

XML と Java を用いるインタラクティブな代謝経路マップ

石塚英弘、宮内智代*

図書館情報大学

*現在は東京大学アメリカ太平洋地域研究センター勤務

本研究の目的は Web リソースに関する RDF(Resource Description Framework)ベースのメタデータをグラフィカルかつインタラクティブに表示することであり、またそれを通じたメタデータの機能拡張に在る。対象は生物学、生物化学で基本的な知識である代謝経路マップとした。RDF に基づいて設計したモデルは6つのクラス：substrate, enzyme, direction, entity, block and map を持つ。そして、XML で書かれた RDF 外部表現を解釈し、代謝経路マップを描画する Java アプレットを開発した。そのマップではマウス・ドラッグによりノードをインタラクティブに移動することが可能であり、本研究は静的な RDF に動的な機能を加えるなど当初の目的を達成した。

An interactive metabolic pathway map with XML and Java

Hidehiro ISHIZUKA, Tomoyo MIYAUCHI*

Univ. Library and Information Science

*Present address: Center Pacific Amer. Studies, Univ. Tokyo

The aims of this study are to display metadata about Web resources graphically and interactively and to extend a function of metadata based on Resource Description Framework (RDF). The target of the study is a metabolic pathway map that is fundamental knowledge in biology and biochemistry. A designed model based on RDF contains 6 classes: substrate, enzyme, direction, entity, block and map. XML is used in the RDF external representation. The authors developed a Java applet that interprets the representation written in XML, and displays a metabolic pathway map. The map is interactively manipulated with mouse drag. The study extended a function of metadata.

