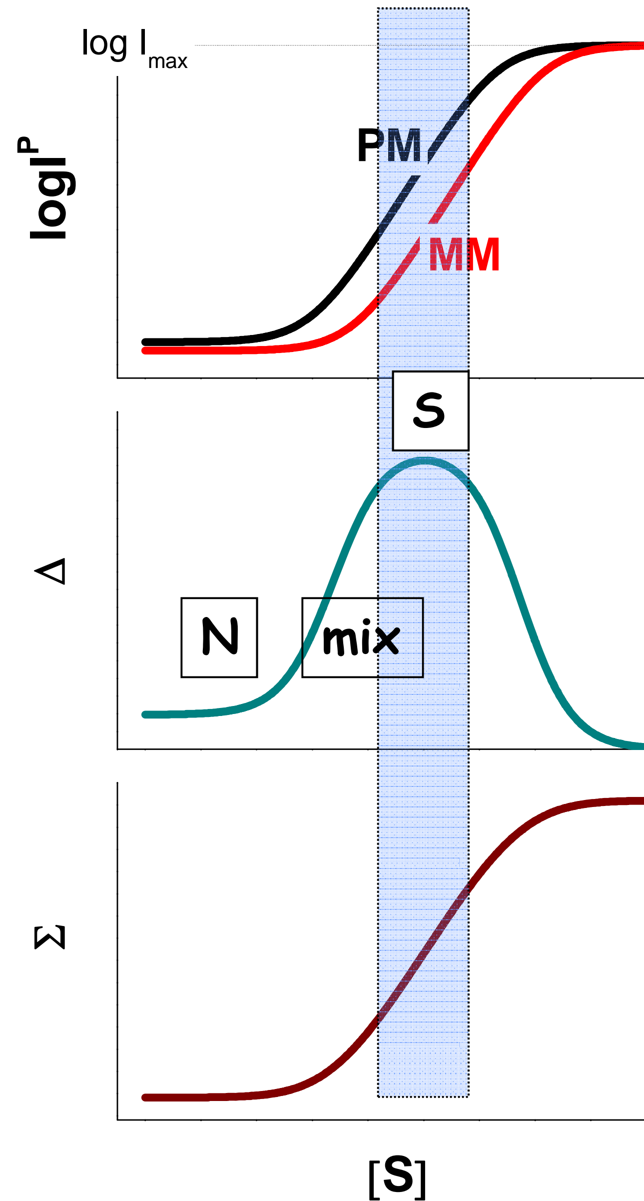


… can be clearly identified:

mix… N- and S-hybridization

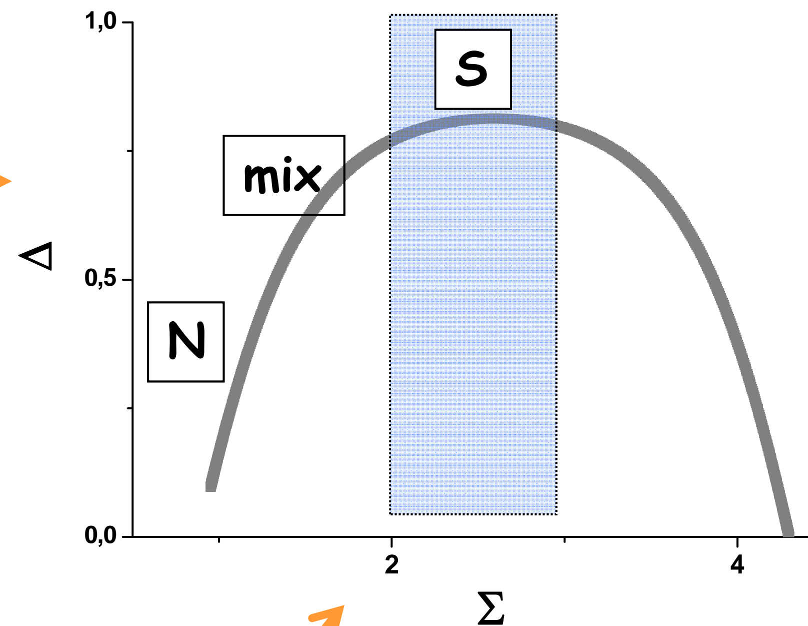


The hybridization regimes

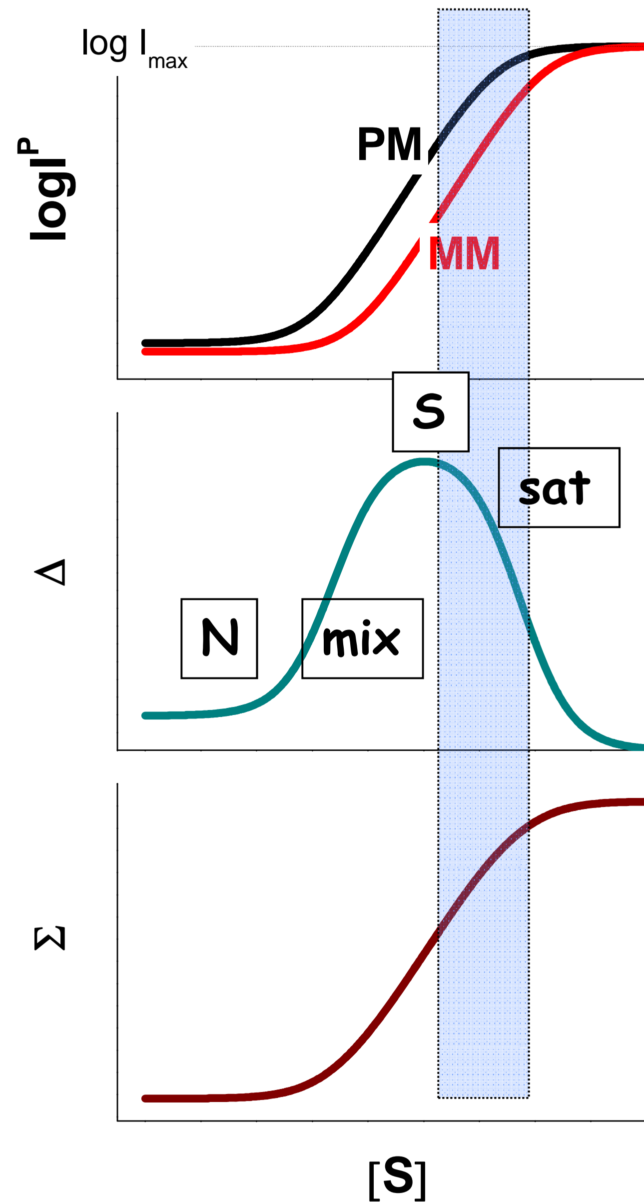


... can be clearly identified:

