

0.1 cDNA binding

New result:

Energy of binding depends on position as well as neighbor context.

Nature Biotechnology 21, 818–821 (2003)

A model of molecular interactions on short oligonucleotide microarrays

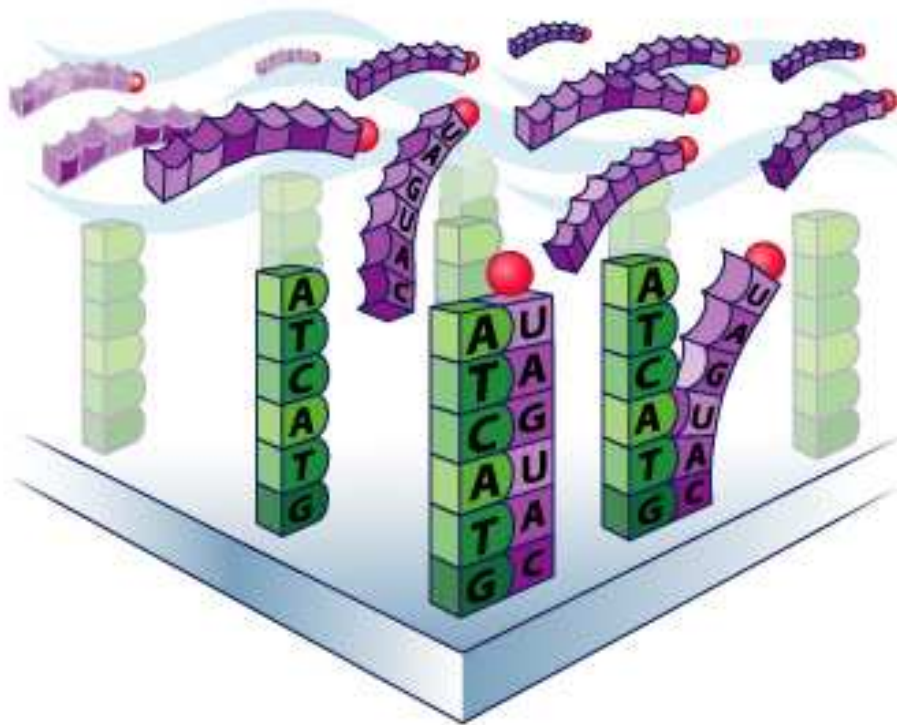
Li Zhang, Michael F Miles & Kenneth D Aldape

PNAS 100, pp. 11237–11242 (2003)

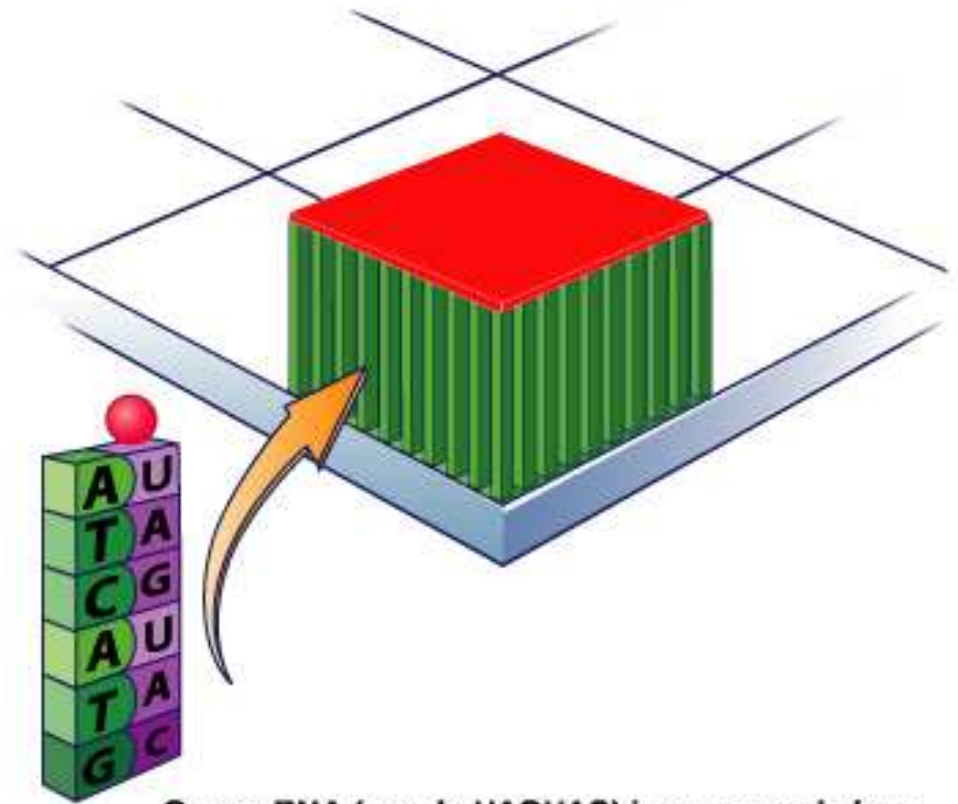
Probe selection for high-density oligonucleotide arrays