1 はじめに

インターネット上で公開される Web リソースの数が膨大かつ多種多様になったため、Web リ

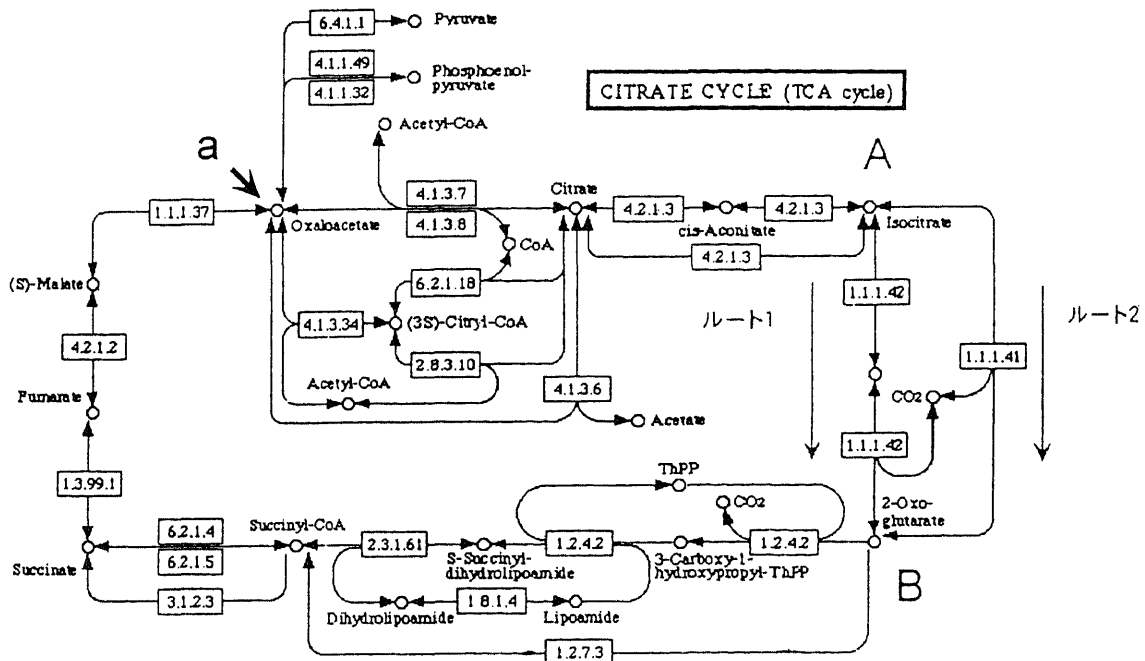


図1 人手による代謝経路マップ (クエン酸サイクル) の描画の例

出典: 京都大学化学研究所の KEGG <URL: <http://www.genome.ad.jp/kegg/pathway/map/>>

ソースに関するデータを記述するメタデータとそのモデルである RDF(Resource Description Language)[RDF]、そして RDF の外部表現に使用される XML(eXtensible Markup Language)[XML]などが注目されている[AI]。しかし、W3C の規格である RDF は未だ仕様確定段階には至っておらず、世界の利用者に対してコメントを求めていることから、なお様々な分野の Web リソースへの適用可能性を検討する必要があると言えよう。

そこで本研究では、複雑な情報構造を持つ代謝経路マップを適用対象に採り上げて検討し、RDF、XML、Java アプレットを用いることによってインタラクティブな代謝経路マップを試作した。

代謝経路マップとは、生体物質の代謝を担う酵素反応群を代謝経路の順に整理してマップ状に表現したものであって、生物学、生物化学の教科書に掲載される基本知識である。経路の形状は直線であることはむしろ少なく、環状の場合が多く、また分岐もあるなど複雑である。例えば、代表的な例であるクエン酸サイクルはクエン酸(citrate)が 9 段の酵素反応によって代謝される環状の経路であり、さらにその中に複数の分岐を含んでいる(図 1 参照)。

Web による代謝経路マップとしては、京都大学化学研究所によって行われているゲノムネットサービスの KEGG(Kyoto Encyclopedia of Genes and Genomes)の中の代謝経路マップが挙げられる。KEGG では代謝経路マップがクリッカブル・マップになっており、マップの中の任意の酵素反応をクリックすれば、その反応を記述した Web ページに跳ぶことができる。しかし、メタデータ、RDF、XML は何れも使用されていない。

Ochs[Ochs91][Ochs96]はコンピュータで代謝マップを扱う研究を行った。個々の酵素反応のデータは関係 DBMS に格納される。そして彼らが開発したプログラムが、そのデータに基づいて接

続する酵素反応を見つけ、代謝マップを木構造で表示する。しかし、木構造では環状の経路が表示できず、分岐の表示に無理があり、また、木構造で表示された代謝経路は教科書に描かれている環状のマップとは異なり見難いなどの欠点がある。

そこで本論文では、ユーザが環状かつ分岐を持つ代謝経路マップの構成要素をディスプレイ上でインタラクティブに移動操作することによって、ユーザの自由な視点で代謝経路を検討可能とすることを研究目的とし、RDF、XML、Java アプレットを用いることによってそれを実現したことを示す。

2 代謝経路マップの特徴と電子化の要件

代表的な代謝経路であるクエン酸サイクルを例として述べる。まず、このサイクルは酵素反応が組み合わさって構成されている。そして、ある酵素反応によって生成される化学物質は次の酵素反応の原料物質になる。出発化学物質：クエン酸(citrate)は酵素：aconitase によって cis-aconitate に変化し、以下次々と酵素反応によって炭酸ガス(CO₂)と高いエネルギーを持つ化学物質を出しながら炭素数の少ない化学物質に変化し、途中の2個の fumarate が酵素：fumarase によって1個の(s)-malate に成る段階を経て、元の citrate に戻るが、サイクルを1回回るごとに1個分の citrate が消えていく。図1では化学物質は○で示されてその横に名前が書かれ、酵素は長方形で示されて酵素名ではなく、ID番号、たとえば、aconitase は 4.2.1.3 と書かれている。

同じ酵素であっても関わる反応が異なることがある。たとえば、aconitase は citrate を cis-aconitate に変化させる反応と、cis-aconitate を isocitrate に変化させる反応の2つに関与するから、この2つの反応を区別する必要がある。

