

KEGG Chemical Function 形式を応用した化学部分構造データの生成と代謝パスウェイ再構築への応用

Generating chemical substructure data using KEGG Chemical Function format and its application to the metabolic pathway reconstruction.

化学生命科学研究領域 小寺正明

背景と目的

目的化合物の代謝パスウェイが未同定であるとき、そのパスウェイに関する仮説を立てるには官能基や化学部分構造を指す語彙を用いて自然言語で記述する方法がよく用いられる。一方で、計算化学や化学情報学の分野では、化合物は化学記述子と呼ばれる数値表現で記述される。この化学記述子は、生化学者がよく用いる官能基などの語彙と対応が付くものは少なく、計算により得られた結果を解釈する際に困難を生じている。私は KEGG Chemical Function (KCF)形式と呼ばれる化学構造記述フォーマットを応用し、生化学者が「官能基」「化学部分構造」として認識する構造をグラフ理論的に模倣する形で自動生成することを行なった。得られた KEGG Chemical Function-and-Substructure (KCF-S) の応用として、化合物集合からの特徴抽出、化合物のクラスタリング、および化学構造からの代謝パスウェイ再構築を行い、良好な結果を得ることが出来た。

検討内容

本研究では、筆者が 2008 年に発表した BiSSCat データベース（参考文献 2）で定義された「部分構造」の定義に対して、服部らが 2003 年に発表した KEGG Chemical Function (KCF)形式（参考文献 3）で定義した KEGG Atom Typing を適用し、化合物中に存在する複数の部分構造を文字列の集合（自然数ベクトル）として表現した。

結果と考察

化合物のデータ集合として、KEGG COMPOUND、KEGG DRUG、KNAPSAcK の 3 つのデータベースを用いて、それぞれのデータベースに特徴的な部分構造を抽出した。3 つのデータベースはそれぞれ化合物データの収集目的が異なり、結果として化合物の特徴も異なる（それぞれ中

心代謝、医薬品化合物、植物二次代謝を主な目的としている)。KEGG COMPOUND に特徴的に多く出現する部分構造としてはチオエステル結合、 β ケトカルボニル構造、イミダゾール環、アデニン環などがあつた。KEGG DRUG に特徴的に多く出現する部分構造としては第三級アミンや、ピペリジン環などがあつた。KNApSACk に特徴的に多く出現する部分構造には 5-アレニルベンゼン 1,2,3-トリオール環や、p-ヒドロキシベンゾエート骨格などがあつた。このような解析は、既存の他の化学記述子においては行なえず、部分構造と対応づけられる KCF-S を用いることの利点のひとつである。

既存の他の化学記述子との比較として、筆者らが開発した化学構造からの代謝パスウェイ再構築実験 (参考文献 1) の性能評価を行なった。手法としては、任意の 2 つの化合物を表すベクトルから、差分だけを考慮したベクトル、共通部分も考慮したベクトルを生成し、それらをサポートベクターマシンで「酵素反応式中に含まれる化合物ペア」と「含まれないペア」に分類する問題を解いた。結果として、化学情報学分野で知名度の高い PubChem fingerprint や MACCS key を上回る性能を得ることが出来た。

以上の結果の詳細は BMC Systems Biology 誌にて発表済みである (発表論文 1)。今後は、薬物・タンパク質相互作用解析や、酵素タンパク質配列との関連解析などに応用する予定である。

発表論文

1. Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Yuki Moriya, Toshiaki Tokimatsu, Minoru Kanehisa and Susumu Goto. **“KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics.”** *BMC Systems Biology*, 7(Suppl 6):S2 (2013).

参考論文

1. Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Toshiaki Tokimatsu and Susumu Goto. **“Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets.”** *Bioinformatics*, 29(13), i135-i144 (2013).
2. Masaaki Kotera, Andrew G. McDonald, Sinéad Boyce, Keith F. Tipton. **“Functional Group and Substructure Searching as a Tool in Metabolomics.”** *PLoS ONE*, 3(2): e1537 (2008). doi:10.1371 / journal.pone.0001537.
3. 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”**, *Journal of the American Chemical Society*, 125, 11853-11865 (2003)