Distributions of Free Energy, Melting Temperature, and Hybridization Propensity for Genomic DNA Oligomers

Ryan T. Koehler
Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.
koehlert@appliedbiosystems.com

Nicolas Peyret
Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.
peyretn@appliedbiosystems.com

1. Introduction

Many molecular biology techniques such as PCR, southern blotting, molecular beacon based assays, and DNA microarrays rely on the ability to design oligonucleotide probes possessing specific thermodynamic properties.

Thermodynamic parameters for DNA duplex formation (melting temperature: Tm, free energy: $\Delta G°_T$, and hybridization extent: Fb) are accurately predicted using the nearest-neighbor model [1] for a range of physical conditions for oligonucleotides up to about 50 bases in length.

The use of thermodynamic quantities is ubiquitous in probe design schemes, but invariably focus on achieving specific values for sequences in hand. This fails to provide general insights about how these quantities depend on sequence composition, length, and experimental conditions. Here we present Tm and Fb distributions calculated for genomic DNA samples of 10 to 50 bases.

2. Methods

Thermodynamic quantities were computed with specially built software using published DNA parameters [1]. 10,000 sample 50-mers were randomly picked from human chromosome 19 [2]. Distributions were tabulated for calculated thermodynamic quantities at each evaluated length then percentiles were identified and plotted.

3. Results And Discussion

Figure 1 shows distributions of calculated Tm and Fb for 10,000 sequences from the human genome. Tm increases smoothly with length, as the distribution plotted in Figure 1 clearly shows. Consideration of limiting Tm values for different sequence lengths is useful. For example, no 10-mer in our sample of 10,000 sequences has a Tm greater than 49° C, 98% (percentiles 1-99) of sampled 20-mers have a Tm between 54° and 71° C, etc. One may also look at the length distribution associated with a Tm. For example, for Tm = 60° C, 80% of sequences have 16 to 22 bases; no sequence is shorter than 13 or longer than 26 bases. Increasing DNA and/or salt concentrations shifts the distribution toward higher Tm with little change of shape.

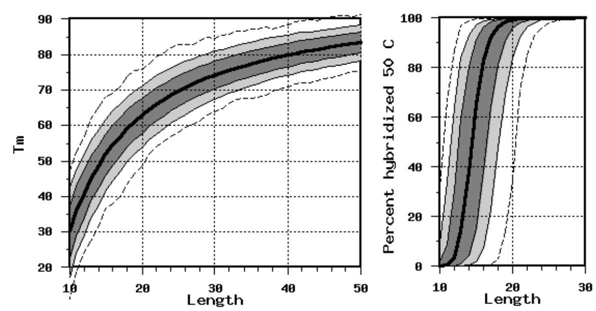

Figure 1. Distributions of Tm (left) and Fb (right). Distributions presented here use [DNA]= 1e-8 M, [cation$^+$]= 0.6 M. Heavy lines show median values, dark and light shadings bound percentiles 10-90 and 1-99, respectively; dashed lines indicate the extreme limiting values obtained in the sampling process.

Hybridization extent is conveyed by Fb, where values close to 0 indicate single stranded DNA, and values close to 100 indicate full hybridization. Tm is defined as Fb = 50. The distribution in Figure 1 show Fb calculated at 50° C. As with Tm, it is informative to look at limiting length values. For example, at 50° C most 10-mers are single stranded while 99% of 24-mers and every 26-mer is hybridized.

Our goal is to provide a general overview of thermodynamic distributions from which rules of thumb may be gleaned and our plots facilitate this. Such data can also be useful for evaluating probe specificity.

4. References

[1] John SantaLucia, Jr., "A unified view of polymer, dumbbell, and oligonucleotide nearest-neighbor thermodynamics" *Proc. Natl. Acad. Sci. USA 95*, 1998, pp 1460-1465.

[2] UCSC Human Genome, hg8, http://genome.ucsc.edu/goldenPath/06aug2001/chromosomes/

Proceedings of the IEEE Computer Society Bioinformatics Conference (CSB'02)
0-7695-1653-X/02 $17.00 © 2002 **IEEE**