また、この酵素反応は可逆反応である。図1で矢印が双方向に付いているのはそれを示している。従って、一つの酵素反応の中でも原料物質、生成物質という属性を付けることは望ましくない。反応の向きが逆になれば、属性が逆になるからである。

さらに経路には枝分れが在る。図1のA点からB点に至るルート1とルート2がその例である。この場合は反応に関与する酵素が異なるが、これは生物種や生体組織に依り異なる酵素が存在するためである。このことは、対象とする生体組織に依り、主たる経路が異なることを示し、その場合は経路の図を変形して検討できると便利であることを示唆している。また、情報の構造モデルは経路の枝分れに対応できることが必要である。また、図1のa点に上から来る矢印はクエン酸サイクルとは異なる代謝経路マップからの接続である。この種の接続関係も表現できることが必要である。

3 代謝経路マップの RDF によるモデルとその XML による外部表現

以下のクラスを設定した。

- **substrate**: 図1の○で表されるノードに対応する。化学物質を示すが、クラスの各インスタンス(以下、リソースという)を識別するため、○一つ毎にIDを設定する。異なる場所に在る○が示す化学物質が同一物質の場合でも、IDを個別に設定する。

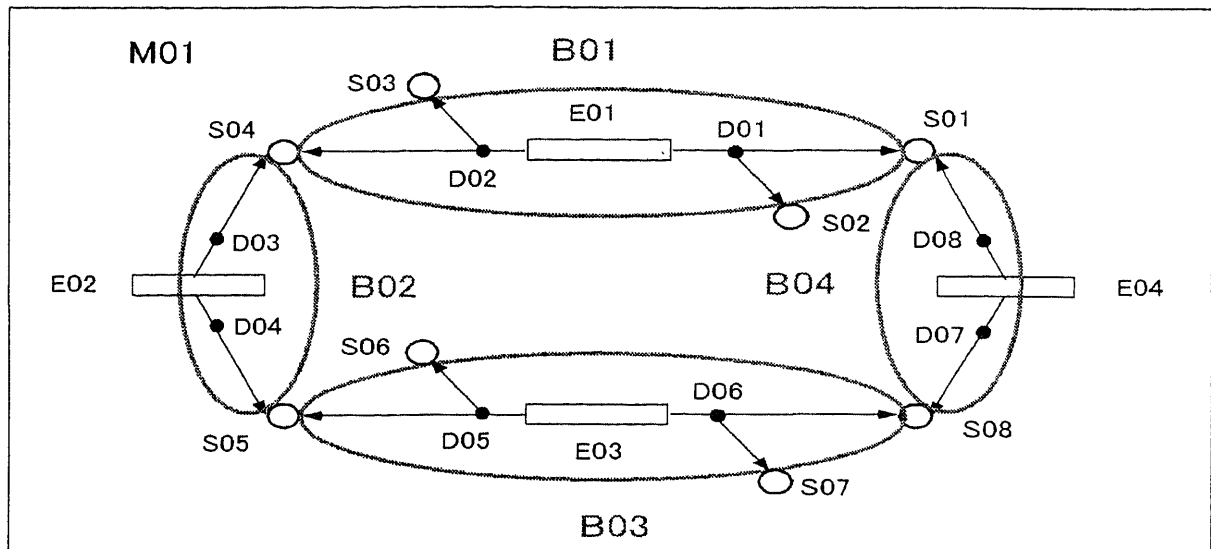


図2 代謝経路マップの構成モデル 仮想のマップM01を例として

- **enzyme:** 図1の長方形で表されるノードに対応する。酵素を示すが、クラスの各リソースを識別するため、長方形一つ毎に ID を設定する。異なる場所に同一酵素が在る場合でも、ID を個別に設定する。
- **direction:** 酵素反応に対して同じ向きに在る化学物質を反対向きにある化学物質と区別するために用いる。たとえば、citrate から cis-aconitate への反応では酵素 ID:4.2.1.3 の citrate 向きと cis-aconitate 向きの2つの direction のリソースを設定し、それらを識別する ID を設定する。
- **entity:** substrate や enzyme の実体を示す。substrate や enzyme はマップ上の位置に対応しているため、entity によって実体を示す。
- **block:** 個々の酵素反応の単位を示す。個々のリソースは ID によって識別する。
- **map:** 代謝経路マップに対応する。連鎖した酵素反応の集まり。個々のリソースは ID によって識別する。