Rui Mei, Earl Hubbell, Stefan Bekiranov, Mike Mittmann, Fred C. Christians, Mei-Mei Shen, Gang Lu, Joy Fang, Wei-Min Liu, Tom Ryder, Paul Kaplan, David Kulp, and Teresa A. Webster (Affymetrix, Inc.)

0.1.1 Microarray tutorial (from Affymetrix)



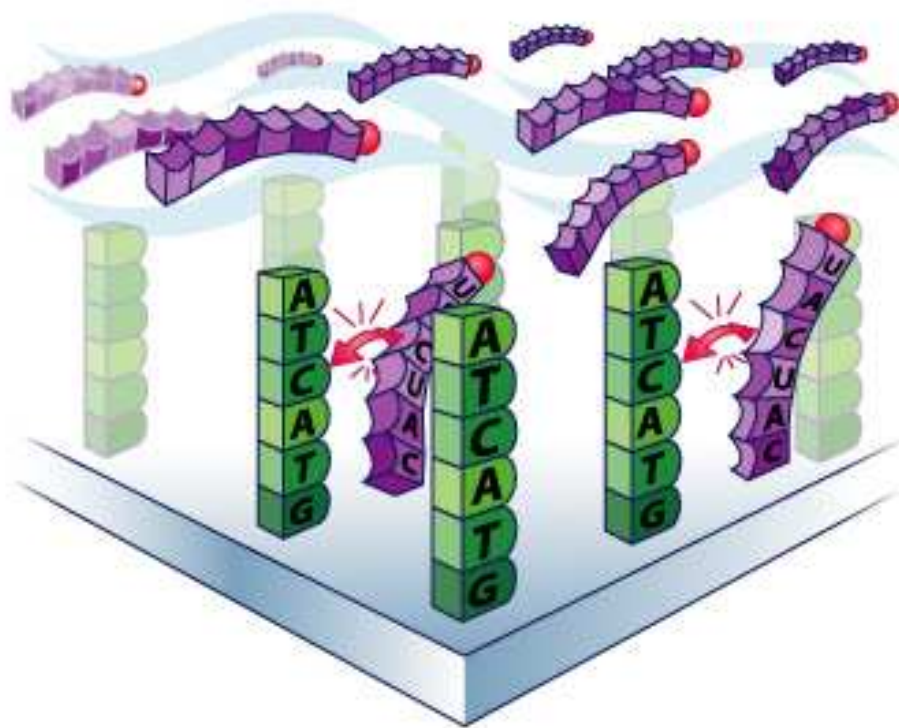
Sample RNA fragments (purple) hybridized to DNA probe array (green)



Goose RNA (purple UAGUAC) in our sample has bound to the goose DNA probe built on the array.

