

（富士ファコムシステム（株）） 佐藤英俊、佐藤利幸
（富士通（株）） 湯田浩太郎*、塩原紀行、野々村辰彦
（味の素（株）） 高原 義之

1. はじめに

化学¹⁾、蛋白およびバイオ²⁾分野において様々な目的のもとに多種多様なデータベースが開発されている。これら多くのデータベースは単一目的/分野での利用を前提として設計されている。このために、複数領域にまたがる境界領域研究や、研究内容的に複数データベースの参照が必要な時には利用効率が極端に低下する。

データベース技術的には複数データベース間を連携して検索する串刺し検索³⁾等も行われるが、これらは同質のデータベース間での連携が基本であり、異質のデータベース間での連携は困難である。

バイオ分野では簡単な形での複数分野のデータベース連携が実現されている。これは互いに独立して構築されたゲノム、蛋白および文献データベースを連携するものであるが、リンク項目が限られており、より高度、且つ細かな連携利用を目指した場合には機能的に不十分である。

演者らは発現プロファイル解析システムと連携する統合データベース（化合物、蛋白、薬理活性を単一のデータベースとしてまとめた）を設計/構築したので報告する。

2. ドラッグインフォマティクスの提案

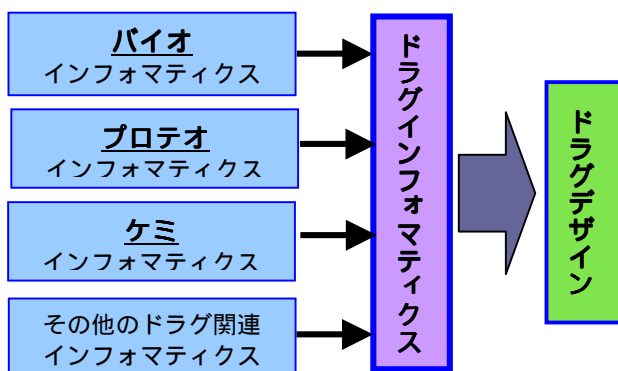


図1 ドラッグインフォマティクスの構成図

現在のドラッグデザイン関連研究はゲノム、蛋白、化合物のそれぞれの分野で急速に拡大している。情報内容的に見るならば、バイオインフォマティクス、

プロテオインフォマティクスおよびケミインフォマティクスが相互に独立して存在し、これらがドラッグデザインの目的に従って適宜利用されている。

上記情報科学群をドラッグデザインを原点として見るならば、個々の情報科学はドラッグデザインを行うに必要な情報の部分集合にしかすぎない。従って、ドラッグデザインという観点から情報を扱う情報科学をドラッグインフォマティクスとし、新たな統一された情報科学として扱うことを提案する。

3. 統合データベースの構築と意義

ドラッグインフォマティクスでは異質のデータ群を扱うことが基本である。さらに、ドラッグデザインという観点で高度に連携し統一して扱うことが求められる。この観点でデータベースを意識するならば、互いに独立する異質のデータベース群は何らかの形で統一した管理を行うことが理想である。

統合データベースの構成

今回開発した統合データベースは化合物、蛋白および薬理活性（毒性を含む）の3種類の情報から構成される。本データベースは発現プロファイル解析を行う研究者が、実験に関連する情報を参照する時に利用することを主目的として開発された。従って、本データベースは発現プロファイル解析の実験支援システムからの呼び出しにより起動する。発現プロファイル実験支援システムからの呼び出し、および3種類の異種データベース間をリンクする基本項目としては化合物情報を用いた。統合データベースの基本構成を図2に示す。

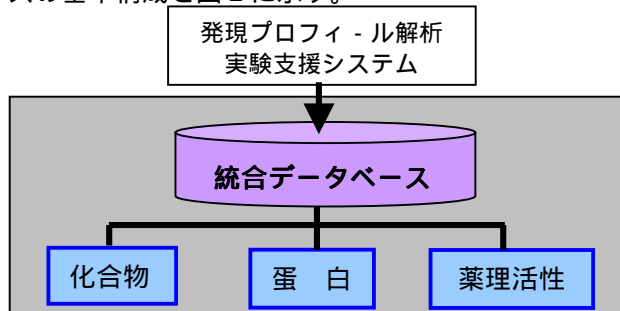


図2 統合データベースの構成図

ドラッグデザインと化合物データベース
化合物データベースのデザインは最終的な使用目的をドラッグデザインに設定して行った。これによ

り、化合物単体としての情報の他に、ドラッグデザインを目指した時に必要となるデータ構成/項目群を優先して設計した。この結果、化合物データベース部分は外部のドラッグデザイン解析システムとの連携が容易である。表1に化合物データベース中に収録されたデータ項目を列記する。

