On the Power of Universal Bases in Sequencing by Hybridization

Franco P. Preparata   Alan M. Frieze   Eli Upfal

Abstract

Sequencing by hybridization is a novel DNA sequencing technique in which an array (SBH chip) of short sequences of nucleotides (*probes*) is brought in contact with a solution of (replicas of) the target DNA sequence. A biochemical method determines the subset of probes that bind to the target sequence (the *spectrum* of the sequence), and a combinatorial method is used to reconstruct the DNA sequence from the spectrum.

Since technology limits the number of probes on the SBH chip, a challenging combinatorial question is the design of a smallest set of probes that can sequence an arbitrary DNA string of a given length. We show in this work that the use of universal bases (bases that bind to any nucleotide [?]) can drastically improve the performance of the SBH process. We present a novel probe design with performance that asymptotically approaches the information-theoretical bound up to a constant factor, and, for any number of probes, is significantly better than previously analyzed probe patterns. Furthermore, the sequencing algorithm we use is substantially simpler than the Eulerian path method used in previous work.