

Cheminformaticsシステムの現状と今後の動向

Present Situation and Future Trend of Cheminformatics

あらまし

“Cheminformatics”とは、化学研究において派生する膨大な各種のデータをどのように解析・加工し、どのように蓄積・データベース化していくかをITの側面から支える技術の総称である。

本稿では、Cheminformaticsシステムがどのように進化・発展してきたかを述べ、現状の技術動向および問題点について明確にするとともに、これらの問題点を踏まえた上で今後のCheminformaticsシステムの方向性・富士通の今後の取組みについて述べる。また、Cheminformaticsと対で使われる“Bioinformatics”「生物情報工学」システムとの統合の重要性についても述べる。

Abstract

Cheminformatics is a generic name for the information technologies (IT) for studying and realizing methods of analyzing and processing vast amounts of the many kinds data derived from chemical research and storing the data in a database. This paper explains how cheminformatics systems have evolved and progressed, describes the present technical trend and problems related to cheminformatics, and forecasts the future directions of cheminformatics systems and Fujitsu's activities in this field. This paper also describes how important it is to combine cheminformatics systems with bioinformatics (biology information engineering) systems, which are often paired with each other in practical applications.



萩原 稔（はぎわら みのる）
ライフサイエンス推進室 所属
現在, Bioinformatics, Cheminformatics
関連システムの企画・開発・拡販・
サポートに従事。

ま え が き

“Cheminformatics”は一般には余り使用されない単語であるが、ライフサイエンス分野においては日常的に使われている単語であり、直訳すれば「化学情報学」ということになるが、一般的には「化学情報管理」と呼ばれている。

Cheminformaticsシステムは、製薬企業や化学企業の基礎研究所において、大量に合成される化合物情報をどのように管理していけばよいかという課題とともに大きく発展してきたと言える。Cheminformaticsにおいて最も重要かつ基礎となる情報ソースは化学構造式である。化合物を合成するという事は、すなわちこの化学構造式から成る物質を作成するということである。この化学構造式の一部が違っただけでその物質の性質・機能が大き

く異なることから、この化学構造式を効率良く管理することは化学研究者にとっては極めて重要である。ダイオキシンの化学構造式の例を図-1に示す。二つの化学構造式は非常に類似しているが、毒性は全く異なる。

また、合成された化合物が当初設計したとおりの化学構造式となっているかを測定したり、この化合物が予測していたとおりの性質・機能を有しているかどうかを試験したりする必要があり、これらの測定・試験から出力されるデータもまた化学構造式とリンクして管理する必要がある。さらに何を合成すべきか、またどうやって合成するかといった合成実験以前に調査・情報収集しておくことも合成実験を効率良く進めていくためには非常に重要であり、この調査・情報収集のためにCheminformaticsシステムの果たす役割は大きい。創薬研究における研究業務フローとCheminformaticsシステムの関連を図-2に示す。

本稿では、Cheminformaticsシステムの現状と今後の動向について述べる。

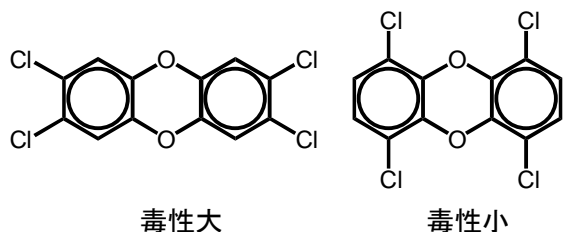


図-1 ダイオキシンの化学構造式比較
Fig.1-Structure comparison of dioxin.

Cheminformaticsシステムの変遷

Cheminformaticsシステムは、他分野のシステムと同様にIT技術の発展に合わせて大きく進歩してきた。1980年代初頭は、データベース(DB)サーバに汎用機を用い、エミュレータ端末からこの汎用機にアクセスす

