

## 研究開発課題別中間評価結果

### 1. 研究開発課題名

2項関係に基づくゲノムと生命システムの機能解読

### 2. 代表研究者名

金久 實 (国立大学法人京都大学化学研究所 教授)

### 3. 研究開発概要

本研究開発では、ゲノムから細胞、個体までの高次生命システムの機能や有用性を解読するために、オントロジーを構成する階層的な2項関係も含め、様々な2項関係を集積したBRITE(Biomolecular Relations in Information Transmission and Expression)データベースを構築するとともに、ゲノムの情報から高次生命システムの機能情報を解読する推論システムの開発を行う。KEGG(Kyoto Encyclopedia of Genes and Genomes)においてパスウェイの階層として表現された遺伝子の機能分類をオントロジー化し、同時に生体内化学反応の情報をオントロジー化して遺伝子機能と融合することにより、分子レベルの物質情報から細胞・個体レベルの機能情報を解読する推論システムを開発する。また、ゲノムの情報を医療や創薬へつなぐことを目的とし、病気のメカニズムに関する知識の集約とそのオントロジー化を行っている。さらに、知識集約のために、文献や専門家の頭の中にある知識をコンピュータ化する支援システムの開発も行う。

### 4. 中間評価

#### 4-1. 研究開発の進捗状況と今後の見込み

KEGGの分子間相互作用ネットワークという実体のあるデータオブジェクトに焦点を絞ってオントロジーを開発している。その要素の1つとして、化学反応の知識を体系化することにより、新たな反応を予測し酵素遺伝子を予測する方法論を開発している。そのため、化学構造比較アルゴリズムを開発し、これを用いて酵素反応の基質と生成物の化学構造変化を網羅的に解析し、反応パターン分類を行っている。

病気については、オントロジーのデザイン、感染症に関する知識の収集 及び知識収集のための支援システムの開発を行っている。また、代謝異常に関与する遺伝子の情報をKEGGの代謝パスウェイに対応づけて蓄積している。

研究者コミュニティと連携した知識集約については、シアノバクテリアや枯草菌などの微生物ゲノムの遺伝子機能アノテーションのためのシステムを開発し、運用を行っている。また、新たに糖鎖研究者コミュニティと連携し、糖鎖構造に関する知識集約を行うアノテーションシステムを開発し、運用を開始している。シアノバクテリアについてはすでに国際的な連携が行われており、糖鎖についても国際的な協力関係ができつつある。当初計画に比べ、目的が明確になり、着実な進捗状況である。さらなる今後の展開に期待する。

#### 4-2. 研究開発成果の現状と今後の見込み

Web公開状況

BRITEデータベース <http://www.genome.ad.jp/brite/>

コミュニティデータベース (CYORF/BSORF/GLYCAN) <http://orf.genome.ad.jp/>

ゲノムネットサービスの月間アクセス件数は約500万件で、その8割が海外からのアクセスである。1日あたりのユニークホスト数は6千人に達している。

本研究開発で、基質・生成物の化学構造からEC(Enzyme Commission)番号を自動アサイメントするシステムが開発され、このシステムはEC番号命名委員会にも提供されている。

BRITEを構成するデータ群としての2項関係ECrel、PPrel、GErelは、KEGGのパスウェイマップ上に存在する分子(主にタンパク質)間の関係で、これを手作業で取得するツールtkmapinfoを開発し、取得した2項関係データは標準化されたXML形式のKEGGパスウェイマップに含めて公開している。一方、手作業での遺伝子機能の階層分類KO(KEGG Orthology)と配列比較による計算による遺伝子機能分類OC(Ortholog cluster)、RC(Reaction Classification)は生体内の化学反応分類で、いずれも国際標準となる分類体系を目指している。また、これらをもとにゲノムから生命システムへの推論を可能とするBRITEのオントロジーを開発している。

さらに、5種類のデータベース構築用ソフトウェアと4種類の一般利用者ソフトウェアが開発されている。

本研究開発の成果はweb上での公開やNature Genetics等の国際雑誌への掲載、様々な学会での発表等、広範に発表されている。着実に成果をだしており、今後にも期待する。

#### 4-3. 今後の研究開発にむけて

オントロジー開発としては、ゲノム情報と化学情報を融合させ、高次生命システムにつながるオントロジーを開発し、代謝系については今後2年での完成を目指している。標準化については、国際標準となる分類体系の確立を目指している。

化学情報をゲノム解読に利用するというユニークな観点で、今後とも低分子化合物、糖鎖、脂質、ペプチド等の構造情報を蓄積していくことが考えられている。また、ゲノム解読の観点から重要であるにもかかわらず、データベース化が遅れている部分にも力を入れ、物質の構造を統合して入力できるツールの開発が計画されている。

分子の2項関係データ入力は、KEGGとの連携体制による対応が考えられている。

疾病データの入力は感染症の他に神経変性疾患、代謝異常、免疫異常、がん等も取り上げて知識集約、2項関係データの入力を行うことが計画されている。

推論システムの開発としては、オントロジーを取り入れた推論システムを新規に開発し、ゲノム情報から代謝系を解読し、代謝異常について示唆を与えるプロトタイプシステムを開発する計画である。ネットワークと生物学的コンセプトの重要性を考えれば、これらについて優先的に、さらに充実させることを望む。二項関係データをさらに蓄積し、マイクロアレイなどに代表される網羅的研究で得られる大量データの解析に資することを期待する。

#### 5. 国内外のデータベース高度化・標準化の動向・状況と本課題の位置づけ

KEGGは既に知識集約型データベースとして国際的な地位を築いており、さらに高度なデータベースを目指しているBRITeは、意義深い総合データベースである。KEGGに集約されたデータベースに、それとは独立に得られる二項関係をあてはめ、推論することにより、代謝マップ上の位置確認と新しいマップ予測が可能になるシステムは、世界初の試みであり、国際的な地位の確立が期待できる。知識体系の内容の標準化には独自の方策が必要であり、具体的には遺伝子機能KOやEC番号付与の自動化を補助する生体内反応分類RC等を開発して、実際に提供している。

#### 6. 研究開発成果の社会への貢献

本研究開発の成果として、ネットワークと二項関係から新たに有用な遺伝子見いだすアプローチが実用化されると、バイオインフォマティクス技術力で国際的に優位に立つことができる。また、遺伝子機能と生体内化学反応の階層が明らかになれば、創薬や環境浄化への応用が可能になり、人類の健康と福祉の向上及び地球環境の維持・保全に貢献が期待できる。

#### 7. 総合的評価

本計画は、すでに国内外で高い評価を得ているKEGGの高度化として着実に進んでおり、成果を出している。当初の計画より対象を絞ったことにより、目的が明確になっている。KEGGをBRITeの技術的手法によって根底から再構成し、高度化しようとする試みの意義は大きく、基礎知識の充実だけでなく、企業などへの研究開発に対する貢献も大いに期待される。今後、目標をふまえながら、生物に特有の細胞の変化に伴う時間軸や、固有の位置を占めるそれぞれの組織や器官、細胞などの性質を表す空間軸の導入など、生物の拘束条件に沿った、意味のあるデータベースの構築に挑戦して貫きたい。また、多数のデータベースとの統合など、更なる高度化についても検討を期待したい。本計画に期待されるテーマの大きさを考えれば、それに則して研究開発体制を充実させ、継続していくことが重要である。