148. Designing DNA-microarray Probes Optimized for Hybridization Probability

Agnes J. Szczepek[1], **Jerzy Tiuryn**[2] **and Bartosz Wilczynski** [2]

Keywords: DNA-microarray, probe design

1 Motivation

DNA-microarray technology permits an effective global approach for transcriptome analysis in biomedical research. Information obtained from DNA microarrays is used in cancer biolog, transplant medicine and other disciplines of biomedical research. The challenge of array creation is how to identify the optimal probes for each gene. Our current knowledge does not allow to compute the exact hybridization free energies which is critical for selecting the optimal probes. Current approach to microarray design [4] includes many in-vitro accuracy tests of the heuristically selected probe sets. This technology is very expensive and for many laboratories creation of new or customized DNA-arrays is simply impossible. Invention of new algorithms based on different models should reduce the use of in-vitro testing and in effect lower the cost and increase the availability of the Microarrays.

2 Model

We develop an algorithm for designing a set of oligonucleotides (probes) for a given set of genes. As the input we take the set of genes $G = \{g_1, \ldots, g_n\}$ with known sequences for which we design probes. We also have another (possibly empty) set of genes $H = \{h_1, \ldots, h_k\}$ that can be present in the RNA solution. Our goal is to create a set of probes $S = \{s_1, \ldots, s_n\}$ that will conform to the set of conditions:

1. $\forall_{i=1\ldots n}\ s_i$ hybridizes with g_i
2. $\exists_{d\in\mathbb{N}}\forall_{i=1\ldots n}\,|s_i| = d$
3. $\exists_{t\in\mathbb{R}}\forall_{i=1\ldots n}\,|\mathcal{T}(s_i) - t| < \delta t$, where δt stands for a acceptable difference of melting temperatures between probes
4. $\forall_{1\leq i,j\leq n, i\neq j}\ s_i$ does not hybridize with g_j and s_j does not hybridize with g_i
5. $\forall_{i=1\ldots n, j=1\ldots k}\ s_i$ does not hybridize with h_j.

We assume that the hybridization is free of nucleotide mismatches.

3 Method

The main concept of our approach is to express the quality of the probe in the terms of probability. We construct the function describing the probability of hybridization between the probe and the target. Instead of saying that the probe hybridizes with the target we now say that the probability of hybridization under certain conditions equals P. In those terms our goal is to find for each gene the probe that maximizes the probability of hybridization with the target gene and minimizes the probability of hybridization with other genes. Our probability function is based on nearest-neighbor thermodynamics [3, 2].

[1] Max-Planck-Institut fuer Infektionsbiologie, Berlin, Germany
[2] Warsaw University, Faculty of Mathematics, Informatics and Mechanics, ul. Banacha 2, Warszawa, Poland, E-mail: tiuryn@mimuw.edu.pl

4 Future work

Our approximation of the true hybridization probability is still rather simple and we will focus on extending it by introducing the thermodynamics of internal DNA mismatches.

References

[1] Li,F. Stormo,G.D. 2001. Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* vol. 17(11) p. 1067-1076

[2] SantaLucia,J.Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. USA* vol.95, p. 1460-1465

[3] Hartemink,A.J. Gifford,D.K. 1997. Thermodynamics simulation of deoxyoligonucleotide hybridization for DNA computation, 3rd Annual DIMACS workshop on DNA-based computers

[4] Lockhart,D.J. Dong,H. Byrne,M.C. Follettie,M.T. Gallo,M.V. Chee,M.S. Mittmann,M. Wang,C. Kobayashi,M. Horton,H. and Brown,E.L. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*, **14**, 1675-1680