

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」

代表研究者 京都大学化学研究所 教授 金久 實

1. 代表研究者による成果概要報告

1-1. 研究開発のねらい

ヒトからバクテリアまで数多くの生物種において全ゲノム配列が続々と決定され、ゲノム情報を基盤に細胞、個体、生態系といった高次生命システムの理解が進み、同時に創薬、医療をはじめとしたゲノム情報の有効利用が活発化している。これまでに決定されたゲノムの配列情報は国際 DNA データベース DDBJ/EMBL/GenBank に登録され、誰でも自由に利用できる形になっている。しかしそれだけでは不十分であり、ゲノムに書かれた生命のはたらきや有用性を見いだすことを可能にするデータベースがなければならない。生物情報データベース高度化の一貫として、本研究開発では新しいタイプの機能情報データベース BRITE を構築する。

機能情報のデータベース化については、大きく 2 つのやり方がある。1 つは配列データベースのアノテーションのように、言葉として記述することである。機能情報はただ利用者が読んで理解できればよいとするのなら別であるが、一般には生物種間の比較をしたり、コンピュータ処理をするために、語彙の標準化を行わなければならない。また、分子、細胞、個体といった生命システムのどのレベルでの機能なのか、機能情報の階層化を行う必要がある。これは狭い意味でのオントロジーの問題である。GO (Gene Ontology) では、異なる生物種での遺伝子アノテーションの標準化と知識の共有のために、Biological process、Cellular component、Molecular function の 3 つの観点（オントロジー）で語彙の階層的な定義を行っている。

もう 1 つは我々が KEGG において提唱し実践しているやり方で、細胞レベルの生命システムの「はたらき」を分子間相互作用ネットワークの「かたち」として表現する。ゲノムの遺伝子の並びからタンパク質同士のつながり方（かたち）を予測し、すなわち KEGG パスウェイを再構築し、そこから例えばリジン合成するはたらきがある、浸透圧変化に応答するシグナル伝達のはたらきがあると判定する。主観的な語彙での表現ではなく客観的な形の表現を用いているところが KEGG の特色である。これは逆の言い方をすると、KEGG パスウェイで表現ができる機能情報とは、分子間ネットワークが解明されたものに限られるわけで、例えばこの遺伝子は細胞周期に関与しているらしいといった手がかり程度では、KEGG では表現ができない。

本来、機能とは曖昧なものであり、大雑把な手がかりであっても有用性はあり得る。そこで本研究開発の BRITE では、上記 GO のやり方と KEGG のやり方を、語彙の 2 項関係で融合する。2 項関係とは 2 つのオブジェクト間の関係情報であり、KEGG パスウェイにおける分子間の関係、オントロジーの階層における親子関係、さらには「かたち」と「はたらき」の関係として、配列や立体構造と分子機能との関係、ネットワーク構造と細胞機能との関係といった様々な関係を含めて考える。2 項関係の集合はグラフであり、語彙で表現された様々なオブジェクトがそのノードとなる一般的なグラフを考えていることになる。本研究開発ではこのように KEGG の高度化・標準化の観点から BRITE データベースを構築する。

1-2. 研究開発の成果

本研究開発開始時の KEGG は、生命システムを構成する部品の情報として遺伝子・タンパク質に関するゲノム情報（GENES データベース）と化合物・化学反応に関するケミカル情報（LIGAND データベース）、部品間の配線図情報として相互作用・反応ネットワークに関するパスウェイ情報（PATHWAY データベース）を統合したデータベースであった。本研究開発により、生命システムに関連した様々な機能階層情報を BRITE データベースとして構築し、KEGG の第 4 の基幹データベース KEGG BRITE