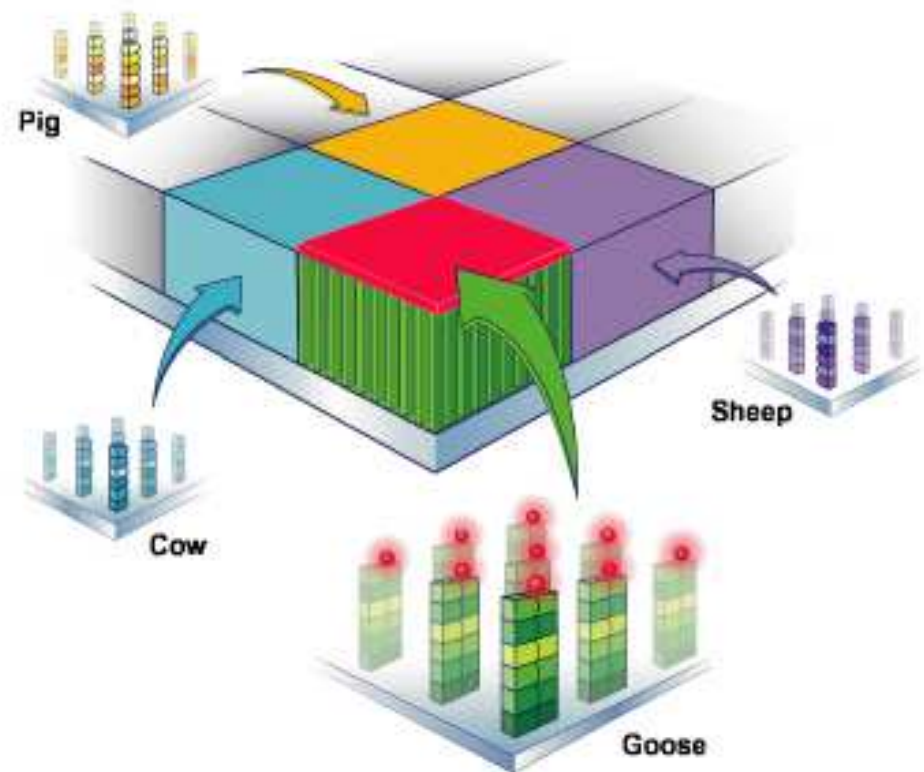
DNA sequences are attached to a slide, and sample RNA is introduced. RNA has fluorescent tags added.

0.1.2 Microarray tutorial (from Affymetrix, continued)



C does not stick to another C,
so no match is made

Shining a laser light on the FoodExpert ID Array causes the tagged RNA fragments that hybridized to glow



Hmmmm. C does not stick to C; seems reasonable, but maybe we should check. What about G binding to G? A to A? T to T?

0.1.3 Models for RNA/DNA binding strength

For a sequence $\sigma = (\sigma_1, \dots, \sigma_n)$ (ignore end effects)

Sequence composition model: $\sum_{i=1}^n w(\sigma_i)$

Basic nearest-neighbor model: $\sum_{i=2}^n W(\sigma_{i-1}, \sigma_i)$

where W is the energy for each pair of letters.

