Sequence reconstruction from nucleic-acid microarray data

Franco P. Preparata
Department of Computer Science
Brown University, USA
franco@cs.brown.edu

Abstract

Sequencing by Hybridization (SBH) has been proposed as an alternative to sequencing by gel electrophoresis over a decade ago. This approach consists of the biochemical acquisition, by complementary hybridization, of all subsequences of a given target sequence (sequence spectrum), followed by the algorithmic reconstruction of the sequence from its spectrum. However, the potential of this method has not materialized due to biochemical difficulties and, more essentially, to the combinatorial inadequacy of the probing scheme. Recently, the latter hurdle has been overcome by a novel scheme, whose probing pattern contains gaps of "don't care" symbols, and whose performance has been demonstrated to approach the information theory bound. This paper reviews the new technology from a combinatorial viewpoint, presents a probabilistic analysis of its failure modes with reference to maximum-entropy random sequences, and explores new sequence reconstruction refinements which have substantially narrowed the gap between the optimal and the achievable.