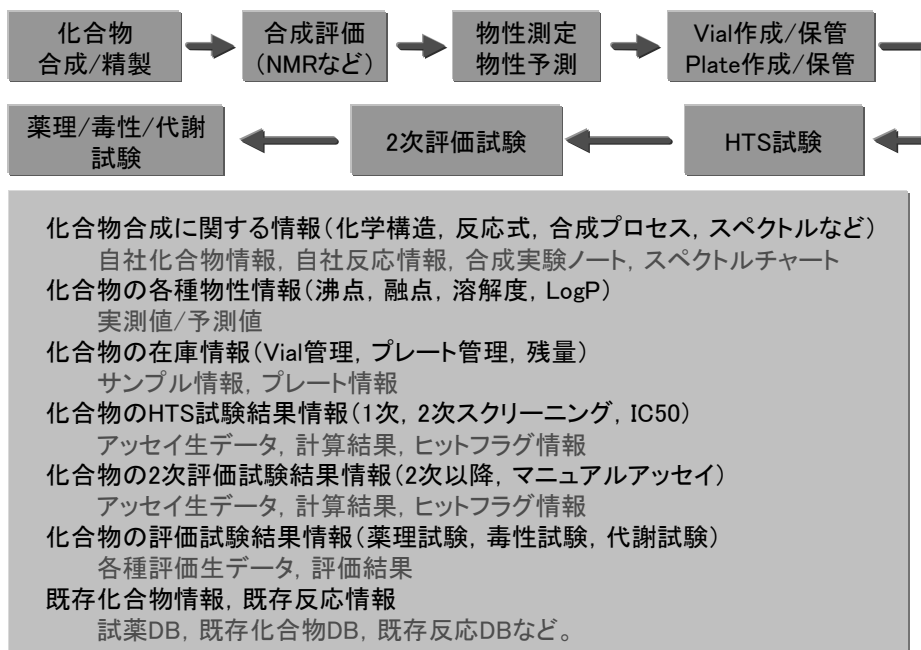


図-2 創薬研究プロセスとCheminformatics関連データ
Fig.2-Drug discovery process and cheminformatics data.

るといった利用形態が一般的であった。当時、化学構造式を取り扱えるCheminformaticsシステムとしては、MDL社⁽¹⁾のMACCS（マックス）やREACCS（リアックス）しかなく、事実上のデファクトスタンダードシステムとして多くの製薬・化学企業がこれらのシステムを導入した。富士通も1983年からMDL社製品の販売・サポートをスタートし、1997年までの15年間にわたりMDL社製品の販売・サポートを行ってきた。

1980年代の後半から1990年代前半になると、DBサーバの主役は汎用機からUNIXワークステーションへとシフトしていった。クライアント側も、WindowsやMacintoshなどのGUIを基本としたパソコンが主流となり、これらGUIベースのサーバアクセス用ツールが開発され、いわゆるクライアント・サーバ型のCheminformaticsシステムが主流となった。この時代においてもシステムの主役はMDL社であり、MACCS/REACCSの後継システムであるクライアント・サーバ型のISIS（アイシス）というシステムをリリースした。ISISもまたこの時代のデファクトスタンダードのCheminformaticsシステムとして、多くの製薬・化学企業がMACCS/REACCSからISISへとリリースした。

この時代までCheminformaticsシステムはMDL社製品の独占状態であったが、これはCheminformaticsシステムで最も重要な情報ソースである化学構造式の管理技術（登録・検索・表示の機能）をMDL社しか持ち合わせていなかったからである。化学構造式を検索条件として検索するためには、化学構造式専用の検索用インデックスが必要であり、表示させる化学構造式についても単なる画像データではなく、化学構造式を構成する原子や結合についてすべての情報をリンクさせておく必要がある。化学構造式をハンドリングするIT技術こそCheminformaticsシステムにとって最も重要である。

1990年代後半に入ると、MDL社以外にも化学構造式をハンドリングする技術を有したソフトベンダが出現してきた。旧Oxford Molecular Group社（現Accelrys社⁽²⁾）のRS³ Discovery（アールエスキューブディスカバリ）やCambridgeSoft社⁽³⁾のChemFinder（ケムファインダ）がこれに当たり、富士通ではこれらの製品をリリースしている。

このことは化学構造式のハンドリング技術が一般化したことを意味しており、化学構造式検索は今や特殊な検索技術ではなくなった。富士通社内においても化学構造

式検索システムを独自に開発しており、その検索性能（スピード）は既存の他社システムよりも優れている。

2000年代になるとDBサーバはPCサーバへとシフトし、クライアントも専用のサーバアクセスツールではなく、ブラウザをベースとしたWebクライアントからアクセスできるシステムに代わりつつある。CambridgeSoft社が提供しているChemOffice WebServer（ケムオフィスウェブサーバ）は、これからのCheminformaticsシステムの形態を先取りしたWebベースのシステムであり、富士通からも既に多くの導入実績がある。MDL社をはじめとする他社もWebをベースとしたシステムを新規に開発したり、既存システムの機能拡張によりWeb対応を急いでいる状況にある。またCheminformaticsシステムのもう一つの大きな技術のトレンドとして、化学構造式をはじめとするデータをすべてOracle⁽⁴⁾上で一元管理できるシステムが主流になりつつある。各種の測定データや試験データは、化学構造式と密接にリンクしており、これらのデータ（化学構造式も含む）を効率的に管理するためには、リレーショナルデータベースシステムを利用することが最も効果的であり、現状ではOracleが事実上のデファクトスタンダードとなっていることから、CheminformaticsシステムをOracleアプリケーションとして開発しているケースが多い。MDL社やCambridgeSoft社、Accelrys社から提供されているCheminformaticsシステムは機能の差はあるが、すべてOracleベースのシステムとなっている。Cheminformaticsシステムの変遷を図-3に示す。