S... specific hybridization

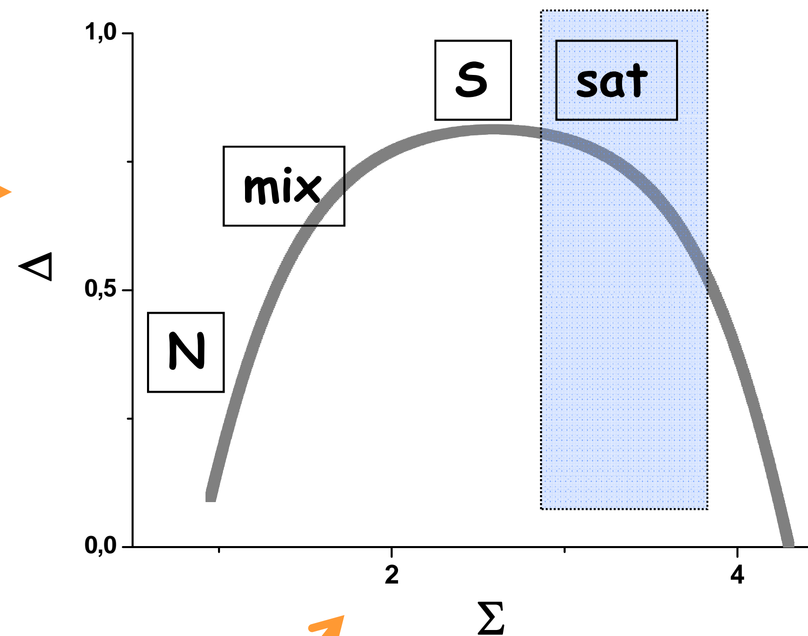


The hybridization regimes

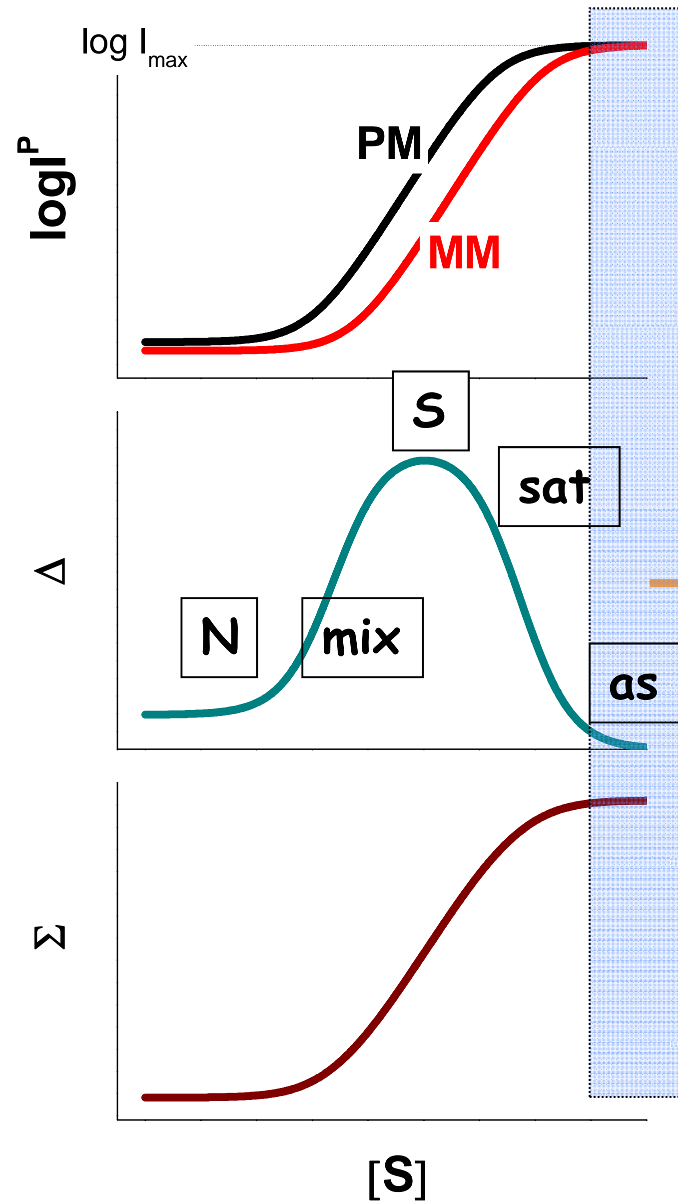


… can be clearly identified:

sat… saturation

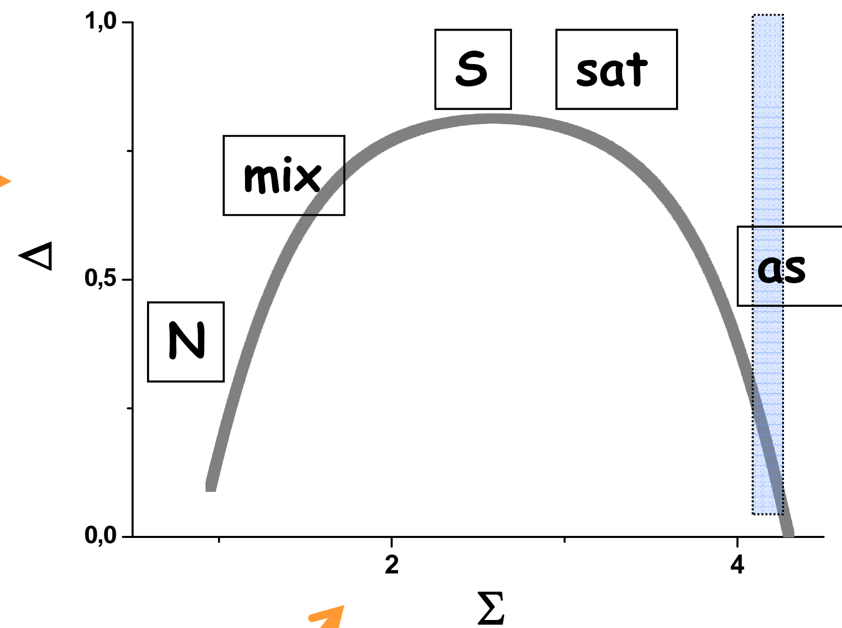


The hybridization regimes

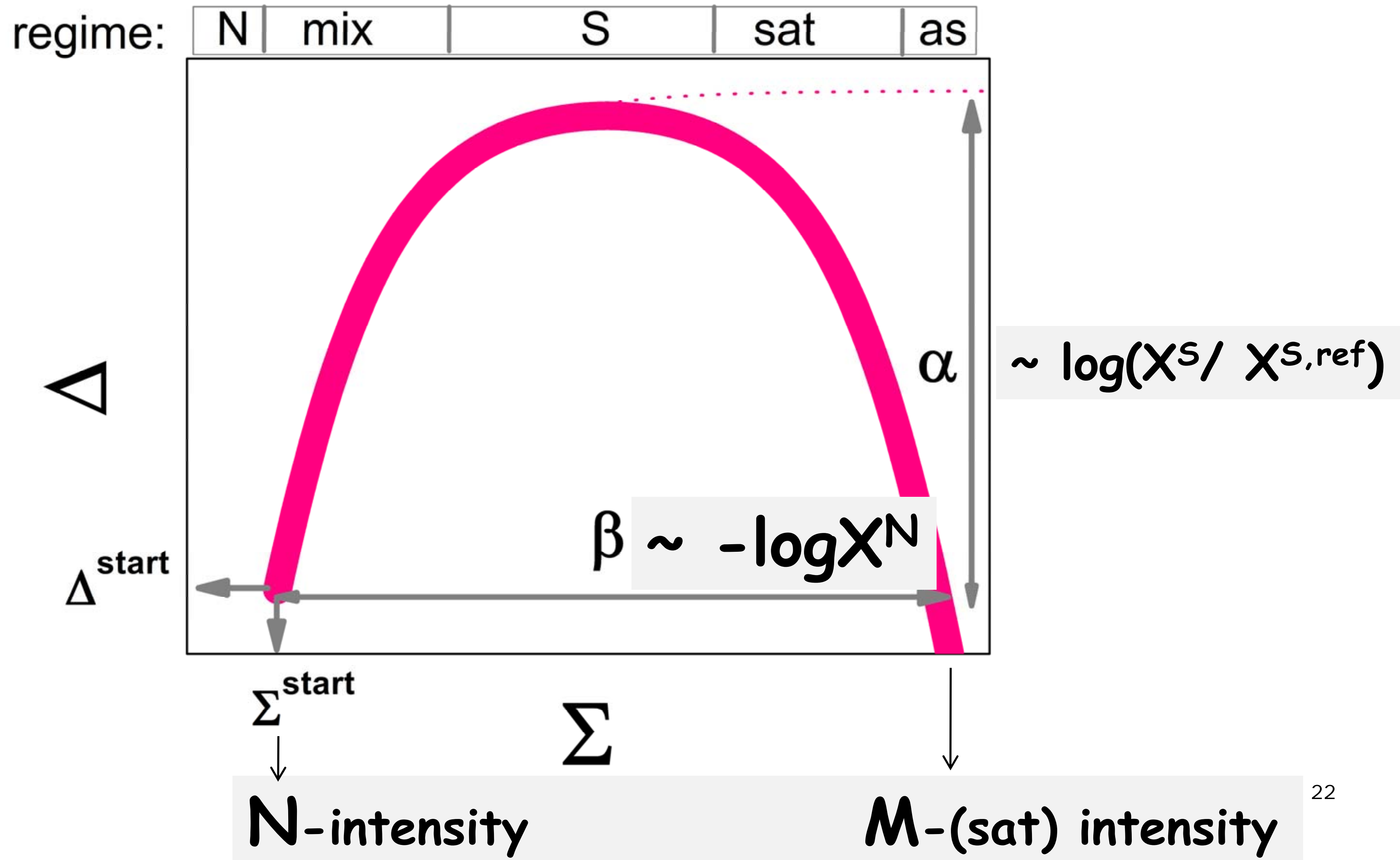


... can be clearly identified:

