

## 論文内容の要旨

博士論文題目 Bioinformatics tool for genomic era: A step towards the  
*in silico* experiments focused on molecular cloning

氏名 大山 彰

### (論文内容の要旨)

本論文では、情報科学と分子生物学の融合を目指し、インビトロ（試験管内）での生物学実験を模倣することを目標としたインシリコ（コンピュータ内）における新規分子生物学実験シミュレーションシステムの構築について述べられている。インシリコ実験にはいくつかの利点があり、たとえば、分子生物学実験の計画に活用できる、実験ノート  
の代用となる、分子生物学実験の教育用ツールとして用いることができることなどが挙げられる。しかし、インシリコ実験を実装するためには、制限酵素による消化切断生成物の末端形状や PCR 生成物などの記録方法などの定式化が必要である。新規 Feature および Qualifier として定義することによる GenBank/EMBL データベース注釈規約拡張により制限酵素処理による切断構造を維持したシステムの開発を進めた。さらに、DNA 配列上に表現された Feature は PCR や制限酵素消化切断などの際に部分的に削られることをも可能にした規約の定式化法を考案した。また、DNA 配列解析ソフトウェアでは、通常興味の対象としては片方の DNA 鎖のみを表示するが、インシリコ実験を実装するソフトウェアにおいてはこの対象としての長鎖 DNA 配列をその上に Feature を保持したままその逆相補鎖と頻繁に切り替える操作が必要であることを示した。これらの定義やデータ記述に従い、「インシリコモレキュラークロニング」というインシリコ実験ソフトウェアを開発し、いくつかの典型的な分子クローニング実験をコンピュータ上で実行し、このシステムの有効性を示した。同時に、このソフトウェアにはゲノム時代に対応した様々な機能を搭載した。ゲノム情報については、現在もなお膨大な量のデータが集積されつつあり、これらの大量情報を高速に参照かつ編集できることが望まれている。本ソフトウェアでは、これらのニーズに対応し、通常の PC 等で動作するソフトウェアとしては最高速のデータ参照・編集速度を実現した。この高速性能を利用して、異種ゲノム間の比較解析機能を充実させた。近年研究が進んでいる網羅的な発現解析への対応機能として、ゲノム全体をタイルのように多い尽くす高密度アレイであるタイリングアレイ用のデータ操作を情報科学の視点で検討し、発現情報プロファイル表示機能を開発した。これにより、アレイ解析速度が大幅に向上し、発現解析研究に貢献することが期待される。一方、より上流の実験である塩基配列シーケンシングを一層容易に遂行

するためのソフトウェアである「メタゲノムギャンプラー」を開発し、配列決定からアノテーションまでの一連の作業の多くを自動化するとともに、配列などの品質評価機能を充実させることにより、データの正確性を向上させている。このソフトウェアはゲノムサイズが巨大であるため、全ゲノム塩基配列決定が困難である植物ゲノムの cDNA 発現解析にも使用できる。

これらのソフトウェアを使用した応用例として、原核生物における塩基組成の非対称性を調べた結果、ある種の原核生物にはこれまで報告されていない転写に共役するGC組成の非対称性を確認した。原核生物の中でも比較的GC含量の低い種に確認されたこの現象は、特に以下の生物種、*Clostridium perfringens*を含む*Clostridium*属、*Fusobacterium nucleatum*などに著しく、*Escherichia coli*や*Bacillus*属にも、その存在が確認された。この発見に基づいた原核生物における転写単位の予測への応用の可能性についても検討した。

氏名	大山 彰
----	------

(論文審査結果の要旨)

平成18年8月10日に開催した公聴会の結果を参考に、平成18年9月7日に本博士論文の審査を実施した。以下に述べる通り、本博士論文は、本学位申請者が、独立した研究者としてバイオインフォマティクスを中心とした分野で研究開発活動を続けていくために必要な素養を備えていることを示すものである。

大山彰は、本博士論文において、情報科学と分子生物学の融合を目指し、インビトロ(試験管内)での生物学実験を模倣することを目標としたインシリコ(コンピュータ内)における新規分子生物学実験シミュレーションシステムの提案し、実際に構築した。その際、制限酵素による切断末端の記述方法を開発し、遺伝子データベースへ適用し、ゲノムが既知の生物からの遺伝子クローニング実験のシミュレーションを効率よく実現することを可能とした。同時に、膨大な量のデータが集積されつつある大量ゲノム情報についても、高速に参照かつ編集できる機能も実装し、通常のPC等で動作する遺伝子・たんぱく質解析ソフトウェアとしては最高速のデータ参照・編集速度を実現した。さらに、開発したシステムの、タイリングチップによる発現プロファイル解析および比較ゲノム解析などの、ゲノムおよびポストゲノム解析への有効性を示した。また、より上流の実験である塩基配列シーケンシングを一層容易に遂行するためのソフトウェアである「メタゲノムガンブラー」を開発し、配列決定からアノテーションまでの一連の作業の多くを自動化した。

本論文で提案され、実現した諸システムは、ゲノム情報に対応した大量情報処理を高速に、かつ、統合的に実現している点で実用的であり、既存の遺伝子解析ソフトウェアに対して優位性と新規性があると評価できる。また、JAVAを用いたマルチプラットフォーム上で高速処理の実現、GenBank/EMBLデータベース注釈規約拡張等、情報処理技術としても新規性があり、情報科学とバイオインフォマティクスの境界領域の発展に貢献するものである。

よって、本論文は、博士(理学)の学位論文としての価値があるものと認める。