Cheminformaticsシステムの問題点

先にも述べたようにCheminformaticsシステムでは化学構造式が最も重要な情報ソースであり、この化学構造式をいかにハンドリングできるかがシステムの品質を左右する。しかし、余りにも化学構造式の管理だけに着目しすぎて、化学構造式に関連する様々な情報とのリンクがシステムの的に十分ではなく、化学構造式に関連する測定情報や試験情報を化学構造式といっしょに検索・参照することが非常に困難な状況にある。

コンビナトリアルケミストリ⁽⁵⁾やHTS（ハイスループットスクリーニング）⁽⁶⁾などの実験手法の確立とこれに付随した測定機器や実験装置の技術革新により、測定データや試験データは急激に増大し、またこれらのデータを管理・解析するための専用のデータ管理システムも必要となってきた。これによって各種のデータはそれぞ

	1980年代	1990年代	2000年～
主要な 化学DBMS	MACCS/REACCS MDL社	ISIS/Host Oxford Molecular Group社	ISIS/RCG2 Accelrys社
		Synopsys社	RS ³ Accord 統合化
	CambridgeSoft社	ChemFinder ChemOffice WebServer	
サーバ	汎用機	ワークステーション	PCサーバ
クライアント	エミュレータ端末	WindowsPC/Mac	ブラウザ
システム形態	ホスト型	クラサーバ型	Web型
その他			Oracleベース

図-3 Cheminformaticsシステムの変遷
Fig.3-Cheminformatics system history.

れ専用のデータ管理システムで管理されるようになったが、これらのシステムから必要な情報を抜き出し、化学構造式と合わせて検索・参照しようとする、それぞれのデータ管理システムが独自の仕様で開発されており、化学構造式をハンドリングしているシステムも独自仕様の部分が多く、これらのシステムを統合検索できるようにはならない。本当に使いやすいCheminformaticsシステムとは、研究者が簡単なGUIの操作で、化学構造式やこれに付随する測定情報、試験情報を容易に検索・参照できるようなシステムであり、著者らはこれを実現してくれるCheminformaticsシステムを提供してきたと考えている。

Cheminformaticsシステムの今後の動向

「Cheminformaticsシステムの変遷」でも述べたが、今後のCheminformaticsシステムの技術トレンドはWebベースおよびOracleベースの二つに集約されていくであろう。これら二つの技術は、先に挙げたCheminformaticsシステムの問題点を根本から解決することが可能であり、長年Cheminformaticsシステム担当者の頭を悩ませていた化学情報の統合化という高いハードルをクリアすることができる。

Oracleベース

Oracleは最も一般的に利用されているリレーショナルデータベース（RDBM）であり、化学研究分野においても標準的に利用されている。化学構造式をはじめとする各種の測定データ・試験データは、それぞれ化学構造式とのリンクを持たせる必要があり、Oracleはこれ

らのデータを管理する上では最も適しているRDBMである。化学構造式もOracleのテーブルに格納することで、すべてのデータをOracle上で管理し、データの検索もすべてSQLコマンドを使って実施することが可能である。化学構造式を検索条件とする構造検索にしても同様であるから、SQLコマンド一つで構造検索を実行することが可能である。また、化学構造式とほかのテーブル上に格納されている測定値や試験結果を合わせて検索・参照することもSQLコマンドのみを実行することで容易に実現させることができる。化学構造式をすべてOracleに格納して、SQLコマンドで検索・参照することが可能なCheminformaticsシステムが最近になりリリースされるようになってきた。CambridgeSoft社が提供しているOracleをベースとしたCheminformaticsソリューションを図-4に示す。このシステムでは化学構造式を含むあらゆる研究情報はOracle上で一元管理され、目的とする研究情報を容易に取り出せるように各業務アプリケーションが用意されている。

また、Oracleはセキュリティという面からも高い信頼性を得ており、機密性の高い基礎研究情報をすべてOracleによって管理できることは、システム管理者にとっては非常に使いやすく効果的である。

Webベース

様々なアプリケーションも今やWebベースのシステムに変わりつつあるが、Cheminformaticsの領域においてもWeb対応のシステムが浸透しつつある。CheminformaticsシステムのWeb化において最も重要な技術は、ここでも化学構造式の検索・表示をどのように