表1 化合物データベース項目リスト

electron	calculation method, total energy, HOMO, LUMO, mean polarizability, total dipole moment, heat of formation, comments
general	melting point, boiling point, viscosity, vapor pressure, molar refractivity, surface area, molecular weight, parachor, others
metabolization	bio concentration, water solubility, aqua toxicity, atmosphere oxidation, aqua hydrolysis, bio degradation, comments
drug-design	actual LogP, calcd LogP, Kappa1 ~ 3, molecular connectivity chi0 ~ 3, Randich, Balaban, Wiener, comments, others
structure	number of atoms(all), halogen, hetero, hydrogen, ring(all), ring(aromatics), ring(aliphatics), moment of inertia(X, Y, Z)
Journal	journal name, volume, p.p., year, authors, keywords, title, abstracts, comments, url
others	linear and non-linear regression equation, 3d-coordination, others, comments,

代謝情報および毒性、副作用情報の取りこみ
 ドラッグデザインという観点では代謝や毒性に関する情報は極めて重要である。今回の統合システムにはこれらの情報に関するデータ項目を設けてある。代謝関連情報は化合物関連データベースに、毒性や副作用関連の情報は薬理活性データベース中に個々のデータ項目を組み入れた。

データベース構成とGUI設計

データベースはクライアントサーバ形式のRDBとし、クライアントとサーバ間はインターネットを利用した。従って、クライアント側のWEBの

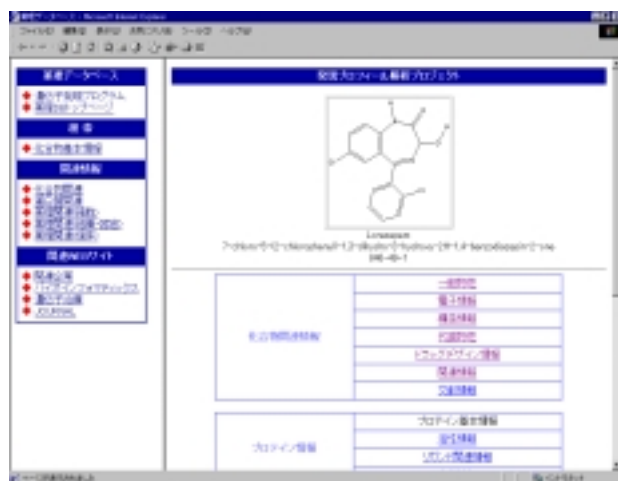


図3 統合データベースのエントリー画面

GUIに関しては化合物、蛋白、薬理活性という異種データの扱いを可能としつつ、且つ統一された設計を目指した。

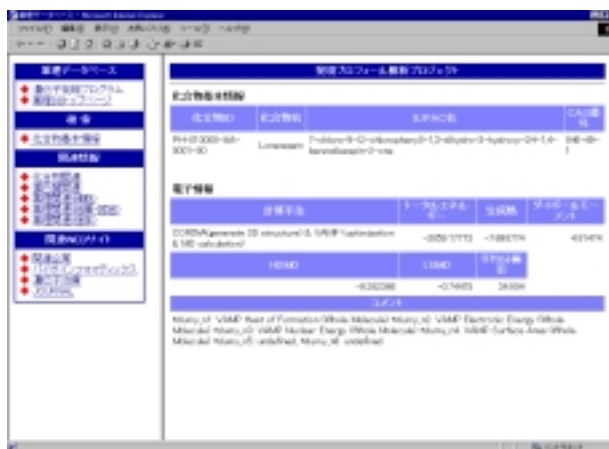


図4 統合データベースの検索結果出力画面

図3および図4に本統合データベースのエントリー画面と、検索結果出力画面の一部を示した。

4. まとめと今後

化合物を相互連携キーとすることで異なる3種類の情報(化合物(代謝を含む)、蛋白、薬理活性(毒性情報を含む))の統合データベースを構築した。これら3種類の異種データ群を統合したデータベースは存在しない。今回開発した統合データベースは“ドラッグインフォマティクス”という新たな概念を実現する最初の事例となる。

今後は化合物(代謝)、蛋白、薬理活性(毒性)のデータベースにおける個々のデータ項目の充実を目指し、コンテンツの充実も目指す。また、3種類の分野以外にも、臨床/治療関連データの取りこみや、ゲノム関連データ(配列、多型、発現プロフィールデータ)の取りこみ、外部ソフトウェア群との連携機能を持たせることで、ゲノム、蛋白、ドラッグ、医療と一貫通貫で情報を見渡せる“ドラッグインフォマティクス”の環境作りを目指したい。

謝辞

本システムの開発に際し、適切なアドバイスを頂きました「遺伝子発現頻度解析(ヒトゲノム機能探索のための遺伝子発現プロファイルデータベース研究開発)」参加研究諸氏に感謝申し上げます。

参考文献

- [1] 化学関連データベースURLリスト;
<http://www.liv.ac.uk/Chemistry/Links/refdatabases.html>
- [2] バイオ関連データベースURLリスト
<http://sullivan.bu.edu/~mfrith/tools.shtml>
- [3] 串刺し検索サイトURL ;
http://www.access.gpo.gov/su_docs/aces/aaces002.html