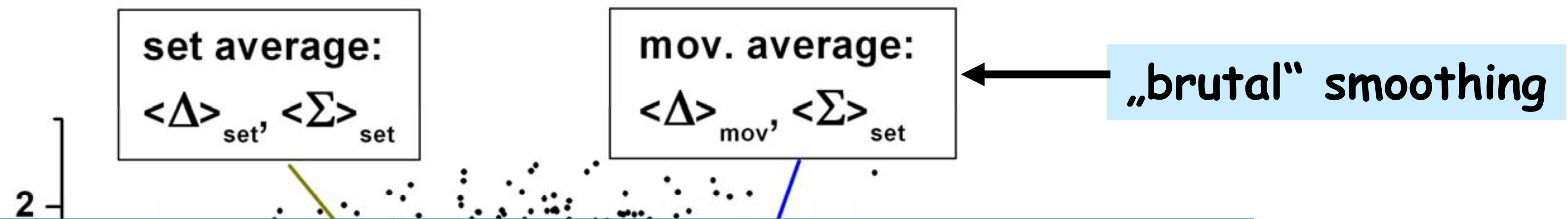
as... asymptotic



Position and dimensions of the hook curve



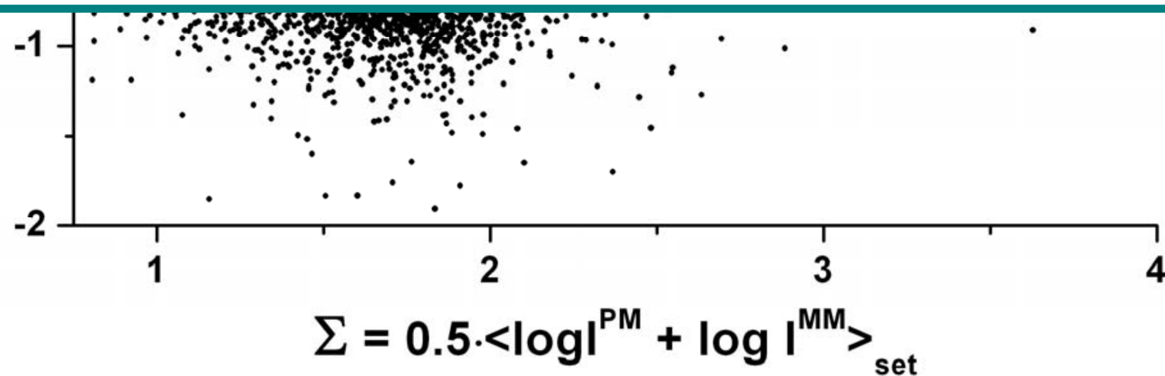
Mean Δ-vs-Σ plot of the chip



We consider all probe-pairs of a chip

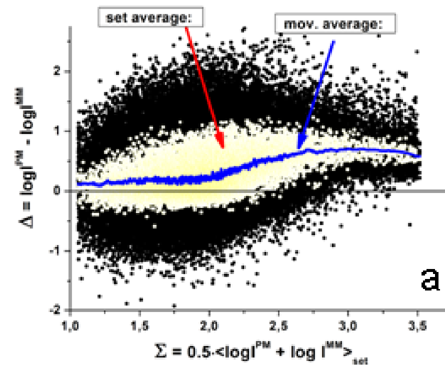
- we calculate the average Δ-vs-Σ trajectory by applying heavy smoothing

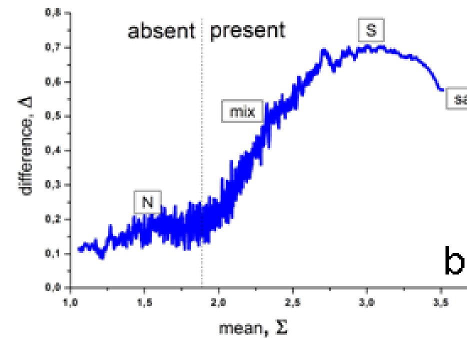
- provides the mean isotherm of a particular hybridization

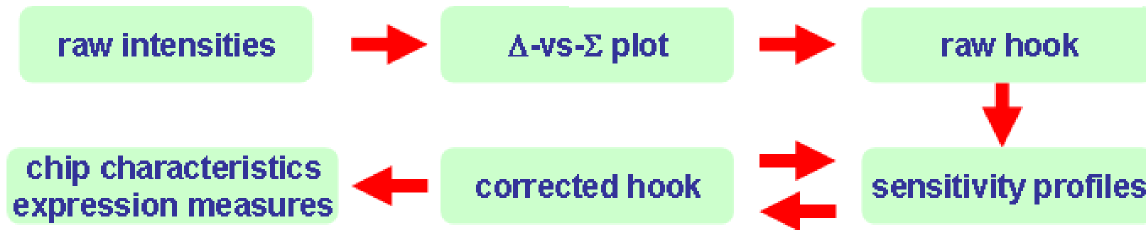


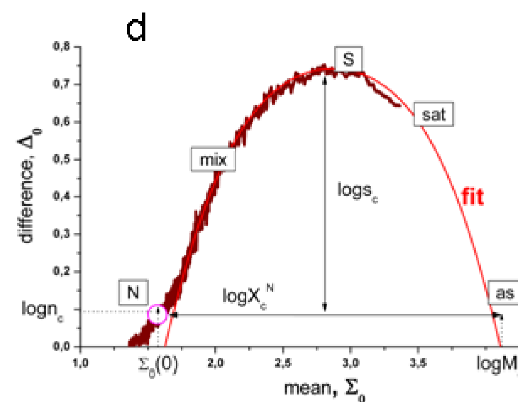
…all probe intensities are corrected for the optical background using the Affymetrix zone-algorithm

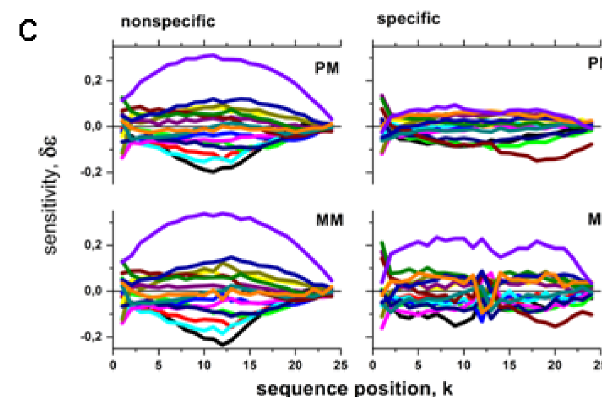
Hook analysis: overview





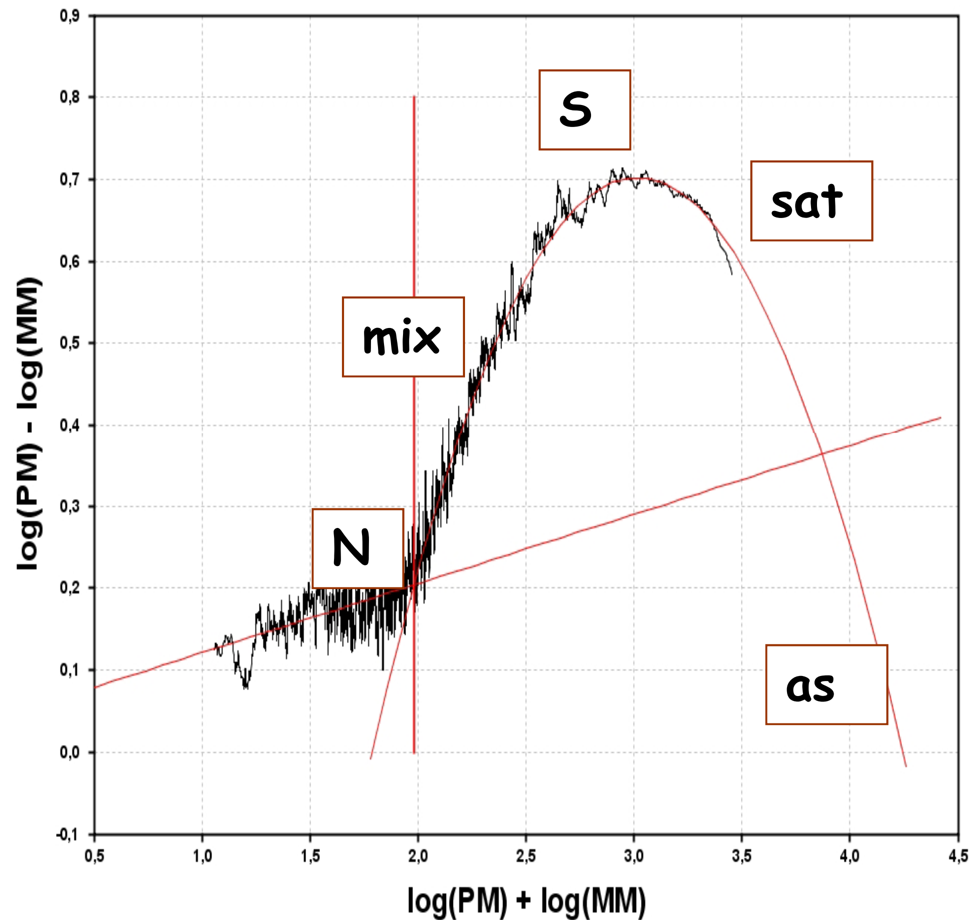






positonal-dependent NN model

the (raw) hook – curve reflects the…



Hybridization regimes:

N…non-specific hybridization dominates

mix…N+S hybridization

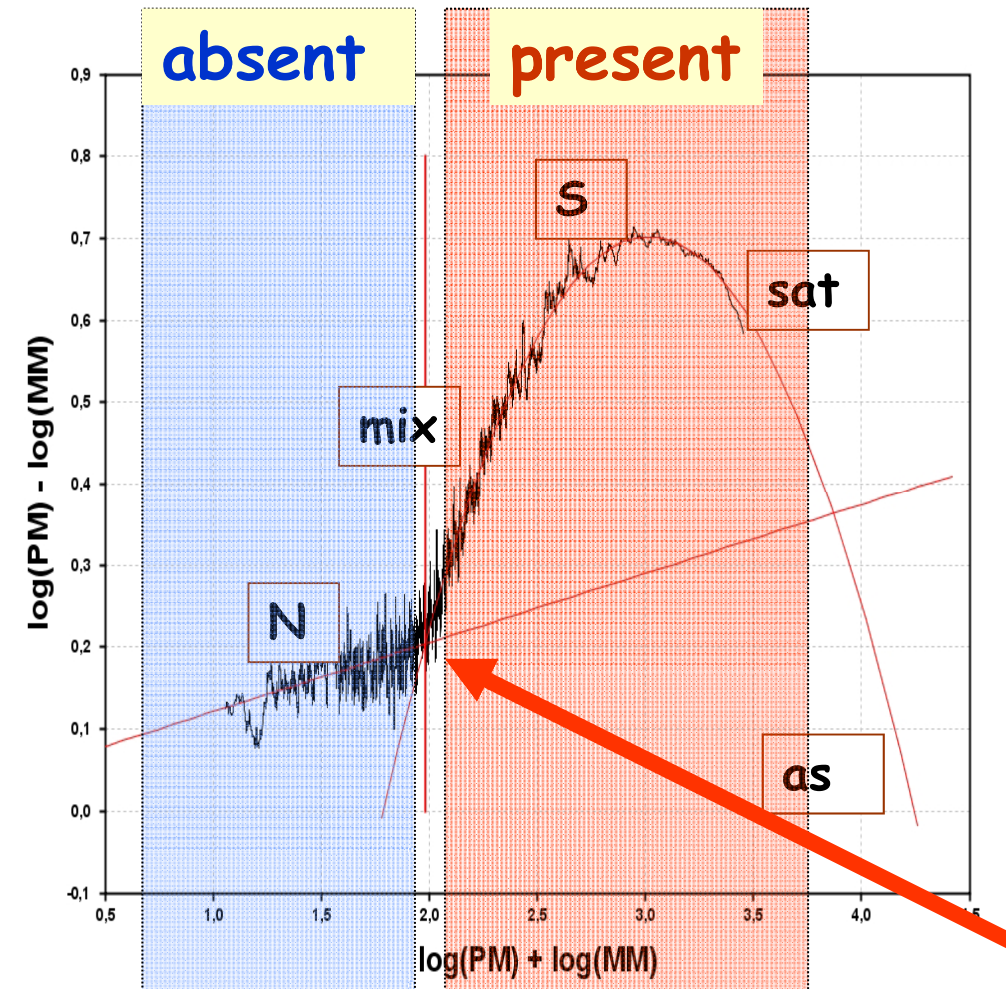
S… specific binding dominates

sat…progressive saturation

(as…asymptotic range, the probes are saturated)

…which we know from the single probes

Absent and present probes…



We suggest:
the initial break can be used to identify „absent" probes,

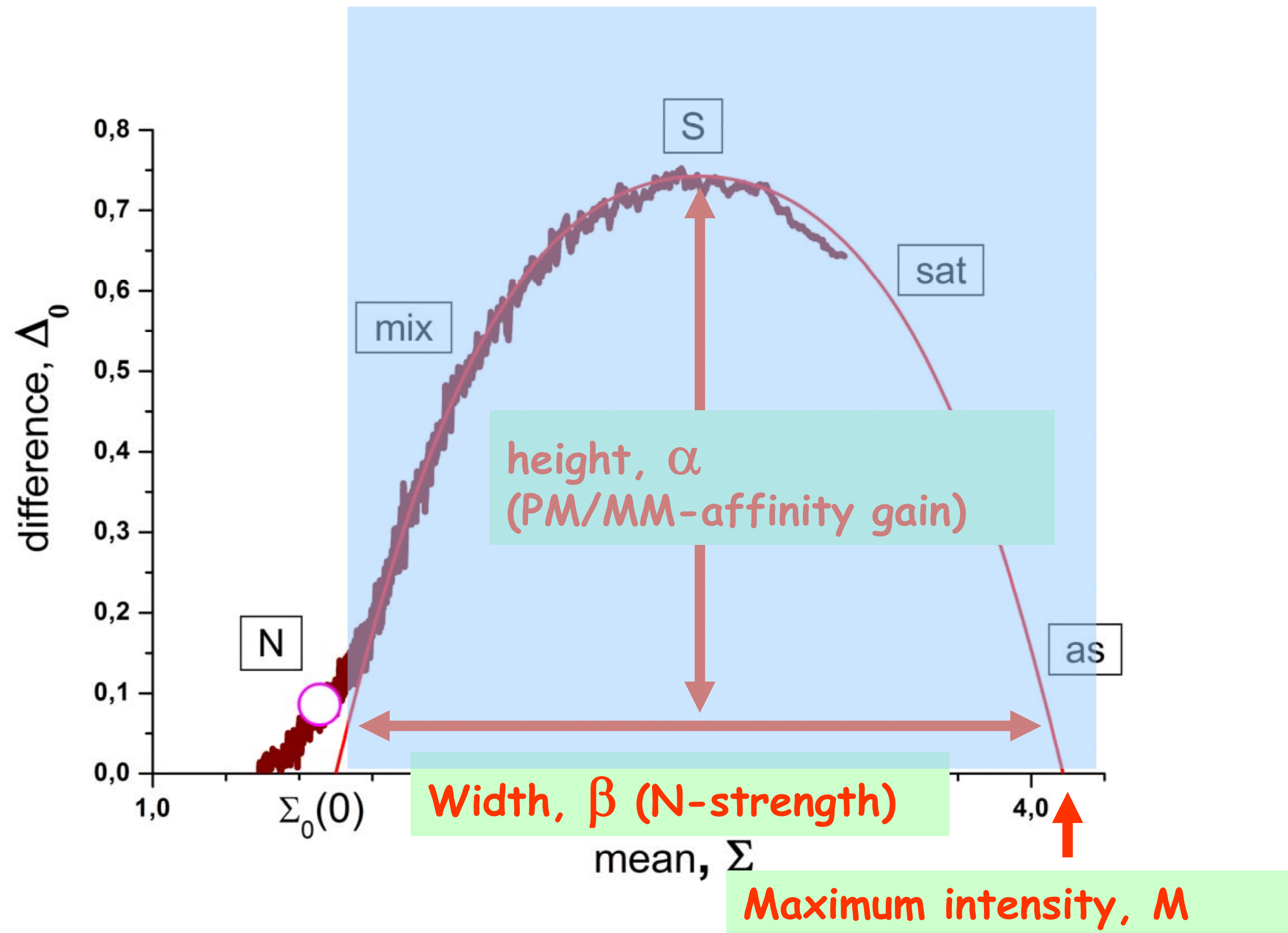
i.e. which predominantly hybridize non-specifically because their specific transscripts are virtually absent in the hybridization solution

slope(N) ~ correlations PM/MM

slope(mix) ~ S-dependence of the isotherm

break

Fit of the hook-equation provides…

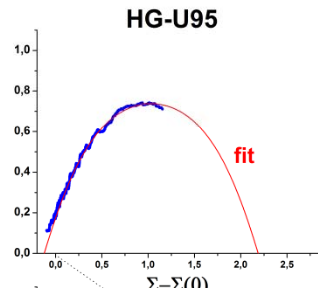


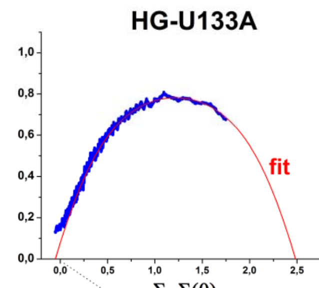
Chip generations (human genome)