図2に例として仮想の map: M01 の RDF モデルに基づく構成を示す。M01 は block: B01, B02, B03, B04 から構成される。そして、B01 は E01, S01, S02, S03, S04, D01, D02 を含む。また、S01 から見れば、S01 は B01 と B04 に繋がりに、B01 における direction は D01、B04 では D08 となる。

個々のリソースは以下に示すプロパティとリテラルを持つ。

- リソース: substrate はプロパティとして、1つ以上の block、2つの direction、1つの entity を持つことができる。対応するリテラルは順に block の ID、direction の ID、化学物質名である。
- リソース: enzyme はプロパティとして、1つの block、2つの direction、1つの entity を持つことができる。対応するリテラルは順に block の ID、direction の ID、酵素 ID である。

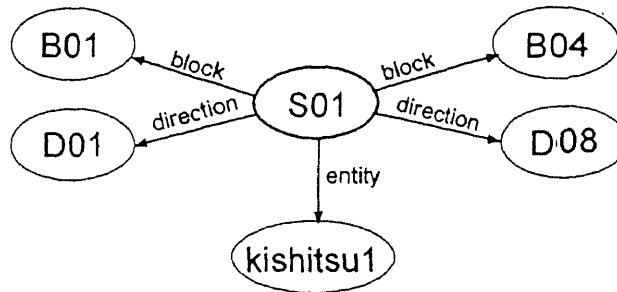


図3 substrate: S01に関するRDF表現

- ・ リソース：direction はプロパティとして、1つの block と、1つの enzyme と、1つ以上の substrate を持つことができる。対応するリテラルは順に block の ID、酵素 ID、substrate の ID である。
- ・ リソース：block はプロパティとして、1つ以上の enzyme と、2つの direction と、2つ以上の substrate を持つことができる。対応するリテラルは順に酵素 ID、direction の ID、substrate の ID である。
- ・ リソース：map はプロパティとして、2つ以上の block を持つことができる。対応するリテラルは block の ID である。

例として substrate: S01 に関する RDF 表現を図3に示す。S01 は2つの block プロパティのリテラルとしてそれぞれ B01 と B04 を持ち、2つの direction プロパティのリテラルとしてそれぞれ D01 と D08 を持ち、1つの entity プロパティのリテラルとして kishitsu1 を持つ。同様に代謝経路マップを構成する各リソースについて RDF 表現が書ける。

図3のRDF表現に対応するXMLによるRDFの外部表現は以下のとおりである。

```

<description about="S01">
  <entity>kishitsu1</entity>
  <direction>D01</direction>
  <direction>D08</direction>
  <block>B01</block>
  <block>B04</block>
</description>

```

同様に代謝経路マップを構成する各リソースについてXMLによるRDF外部表現が書ける。そして、これらのRDF外部表現を並べることによって代謝経路マップ全体を記述する。

4 Java アプレットを用いたインタラクティブな代謝経路マップの表現

前述のXMLによるRDF外部表現を解釈し、それをもとに代謝経路マップをインタラクティブに再構成できるJavaアプレットを作成した。このアプレットを動かせば、ユーザはマウスを使ってディスプレイ上のノードをドラッグすることによって、代謝経路マップを自由に変形することができる。