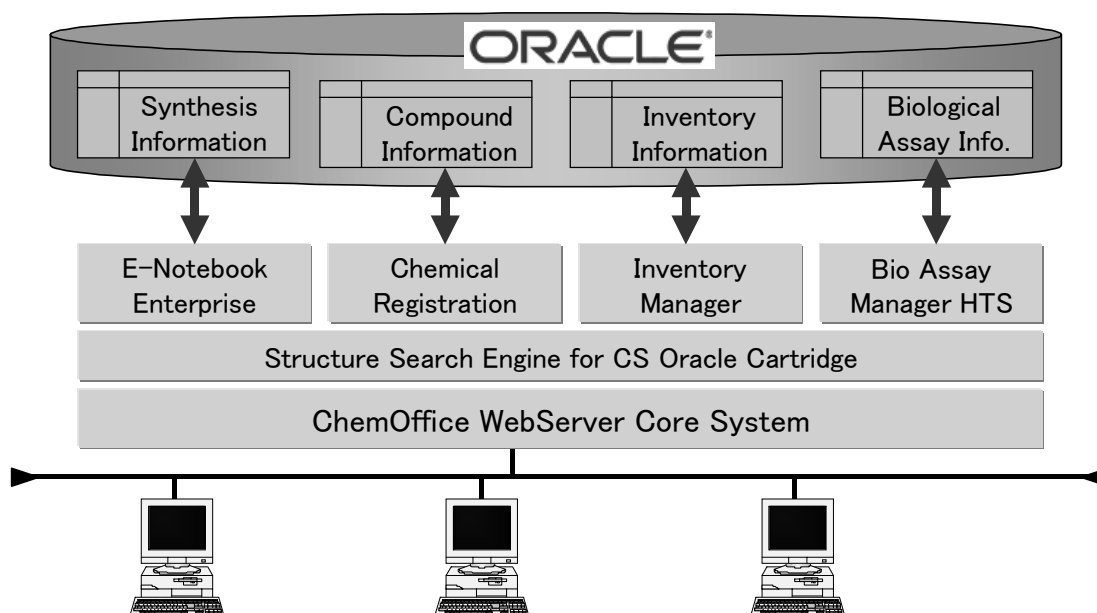


図-4 ChemOffice WebServerエンタープライズソリューションマップ
Fig.4-ChemOffice enterprise solutions relation map.

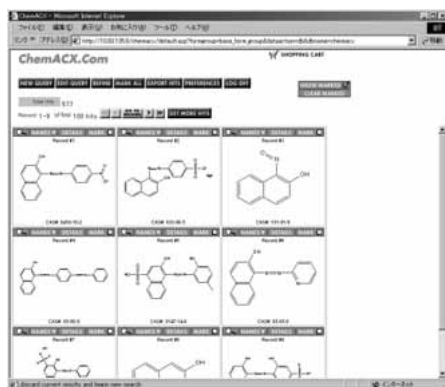


図-5 ChemACX (試薬DB) 表示例
Fig.5-ChemACX matrix view example.

実現するかということである。最近では化学構造式の作画や表示が可能なPluginやActiveX Controlを使ったソフトが開発されるようになり、ブラウザ上で化学構造式を容易にハンドリングできるようになってきた。ブラウザによる化学構造式の検索・参照の例を図-5に示す。これは ChemOffice WebServer で運用されている ChemACX という試薬データベースの検索・参照例である。化学構造式に付随する各種のデータをブラウザ上に表示させることは、一般のWeb技術で容易に実現することができる。つまり、化学構造式と関連する測定値や試験情報を合わせて表示させることは非常に容易である。

「化学構造式を含むすべての化学研究情報がOracle上で一元管理され、これらの情報をすべてブラウザから

アクセスする」、今後のCheminformaticsシステムのあり方がこの一言に集約されており、著者らはこの基本コンセプトに従ったCheminformaticsシステムを提供してきたと考えている。既存の優れたアプリケーションを利用したり、あるいは顧客ニーズに基づくシステムを新規に開発したりすることによって、顧客にとってのベストソリューションを提供していくことが重要である。

Bioinformaticsシステムとの統合

ヒトゲノムプロジェクトに代表されるバイオ研究はここ数年で飛躍的に発展してきたが、この発展を大きく支えてきた一つの技術としてBioinformatics (生物情報工学) は極めて重要であり、今後もBioinformatics技術なくしてバイオ研究の発展はない。遺伝子レベルからの機能解明が可能になった今日、創薬研究はバイオ研究から得られたアウトプットをもとに化学研究を行う時代になってきた。すなわちBioinformaticsシステムとCheminformaticsシステムは別々に存在するのではなく、これらのシステムは統合されなくてはならない。ここで重要なことは、WebやXMLの共通技術を駆使したシステム構築を目指すことであり、著者らはこれらを実現するための各種ツールを提供している。またバイオ研究と化学研究とが密接に関わり合う蛋白質の構造解析の領域にも取り組んでおり、統合的なゲノム創薬研究支援システムの構築を目指している。現在、富士通が提供してい