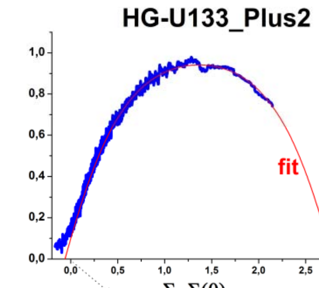
fit

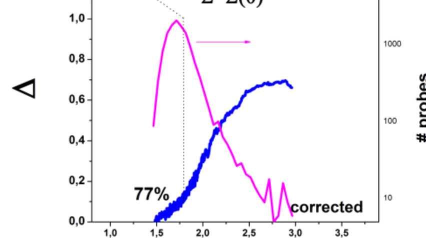
cor-
rec-
ted

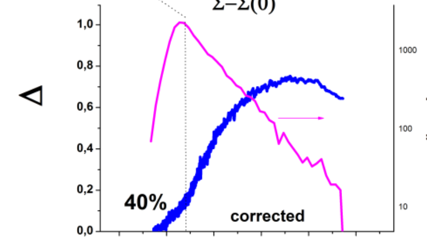
raw

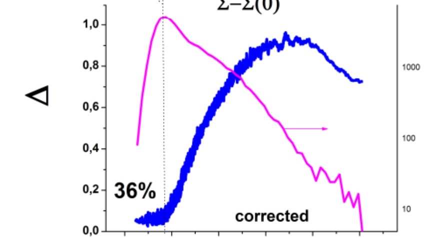


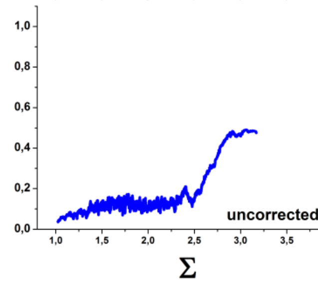


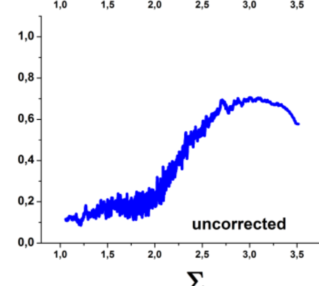


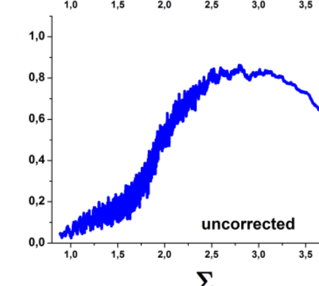










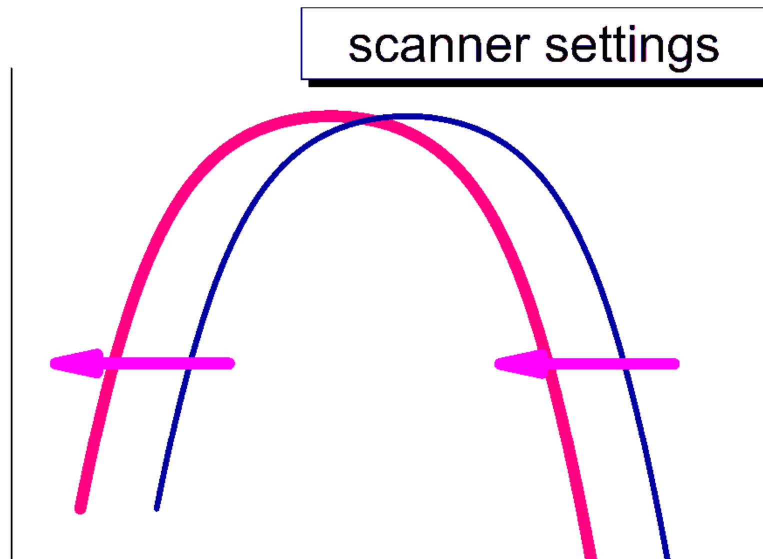


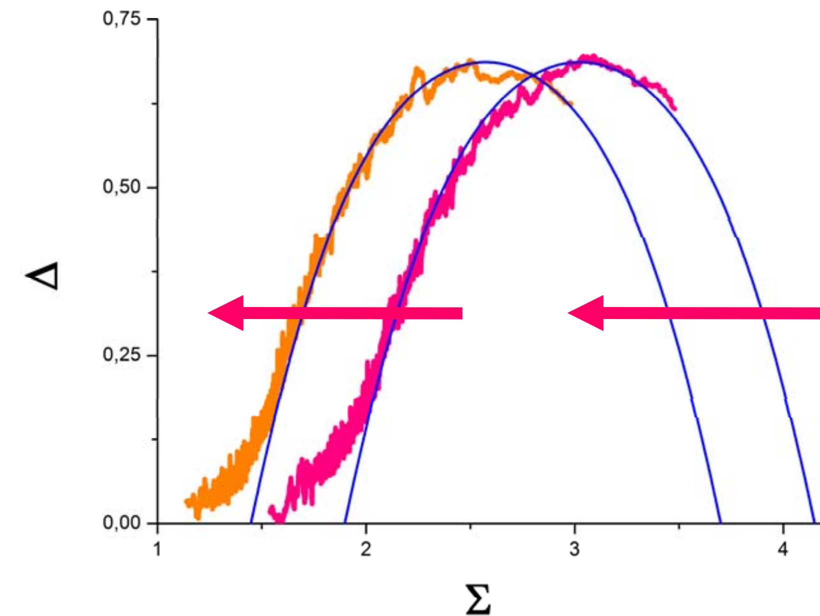
Spot size:	20 μM	18 μM	11 μM
# probe sets:	16,000	22,000	54,000

Improvements: probe selection: genomic, thermodynamic criteria

hybridization, labeling kits, scanner etc.

