

## 論文の内容の要旨

論文題目     A computer aided design of large scale  
                  biological experiments based on subsequence properties

(部分文字列の性質に基づく計算機援用大規模生物実験設計)

氏名     土井晃一郎

ヒトゲノム計画が1980年代に始まって以来、現在に至るまでゲノム解析が活発に行われている。多くの生物実験が、配列決定、遺伝子発現パターンの解析などを目的に数多く行われている。そして、膨大な生物実験から処理しなければならない膨大なデータが生まれ、情報科学の研究者の計算生物学という分野の研究も数多くなされている。本研究では生物学的に大変重要なDNA配列決定、遺伝子発現パターンの解析といった目的の実験デザインとその周辺の問題を情報科学的な問題と捕らえ直し、効率的で実用的なデザインとその情報科学的性質を明らかにすることをを行う。

はじめに、配列決定アルゴリズムを扱う。DNA配列を決定することはゲノムサイエンスにおいて大変重要な問題である。配列を決定する方法は様々な方法が考えられている。ここではシーケンシングバイハイブリダイゼーション(SBH)と呼ばれるDNAチップによって得られる部分文字列の有無のデータによる配列決定の方法について扱う。このSBHは理論的にはさまざまな研究がなされているが、実用上大切なエラーに対する対処があまり考えられていない。本論文では、DNAチップにおけるハイブリダイゼーションに誤りがある場合に対しても効率的に動くようなアルゴリズムを提案する。そして、エラーを入れたランダムデータに対して適用した。

それにより、ハイブリダイゼーションデータが欠損している場合の影響が過剰にハイブリダイゼーションしているときより大きいものとなっていることを示した。また少量のエラーに対するこの手法の有用性を示した。

次に、PCR 実験におけるプライマー選択を行う。プライマー選択は PCR 実験の効率を大きく左右するものである。今までに作られてきたプライマー選択に対するプログラムは 1 つの DNA 配列に対して 1 組のプライマーを選択するものであった。本論文では、プライマーが複数の箇所につくことを許すことにより全ての DNA 配列を増幅するような少ないプライマーの集合を選択することを考える。このプライマー選択問題を定式化して、その問題の計算量的性質を解析する。それにより、増幅の長さを異なるようにするのが計算量的に難しいということを示した。そして、単純な定式化に対する貪欲アルゴリズムに禁止条件を作って、生物学的制約を満たすようにした。このアルゴリズムを酵母などの配列が分かっている DNA 配列に対して適用して、よいプライマーセットを得、複数実験に対するプライマーセットも得ている。

SBH の解析というのはランダムな配列に対してなされている。しかし、実際には DNA 配列に適用するものであるし、また、DNA 配列はランダムな配列でないのは明らかである。そこでランダム配列と DNA 配列との違いを SBH のアルゴリズムを適用することによって解析する。また、SBH によって解が一意に定まらない場合について理論的な考察を行う。