

(論文内容の要旨)

膨大な量のゲノムデータが蓄積されるのに伴い、核酸・タンパク質の配列情報を解析するための様々なアルゴリズムや手法が開発されてきた。その一方で、これらと並ぶ重要な生体高分子である糖鎖や脂質に対しては、計算機による情報解析手法の開発はそれほど進んでいなかった。本論文において申請者は、糖鎖・脂質へのバイオインフォマティクスによるアプローチを可能にし、生物学的に意味のある情報を引き出すことを目的とした解析を行っている。

申請者は、まず糖鎖構造に関するデータのリソース KEGG GLYCAN を用いて解析を行った。KEGG GLYCAN には約 1 万の糖鎖構造が含まれており、このような大量の糖鎖構造を計算機によって取り扱う手法の開発が本論文の前半の内容である。具体的には、多数の糖鎖構造から生物学的機能に関連のある部分構造を効率的に抽出するという問題を設定し、アルゴリズムの開発を行っている。申請者はこの問題に対し、データマイニングの分野で提案されている手法を改良し、新たな木構造マイニング法 (α -closed 法) を開発した。一方、得られた部分構造の有意性を検定する手法も同時に開発している。この検定は、探索の対象となるデータセットに加えてコントロールデータセットを作成し、2 つのデータセットに特定の部分構造がいくつ含まれるかを数え、それによって得られた 2x2 分割表を用いて行われる。これら α -closed 法と有意性検定を組み合わせた手法を、KEGG GLYCAN に登録されている実際の糖鎖構造に適用したところ、Lewis X、Lewis A などの部分構造や、糖タンパク質、糖脂質のコア構造といった機能的に重要と考えられる部分構造が抽出された。

本論文の後半では、申請者は脂肪酸関連酵素の解析と脂肪酸構造の予測を行っている。具体的には、脂肪酸不飽和化酵素である *desaturase* と脂肪酸伸張酵素である *elongase* の組合せから、ゲノム配列の決まった 56 の真核生物が合成できる脂肪酸レパートリーを明らかにすることを目的とした解析を行った。その手法として、まず、PSI-BLAST による探索と階層型クラスタリングにより、*desaturase* 及び *elongase* のホモログを、対象とする生物種のゲノムから取得した。次に、系統解析によって *desaturase* は 4 つの、*elongase* は 2 つの、機能の異なるサブファミリーに分類できることを示した。さらに、各生物種がどのサブファミリーの酵素を持つかを明らかにし、それらをパスウェイにマッピングすることで、主要な 7 つの脂肪酸について合成能力を予測した。その結果、生物界の間で合成できる脂肪酸に違いが見られるだけでなく、近縁種の間でも違いがあることが示唆された。この理由として申請者は、*desaturase* と *elongase* が 2 段階の多様化を経て進化してきたと考察している。すなわち、最初の多様化によってこれらの酵素は異なる機能をもつサブファミリーに分かれ、次の多様化で細かい機能分化が起こったということである。このことは、系統樹やサブファミリー毎のモチーフの違いとも一致している。最後に、このような酵素の機能分化によって獲得された脂肪酸は、環境への適応や新しい代謝物の合成に役立っていることが、いくつかの実例によって示唆された。

(論文審査の結果の要旨)

本論文は糖鎖と脂質に関する2つの研究成果が述べられており、まずそれぞれについて評価を行った。糖鎖構造データベースから部分構造を抽出する問題は、木構造から頻出する部分構造を抽出するという問題であり、データマイニングの分野で既に行われてきたことであるが、本研究ではこれを改良し、生物学的に意味のある手法として確立したことが評価される。糖鎖分子に適用する際の最大の問題は、抽出される部分構造の数が多すぎるために、機能的に意味のある結果を得られないという点であった。開発された α -closed 法では、冗長な構造を取り除くことができるアルゴリズムを実装している。実際の糖鎖構造を用いた解析においても、パラメータ α の調節によって冗長な部分構造の数を大幅に減少できることが分かり、本手法の有効性が示された。さらに有意性検定においては、探索対象のデータセットとコントロールデータセット、ある構造を含む場合と含まない場合、という2つの因子によって分割された表を使い、2因子の独立性をフィッシャーの正確確率検定によって行う方法を提案している。これを用いて計算される P 値でランク付けを行い、頻出構造の中から統計的により意味のある部分構造を抽出することを可能とした。実際の糖鎖構造を用いた解析において、生物学的な観点からこの検定の有効性が示されている。

一方、脂質の研究では、新しい手法の開発ではなく生物学的な解析に焦点をあて、脂肪酸修飾酵素の網羅的な機能解析と脂肪酸構造の予測を行っている。まず、PSI-BLAST やクラスタリングといった既に確立されている配列解析の手法を組み合わせることによって、多数のゲノムから信頼性の高いホモログセットを得ることに成功している。ここでは、得られた配列群を機能の異なるサブファミリーに分類し、サブファミリー毎に明確に保存された異なる種類のアミノ酸を見出した点が重要である。この発見により、配列だけからサブファミリーレベルの機能の違いを区別することが可能になり、実配列を用いたテストの結果からもその判別能力の高さが示された。さらに、これを応用すれば、与えられたゲノム配列から含まれる desaturase と elongase のセットを直ちに明らかにすることができ、今後決定されるゲノムに容易に適用できると考えられる。また、酵素のセットから予測された合成可能な不飽和脂肪酸に関しては、この予測結果が妥当であることが示されただけでなく、酵素の機能分化によって獲得された脂肪酸は、環境への適応や新しい代謝物の合成に役立っていることも示唆され、生物学的に非常に興味深い結果であるといえる。

核酸・タンパク質と比較すると、糖鎖・脂質の情報については、アルゴリズム等の新しい情報技術開発の面でも、またゲノム等大量データからの知識発見の面でも、バイオインフォマティクス研究の中で未開拓の分野であった。本論文ではこうした分子群に対する計算機的アプローチの可能性を示すことに成功している。よって、本論文は博士(理学)の学位論文として価値あるものと認められる。また、論文内容とそれに関連した試問の結果、合格と認めた。