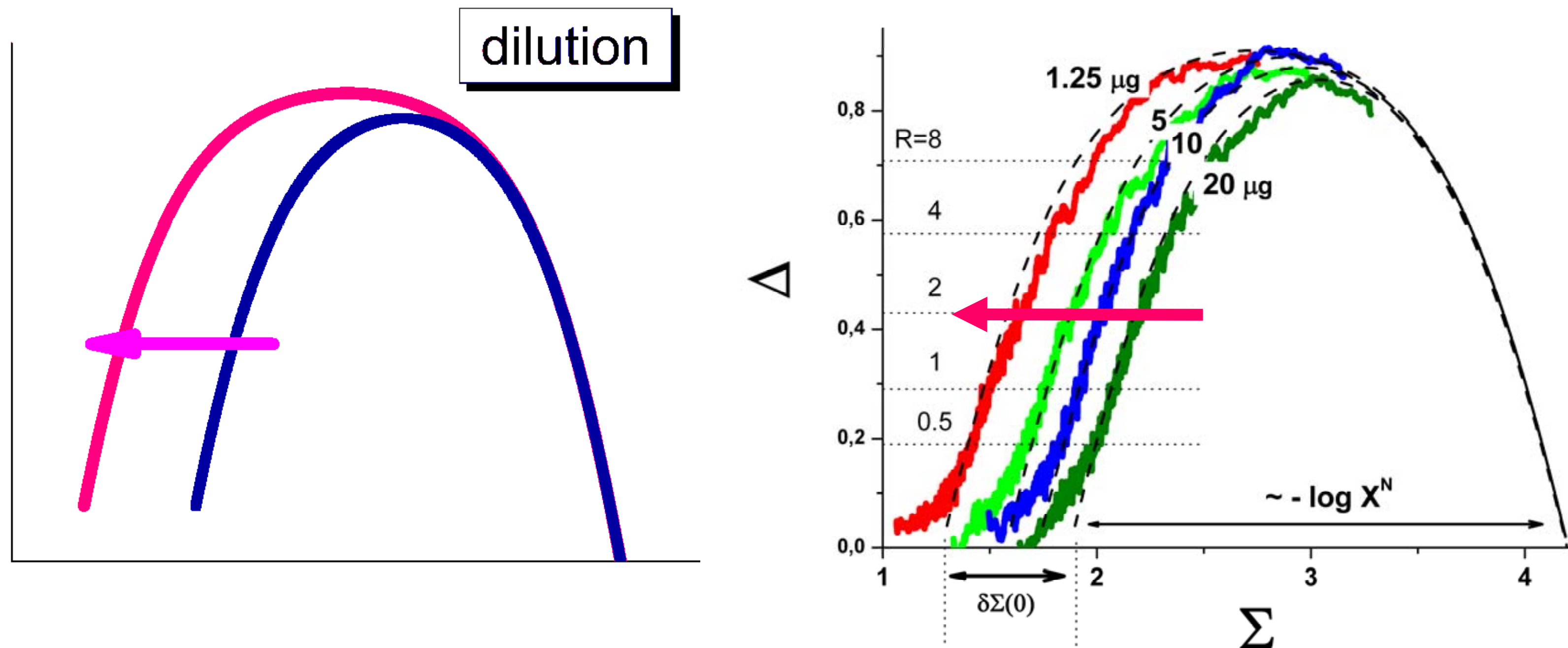
Hook-pattern I: shift





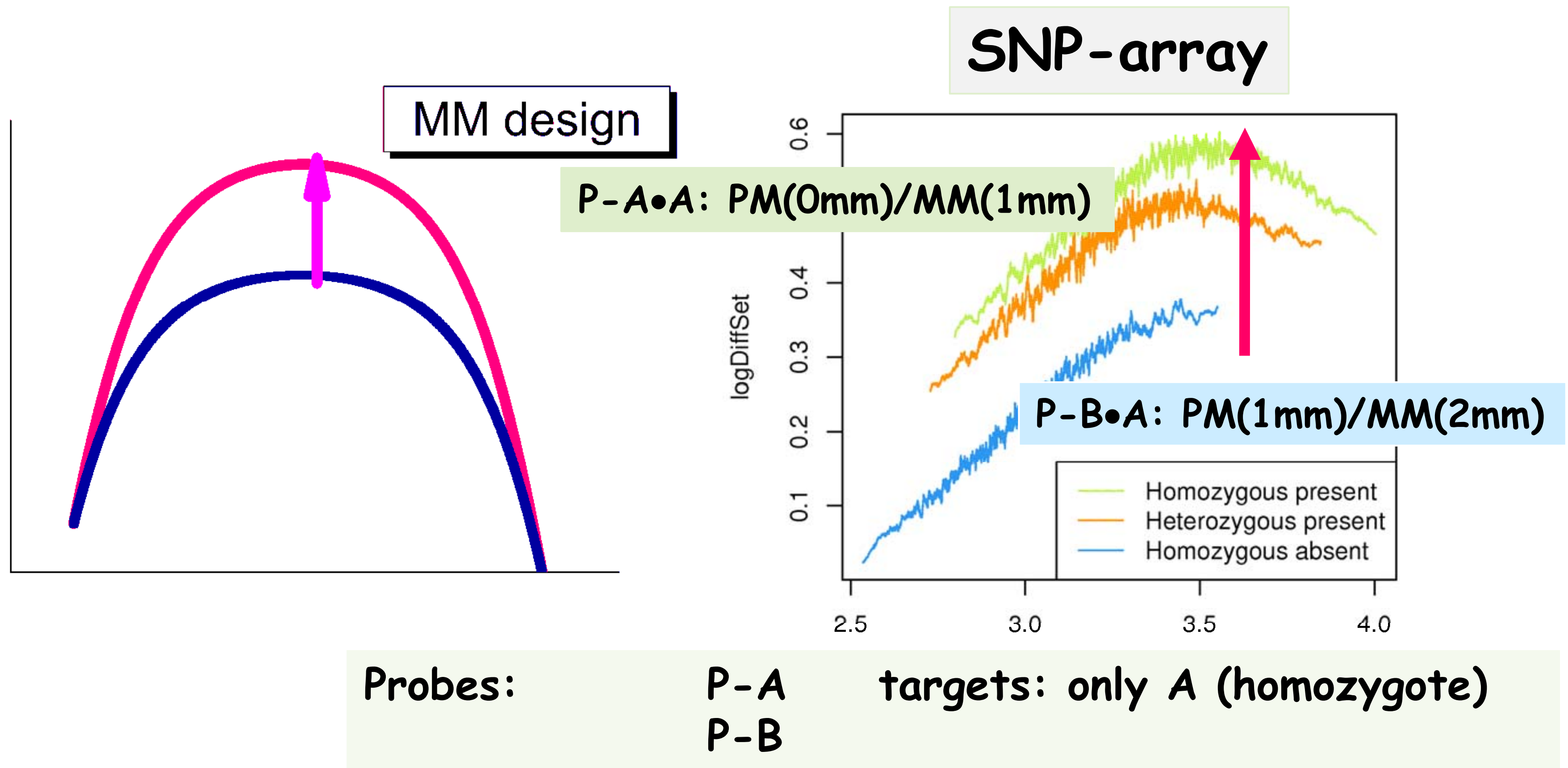
changes of the sensitivity-settings (scanner, amount of labelling) shift the whole curve in horizontal direction

Hook pattern II: widening



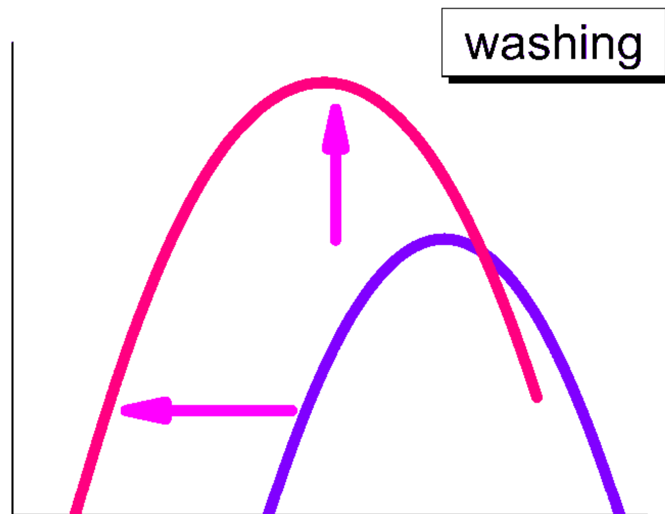
dilution (less total RNA) increases the width of the hook

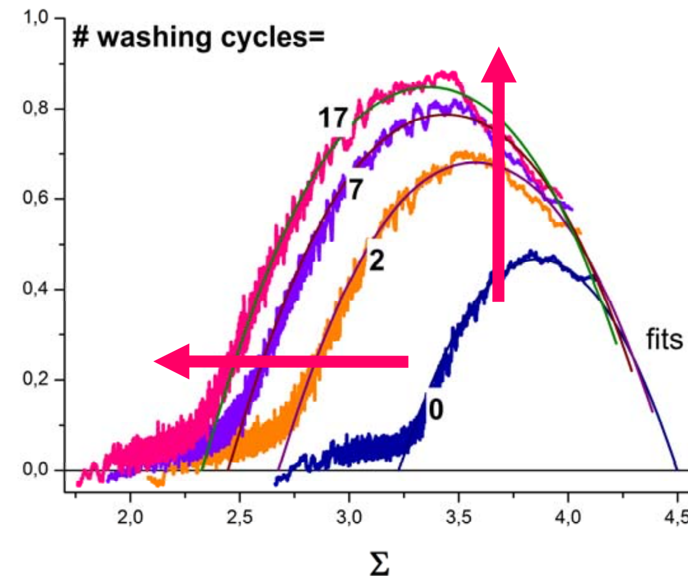
Hook pattern III: vertical stretch



PM/MM-gain modifies the height of the hook

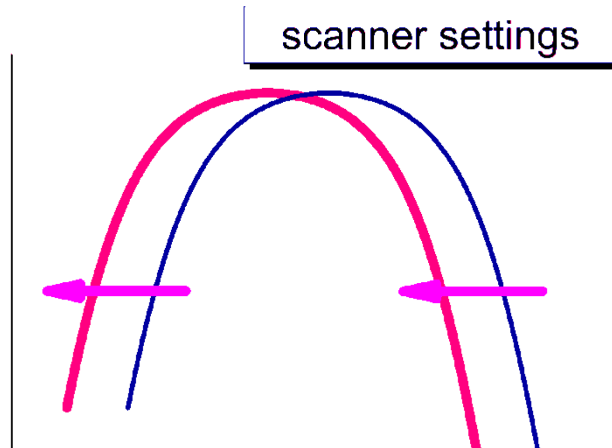
Hook pattern IV: „blow up“

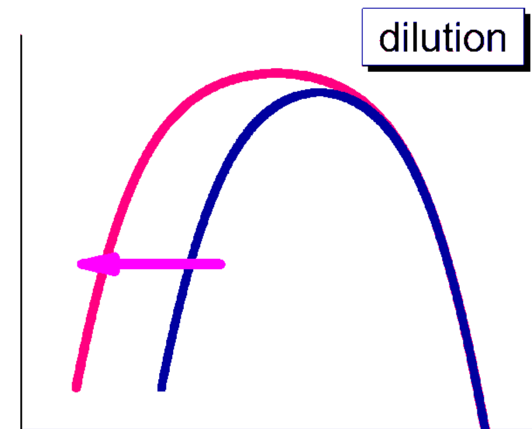


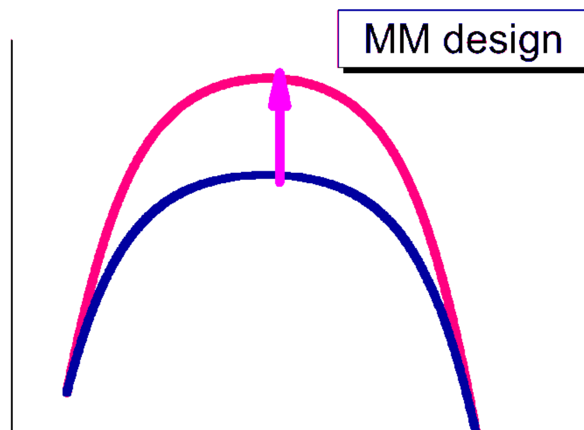


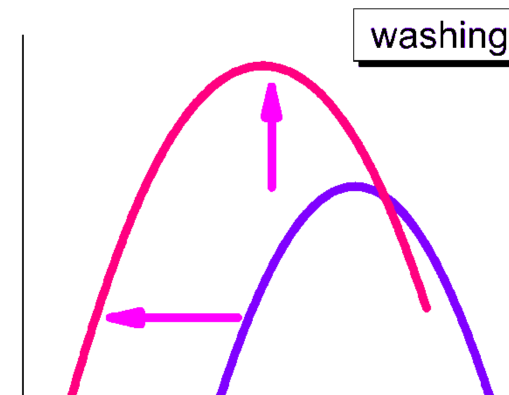
washing increases the width AND the heigth of the hook