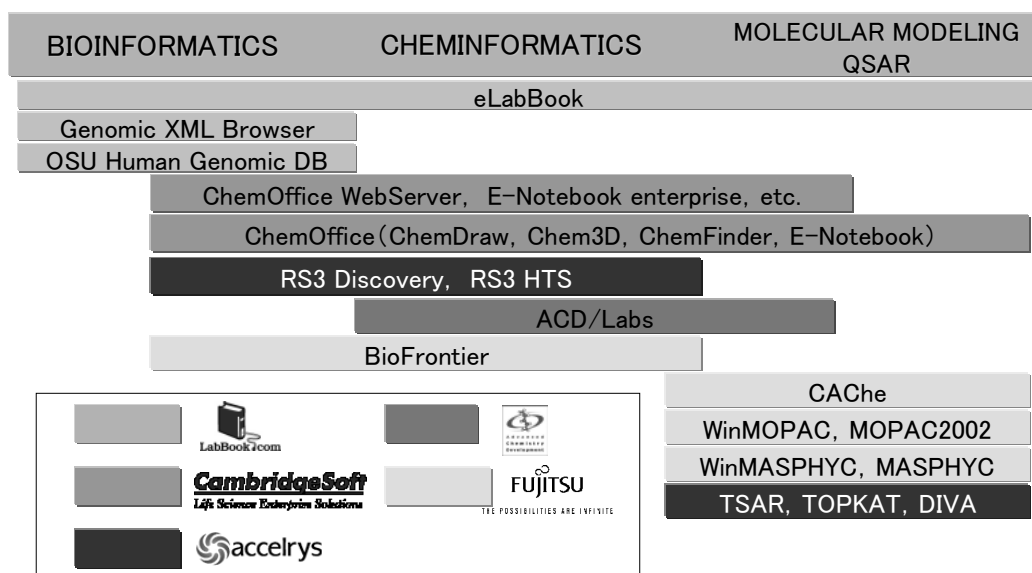


図-6 富士通アプリケーションマップ
Fig.6-Application map for life science area.

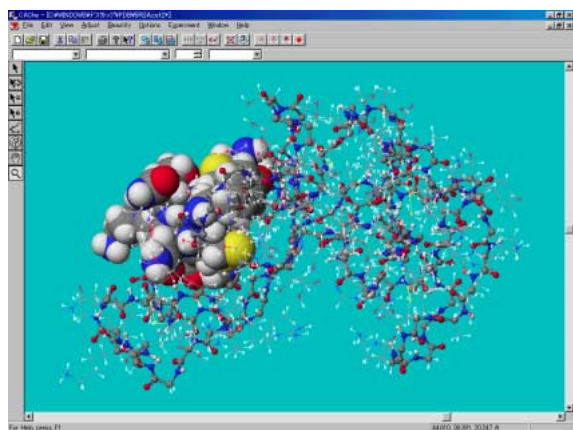


図-7 BioMed CACHeによる蛋白質立体構造の解析
Fig.7-Protein structure analysis using BioMed CACHe.

るCheminformaticsおよびBioinformatics関連のパッケージソフトウェア一覧を図-6に示す。またCACHeシステムによる蛋白質の立体構造解析例を図-7に示す。

む す び

Cheminformaticsシステムの変遷ならびに現状の問題点、さらには今後のCheminformaticsシステムの方向性

や創薬研究におけるBioinformaticsシステムとの統合にまで触れてきた。研究者の方々のニーズにマッチするCheminformaticsシステムを提供していくためには、まだまだクリアしなければならない課題もあるが、著者が提供するCheminformaticsシステムを利用して頂き、創薬研究や材料開発研究の作業の効率化・スピードアップが実現できることを期待している。

参 考 文 献

- (1) MDL Information Systems Inc.ホームページ：
<http://www.mdli.com>
- (2) Accelrys Inc.ホームページ：
<http://www.accelrys.com>
- (3) CambridgeSoft Corporationホームページ：
<http://www.cambridgesoft.com>
- (4) Oracleホームページ：
<http://www.oracle.com>
- (5) コンピナトリアルケミストリー研究会：コンピナトリアルケミストリー．化学同人，p.2-22．