Distance-dependent nearest-neighbor model

$$\sum_{i=2}^n d_i W(\sigma_{i-1}, \sigma_i)$$

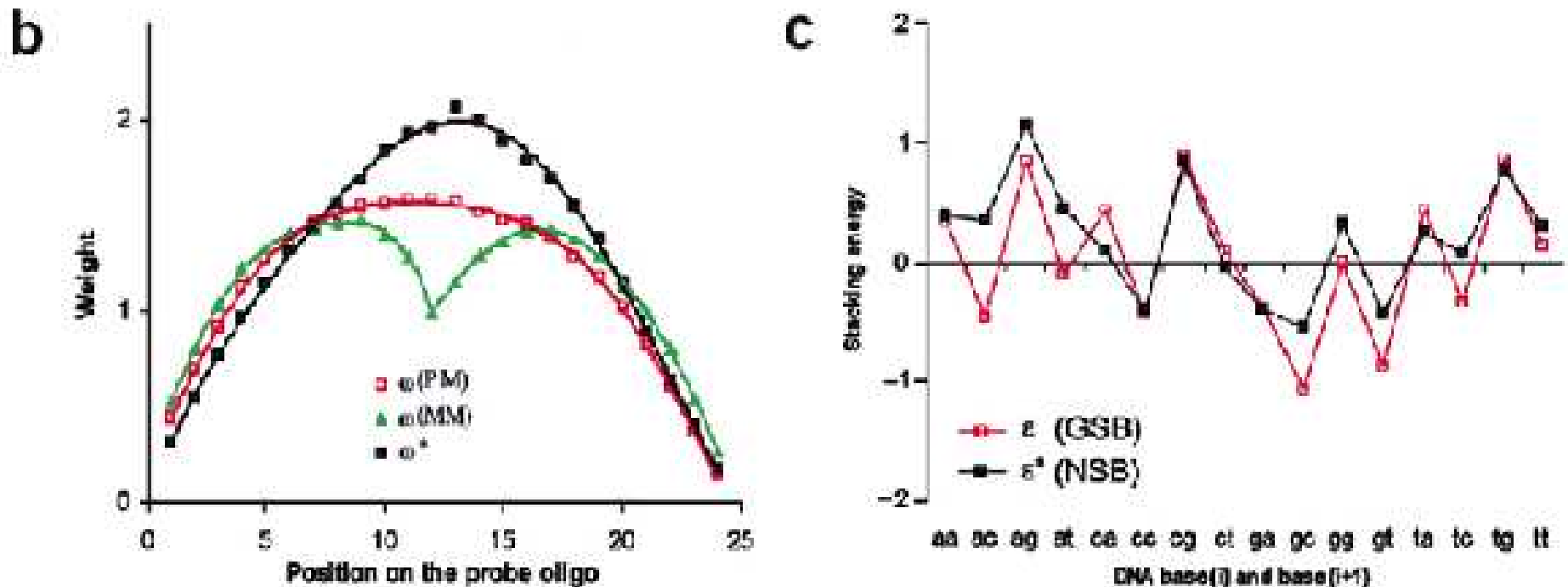
where d_i depends on the position in the sequence.

Another distance-dependent model: $\sum_{i=1}^n d_i w(\sigma_i)$

depending only on the sequence composition, **not the context.**

0.1.4 Using Affymetrix to measure binding

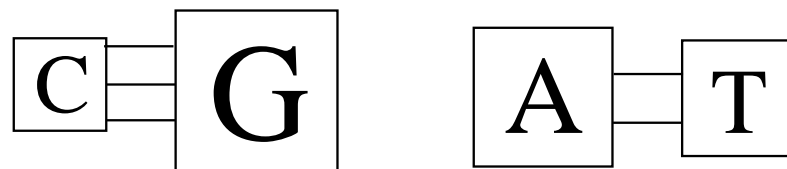
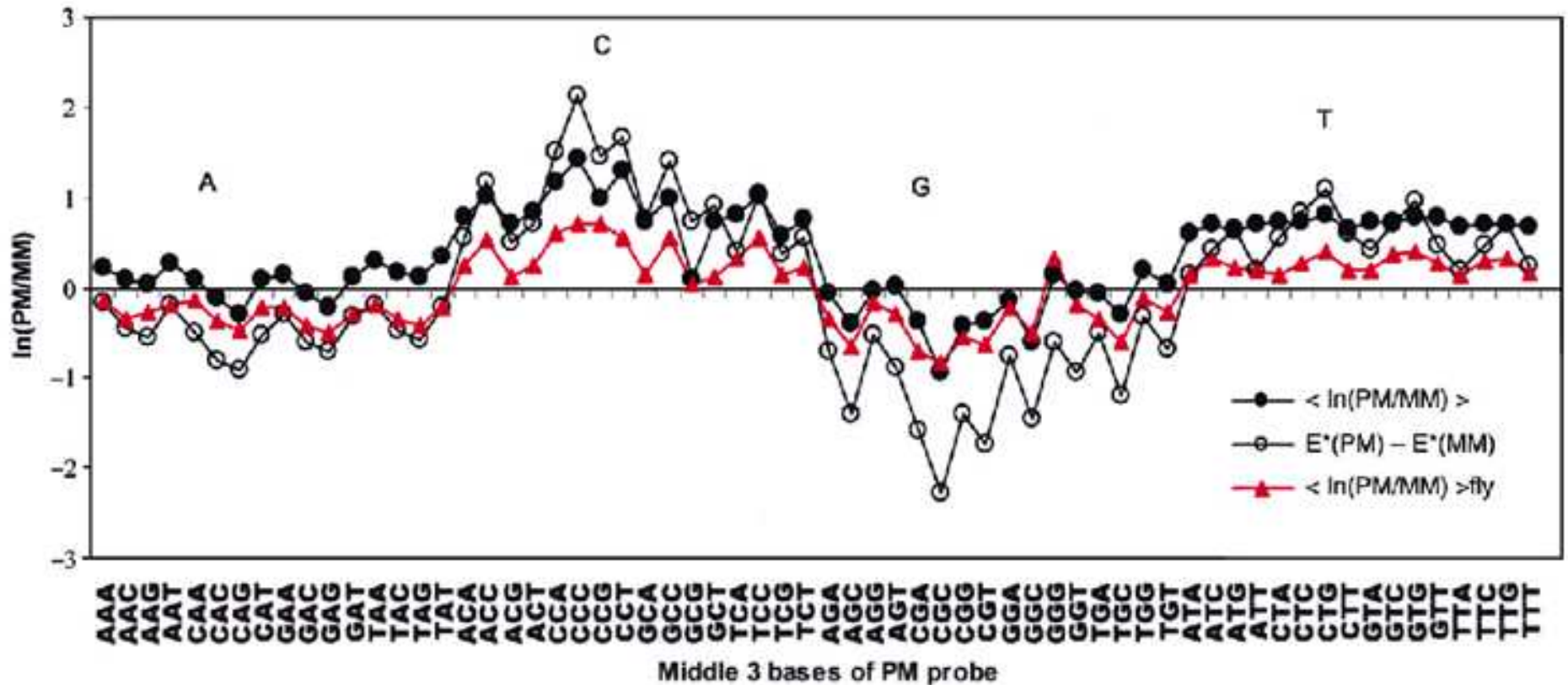
From Nature Biotechnology 21, 818–821 (2003)