Simple rules reveal underlying effects…



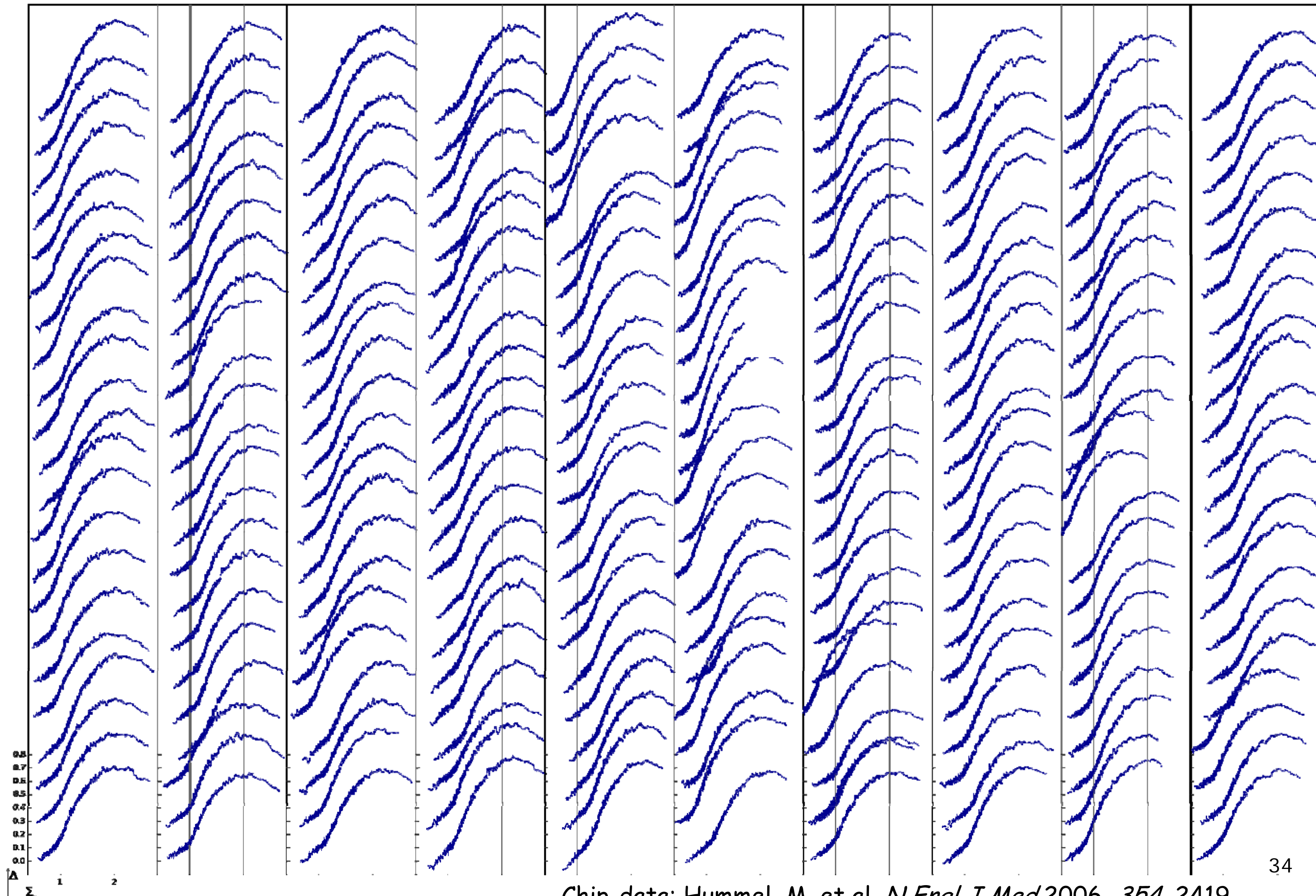






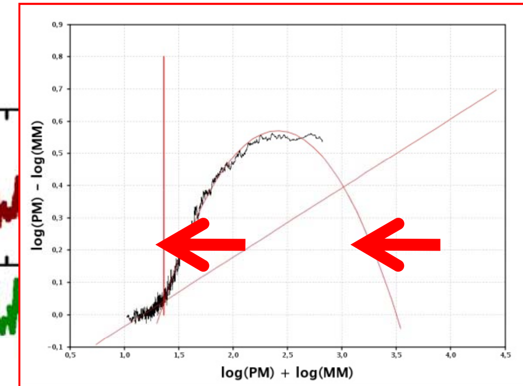
IZBI
University Leipzig

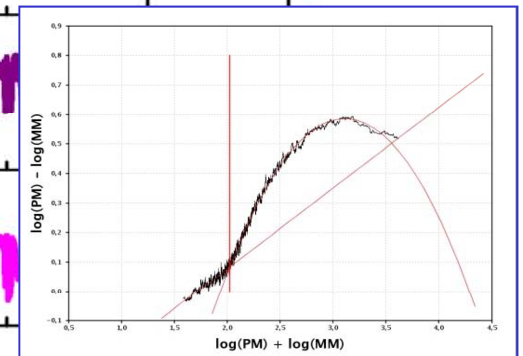
Clinical trial: Burkitts Lymphoma: 221 patients



Chip-data: Hummel, M. et al. *N Engl J Med* 2006, *354*, 2419.

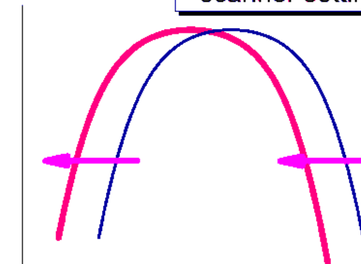
Scanner sensitivity





more

scanner settings



Dilution (less RNA (30%))