(<http://www.genome.jp/kegg/brite.html>) として統合した（図 1）。

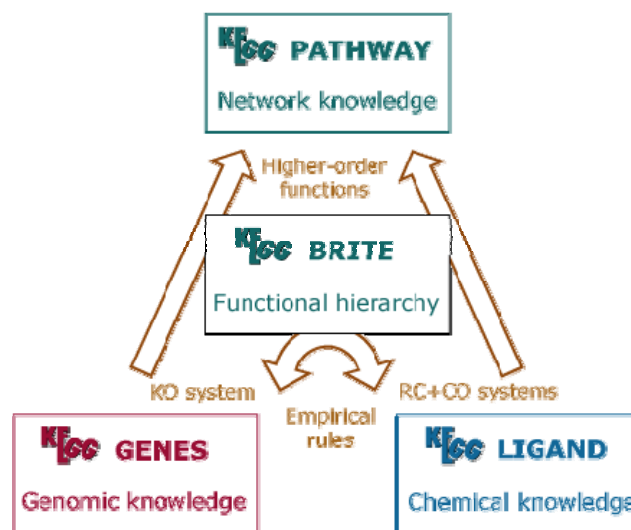


図 1.KEGG 高度化・標準化としての BRITE

BRITE の大きな目標はゲノムから高次生命システムの機能と有用性を解読するプロセスを自動化することであった。そのためにまずデータベースの内容として、パスウェイマップで表現された「かたち」の情報に語彙で表現された「はたらき」の情報を付与することと、パスウェイマップでは表現できない幅広い機能情報を語彙の体系及びそれに伴う分子の体系としてコンピュータ化することを行った。BRITE データベースは階層テキストファイルと呼ぶファイルの集合で、その内容は表 1 の通りである。この中で KO (KEGG Orthology) は、もともと KEGG パスウェイに対応したオーソログ遺伝子グループの機能階層分類として始めたものであるが、パスウェイが既知という大きな制約のため、ゲノム中でカバーされる遺伝子数が限られるという大きな問題があった。本研究開発においてタンパク質ファミリーの知識を KO システムに反映させ、また独立に作成した化合物の分類、化学反応の分類、あるいは薬の薬効分類などとも、低分子とタンパク質の相互作用、反応とそれを触媒する酵素の関係、薬とターゲットの関係を通じて KO システムと関連づけを行った。

表 1. BRITE データベースの内容 (2006 年 2 月末現在)

<p>Genes and Proteins</p> <ul style="list-style-type: none"> Network hierarchy KO Protein families Enzymes Cytochrome P450 Transcription factors Ribosome Translation factors ABC transporters G-protein coupled receptors GTP-binding proteins Ion channels Cytokines Cytokine receptors Cell adhesion molecules (CAMs) CAM ligands CD molecules Bacterial motility proteins 	<p>Compounds and Reactions</p> <ul style="list-style-type: none"> Compounds Compounds with biological roles Lipids Phytochemical compounds Compound interactions Ion channel agonists/antagonists Cytochrome P450 substrates <p>Drugs and Diseases</p> <ul style="list-style-type: none"> Drugs Therapeutic category of drugs Drug classification Diseases Infectious diseases <p>Cells and Organisms</p> <ul style="list-style-type: none"> Organisms KEGG organisms
---	---

一方、自動化を実現するプログラムとしては、KAAS (KEGG Automatic Annotation Server) システムを開発した。このプログラムは、ゲノム配列中の各遺伝子や大量の EST から生成されたコンセンサスコンティグに対して自動的に KO づけを行い、これを KEGG のパスウェイ上の KO や BRITE の機能階層中の KO にマッピングすることで、特定のパスウェイや特定の機能階層が浮かび上がる (色づけされる)。ここまでのプロセスは自動化され、機能解釈の自動化は技術的には達成された。今後は BRITE データベースを充実させること、とくに人体あるいは病気といった高次レベルの知識の階層化と KO システムとの関連づけを行うことで、さらに高次の機能解釈が可能になると考えている。