(b) Distance coefficients. (c) Nearest-neighbor stacking energy.

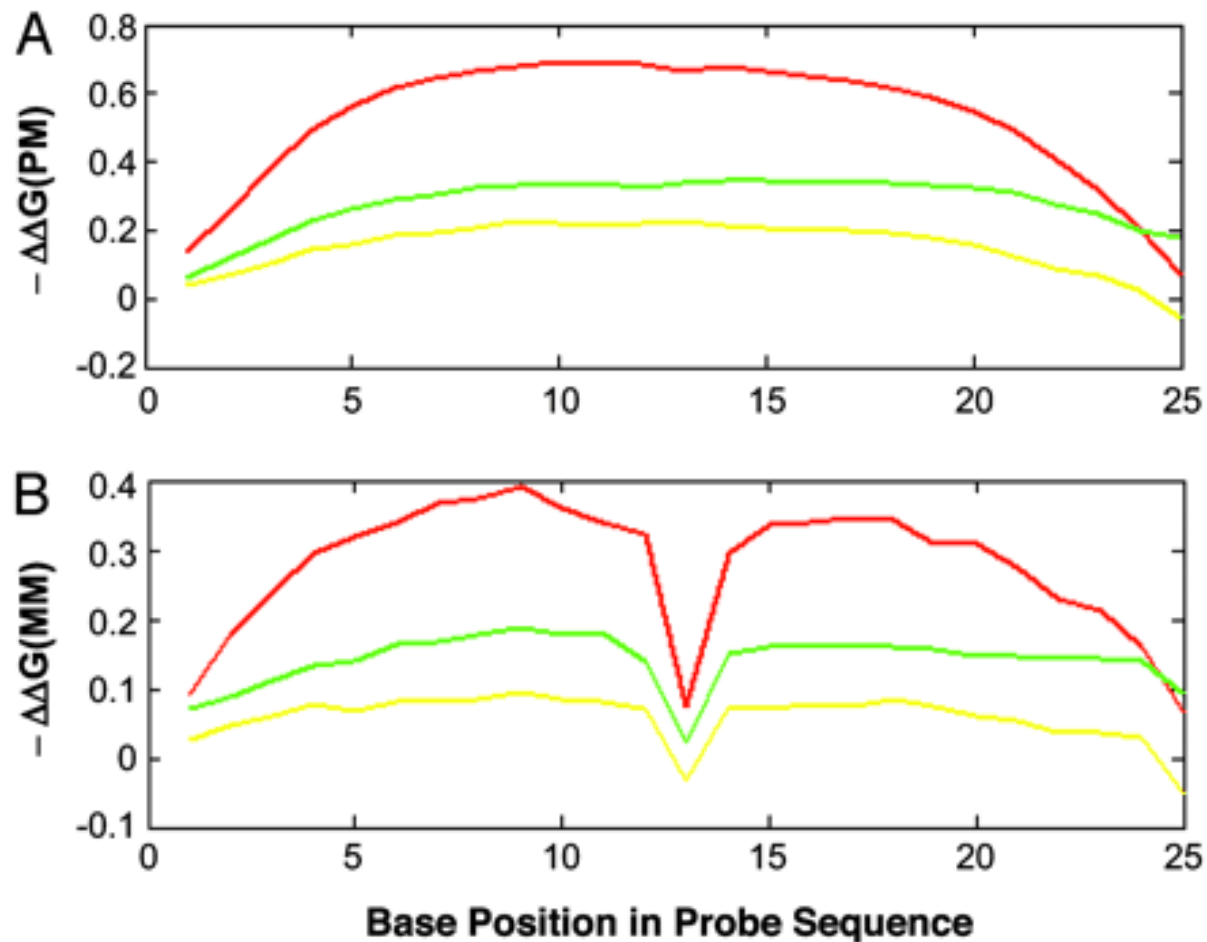
These stacking energies weakly correlated ($r = 0.6$) with that found in aqueous solution, and are smaller in magnitude.