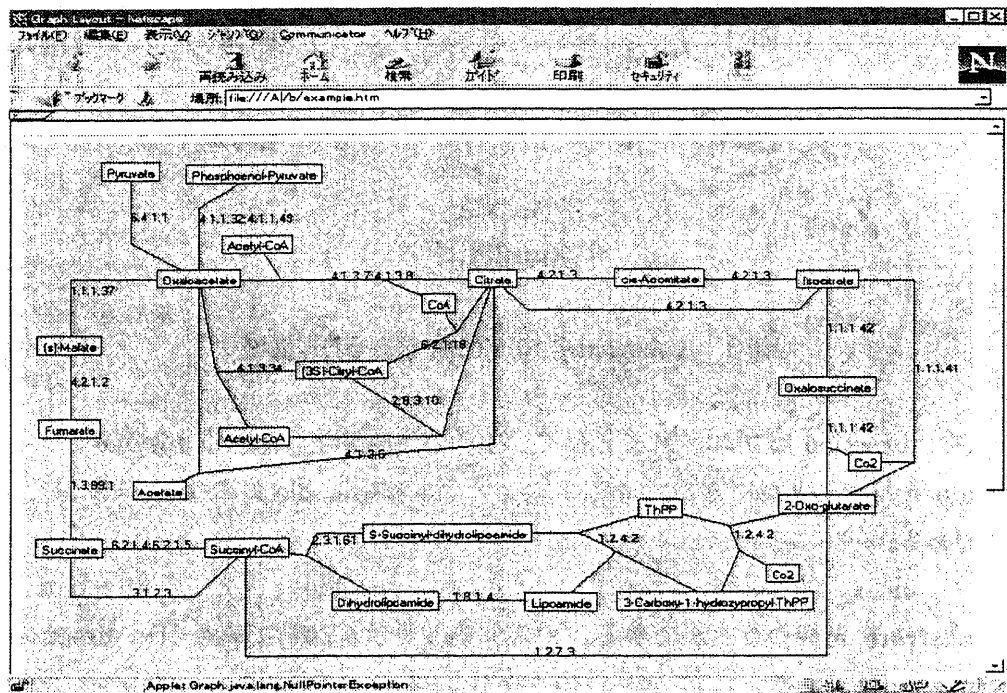


図4 代謝経路マップ(クエン酸サイクル)のインタラクティブな変形の結果 その1

この Java アプレットをクエン酸サイクルに適用した例を図4に示す。まず、クエン酸サイクルを構成する各リソースについて、XMLによるRDF外部表現で記述した。次いで、そのXMLデータを前述のJavaアプレットが解釈し、クエン酸サイクルを描画する。描画された代謝経路マップの構成要素であるsubstrateやenzymeなどのノードをマウス・ドラッグすることによって代謝経路マップを変形することができる。こうして図1に近い形にした代謝経路マップを図4に示す。図4は図1の表現と基本的に同じである。異なる点は、化学物質は○ではなく、長方形で示されていること、反応の矢印は省略されていること、酵素IDは長方形で囲まれてはならず、そのまま書かれていることなどに限られる。

描画された代謝経路マップは自由に変形できることを示すために、図4とは左右逆の対称の形に変形したマップを図5に示す。

5 考察

map, block, substrateは一見、階層関係にあるように見えるが、substrateが隣り合う2つのblockの双方に属するため厳密な階層関係にはない。これはsubstrateが或るblockでは生成物質であると同時に隣のblockでは原料物質だからである。このような現象は代謝経路に限らず内部構造を持つものが経路を構成する場合には起きる現象である。この場合には階層関係として記述するのではなく、構成要素(代謝経路マップの場合はsubstrate)が同時に複数の構成単位(今の場合はblock)に属することを記述する必要がある。また、この記述により、substrateが複数のblockを接続していることが明示され、その結果、経路の描画が可能になる。