2. 事後評価結果

2-1. 当初計画の達成度

研究開発計画の個々の項目の実施という観点から見ればいくつか変更があったことから、当初計画に記載したことがすべて実現されているわけではない。しかし、BRITE の位置づけとしてオントロジーのような機能階層に焦点を絞ったこと、つまり BRITE に KEGG を取り入れるのではなく、BRITE を KEGG の一部としたことで、本研究開発の目的が明確にされた。KEGG に GO のような語彙の階層表現を取り入れ、KEGG BRITE として完成させたことから、大きな枠組みの中で当初計画は充分達成されたと評価できる。

2-2.知的財産権、外部発表(論文等)等研究開発成果の状況

KEGG BRITE : <http://www.genome.jp/kegg/brite.html>

KEGG を中心としたゲノムネット (www.genome.jp) への月間アクセス数は1千万件に達し、1日あたりのユーザー数(ユニークホスト数)は約8千人で、その9割近くが海外からである。原著論文発表、招待・口頭講演も海外におけるものが多く、国際的な活躍が認められる。論文への引用も多い。

2-3.研究開発成果の公開による波及効果

KEGG は今や米国の NCBI や欧州の EBI、SwissProt と並び、日本を代表する国際的な生物情報データベースである。NCBI との緊密な連携、また米国糖鎖コンソーシアム、国際生化学命名委員会、シアノバクテリア研究コミュニティ、その他との連携も進んでおり、大きな影響を与えている。KEGG のパスウェイ情報やケミカル情報は、新しいポストゲノム研究を開拓し、新しいデータベースの考え方を提供してきたが、これだけでは KEGG の先駆性を維持することはできない状況でもあった。本研究開発により、階層ファイルによる知識の統合、KegHier 等のデスクトップアプリケーションの利用といった大きな可能性が示されたことは非常に意義がある。

2-4.成果の実用化の可能性及び成果から予想される波及効果

本研究開発で行われている薬と病気に関する知識集約は、日常生活に直結する部分であり、ゲノム研究・ポストゲノム研究の成果、バイオインフォマティクスの成果を直ちに社会へ還元できる部分である。誰でも自由に利用でき、1つ1つのデータを閲覧することよりも系統的なコンピュータ処理に重点を置いている KEGG に、薬に関する情報がゲノムその他の大量データと統合されていることは大変意義深い。これらのデータを、従来とは異なるタイプの研究者が異なる視点で解析することにより、新たな創薬・医療への応用が大いに期待される。また、KEGG は研究基盤としてだけでなく、日本語化や KegHier で学生の教育あるいは一般の啓蒙にも有用な基盤情報であり、薬や病気を通してゲノムへの理解を深めることが非常に期待できる。

2-5.総合評価

すでに日本を代表する国際的なデータベースとなっている KEGG の高度化・標準化の観点から着実な成果を出している。BRITE データベースを構築し、KEGG のネットワーク構造の階層に基づく高次機能の推論と、KEGG パスウェイでは表現できない知識のコンピュータ化を、階層テキストファイルという語彙の体系で実現したことは非常に意義が大きい。基礎知識の充実のみでなく、産業界への貢献も大いに期待できる。今後さらに研究開発体制を充実させ、日本を代表する国際的データベースとしてデータの信頼性の向上などを含めたさらなる発展を期待する。本研究開発の成果は国家戦略的に重要であり、継続して研究開発を進めていくべきものである。

3.主な論文発表

- 1)Kanehisa, M. and Bork, P.; Bioinformatics in the post-sequence era. *Nature Genetics* 33, 305-310 (2003).
- 2)Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M.; Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* 125, 11853-11865 (2003).
- 3)Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M.; Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.* 126, 16487-16498 (2004).
- 4)Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., and Kanehisa, M.; KEGG as a glycome informatics resource. *Glycobiology*, in press (2005).
- 5)Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M.; From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34, D354-357 (2006).

参考

論文発表 国内 0 件、海外 25 件

口頭発表 国内 13 件、海外 33 件

ポスター発表 国内 120 件、海外 7 件