

CpG Island Search Tool Application Development

ABSTRACT

The frequency of CG dinucleotides is lower than expected in vertebrate DNA due to the frequent methylation of cytosine, which results in a mutation to thymine, producing TG dinucleotides. There are however areas of DNA with CG frequency greater than the expected amount. These regions are referred to as CpG islands (CGIs), are often found near promoters (2,4,5,9). Because of this, methylation of CpG islands can affect gene regulation, and has been implicated in the initiation of cancer (10,13). CpG islands located in the promoter regions of genes can also play important roles in gene silencing (8,12,14). It is therefore important to find these islands in order to study the effects of methylation in these regions.

Since new algorithms for finding these islands are used and tested in the literature often, a software application to perform these searches should be flexible enough to keep up with the latest CpG island definitions. A rewrite of an existing application with hardcoded parameters is presented here, along with considerations for generically defining the parameters used in a CGI search. Software engineering issues, such as maintenance, portability, and running time, are also discussed. This internship project was completed as an employee of Codon Solutions LLC, working with Ribomed Biotechnologies Inc.