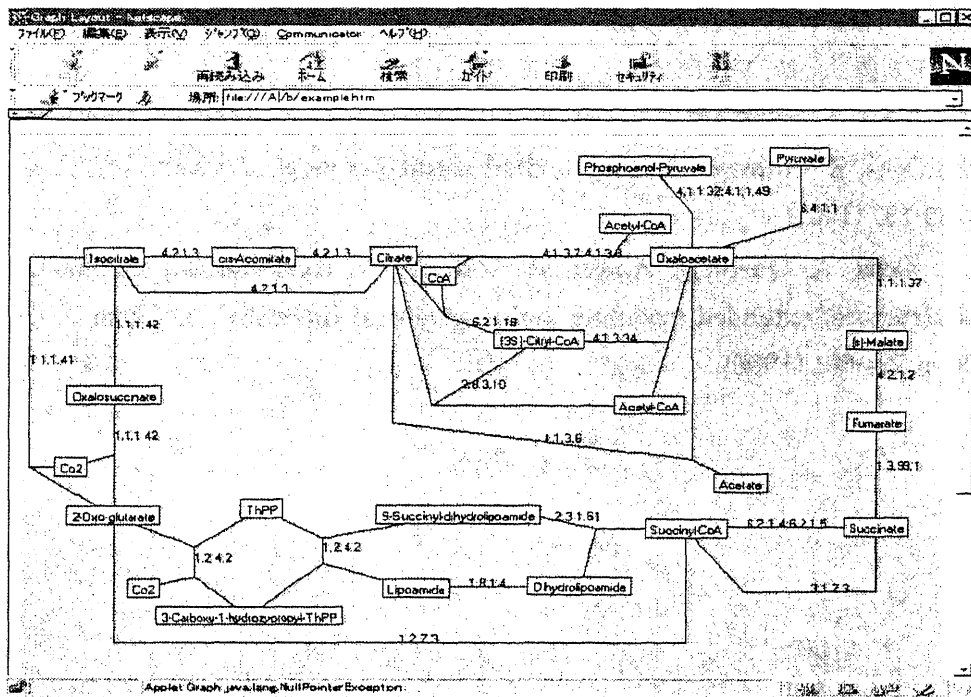


図5 代謝経路マップ(クエン酸サイクル)のインタラクティブな変形の結果 その2

代謝経路マップを自由に変形することによって、経路の中でユーザが注目している部分を強調したマップにすることができる。また、そうすることによって複雑な代謝経路の理解を深められる。さらには何かアイデアを思いつくことを助ける可能性もある。

RDF 自体は静的な表現であるが、Java アプレットと組み合わせることにより動的かつインタラクティブな表現になりうることを本研究は示した。

6 おわりに

以上、代謝経路マップを記述する RDF 表現を考案して、XML による RDF 外部表現を記述し、それを解釈してインタラクティブに変形可能な代謝経路マップを描画する Java アプレットを作成した。

参考文献

- [RDF]O. Lassila, R. Swick: "Resource Description Framework (RDF) model and syntax specification", W3C recommendation 22 Feb. 1999; <URL:http://www.w3.org/TR/REC-rdf-syntax>; D. Brickley, R. Guha: "Resource Description Framework (RDF) schema specification 1.0", W3C candidate recommendation 27 Mar. 2000; <URL:http://www.w3.org/TR/rdf-schema>.
- [XML]Extensible Markup Language (XML) 1.0, W3C recommendation, 10 Feb. 1998; <URL:http://www.w3.org/TR/1998/REC-xml-19980210>

[AI] 浦本直彦, 村田真, 田村健人, 武田浩一: 小特集「XML: インターネット上での情報の記述と交換」, 人工知能学会誌, Vol.13, No.4, p.506-527 (1998)

[KEGG] ゲノムネットの代謝経路マップ <URL: <http://www.genome.ad.jp/kegg/pathway/map/>>

[Ochs91] R. Ochs, K. Conrow: "A computerized metabolic map", J. Chem. Inf. Comput. Sci., Vol.31, p.132-137 (1991)

[Ochs96] R. Ochs, A. Qureschi, A. Sycz, J. Vorbach: "A computerized metabolic map. 2. Relational structure, extended modeling and a graphical interface", J. Chem. Inf. Comput. Sci., Vol.36, p.594-601 (1996)