IZBI
University Leipzig

0 200 400 600

α 1,0 0,8

β 3,0 2,5 2,0

logM 5,0 4,5 4,0

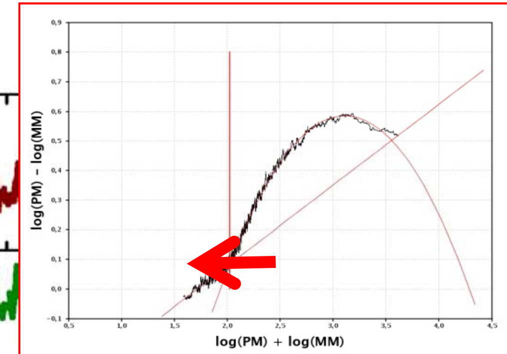
logN 2,5 2,0 1,5

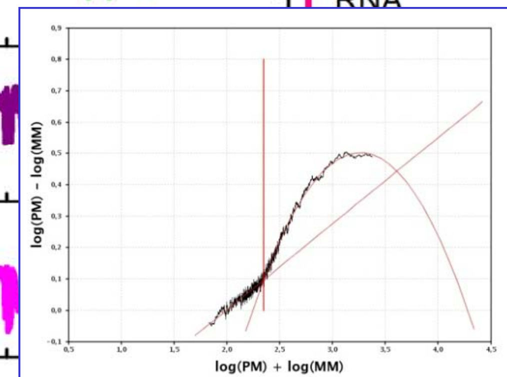
abs 75 50 25

3'/5'-bias 0,8 0,6 0,4

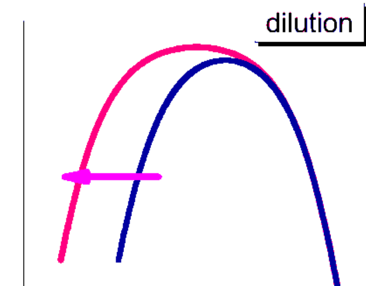
0 200 400 600

probe # ~ time (3 years), laboratories, sampl

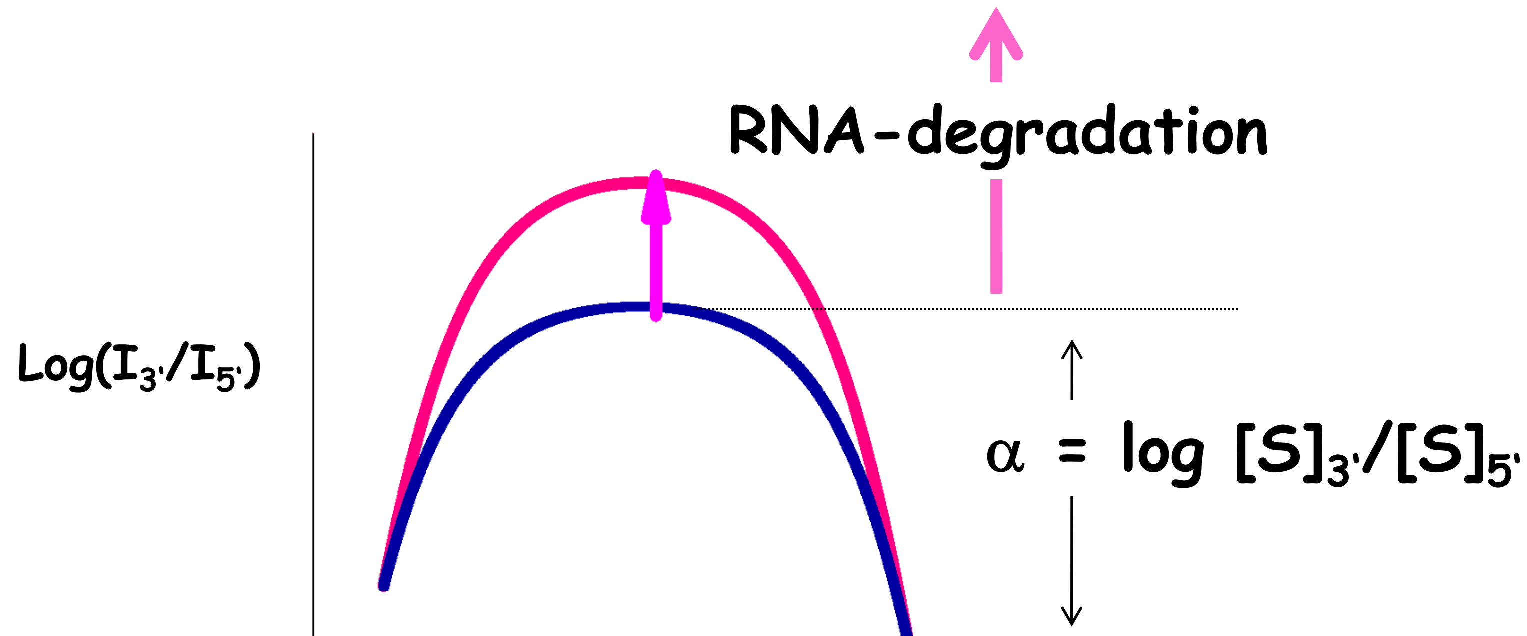




more

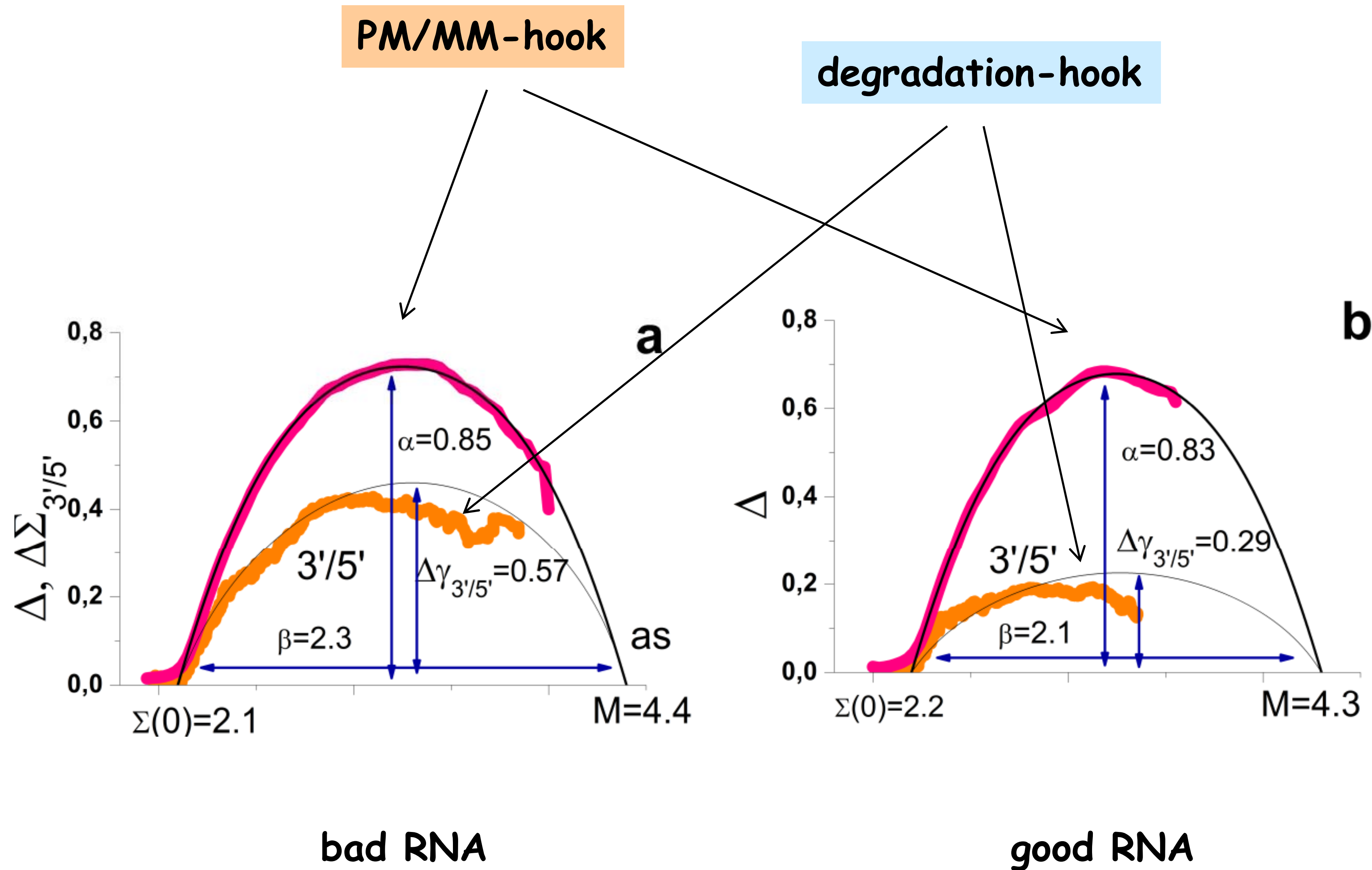


degradation hook



Depletion of targets with increasing distance from 3'-end defines the heigth of the degradation hook

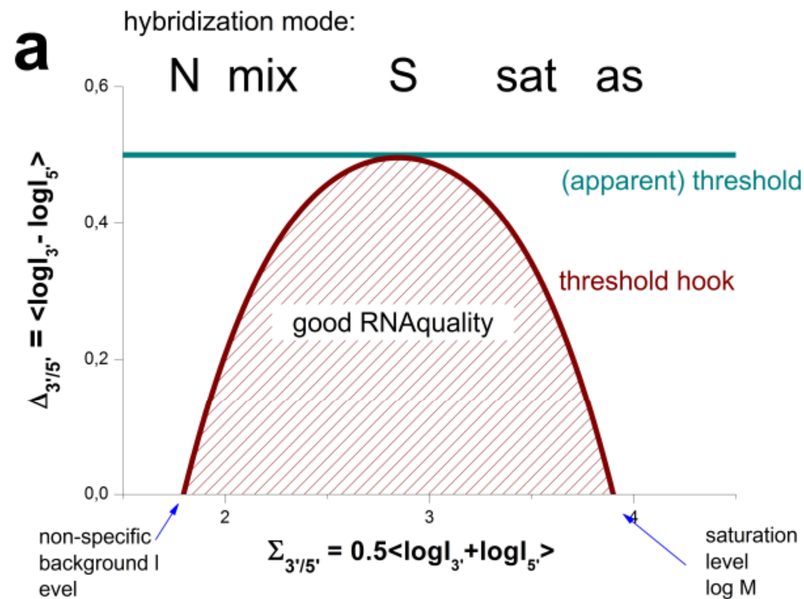
degradation hook: example



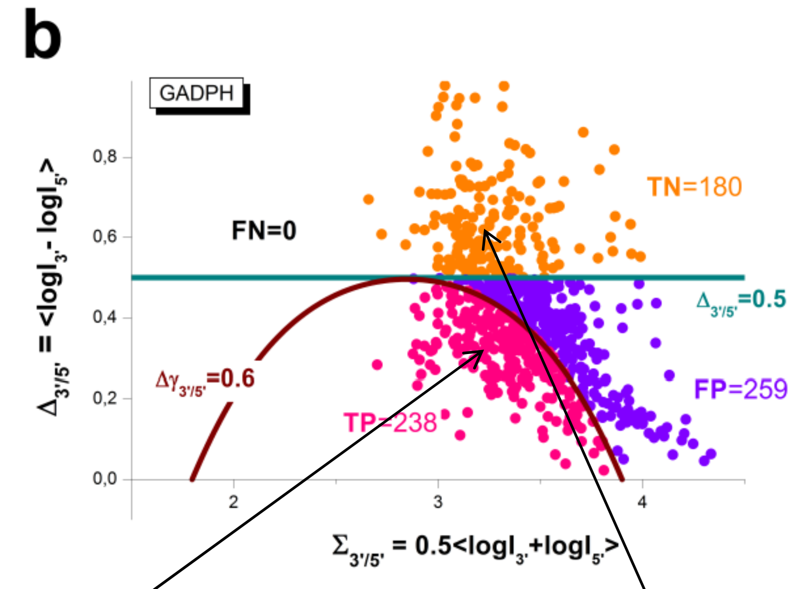
IZBI University Leipzig

Degradation threshold

Threshold hook



Human tissue data



PositivePredictiveValue=TP/(TP+FP) = 0.5

Specificity=TN/(FP+TN) = 0.8

Thank you !

Papers see

http://www.izbi.uni-leipzig.de/izbi/mitarbeiter/Binder/binderpubl.php