Chapter 3
Computational Sequence Design Techniques for DNA Microarray Technologies

Dan Tulpan
National Research Council of Canada, Canada

Athos Ghiggi
University of Lugano (USI), Switzerland

Roberto Montemanni
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Switzerland

ABSTRACT

In systems biology and biomedical research, microarray technology is a method of choice that enables the complete quantitative and qualitative ascertainment of gene expression patterns for whole genomes. The selection of high quality oligonucleotide sequences that behave consistently across multiple experiments is a key step in the design, fabrication and experimental performance of DNA microarrays. The aim of this chapter is to outline recent algorithmic developments in microarray probe design, evaluate existing probe sequences used in commercial arrays, and suggest methodologies that have the potential to improve on existing design techniques.

INTRODUCTION

The design of DNA oligos is a key step in the manufacturing process of modern microarrays – biotechnology tools that allow the parallel qualification and quantification of large numbers of genes. Areas that have benefited from the use of microarrays include gene discovery (Andrews *et al.*, 2000; Yano, Imai, Shimizu, & Hanashita, 2006), disease diagnosis (Yoo, Choi, Lee, & Yoo, 2009), species identification (Pasquer, Pelludat, Duffy, & Frey, 2010; Teletchea1, Bernillon, Duffraisse, Laudet, & Hänni, 2008) and toxicogenomics (Jang, Nde, Toghrol, & Bentley, 2008; Neumanna and Galvez, 2002).

Microarrays consist of plastic or glass slides, to which a large number of short DNA sequences (probes) are affixed at known positions in a matrix

DOI: 10.4018/978-1-61350-435-2.ch003

Copyright © 2012, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

pattern. A probe is a relatively short DNA sequence (20-70 bases) representing the complement of a contiguous sequence of bases from a target that acts as its fingerprint. The purpose of each probe is to uniquely identify and bind a target via a process called hybridization. Nevertheless, in practice probes could bind to more than one target via a process called cross-hybridization.

While microarrays could be used for a variety of applications like transcription factor binding site identification (Hanlon & Lieb, 2004), eukaryotic DNA replication (MacAlpine & Bell, 2005), and array comparative genomics hybridization (Pinkel & Albertson, 2005), their main use remains gene transcript expression profiling (Schena, Shalon, Davis, & Brown, 1995; Ross *et al.*, 2000; Aarhus, Helland, Lund-Johansen, Wester & Knappskog, 2010). However, at present, the fundamental understandings of the bio-chemo-physical mechanisms that power this technology are poorly understood (Pozhitkov, Tautz, & Noble, 2007), thus leading to hybridization signal levels that are still not accurately correlated with exact amounts of target transcripts. While most of the microarray research work carried today focuses on the development of reliable and fault-tolerant statistical techniques that could pre-process large data sets (Holloway, van Laar, Tothill, & Bowtell, 2002; Irizarry *et al.*, 2003; Quackenbush, 2002; Yang *et al.*, 2002; Zhao, Li, & Simon, 2005) and identify significant factors relevant to each particular study (Chu, Ghahramani, Falciani, & Wild, 2005; Harris & Ghaffari, 2008; Leung & Hung, 2010; Peng, Li, & Liu, 2007; Zou, Yang, & Zhu, 2006), more work needs to be done on improving the infrastructural aspects of microarray technology, thus reducing the amount of noise earlier rather than later in an experiment based on microarrays data.

Thus, one of the greatest challenges in DNA microarray design resides in how to select large sets of unique probes that distinguish among specific sequences from complex samples consisting of thousands of closely similar targets. The daunting task of designing such large sets of probes is hampered by the computational costs associated with probe efficacy evaluations. Various design strategies are presented that employ the utilization of intricate probe evaluation criteria. Some of these strategies were inspired from design techniques employed for solving similar problems that arise in coding theory (Bogdanova, Brouwer, Kapralov, & Östergård, 2001; Gaborit & King, 2005; Gamal, Hemachandra, Shperling, & Wei, 1987), bio-molecular computing (Feldkamp, Banzhaf, & Rauhe, 2000; Frutos *et al.*, 1997), molecular tagging (Braich *et al.*, 2003; Brenner & Lerner, 1992) and nano-structure design (Reif, Labean, & Seeman, 2001; Yurke, Turberfield, Mills, Simmel, & Neumann, 2000).

In recent years there has been considerable interest in the application of meta-heuristic algorithms for the design of DNA strands to be used in microarray technologies. Most of the proposed algorithms deal with combinatorial constraints only (Frutos *et al.*, 1997; Marathe, Condon, & Corn, 2001; Kobayashi, Kondo, & Arita, 2003), in order to increase the tractability of the problem from a computational point of view. On the other hand, because of this simplification, the strands obtained do not always have the desired characteristics when used in experimental settings. Therefore, thermodynamic constraints are typically employed to have satisfactory results in practice, notwithstanding they tend to destroy most of the combinatorial structures exploited by the algorithms themselves. A summary of state-of-the-art combinatorial algorithms is presented here. We then describe how two algorithms originally developed for combinatorial constraints can be extended to efficiently deal with thermodynamic constraints.

The first approach is a Stochastic Local Search method originally described in (Tulpan, Hoos, & Condon, 2002; Tulpan & Hoos, 2003; Tulpan, 2006), which works on the search space of infeasible solutions. This means it starts with a given number of random strands, violating many

33 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/computational-sequence-design-techniques-dna/60828

Related Content

Alzheimer's and Parkinson's Disease Novel Therapeutic Target: The Mitochondrial Pyruvate Carrier - Ligand Docking to Screen Natural Compounds Related to Classic Inhibitors
Allen K. Bourdon, Greg Villareal, George Perryand Clyde F. Phelix (2017). *International Journal of Knowledge Discovery in Bioinformatics (pp. 68-82).*
www.irma-international.org/article/alzheimers-and-parkinsons-disease-novel-therapeutic-target/190793

Social Network Sites and Their Role in the Sharing of Health Information
Prajesh Chhanabhai (2012). *Pharmacoinformatics and Drug Discovery Technologies: Theories and Applications (pp. 236-247).*
www.irma-international.org/chapter/social-network-sites-their-role/64075

MSDC-0160 and MSDC-0602 Binding with Human Mitochondrial Pyruvate Carrier (MPC) 1 and 2 Heterodimer: PPAR Activating and Sparing TZDs as Therapeutics
Clyde F. Phelix, Allen K. Bourdon, Jason L. Dugan, Greg Villarealand George Perry (2017). *International Journal of Knowledge Discovery in Bioinformatics (pp. 43-67).*
www.irma-international.org/article/msdc-0160-and-msdc-0602-binding-with-human-mitochondrial-pyruvate-carrier-mpc-1-and-2-heterodimer/190792

Identification Methods of G Protein-Coupled Receptors
Meriem Zekri, Karima Alemand Labiba Souici-Meslati (2011). *International Journal of Knowledge Discovery in Bioinformatics (pp. 35-52).*
www.irma-international.org/article/identification-methods-protein-coupled-receptors/73910

Mining of Medical Trends Using Social Networks
Shruti Kohliand Sonia Saini (2016). *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes (pp. 164-182).*
www.irma-international.org/chapter/mining-of-medical-trends-using-social-networks/140489