Mismatch signals ($C \leftrightarrow G$, $A \leftrightarrow T$) are stronger with certain triplets for non-specific binding (NSB).



DNA pairs differ in **size** and binding strength: removing bulky A or G increases signal.

From PNAS 100, pp. 11237–11242 (2003): model based on bases and locations



The effective $\Delta\Delta G$ values for the 25 probe base positions. The fitted weights ω_{xi} are the effective values for the bases: $x = \text{C}$ (red curve), G (green curve), and T (yellow curve) in each sequence position, i ($i = 1, \dots, 25$ from the 3' end of the probe), relative to the reference base, A , in the same position.

Mismatch energies were measured in solution in

Biochemistry. 1999 Mar 23;38(12):3468-77.

Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches.

Peyret N, Seneviratne PA, Allawi HT, SantaLucia J Jr.

Excerpt of abstract: Thermodynamic measurements are reported for 51 DNA duplexes with A.A, C.C, G.G, and T.T single mismatches in all possible Watson-Crick contexts. These measurements were used to test the applicability of the nearest-neighbor model and to calculate the 16 unique nearest-neighbor parameters for the 4 single like with like base mismatches next to a Watson-Crick pair. The observed trend in stabilities of mismatches at 37 degrees C is $G.G > T.T \approx A.A > C.C$ The mismatch contribution to duplex stability ranges from -2.22 kcal/mol for GGC.GGC [stabilizing] to +2.66 kcal/mol for ACT.ACT. [destabilizing]

0.2 Multiple probes per gene

Affymetrix uses multiple DNA sequence probes

actcatatactagagtacttagact	ctcatatactagagtacttagactt
tcatatactagagtacttagactta	catatactagagtacttagacttat
atatactagagtacttagacttata	tatactagagtacttagacttatac
atactagagtacttagacttatact	tactagagtacttagacttatacta
actagagtacttagacttatactag	ctagagtacttagacttatactaga
tagagtacttagacttatactagag	<u>agagtacttagacttatactagagc</u>
gagtacttagacttatactagagca	agtacttagacttatactagagcat

per gene:

actcatatactagagtacttagacttatactagagcattacttagat

These provide substantial data to assess various binding models.